JSLHR

Research Article

The Influence of Misarticulations on Children's Word Identification and Processing

Breanna I. Krueger,[a] Holly L. Storkel,[a] and Utako Minai[a]

Purpose: The purpose of the present studies was to determine how children's identification and processing of misarticulated words was influenced by substitution commonness.

Method: Sixty-one typically developing preschoolers across 3 experiments heard accurate productions of words (e.g., "leaf"), words containing common substitutions (e.g., "weaf"), and words containing uncommon substitutions (e.g., "yeaf"). On each trial, preschoolers chose between a real object picture (e.g., a leaf) and a nonobject (e.g., an anomalous line drawing). Accuracy and processing were measured using MouseTracker and eye tracking.

Results: Overall, children chose real objects significantly more when presented with accurate productions (e.g., "leaf") than misarticulated productions (e.g., "weaf" or "yeaf"). Within misarticulation conditions, children chose real objects significantly more when hearing common misarticulations (e.g., "weaf") than uncommon misarticulations (e.g., "yeaf"). Preschoolers identified words significantly faster and with greater certainty in accurate conditions than misarticulated conditions.

Conclusions: The results of the present studies indicate that the commonness of substitutions influences children's identification of misarticulated words. Children hear common substitutions more frequently and therefore were supported in their identification of these words as real objects. The presence of substitutions, however, slowed reaction time when compared with accurate productions.

Supplemental Material: https://doi.org/10.23641/asha.5965510

Every spoken utterance is characterized by elements of variability caused by a variety of sources. These sources range from subtle, mechanical differences in oral motor movement (e.g., phonetic variability) to the substitution of entire phonemes as a result of accent or developmental misarticulations (e.g., phonemic variability). The ability to attend to the intended speech target despite such inconsistency plays a key role in children's language acquisition. Children are capable of and, in fact, quite good at understanding words from the input they receive; however, different types of variability seem to influence the perception and processing of words in different ways.

Speech Variability

There are two types of variability produced by speakers that may impact speech perception. Phonetic variability occurs in every utterance of a phoneme—each production presents with unique spectral characteristics. Phonetic variability is variability that does not trigger a change in meaning. For example, the [tʰ] in "top" is different from the [t] in "stop" in terms of aspiration, yet both are perceived as the underlying phoneme /t/ in English. The phonetic context of the cluster /st/ in "stop" reduces the degree of aspiration and voice onset time as a result of coarticulation, which produces a source of phonetic variability (Klatt, 1975). Although this type of variability is predictable from context and arrives from many sources (e.g., phonotactics, accent), children still acquire words from their environment, thanks to the ability to encode acoustic information into phonemic categories (Liberman, Harris, Hoffman, & Griffith, 1957; Liberman, Ingemann, Lisker, Delattre, & Cooper, 1959).

A second form of speech variability is phonemic variability. Phonemic variability occurs when a single phoneme is substituted for another. Phonemic variability occurs in cases where the speaker is, for whatever reason, unable to produce the target phoneme in the phonological system. In some cases, the phoneme may not occur in the speaker's first language; thus, a substitution is made (i.e., Arabic does not have a /p/ as a phoneme; thus, a /b/ is

[a]University of Kansas, Lawrence

Correspondence to Breanna I. Krueger: bkrueger@uwyo.edu

Editor: Julie Liss

Associate Editor: Maria Grigos

Received September 26, 2016

Revision received April 9, 2017

Accepted November 21, 2017

https://doi.org/10.1044/2017_JSLHR-S-16-0379

Disclosure: The authors have declared that no competing interests existed at the time of publication.

often used in its place). Depending on the type of change, phoneme changes may trigger a change in meaning. For example, when a child produces a developmental misarticulation such as /w/ for /l/, the intended message may be misunderstood. If a child says "wight" instead of "light," this triggers a change in meaning for the listener, because the substitution is a different real word, "white."

Perception of Variability

The perception of speech variability, particularly phonetic variability, has been well studied in adults. Generally, adult listeners process phonetic variability with little difficulty. The ease of interpretation can likely be attributed to having mature phonological representations and lifelong experience with many forms of variability. Clopper and Pisoni (2004) found that adults who were exposed to a variety of dialects as children accurately categorized the dialects into the region in which the dialect is commonly found as adult listeners. This finding suggests that adults draw upon experiences with speech variability from experiences as children to identify regionally specific locations of dialect types and use that experience to comprehend and understand words with phonetic variability. This finding may suggest that children have the capability to be flexible in their interpretation of speech sound differences and may use top–down information, such as knowledge of dialect, to assist in their interpretation. In addition to drawing upon these experiences, evidence has shown that adults use top–down lexical and prelexical cues to assist with accommodation of variability, as in the case of ambiguous or degraded productions (McClelland, Mirman, & Holt, 2006). Adults use word level information to infer the intended target, which allows them to categorize the different phonetic variations of each phoneme. Each of these processes involves online learning and adaptation on the part of the perceiver, which slows processing time considerably (McClelland et al., 2006).

Previous research investigating children's perception of phonological and phonetic variation are largely focused on how these types of variability impact children's identification of words and how that process is attuned to the word learning process. This research suggests that very young children perceive words with a level of phonological detail and segmentation ability that assists them in their recognition of words (Swingley, 2005, 2009). At approximately 19 months, toddlers acquire the ability to accommodate different dialects of English, such as Jamaican-accented speech (Best, Tyler, Gooding, Orlando, & Quann, 2009). According to Best et al. (2009), this skill appears to emerge sometime after the age of 15 months, as younger children produce more phonetic approximations of auditory stimuli than older children; older children, on the other hand, identify the words as variable productions and produce them in their own dialect. This level of perceptual ability assists children in learning new words from their environment, and from a variety of sources, but also points to a level of meta-awareness of the speech sound system itself. Other research has suggested that, by the age of 4 years, children phonetically imitate different dialects and accents when given auditory stimuli, but by the age of 7 years, children produced the accented productions in their own dialects, rather than making a phoneme-by-phoneme imitation (Bent, 2014). In each of these studies, the researchers suggested that their findings may be the result of 4-year-olds being in a period of lexical acquisition, and this flexibility is necessarily a requisite to word learning (Bent, 2014; Best et al., 2009). Children must be able to segment words into phonemes, evaluate whether they are part of a new word, and then access possibilities that match what is already existing in the lexicon to make their decision. This ability appears to be involved in the process of word learning. In fact, Altvater-Mackensen and Mani (2013) found that misarticulated words produce cascaded activation of semantically related words in 2-year-olds. That is, a misproduction of the word "nurse" would produce cascaded lexical activation of the semantically related word "doctor," suggesting that, even at the age of 2 years, children are flexible in their acceptance of variable speech. The findings of these studies show that children across the preschool years are sensitive to both phonetic and phonemic variability and that their awareness is a key element in the process of word and sound learning.

In the acquisition of their native language, children must be able to segment words into phonemic units to establish whether a word is known or unknown; however, these units must have some flexibility in terms of children's acceptance to accommodate speech variability that they are inevitably exposed to whether it is the result of dialect, accent, or misarticulation. Creel (2012) examined children's identification and processing of words containing phoneme substitutions. This study targeted children's identification and processing of words containing phoneme substitutions as either real words or new words. The substitutions were created by shifting the typically produced phoneme by phonological features. For example, "fish" is shifted to "fesh" where the vowel /ɪ/ is shifted downward one step in the vowel space to /ɛ/, and the /g/ in "grape" is shifted to "krape" where there is a change in voicing from /g/ to /k/. These substitutions varied in terms of feature distance from the target (e.g., /greɪp/ → /kreɪp/ is "close" and /greɪp/ to /treɪp/ is "far") to determine if the feature distance was an influential factor in children's identification and processing of these words. The results showed that, as feature distance increased, children identified words as real objects significantly less often than "close" or accurate productions. Furthermore, any presence of variability slowed children's processing. Creel also included an analysis of children's eye gaze and found that fixations on the target occurred later when hearing altered productions of words, but there was no difference in online eye-gaze fixation latency between "close" feature changes and "far" feature changes within shifted stimuli conditions. Both atypical productions yielded slowed and reduced looks to the target object. These results suggest that children are largely biased toward selecting real objects, but their processing is adversely affected, and they seem to consider other objects longer when hearing shifted stimuli than accurately produced stimuli.

The results of Creel (2012) suggest that children are sensitive to phonemic variability but processing is slowed. However, the stimuli in that study were created by altering phonological features of the phonemes, such as voicing and place, systematically. In other words, these changes were not developed to mimic any real-world model with which a child may have experience. However, children do have experience with phonemic variability: the misarticulations they hear from their peers and produce themselves. Most children will produce phonemic substitutions as they develop their ability to produce the sounds of their first language. These substitutions are typically the same across children, according to developmental normative data (Smit, Hand, Freilinger, Bernthal, & Bird, 1990). This raises the question of whether common misarticulations that are routinely heard in the environment (e.g., "wight" for "light") are processed differently than uncommon misarticulations that are rarely, if ever, encountered in the environment (e.g., "yight" for "light") as well as whether these misarticulated productions are processed differently than accurate productions (e.g., "light" for "light"). The investigation of this question allows for consideration of misarticulations as part of the ambient linguistic input that may interact with children's perception and processing of words, which in turn allows for an examination of the role of linguistic input impacting the development of word perception and processing.

The present series of experiments investigated whether children's experience with substitution types impacted their identification and processing of words. Experiment 1 used MouseTracker software (Freeman & Ambady, 2010) and a two-alternative forced-choice selection task to examine whether preschool-age children identify and process words with common (e.g., "weaf" for "leaf") or uncommon misarticulations (e.g., "yeaf" for "leaf") in the same way as accurate productions (e.g., "leaf" for "leaf"). In Experiment 1, the speaker of the stimuli was an adult, to control for substitution type and duration of the tokens and to ensure high-quality, consistent recording conditions. Experiment 2 is a replication of Experiment 1, but with tokens produced by children who misarticulated naturally. This experiment only included accurate productions and common misarticulations because of the rarity of uncommon misarticulations in naturally produced speech. Finally, Experiment 3 is a replication of Experiment 1 using an eye tracker for a closer examination of children's processing of accurate and misarticulated words. Eye tracking allowed for a more fine-grained analysis of children's online processing and eye-gaze behavior as well as a comparison between the MouseTracking and eye-tracking technologies. Across experiments, the dependent variables were proportion of real-object selections (when given a choice between a real object and a novel object), reaction time, and area under the curve (AUC) for mouse trajectories (Experiments 1 and 2) or looking patterns for eye gaze (Experiment 3). Independent variables were word types: accurate production, common misarticulation, and uncommon misarticulation.

Experiment 1

Research Questions

This study investigated children's identification and processing of words containing misarticulations to determine if frequent exposure to this form of variability influenced children's real-object responses. To this end, we manipulated the commonness of the substitution type and compared children's identification of these words as real objects or novel objects. Our research questions were the following:

1. Do children select real objects more often for accurately produced words than words containing substitutions?
2. Do children select the real objects for accurately produced words more quickly and with more direct mouse trajectories than for misarticulated words?
3. When words are misarticulated, do children select real objects for common misarticulations more frequently than for those with uncommon misarticulations?
4. When words are misarticulated, do children select real objects for common misarticulations more quickly and with more direct eye trajectories than those with uncommon misarticulations?

Method

Participants

Twenty-three monolingual children (age: *M* = 5;0 years; months, *SD* = 0;3 years;months, range = 4;0–5;11 years; months; nine girls, 14 boys) were recruited from local preschools to participate in Experiment 1. Twenty children were White, two children were multiracial, and one child was Asian or Pacific Islander. Three participants were excluded (two for failing the hearing screening and one for scheduling conflicts). Parents signed informed consent statements and completed case history questionnaires that included developmental and health questions to be answered about their child. In these questionnaires, parents were asked to complete a rating of their child's frequency around other children who misarticulate. We used a visual analog scale for this estimation (McCormack, Horne, & Sheather, 1988; Wewers & Lowe, 1990; Wuyts, De Bodt, & Van de Heyning, 1999). Parents marked along a 100-mm line how much their children were around other children who misarticulate. Their marks were then measured to give a value out of 100. Parents rated their children's exposure to children who misarticulate to be 58 on average (*SD* = 25, range = 23–100). In addition, by recruiting from preschools, we ensured that participants had exposure to other children of their age, increasing the likelihood that they frequently heard misarticulated speech that occurs in typical development. To participate in the study, children were required to be monolingual English-speaking children, who were typical in terms of hearing, receptive language, and articulation. Hearing was assessed using the American Speech-Language-Hearing Association (ASHA, 1997) audiologic screening protocol. Each child completed the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4) to measure his or her receptive vocabulary (Dunn & Dunn, 2007). Participants' standard scores were 114 on average (*SD* = 9, range = 96–129), indicating typical receptive language development (standard scores between 85 and 115 were considered to be within the expected range).

For articulation, participants were assessed using the Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2; Goldman & Fristoe, 2000). All participants were typically developing in terms of articulation, with all standard scores falling within the expected range of 85–115 (*M* = 111, *SD* = 4, range = 99–118). The stimuli for this study were constructed using late-acquired sounds and consonant clusters so children had a greater likelihood of having heard these sounds misarticulated. Previous research identifies "late" sounds as those that are the last to be acquired in typical phonological development (Bleile, 2006; Shriberg, Gruber, & Kwiatkowski, 1994; Smit et al., 1990). Thus, each child's production of late-acquired sounds (/θ, ð, s, z, l, ɹ, ʃ, tʃ/) was examined in five additional words selected from the Phonological Knowledge Protocol (Dinnsen & Gierut, 2008). The late-acquired sounds were elicited in word-initial position in five different words that varied the following vowel. The procedure was the same as the GFTA-2, where children orally respond to a series of pictures. Their responses were phonetically transcribed, and any errors were noted. The phonemes targeted in the present experiments were examined for the participants overall. These phonemes are typically acquired later, on average, but can be acquired earlier as shown in our participants. The results indicated that subjects produced some errors on experimental targets but overall were highly accurate (*M* = 91%, *SD* = 8%, range = 70%–100%). This level of accuracy was expected given the age and articulation skills of the participants. Although some children produced articulation errors, an analysis of the experimental data from the children who produced articulation errors showed no significant differences from the children who produced sounds accurately.

Materials

Auditory Stimuli

Auditory stimuli consisted of 12 target words selected to be picturable and frequent in the child's lexicon, as shown in Table 1 (Storkel & Hoover, 2010). The real words had a log frequency at or above 1 and ranged between 1 and 4 (*M* = 2.64, *SD* = 0.65). Target words were selected to begin with late-acquired sounds (Smit et al., 1990) to ensure that children had maximal exposure to misarticulations associated with these phonemes. Common misarticulations (see Table 1) were selected by examining a cross-sectional study of children's substitution types from Smit (1993) in the 4.5- to 5-year age range (Smit, 1993; Smit et al., 1990). A phoneme was chosen as a common substitute if it was the most common substitute for the target phoneme (e.g., the most common substitute for target phoneme /l/ in initial position was /w/) from the tables in Smit for the 4- to 5-year age range. The most common substitute was selected only if the resulting word was not a real word. In this case, the next most common substitute would be selected.

Uncommon substitutes (Table 1) were chosen by examining both the number and type of features (Chomsky & Halle, 1968). First, the change from accurate to common substitute was compared to determine the number of distinctive features they differed by. Then, the type of place–voice–manner features was noted. This was then used to generate other substitutes that were similar in the number and type of feature differences. These candidates were compared with the typical substitution patterns as reported by Smit (1993). Substitutes that never occurred or that were classified as "rare" were selected for the uncommon substitute. This method allowed for the number and type of feature differences to be matched across common and uncommon substitutes. This matching was necessary because Creel (2012) already demonstrated that the number of feature differences impacted processing. In this way, our common and uncommon substitutes differed only in the frequency of the substitution pattern and not in their closeness to the accurate target.

An additional factor that potentially could influence children's selections was the possibility that, although features were controlled, the two substitutes were phonetically similar to the accurate phonemes. To ensure that this was not a factor, we determined the confusability between the accurate initial phoneme and each misarticulated phoneme, in accordance with the procedures in Han, Storkel, Lee, and Cox (2016). From the tables in Wang and Bilger (1973), we found the proportion of times each accurate phoneme was confused with the misarticulated phoneme. These values are shown as a proportion in Table 1. In terms of the mean proportions, it appears that common substitutes were more often confused with the accurate phoneme than uncommon substitutes; however, a paired samples *t* test between common and uncommon word lists showed no significant difference, $t(10) = 1.65$, $p = .13$, between the two word lists, which suggested that confusability between common and uncommon word lists was controlled.

Auditory stimuli were recorded by the first author, a native speaker of a Midwestern dialect of English, in an anechoic chamber using a Marantz PMD671 solid-state digital recorder. Words were recorded three times each in the carrier phrase "Look at the _____" to ensure similar intonation and to control for listing effects. The words were then extracted from the carrier phrase using Praat sound editing software, with 250 ms of silence embedded at the onset and offset of each word (Boersma & Weenink,

Table 1. Characteristics of stimulus words.

Accurate word	Log freq.	Common substitute	Confusability with canonical	Place–voice–manner differences	Feature distance of common substitute from accurate	Uncommon substitute	Confusability with canonical	Place–voice–manner differences	Feature distance of uncommon substitute from accurate	Difference of accurate–common and accurate–uncommon
"chick"	1.48	ʃɪk	12.92%	Manner	1	fɪk	2.69%	Place, manner	4	−3
"leaf"	2.04	wif	3.30%	Place	6	jif	1.80%	Place	5	1
"thumb"	2.11	fʌm	38.96%	Place	2	ʃʌm	1.69%	Place	3	−1
"comb"	2.36	toʊm	14.51%	Place	4	poʊm	19.03%	Place	3	1
"jar"	2.49	dɑɹ	4.29%	Place, manner	4	gɑɹ	8.93%	Place, manner	4	0
"safe"	2.57	teɪf	2.22%	Manner	3	peɪf	2.57%	Place, manner	4	−1
"van"	2.58	bæn	11.99%	Place, manner	3	dæn	2.52%	Place, manner	4	−1
"shirt"	2.81	sɝt	4.94%	Place	2	fɝt	1.53%	Place	3	−1
"clock"	2.91	kwɔk	—	Place	6	kjɔk	—	Place	5	1
"rope"	3.08	woʊp	3.82%	Place	3	joʊp	1.79%	Place	2	1
"fish"	3.34	pɪʃ	10.17%	Place, manner	3	tɪʃ	5.17%	Place, manner	4	−1
"girl"	3.95	dɝl	8.64%	Place	4	bɝl	4.83%	Place	3	1
M	2.64		10.52%		3.4		4.78%		3.7	−0.4
SD	0.62		10.35%		1.4		5.22%		0.9	1.2

Note. Log freq. = log frequency. Dashes indicate confusability scores were not available for that sound. Accurate stimuli were chosen based on frequency, picturability, and age of acquisition of onset phoneme. Common and uncommon substitutes were chosen based on normative data and feature distance from the accurate production. International Phonetic Alphabet transcriptions are provided to illustrate phonemic differences.

2013). The mean intensity of all items together was measured to be 65 dB, and then each word was scaled to a peak intensity of 65 dB to match. The duration of the words was measured and compared to find tokens that were similar in duration across the three stimuli conditions (see Table S1, Supplemental Material S1).

Visual Stimuli

Twelve black-and-white line drawings were selected from Microsoft Clipart and the Snodgrass and Vanderwart (1980) standardized set of pictures to match each production of the accurate words. Twenty-four pictures were black-and-white line drawings of anomalous objects from Kroll and Potter (1984; see Table S2, Supplemental Material S1). Each misarticulated word was matched with a unique nonobject to reduce the possibility of a mutual exclusivity violation (Markman & Wachtel, 1988; Merriman, Bowman, & MacWhinney, 1989). In other words, if children identified a "weaf" as being the nonobject, they may not also call a "yeaf" as that same novel object. This control reduced the likelihood of introducing a potential experimental confound.

Characteristics of the nonobjects were available from Kroll and Potter (1984) and Storkel and Adlof (2009). Storkel and Adlof collected data from adults for all of the nonobjects and from children for only a subset of the nonobjects. Because the current study was conducted with children, only the subset of nonobjects with child data were used (see Table S2, Supplemental Material S1). From this subset, nonobjects were eliminated based on adult object-likeness ratings from Kroll and Potter. This value was obtained by asking adults and children to rate on a 7-point scale the degree to which the nonobject resembled a real object (where 1 indicated "nothing like a real object" and 7 indicated "looks like a real object"; Kroll & Potter, 1984). Nonobjects with ratings of 6 and above were removed so that the remaining nonobjects would be less likely to be confused with real objects.

Child semantic set size and child strength of the first neighbor from Storkel and Adlof (2009) were then used to further reduce the pool and create two sets of matched nonobjects. Child semantic set size was the number of different neighbors reported by at least two child participants (Storkel & Adlof, 2009). The child strength of the first neighbor was calculated by dividing the number of children who responded with this particular neighbor by the total number of child participants. Nonobjects with many semantic neighbors (i.e., nine or more) and relatively weak first neighbors (i.e., strength of 0.20 or less) were selected so that children would not have a strong real-object interpretation of the nonobjects. The 24 selected nonobjects were then divided into two sets of 12 that were matched in terms of child set size, strength of the first semantic neighbor, and object likeness. Each of these pictures was cropped and resized to 144 × 144 pixels to control for the size of the selection area in the mouse-tracking software. Each set of pictures was paired with the common or uncommon substitutions, with the exact pairing varying across participants. This matching was done to ensure that children had a unique, free choice for each word they heard. By having an individual possibility for each word, we allowed for children to identify that each misarticulated word was potentially a novel object without violating principles of mutual exclusivity (Markman & Wachtel, 1988; Merriman et al., 1989). If only one novel object per set were used, children may have been influenced to choose the real object, when they otherwise would not have done so, because they already identified another misarticulated word in the set as the novel object.

Experimental Design

Experimental Control

Each pair of audio tokens (accurate and common substitute, accurate and uncommon substitute) was matched with a visual display of a real object and a nonobject (e.g., chick and "nonobject 75," chick and "nonobject 79") to create four trials for a given stimulus triplet (chick–fick–shick). The four trials were distributed across two blocks so that the two repetitions of the accurate production (chick) occurred in different blocks and the repetition of the visual display (e.g., chick and "nonobject 75") occurred in different blocks (see Table S3, Supplemental Material S1, for details).

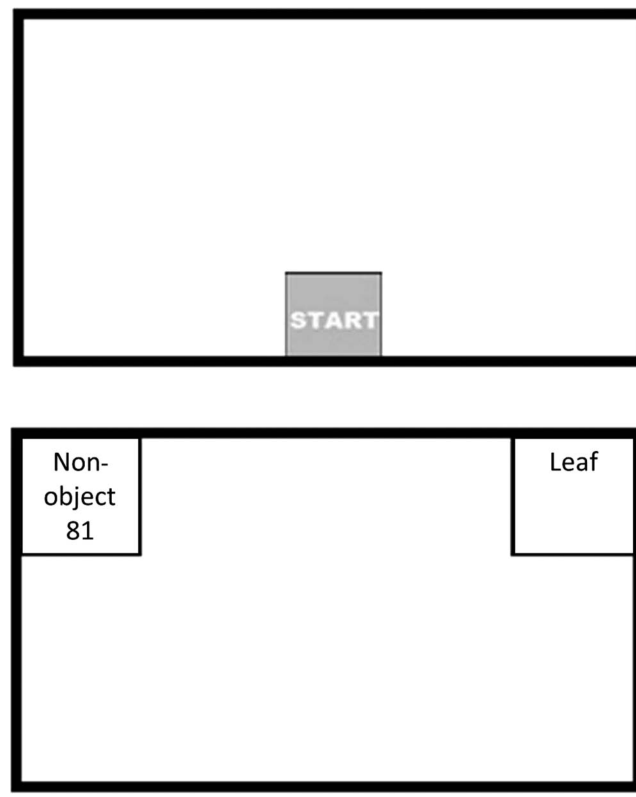
Procedures

Testing took place in a quiet room within the preschools that the participants attended in one or two sessions, depending on the attention of each child. Stimuli were presented on a Dell Latitude D610 PC laptop with external speakers, a 17.25″ × 13.25″ mouse pad, and a single-button optical mouse. The picture and audio stimuli were presented using MouseTracker software (Freeman & Ambady, 2010). MouseTracker was used because previous research suggests that it has the capability to capture processing differences in adult (Freeman, Dale, & Farmer, 2011) and child (Berteletti, Lucangeli, & Zorzi, 2012; Cargill, Farmer, Schwade, Goldstein, & Spivey, 2004) populations that are similar to those found in eye-tracking studies. In addition, MouseTracker required minimal setup and was as portable as a laptop computer, thus allowing for mobile data collection at preschool sites. Although this technology has not been used extensively in children, our pilot data suggested that children as young as 4 years old were capable of acclimating to mouse use in a very short time—within five trials. We ensured that the task was attainable by children by providing large icons placed at a maximal distance from one another (Hourcade, Bederson, & Druin, 2004).

Parents provided informed consent statements, and verbal assent was obtained from preschoolers. Once children assented to the procedure, they were evaluated for hearing, language, and articulation. The experimental study included a brief training session of five trials to acclimate the child to the experimental task in which children were instructed to "click on the picture that matches what you hear." Then, testing proceeded with the experimental trials. If the child was slow or hesitant in their movement of the mouse, he or she was instructed: "Be sure to move the mouse straight and fast!" Once the experimental trials began, all corrective feedback ceased.

Children were shown a screen with a "start" button (see Figure 1). They were instructed to click on the start button to hear the word. After clicking the start button, a 250-ms interstimulus interval of silence was heard, and then the word was presented aurally with two pictures. One picture was a nonobject, and the other was a real-object picture. The child's task was to click the mouse cursor on the picture that corresponded to the word. After the child clicked the picture, a new start button would appear for the next trial. On average, children completed the task in approximately 10 min.

Figure 1. Participants first see the "start" screen (top). After clicking the "start" button, the screen changed to present the object choices (bottom) and the auditory stimulus (e.g., "leaf" or "weaf"). Note that the actual pictures are not shown because of copyright.



Data Analysis

Each output file was processed using the Mouse-Tracker Analyzer software (Freeman & Ambady, 2010). We extracted the proportion of real-object choices, the mean reaction time to the real object, and the mean AUC when selecting the real object. These measurements were separated by independent variable types: common speech sound substitutes (e.g., "weaf"), uncommon speech sound substitutes (e.g., "yeaf"), and accurate production (e.g., "leaf"). Accurate productions were responded to twice: once with the visual display used for the common substitute and once with the visual display used for the uncommon substitute. Thus, dependent variables were computed separately for each display.

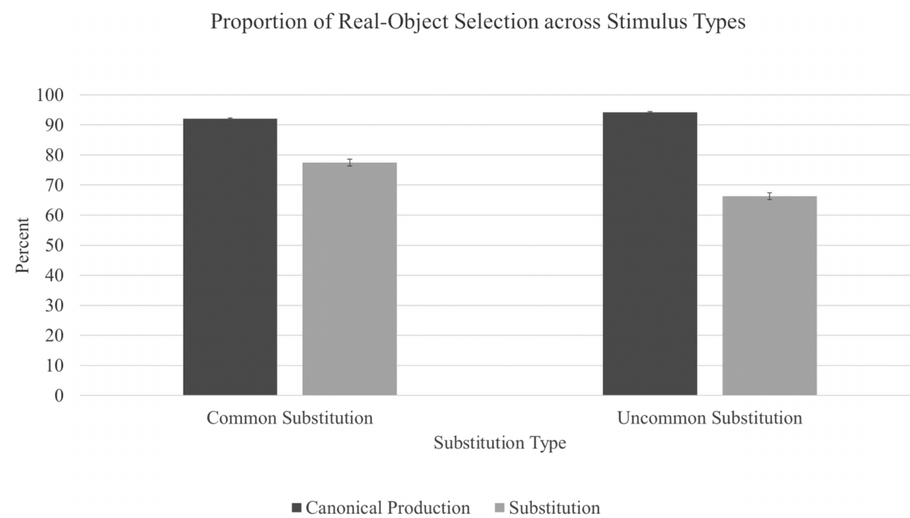
The proportion of real-object selections was calculated for each condition by dividing the number of real-object selections by the total number of trials (12) for each condition. Reaction time was measured by the software as the time between the "start" click and the response click of the mouse. The mean AUC was calculated by considering each mouse path as a line on a coordinate grid. A line of best fit was calculated to find the mean line of all the mouse paths for a condition. Then, the area under this line was calculated. A larger area indicates a more curved line, suggesting greater uncertainty than a smaller area, or more direct line from start to picture selection. Mean AUC provided information about the participants' degree of attraction to the unselected alternative (Freeman & Ambady, 2010; Quétard et al., 2016). A higher AUC measurement suggested that mouse behavior was driven toward the alternate before ultimately making a selection. Reaction time and AUC were only analyzed for trials where the child selected the real object. Real-object responses were analyzed because children, overall, selected the real object for most trials in each condition regardless of the auditory token they heard. Furthermore, the measures of processing were more important on real-object selection trials because they reflected any indecision before finally making the targeted response. Restricting analysis to real-object choices allowed for a comparison of speed and mouse trajectory with the same response outcome. Nonobject selections were examined in a separate analysis but yielded nonsignificant results and have henceforth been excluded from the present discussion.

Results

The proportion of real-object selections was compared using a 2 (Word Type: Accurate vs. Substitute) × 2 (Substitute Type: Common vs. Uncommon) repeated-measures analysis of variance (ANOVA; Figure 2). A significant main effect of Word Type was found, $F(1, 19) = 20.43$, $p < .001$, $\eta_p^2 = .52$, with children choosing real objects significantly more for accurate productions (M = 93%, SD = 1%, range = 67%–100%) than for substitutions (M = 72%, SD = 5%, range = 17%–100%). Furthermore, a significant main effect of Substitution Type was found, $F(1, 19) = 4.43$, $p = .05$, $\eta_p^2 = .19$. However, this main effect was qualified by a significant interaction between Word Type and Substitution Type, $F(1, 19) = 6.20$ $p = .02$ $\eta_p^2 = .25$.

Four paired samples t tests were conducted to investigate the differences for the proportion of real-object selections between accurate productions and misarticulations. A Bonferroni correction for multiple comparisons

Figure 2. Mean proportion of children's real-object selections when hearing accurate productions and common and uncommon substitutions.



yielded a critical p value of .0125 for significance for each comparison. Recall that each real word was paired with two different displays: (a) the real object (e.g., leaf) and the nonobject from the common substitute condition (e.g., nonobject 75) and (b) the real object (e.g., leaf) and the nonobject from the uncommon substitute condition (e.g., nonobject 31). As expected, preschoolers chose real objects equally as often with accurate productions matched with common misarticulation visual displays (M = 92%, SD = 8%, range = 75%–100%) as accurate productions matched with uncommon misarticulation visual displays (M = 94%, SD = 9%, range = 67%–100%), $t(19) = -0.77$, $p = .45$. In contrast, children chose real objects significantly more with accurate productions (M = 92%, SD = 8%, range = 75%–100%) than with common misarticulations (M = 78%, SD = 22%, range = 17%–100%), $t(19) = 3.68$, $p = .002$. Likewise, children chose real objects significantly more when hearing accurate productions (M = 94%, SD = 9%) than when hearing words with uncommon misarticulations (M = 66%, SD = 28%), $t(19) = 4.26$, $p < .001$. Finally, children chose real objects significantly more with common substitutes (M = 78%, SD = 22%) than with uncommon substitutes (M = 66%, SD = 28%), $t(19) = 2.77$, $p = .012$. This finding suggests that children's identification of real objects was impacted by the commonness of the substitution type.

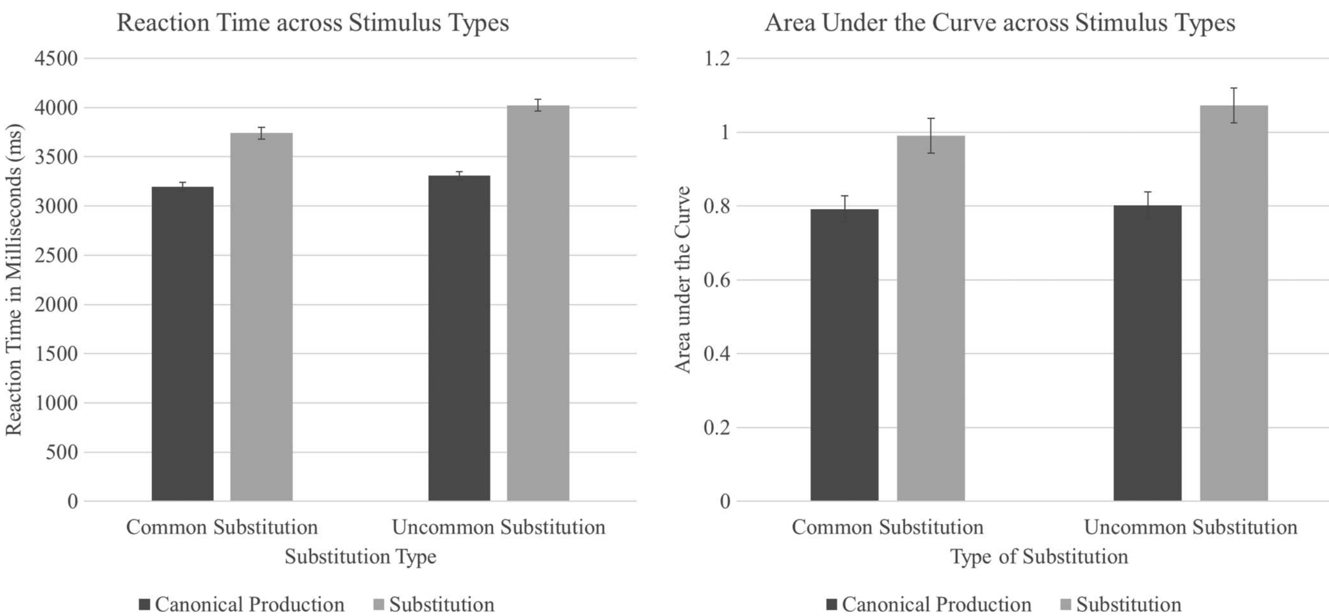
Findings associated with mean reaction time and AUC are shown in Figure 3. Mean reaction time (left chart) was compared using a 2 (Word Type: Accurate vs. Substitute) × 2 (Substitute Type: Common vs. Uncommon) repeated-measures ANOVA. A significant main effect of Word Type (accurate vs. substitute) was found, $F(1, 19) = 30.61$, $p < .001$, $\eta_p^2 = .62$. Children responded more quickly to accurate productions (M = 3250, SD = 942, range = 1814–6196) than to misarticulated productions (M = 3883, SD = 1246, range = 1866–7544), and this was apparent for both the common and uncommon visual displays and substitutes, as shown in Figure 3. A significant main effect of Substitute Type was found, $F(1, 19) = 4.60$, $p < .05$, $\eta_p^2 = .20$. Children responded to words faster in the common condition (M = 3741, SD = 1083, range = 1807–6192) than in the uncommon condition (M = 4024, SD = 1409, range = 2257–7983). No significant interaction between Word Type and Substitution Type was observed, $F(1, 19) = 0.44$, $p = .51$, $\eta_p^2 = .02$. There was no significant interaction, so post hoc testing was not conducted.

AUC was also compared using a 2 (Word Type: Accurate vs. Substitute) × 2 (Substitute Type: Common vs. Uncommon) repeated-measures ANOVA (see Figure 3, right chart). A significant main effect of Word Type was found, $F(1, 19) = 5.11$, $p = .04$, $\eta_p^2 = .21$. These results indicated that the AUC was significantly lower on accurate word trials (Figure 3, right chart, dark bars; M = 0.8, SD = 0.16, range = 0.003–2.65) than on trials containing substitutions (Figure 3, right chart, light bars; M = 1.03, SD = 0.21, range = −0.40 to 4.66). This finding indicates that the mouse trajectory for accurately produced words was more direct than for misarticulated words. No significant main effect of Substitute Type was found, $F(1, 19) = 0.13$, $p = .73$, $\eta_p^2 = .01$. Moreover, no significant interaction between Word Type and Substitution Type was observed, $F(1, 19) = 0.12$, $p = .73$, $\eta_p^2 = .01$. Because no significant interaction between Word Type and Substitution Type was observed, post hoc testing was not conducted.

Experiment 1: Discussion

The results of Experiment 1 showed that children overwhelmingly chose real objects over nonobjects, which is a finding consistent with Creel (2012). These responses suggest that children were generally tolerant of variability

Figure 3. Processing measures for Experiment 1. (Left) Mean reaction time for mouse clicks by condition. (Right) Mean area under the curve for mouse trajectories. This measurement is the area of the sum of the lines created by all of the mouse trajectories by condition. Large values indicate a less direct mouse path and suggest more uncertainty in selection.



in recognizing words. However, children clearly differentiated accurate and misarticulated productions. Specifically, children chose real objects more often, more quickly, and with more direct mouse trajectories for accurate productions than for misarticulated productions. This finding is consistent with the findings of Creel. Furthermore, children were sensitive to whether the misarticulated production was one that they potentially had heard before, as opposed to one that they likely had not heard before. Children selected more real objects for words with common misarticulations than uncommon misarticulations. This extends the work of Creel by showing that children's experience with substitution commonness influences their responding and parallels findings from adults showing that having experience with a variety of different dialects influences responding (Clopper & Pisoni, 2004). In Experiment 1, no difference between common and uncommon was found for reaction time and AUC, suggesting that the commonness of the substitute does not support children in making selections more decisively.

The apparent bias toward selecting real objects may be the result of other untested linguistic factors, rather than an inherent draw toward choosing real objects over novel objects. In adult speech, the fundamental frequency (F0) is lower than that of a child. This supralinguistic factor may influence children's processing because it is rare to hear an adult produce misarticulations. Furthermore, children's processing may be slowed as a result of this exact point, but as viewed through a sociolinguistic lens: Adults do not tend to produce misarticulations. In addition to supralinguistic and sociolinguistic factors, we also considered the possibility that children attend to covert contrasts that children produce in typical phonological development. Covert contrasts are subphonemic contrasts produced by children who are learning new sounds (Scobbie, Gibbon, Hardcastle, & Fletcher, 2000). The productions fall within the typical acoustic values for the sound (e.g., /w/), but the acoustic values for the sound produced as target ([w] for /w/ in white) may consistently differ from the acoustic values for the sound produced as a substitute ([w] for /r/ in right). In fact, in some cases, adults can perceptually distinguish the two different productions. Thus, in naturally produced misarticulations, there may be covert acoustic cues present to indicate whether a child's production should be interpreted as being a misarticulated production of a target word. Moreover, this covert cue could facilitate processing of misarticulated productions, leading to more real-object selections as well as faster and more direct real-object selections for misarticulated words. These cues would likely not be present when an adult is recording a list of real words and nonwords. Therefore, Experiment 2 used naturalistic misarticulations from children with speech sound disorders to determine whether this would decrease the difference between accurately produced and misarticulated productions.

Experiment 2

Experiment 2 contained accurate and misarticulated stimuli produced by children. Because uncommon substitutes are rare in typically developing speech, only commonly produced substitutes and accurately produced words were used as stimuli in this experiment.

1. Do children select real objects more often than nonobjects for accurately produced words than words containing common substitutions?
2. Do children select real objects for accurately produced words more quickly and with more direct mouse trajectories than real objects for commonly misarticulated words?

Method

Participants

Participants were 17 monolingual children recruited from local preschools. Fourteen children were White, one child was multiracial, one child was Native American, and one child was African American. Two were excluded from participation because of failing hearing screenings, and one was excluded for scoring too low on the GFTA-2. Therefore, 14 monolingual preschoolers (age: M = 4;6 years; months, SD = 6 months, range = 3;11–5;4 years;months) were included in the experimental protocol. As in Experiment 1, parents rated their children's exposure to misarticulating children using a visual analog scale. From a possible 0-to-100 scale, parents rated their children's exposure to misarticulating children to be 47 (SD = 24, range = 11–79). These children were typical in terms of standard scores on receptive vocabulary (PPVT-4; M = 114, SD = 11, range = 88–129) and articulation (GFTA-2; M = 113, SD = 8, range = 95–122) and passed a hearing screening at 20 dB for 1000, 2000, and 4000 Hz.

Materials

Auditory Stimuli

Stimuli were selected from recordings taken in previous studies, unrelated to the present research (Storkel & Hoover, 2010; Storkel, Maekawa, & Hoover, 2010). These stimuli were produced by multiple children because no single child produced each of the common misarticulations needed and because typical productions were required as well for the accurate condition. Because of the rarity of uncommon substitutions in typical phonological development, only words containing common substitutions and accurate child productions of words were taken from the recordings and used as stimuli in Experiment 2. These misarticulations were produced by children with phonological disorders, who produced natural misarticulations in a speech probe. Because words were taken from existing recordings, the words from Experiment 1 could not be replicated. Therefore, each of the words in Experiment 2 was different from those used in Experiment 1 (see Table S4, Supplemental Material S1). Recordings were edited in Praat to 65-dB scale peak intensity and to include a 500-ms interstimulus interval, just as in Experiment 1.

Visual Stimuli

Visual stimuli were selected through the same methods as in Experiment 1. Where the same pictures could be used, they were. For words not included in Experiment 1, real-object pictures were selected from the same databases in Experiment 1. Nonobjects were the same as those in Experiment 1. Because there were no uncommon substitutes in this experiment, only 12 nonobject pictures were required. These were selected and assigned in the same manner as in Experiment 1.

Experimental Design

The experimental design was counterbalanced in the same way as Experiment 1, except that there were half as many versions because there was no uncommon substitute for which to control.

Procedures

Experimental procedures and technology were the same as those used in Experiment 1.

Data Analysis

The dependent variables were the same as in Experiment 1, but the independent variables were reduced to only accurate productions versus common misarticulations. The differences between accurate and common misarticulations were analyzed using paired samples t tests for each dependent variable.

Results

The proportion of real-object selections was analyzed using a paired samples t test (accurate vs. common substitution) and is displayed in Figure 4. The results revealed that children chose real objects significantly more when hearing accurate productions (M = 93%, SD = 7%) than common substitutions (M = 87%, SD = 8%), $t(13) = -2.75$, p = .02.

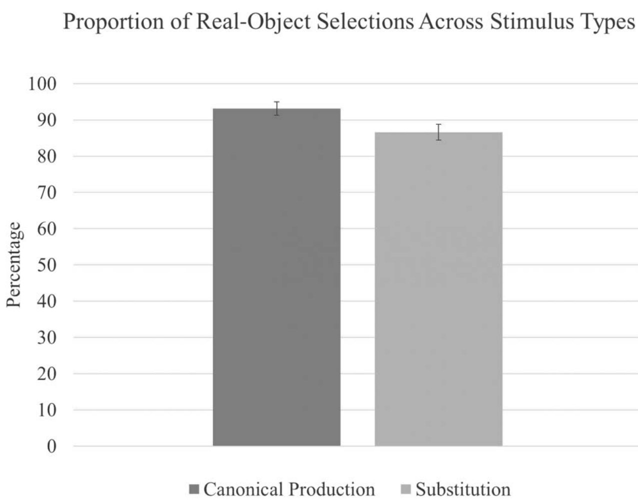
Reaction time on real-object selections was shown in the left panel of Figure 5 and was analyzed using a paired samples t test (accurate vs. common substitution). The results of this comparison showed that children chose real objects significantly faster when hearing accurate productions (M = 4437, SD = 1207) than when hearing common misarticulations (M = 5137, SD = 1237), $t(13)$ = 2.17, p = .05.

In terms of mouse trajectory (see the right panel of Figure 5), children's responses reflected the same pattern as the accuracy and reaction time data. Children's mouse movements were more direct when hearing accurate productions (M = 0.50, SD = 0.34) than common substitutions (M = 0.97, SD = 0.98), $t(13)$ = 2.52, p = .03.

Experiment 2: Discussion

The results of this study mirrored those of Experiment 1 and were consistent with findings from Creel (2012). Preschoolers identified words containing speech sound substitutions as real objects a large proportion of the time but experienced slower processing in the presence of

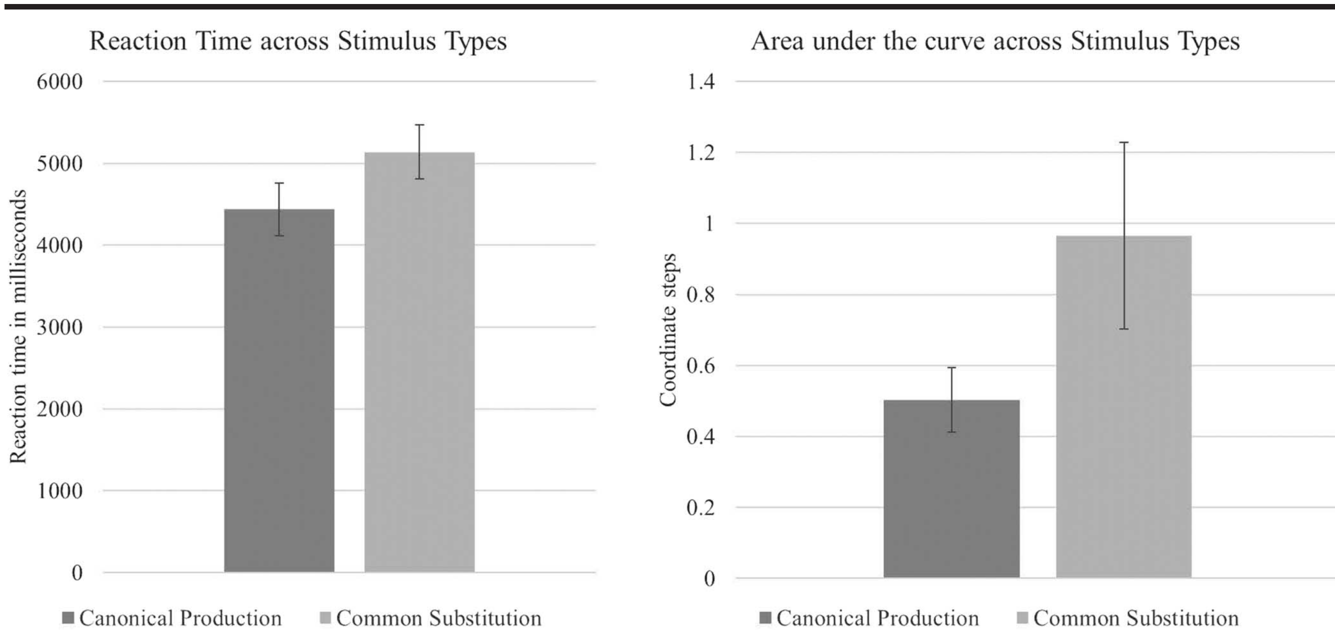
Figure 4. Mean proportion of children's real-object selections when hearing accurate productions and common and uncommon substitutions.



misarticulations. These results suggest that misarticulations influence children's identification and processing of words, not spectral or sociolinguistic factors associated with the incongruence of hearing misarticulations found in adult speech. Using naturally misarticulated speech did not provide additional subphonemic cues that facilitated processing. Therefore, if covert contrasts in the common condition were present, there was no apparent facilitation in terms of real-object selection, because the difference between the accurate and common conditions remained the same as the difference found in Experiment 1. Thus, regardless of whether words were produced by a child or an adult, children selected more real objects, more quickly, and with more direct mouse trajectories for accurate productions than for misarticulated productions.

Previous research using MouseTracker has shown that, in adults, this technology can mirror the results of eye-tracking technology by measuring continuous streams of hand movement (Freeman et al., 2011; Magnuson, 2005). However, this study was one of the first to use hand tracking with children in a psycholinguistic paradigm. Because the validity of hand tracking in children has not been empirically tested and previous research, such as Creel (2012), used eye tracking, it was important to replicate this study using eye tracking to test our results against an established measurement tool. One particular area of concern was the large amount of variability in the mouse trajectory for the AUC measurements in each experiment. These large standard deviations suggest that the mouse trajectory component of these results may not be as sensitive as eye gaze. Although differences between common and uncommon misarticulation conditions were not detected using this measurement, it is possible that differences do exist, and these may be revealed by measuring variability more precisely through the use of children's eye gaze. Because Experiments 1 and 2 showed a delay in reaction time for all substitutes, the analysis of children's gaze over time may reveal what occurs during this delay and show if there were any behavioral differences during this delay that would account for the differences within substitution types found in accuracy data.

Figure 5. Processing measures for Experiment 2. (Left) Mean reaction time by condition with standard error bars for mouse-click responses. (Right) Mean area under the curve for mouse trajectories by stimulus condition.



Experiment 3

The purpose of Experiment 3 was to replicate Experiment 1 as closely as possible to draw a comparison between the two technological methods and to examine children's online processing of misarticulated words. Experiment 3 addressed the following questions:

1. Do children look to real objects more than novel objects in the presence of accurate productions of words than in the presence of misarticulated productions of words?
2. When words are misarticulated, do children look to real objects more in the presence of common substitutions than in the presence of uncommon substitutions?

Method

Participants

Participants were 24 monolingual English-speaking preschoolers (age: M = 5;0 years;months, SD = 6 months, range = 4;0–5;9 years;months) recruited through word-of-mouth and flyers distributed at local preschools. Sixteen were boys, and eight were girls. Twenty children were White, one child was African American, one child was Native American, one child was Asian or Pacific Islander, and one child was of undisclosed heritage. Of these, nine were excluded from analysis: seven because of calibration failure, one because of having a previous diagnosis, and one because of scoring below the 16th percentile on the GFTA-2. Parents rated their children's exposure to misarticulations as 54 on average (SD = 24, range = 15–100). Participants all passed a pure-tone threshold hearing screening at 20 dB HL for 1000, 2000, and 4000 Hz, binaurally. Participants were typical in terms of receptive vocabulary as measured by the PPVT-4, with standard scores falling within the expected range of 85–115 (M = 119, SD = 12, range = 97–142), and typical in terms of articulation as measured by the GFTA-2, with all standard scores falling within the range of 85–115 (M = 105, SD = 9, range = 89–117). Parental consent and child assent were obtained in the same manners as in Experiments 1 and 2.

Materials

Auditory Stimuli

Stimuli were the same as used in Experiment 1. The adult productions were selected for Experiment 3 because they were recorded in an anechoic chamber and the words were more carefully controlled in terms of frequency, feature difference, and duration (see Table 1 and Table S1, Supplemental Material S1). Furthermore, the differences between common and uncommon substitutions could be explored. Because no significant difference was identified between Experiment 2 and Experiment 1, the adult-recorded stimuli were deemed acceptable for further use in Experiment 3. In Experiments 1 and 2, 250 ms of silence were embedded at the beginning and end of the sound files. To ensure that a full visual scan of the display was allowed, 500 ms of silence were inserted before the presentation of each word in accordance with the methods of Creel (2012). As in Experiments 1 and 2, preschoolers heard stimulus words over external speakers.

Visual Stimuli

Visual stimuli were the same as in Experiment 1 in that the two stimulus pictures were located at each top corner of the display; however, in place of a "start" button, children's gaze initiated the trial by fixating on a centralized cross. Typical visual world paradigms employ the use of distractor pictures and arrays of four or more objects (Allopenna, Magnuson, & Tanenhaus, 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). These include objects related to the stimulus as well as filler items. This provides competitors that subjects must ignore to arrive at the targeted object. However, because replication was the key motivator for conducting Experiment 3, the simple array of one nonobject and one real object was used for this study. Therefore, our areas of measurement for eye gaze were limited to the upper right- and left-hand corners of the screen.

Procedures

Children were tested using an Eyelink 1000 (SR Research) arm-mounted LCD eye tracker in a dedicated laboratory. The experiment was designed using Experiment Builder Software (SR Research) to mirror the design of Experiment 1. After speech-language evaluations and hearing screening were conducted, children were seated in front of a computer, and a "target" sticker was placed on their forehead to allow for location of the eye by the eye tracker. Then, children underwent a calibration procedure that varied from 5 to 7 min before beginning the experimental trials. The calibration procedure was a manual 9-point grid calibration and validation on a 1024 × 768 resolution 17-in. LCD display. Children were required to fixate on each of these 9 points as they appeared, one at a time, at regular locations across the monitor screen. Fixations were required to occur twice at each location (once for identification and once for validation) before the experimental protocol was initiated. Once the experiment began, children were required to fixate on a cross in the center of the screen for 1000 ms to initiate each trial. This cross was the same visual cross that was used for calibration and served as a drift correction point before each trial. Visual stimuli were presented for 500 ms to allow for a visual scan before the presentation of the auditory stimuli.

As in Experiments 1 and 2, children were provided with five practice trials to assimilate them with the experimental procedure. Unlike the previous experiments, responses were recorded using a button box. Children were asked to press the left-hand button to select the left-hand picture on the display and to press the right-hand button to select the right-hand picture on the display. A button box was used instead of a mouse for this experiment because the purpose of Experiment 3 was to determine if there was a difference in children's processing within substitution types. Because processing measures were taken through eye gaze, we did not want to risk data loss by using a mouse, which may cause children to look down at their hand to respond. The button box reduced the likelihood that children would look to their hands, because they rested each hand on a button and no movement was required. Because the purpose of Experiment 3 was to examine children's eye-gaze behavior in response to misarticulations, the results of the button box data are included in Supplemental Material S1. Furthermore, methodological considerations between MouseTracker and eye tracking with a button box are discussed in Supplemental Material S1.

All other procedures (evaluations, feedback, counterbalancing) are the same as in Experiment 1.

Data Analysis

Time course eye-gaze fixations were analyzed on all trials to examine children's eye gazes as the word unfolds; this measure was in contrast to the AUC measures reported in Experiments 1 and 2. The average duration of the initial phoneme, the word offset, and the average response time were identified and coded into the experiment as variables, which allowed for an examination of these windows of time. In accordance with typical conventions in visual world paradigm studies, the time windows were shifted 200 ms to account for the initiation of eye movements (Hallett, 1986). Data were extracted and processed using Eyelink (SR Research) Data Viewer software (version 2.5.0.14) package. Once data were processed, the results were binned in 20-ms intervals to prepare for analysis.

Results

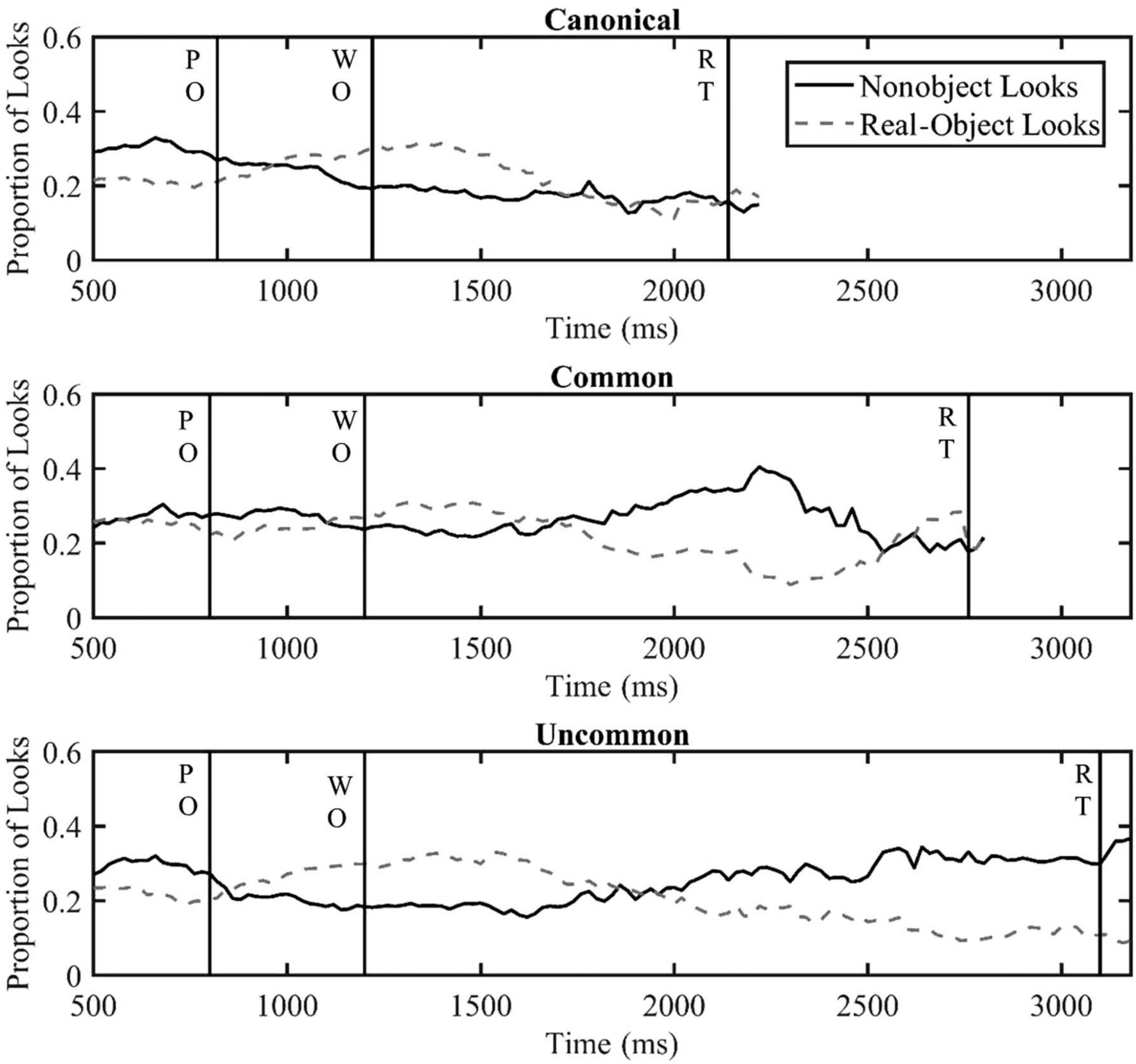
Results of real-object selections and reaction time are located in the Supplemental Material S1. Here, we focused exclusively on analysis of eye gaze to address the research questions.

Children's proportion of time course fixations was analyzed by plotting the average of the mean proportion of fixations to target and novel objects in each condition (Figure 6). For each condition, the average offset of the initial phoneme, the average word offset, and the average response time, respectively, are demarcated at time points along the *x*-axis, identifying the time at which these events occurred on average (in milliseconds), and shifted 200 ms to allow for the initiation of the eye movement, as described above. These time windows were selected because they may reveal at what point during the trials children identify words as being accurate or misarticulated and therefore may reveal more about children's processing of these words by allowing for the identification of how long it took for children to make a statistically meaningful decision about the words they heard. A 3 (Word Type: Accurate, Common, Uncommon) × 2 (Picture Type Looks: Real Object vs. Nonobject) repeated-measures ANOVA was conducted for each of the three time windows.

In the initial phoneme window, a significant main effect of Picture Type Looks was found, $F(1, 6) = 200.46$, $p < .001$, $\eta_p^2 = .97$, where participants across conditions looked to real objects significantly less (M = 21%, SD = 12%) than nonobjects (M = 29%, SD = 16%). No significant effect of Word Type was found, $F(2, 12) = 0.71$, $p = .43$, $\eta_p^2 = .11$, and the interaction between Word Type and Picture Type Looks was nonsignificant, $F(2, 12) = 1.12$, $p = .33$, $\eta_p^2 = .16$. For Picture Type Looks, a planned, post hoc, paired samples *t* test revealed that, within the time window of the initial phoneme, children fixated on nonobjects significantly more than real objects in each condition: accurate, $t(6) = -14.16$, $p < .001$; common, $t(6) = -25.24$, $p < .001$; and uncommon, $t(6) = -7.93$, $p < .001$. This finding was likely due to the novelty of the nonobjects, thus yielding a longer initial fixation on these pictures than real objects.

In the time window (central window, Figure 6) spanning the offset of the initial phoneme to the word offset (encompassing the rime), a significant interaction between Word Type and Picture Type Looks was found, $F(2, 38) = 17.32$, $p < .001$, $\eta_p^2 = .48$. Furthermore, the main effects of Word Type, $F(2, 38) = 84.50$, $p < .001$, $\eta_p^2 = .82$, and Picture Type Looks, $F(1, 19) = 5.34$, $p = .03$, $\eta_p^2 = .22$, were statistically significant. Planned, post hoc, paired samples *t* tests were conducted to evaluate the comparisons that may be causing the interaction between Word Type and Picture Type Looks. As seen in the top panel of Figure 6, in accurate conditions, children looked to real objects significantly more in this time window (M = 26%, SD = 3%) than nonobjects (M = 24%, SD = 2.6%), $t(19) = 2.31$, $p = .03$. In contrast, in the common misarticulation conditions (center panel of Figure 6), the proportion of looks between real objects (M = 25%, SD = 2%) was not significantly different from the proportion of looks to nonobjects (M = 24%, SD = 1.9%), $t(19) = 0.76$, $p = .46$. Finally, in uncommon conditions (bottom panel of Figure 6), children looked at real objects significantly more (M = 28%, SD = 3%) than nonobjects (M = 24%, SD = 3%), $t(19) = 2.97$, $p = .01$. The results of these comparisons showed that the interaction was driven by differences in the behavior observed in the accurate and uncommon conditions compared with the common misarticulation condition. In terms of their selections, recall that children chose real objects more than novel objects across conditions but appear to do so in this time window for the accurate condition but remain undecided for the common condition (no effect). In terms of the uncommon condition (Figure 6, bottom), looking behavior was similar to what was observed in the accurate production condition but converse to the selection results. It is possible that, because the uncommon reaction time was much longer than that of the accurate condition—from phoneme offset to word offset—children were, at this point, considering the real object as a possibility for the uncommon misarticulation. In fact, they ultimately made this selection in most trials; they did, however,

Figure 6. Time course fixations for each condition: accurate substitutions (top), common substitutions (center), and uncommon substitutions (bottom). Initial phoneme offset (PO), word offset (WO), and average response time (RT) are indicated by the vertical bars and labels along the *x*-axis of each chart, respectively.



take longer to do so, as suggested by the response time and eye fixation results.

Within the time window spanning the average word offset through the average response time (final time window, Figure 6), a significant interaction between Word Type and Picture Type Looks was observed, $F(2, 88) = 142.80$, $p < .001$, $\eta_p^2 = .76$. In addition, significant main effects of Word Type, $F(2, 88) = 16.50$, $p < .001$, $\eta_p^2 = .27$, and Picture Type Looks, $F(1, 44) = 22.73$, $p < .001$, $\eta_p^2 = .34$, were found. Planned, post hoc *t* tests were conducted to further examine the influential factors associated with the significant interaction between Word Type and Picture Type Looks. In accurate conditions (Figure 6, top), children looked to real objects significantly more (*M* = 21%, *SD* = 7%) than nonobjects (*M* = 18%, *SD* = 3%), $t(45) = 4.74$, $p < .001$. In addition, children selected the real object on most trials in this condition (*M* = 78%, *SD* = 20%). In common misarticulation conditions (Figure 6, center), children looked to real objects (*M* = 20%, *SD* = 7%) about the same as novel objects (*M* = 22% *SD* = 3%), $t(45) = -1.86$, $p = .07$, but they tended to select the real object on most trials (*M* = 67%, *SD* = 24%). In uncommon misarticulation conditions (Figure 6, bottom), children looked to real objects significantly more (*M* = 24%, *SD* = 7%) than nonobjects (*M* = 13%, *SD* = 4.3%), $t(45) = 17.78$, $p < .001$. However, toward the end of the time window, through a visual inspection of the bottom panel, children's eye fixation appears to return to the nonobject. Interestingly, children selected the real object (*M* = 58%, *SD* = 23%) only slightly more frequently than the nonobject. Therefore, this "jump" in fixation to the nonobject may be the result of children "checking back" toward the nonobject before making their selection on longer trials and a result of the high level of variability in this condition. The results of these comparisons revealed that the interaction was the result of significant differences in accurate and uncommon conditions compared with common misarticulation conditions.

During the initial time window, children appeared to be examining the visual display in a similar manner regardless of the auditory stimuli. That is, children focused on the more salient novel object rather than the familiar real object. For the remaining two time windows, children began to focus differentially on the visual display based on the auditory stimuli heard. For the accurate and uncommon misarticulation conditions, children focused more on the real object than the novel object. For the common misarticulation condition, children looked back and forth between both objects.

Overall, the proportion of looks found in these results was quite low as compared with previous eye-tracking studies. Because we intended to replicate our findings from Experiments 1 and 2, we only used two pictures, rather than the traditional array of four. Therefore, this limits our measurements to only two areas of interest in the upper right and left corners of the screen. If children looked elsewhere on the screen or off-screen, their gazes were not included in the statistical analysis for Experiment 3. However, we calculated the overall proportion of looks to other areas of the screen (“on-screen elsewhere”) and proportion of looks to the hands or blinking (“off-screen”) to determine whether children’s lowered accuracy and low proportion of looks to the areas of interest were potentially related. This calculation was conducted because it explains the nature of children’s eye-gaze behavior and provides insight into behavioral factors that contributed to the findings of Experiment 3. In the accurate condition, across the entire time window, children looked “on-screen elsewhere” 63% of the time (*SD* = 9%, range = 45%–79%) and to the hands or blinking (off-screen) 19% of the time (*SD* = 15%, range = 2%–53%). In the common misarticulation condition, across the entire time window, children looked “on-screen elsewhere” 60% of the time (*SD* = 7%, range = 51%–75%) and off-screen 19% of the time (*SD* = 12%, range = 0%–40%). In the uncommon misarticulation condition, across the entire time window, children looked “on-screen elsewhere” 58% of the time (*SD* = 10%, range = 43%–73%) and off-screen 23% of the time (*SD* = 15%, range = 4%–56%). These findings suggest that children spent approximately 20% of the time looking to their hands, and the range of off-screen looks was quite high. This heightened level of variability between subjects was consistent with the accuracy and reaction time findings for Experiment 3 (Supplemental Material S1).

Experiment 3: Discussion

The results of Experiment 3 confirmed the results of Experiments 1 and 2: Preschoolers identified and processed words with misarticulations differently than those that were accurately produced. When hearing accurate productions, children selected the real object on most trials and tended to look more at the real object during presentation of the rime and during the poststimulus window. This suggests that they quickly and efficiently identified the referent and were confident in their choice. In contrast, when hearing the common misarticulation, children selected real objects on most trials but looked equally at the real object and the nonobject, suggesting that they were uncertain of their interpretation of the word. Finally, in the uncommon misarticulation condition, children selected the real object almost as frequently as they selected the nonobject. Interestingly, they tended to look at the real object more frequently than the nonobject, suggesting that, regardless of which object they selected (real object vs. nonobject: 58% vs. 42% of responses, respectively), their selection was less certain and they were still considering whether the real object was a viable alternative to the nonobject, perhaps due to the rime and final consonant matching to that of the real object referent.

An additional purpose of Experiment 3 was to examine the similarities in processing measures between MouseTracker and eye tracking. Our results suggest that MouseTracker is an appropriate tool for use with preschool children and more easily implemented because of the low cost, portability of the equipment, and lack of calibration required. If reaction time and accuracy data are the only desired measurements, then MouseTracker is a tool that is appropriate for obtaining those. However, if more detailed information about moment-to-moment looking behavior is required, eye tracker provides a greater level of detail for that purpose.

Discussion

Overall, these results mirror the input that these children receive in terms of frequency. Children hear accurate productions of words most frequently in their day-to-day lives. Consequently, when children hear an accurate production, whether by an adult or a child, they quickly select the real-object referent. Although children hear misarticulated words less frequently than accurate productions, they do hear common misarticulations. Thus, when they hear a common misarticulation, whether spoken by a child or an adult, they are willing to accept it as a label for a real object, but they are slower and less certain of this pairing. Finally, children rarely encounter uncommon substitutes. Therefore, when they hear an uncommon misarticulation, produced by either an adult or a child, they are less willing to accept it as a label for a real object and are thus slower and more uncertain of their selection.

These results coincide with the findings of Creel (2012). In Creel, children identified words with a close feature distance as real objects significantly more than words that were far. Furthermore, Creel observed slowed response times in the presence of any phonemic substitution regardless of distance—a finding that is synonymous with the results of this study. Unlike Creel, feature distance in this study was controlled between common and uncommon substitution conditions, and a difference was found in terms of accuracy and in eye-gaze behavior. This finding suggests that, although children used feature distance in previous studies (Creel, 2012) to determine if a word is known or novel, they must additionally engage their own experience in making these judgments. This is especially true when considering that responses to common misarticulations were different than those to uncommon misarticulations, despite the control of feature distance.

Overall, children's reaction times were slowed in the presence of substitutions—a finding that is consistent with those found in previous studies of both child (Creel, 2012) and adult (Munro & Derwing, 1995) studies. The results of processing measures in this study support the idea that children are sensitive to phonological information in their native language, but they may be more flexible in their interpretation of the message. Although children are slowed, they still retrieve an appropriate real object to match with the variability and draw upon their knowledge of the likelihood of those phonemic differences. However, when paired with the real-object selection data, it appears that, despite the additional processing time needed, children's representations of words are flexible and can accommodate variability without destroying the message.

The findings of these studies provide evidence for children's word identification behavior in a carefully controlled environment. This evidence would likely change in the context of a real-world situation. In a preschool setting, for example, children are likely to have a meaningful semantic context that would influence their judgment of words containing misarticulations. Furthermore, this study did not provide any forms of feedback. In a social interaction, children receive feedback from their communication partners that would assist in confirming or denying the listener's understanding of the intended message. For example, if a child requests a toy using a speech sound substitution and receives the wrong toy in response, there is likely to be additional interaction. Whether the child requests the toy again, replies "no," or gets the targeted toy himself or herself, the listener receives negative feedback and may revise their interpretation of the message for the next time. The results of this study show that, in the absence of semantic context and in the absence of positive or negative feedback, children are largely able to understand speech sound variability and accept these phonemic substitutions as variants of real words.

Conclusion

The results of this study showed that children's experience with common misarticulations positively impacted their word identification; however, processing measures, such as reaction time, were slowed as a result. Overall, the results of this study coincided with previous research on children's perception of speech variability (Creel, 2012), suggesting that, at the age of 4–6 years, phonological representations of words in the mental lexicon are flexible enough to accommodate multiple phonological forms for a singular referent.

Acknowledgments

The first author was a doctoral trainee whose research was funded by a National Institute on Deafness and Other Communication Disorders training grant, #5 T32 DC000052-17. We would like to express our gratitude to the families who participated in this research, to the cooperating preschool sites, and to our research team.

This research was conducted at the University of Kansas, as part of the requirements for the master's degree of the first author.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439.

Altvater-Mackensen, N., & Mani, N. (2013). The impact of mispronunciations on toddler word recognition: Evidence for cascaded activation of semantically related words from mispronunciations of familiar words. *Infancy, 18*(6), 1030–1052.

American Speech-Language-Hearing Association. (1997). *Guidelines for audiologic screening*. Rockville, MD: Author.

Bent, T. (2014). Children's perception of foreign-accented words. *Journal of Child Language, 41*(6), 1334–1355.

Berteletti, I., Lucangeli, D., & Zorzi, M. (2012). Representation of numerical and non-numerical order in children. *Cognition, 124*(3), 304–313.

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words: Research report. *Psychological Science, 20*(5), 539–542.

Bleile, K. M. (2006). *The late eight*. San Diego, CA: Plural Publishing.

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer*. Available from http://www.praat.org

Cargill, S. A., Farmer, T. A., Schwade, J. A., Goldstein, M. H., & Spivey, M. J. (2004). Children's online processing of complex sentences: New evidence from a new technique. *Proceedings of the Cognitive Science Society, 29,* 143–148.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Cambridge, MA: MIT Press.

Clopper, C. G., & Pisoni, D. B. (2004). Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change, 16*(1), 31–48.

Creel, S. (2012). Phonological similarity and mutual exclusivity: On-line recognition of atypical pronunciations in 3–5-year-olds. *Developmental Science, 15*(5), 697–713.

Dinnsen, D. A., & Gierut, J. A. (Eds.). (2008). *Optimality theory, phonological acquisition and disorders*. Sheffield, United Kingdom: Equinox Publishing.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4)*. San Antonio, TX: Pearson Assessments.

Freeman, J., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*(1), 226–241.

Freeman, J., Dale, R., & Farmer, T. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology, 2,* 59.

Goldman, R., & Fristoe, M. (2000). *Goldman Fristoe Test of Articulation–Second Edition*. Circle Pines, MN: American Guidance Service.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.11–10.112). New York, NY: Wiley.

Han, M. K., Storkel, H. L., Lee, J., & Cox, C. (2016). The effects of phonotactic probability and neighborhood density on adults' word learning in noisy conditions. *American Journal of Speech-Language Pathology, 25*(4), 547–560.

Hourcade, J. P., Bederson, B. B., & Druin, A. (2004). Differences in pointing task performance between preschool children and adults using mice. *ACM Transactions on Computer-Human Interaction, 11*(4), 357–386.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research, 18*(4), 686–706.

Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior, 23*(1), 39–66.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358–368.

Liberman, A. M., Ingemann, F., Lisker, L., Delattre, P., & Cooper, F. S. (1959). Minimal rules for synthesizing speech. *The Journal of the Acoustical Society of America, 31,* 1490.

Magnuson, J. S. (2005). Moving hand reveals dynamics of thought. *Proceedings of the National Academy of Sciences of the United States of America, 102*(29), 9995–9996.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20*(2), 121–157.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences, 10*(8), 363–369.

McCormack, H. M., Horne, D. J., & Sheather, S. (1988). Clinical applications of visual analogue scales: A critical review. *Psychological Medicine, 18*(4), 1007–1019.

Merriman, W. E., Bowman, L. L., & MacWhinney, B. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development, 54,* 1–129.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289–306.

Quétard, B., Quinton, J. C., Mermillod, M., Barca, L., Pezzulo, G., Colomb, M., & Izaute, M. (2016). Differential effects of visual uncertainty and contextual guidance on perceptual decisions: Evidence from eye and mouse tracking in visual search. *Journal of Vision, 16*(11), 28.

Scobbie, J. E., Gibbon, F., Hardcastle, W. J., & Fletcher, P. (2000). Covert contrast as a stage in the acquisition of phonetics and phonology. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Language acquisition and the lexicon* (Vol. 5, pp. 194–207). Cambridge, United Kingdom: Cambridge University Press.

Shriberg, L. D., Gruber, F. A., & Kwiatkowski, J. (1994). Developmental phonological disorders. III: Long-term speech-sound normalization. *Journal of Speech and Hearing Research, 37*(5), 1151–1177.

Smit, A. B. (1993). Phonologic error distributions in the Iowa-Nebraska articulation norms project: Consonant singletons. *Journal of Speech and Hearing Research, 36*(3), 533–547.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*(4), 779–798.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174–215.

Storkel, H. L., & Adlof, S. M. (2009). The effect of semantic set size on word learning by preschool children. *Journal of Speech, Language, and Hearing Research, 52*(2), 306–320.

Storkel, H. L., & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods, 42*(2), 497–506.

Storkel, H. L., Maekawa, J., & Hoover, J. R. (2010). Differentiating the effects of phonotactic probability and neighborhood density on vocabulary comprehension and production: A comparison of preschool children with versus without phonological delays. *Journal of Speech, Language, and Hearing Research, 53*(4), 933–949.

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science, 8*(5), 432–443.

Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language, 60*(2), 252–269.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634.

Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *The Journal of the Acoustical Society of America, 54*(5), 1248–1266.

Wewers, M. E., & Lowe, N. K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health, 13*(4), 227–236.

Wuyts, F. L., De Bodt, M. S., & Van de Heyning, P. H. (1999). Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *Journal of Voice, 13*(4), 508–517.