Language and Speech

Article

Language *and* Speech
2018, Vol. 61(3) 466–479

DOI: 10.1177/0023830917737108
journals.sagepub.com/home/las
SAGE

Token Frequency Effects in Homophone Production: An Elicitation Study

Erin Conwell
North Dakota State University, USA

Abstract

In natural production, adults differentiate homophones prosodically as a function of the frequency of their intended meaning. This study compares adult and child productions of homophones to determine whether prosodic differentiation of homophones changes over development. Using a picture-based story-completion paradigm, isolated tokens of homophones were elicited from English-learning children and adult native English speakers. These tokens were measured for duration, vowel duration, pitch, pitch range, and vowel quality. Results indicate that less frequent meanings of homophones are longer in duration than their more frequent counterparts in both adults and children. No other measurement differed as a function of meaning frequency. As speakers of all ages produce longer tokens of lower frequency homophones, homophone differentiation does not change over development, but is included in children's early lexicons. These findings indicate that production planning processes alone may not fully account for differences in homophone duration, but rather that the differences could be learned and represented from experience even in the early stages of lexical acquisition.

Keywords

Frequency effects, homophones, lexical acquisition, speech production

Introduction

One challenge to children as they learn the words of their first language(s) is that the referential problem (i.e., what entity a given word refers to) is very ambiguous. In any natural word learning context, the number of potential referents for an unfamiliar word is large and yet children attach words to appropriate referents as young as six months of age (Bergelson & Swingley, 2015). To account for this rapid learning, theories of lexical acquisition tend to posit that children are equipped with a set of word learning constraints or principles that allow them to narrow the possible range of word meanings and reduce referential ambiguity. In nearly every account of word learning, one of these constraints is the assumption that each word will have a single referent and each referent will have a single word, an assumption referred to as one-form/one-function (Slobin, 1973), the

Corresponding author:
Erin Conwell, Psychology, North Dakota State University, PO Box 6050, Dept 2765, Fargo, ND 58108-6050, USA.
Email: econwell@gmail.com

taxonomic constraint (Markman, 1990), or mutual exclusivity (Markman & Wachtel, 1988), among other terms. This assumption works well for many of the words that children are likely to encounter, but it fails to address how children learn words that do not conform to this expectation, such as homophones.

Homophones, words that have two or more distinct meanings, should be particularly problematic for word learners who assume that every word has a single referent. When learners who make that assumption encounter a familiar word, what would motivate them to entertain the possibility that it might have a second meaning? Indeed, in lab tasks, children struggle to assign a novel meaning to a familiar word (Beveridge & Marsh, 1991; Doherty, 2004; Mazzocco, 1997; but see Storkel & Maekawa, 2005, for an important exception), although they can succeed with sufficient syntactic or semantic support for meaning differentiation (Backscheider, Gelman, Martinez & Kowieski, 1999; Casenhiser, 2005; Dautriche, Fibla & Christophe, 2015). It is not the case, however, that homophones are absent from children's early language experience or from their vocabularies. Parents use homophones when speaking to children (Nelson, 1995; Conwell & Morgan, 2012) and children's early vocabularies do contain homophone pairs (Backscheider & Gelman, 1995; Conwell & Morgan, 2012). In spite of violating the one-to-one mapping assumption, homophones do not appear to pose a particular problem to children in natural language learning settings. Because children succeed at learning homophones in natural settings but not in lab settings, natural learning environments must contain information that supports homophone learning, possibly by differentiating homophone meanings in a way that indicates to language learners that a second meaning should be considered for a known word.

A growing body of research suggests that one way in which homophones might be disambiguated is through consistent prosodic differences in their production as a function of the intended meaning. In studies of adult-directed speech, adults use duration as a disambiguating feature in their homophone production, with more frequent homophone meanings exhibiting shorter durations than their less frequent counterparts. These effects were first reported in reading tasks (Cohn, Brugman, Crawford & Joseph, 2005; Guion, 1995; Whalen, 1991), although some of those studies reported durational differences only in sentence-medial position and not in generic carrier phrases (Guion, 1995), while others found that contrastive use and sentence-final prosody to be necessary for eliciting durational differences in homophones (Cohn, et al., 2005). Reading tasks allow researchers to control for many factors that may affect production, such as sentence position, but the prosody of read speech is distinct from that of natural speech (Howell & Kadi-Hanifi, 1991). However, studies of natural speech corpora also suggest that homophones may be acoustically distinct as a function of meaning frequency. Jurafsky, Bell, and Girard (2002) examined four function words with multiple meanings in a large corpus of adult-directed English and found that only some of them showed consistent durational differences as a function of meaning frequency. However, in a corpus study that included all homophones in the corpus, Gahl (2008) reported longer durations for lower frequency homophones, even when speech rate, syntactic category, predictability, and orthography were controlled for. Beyond durational cues, adults use vowel neutralization and syllabic stress as a cue to homophone meaning in disyllabic homophones (Kelly, 1992; Kelly & Bock, 1988; Sereno & Jongman, 1995) and lexical frequency also affects vowel dispersion (Munson & Solomon, 2004).

As the prosody of child-directed speech is somewhat exaggerated relative to that of adult-directed speech (Bernstein Ratner, 1984; Ferguson, 1964; Fernald, 1989), further work has examined whether the same effects are seen in speech to children. In a corpus study of child-directed speech, Conwell (2017a) reported that adults differentiate homophone meanings as a function of frequency when speaking to children, but only in utterance-final positions, which is consistent with the findings from Cohn and colleagues (2005). Additionally, Conwell (2017a) found that higher frequency homophone meanings in child-directed speech had higher mean pitch than their lower frequency counterparts. Child-directed speech also contains exaggerated vowel spaces for verb tokens of noun/verb homophones (Conwell & Morgan, 2012; Conwell, 2017b), but does not appear to show vowel space differences related to meaning frequency for homophones more generally (Conwell, 2017a).

Taken together, the previous research on acoustic distinctions between homophone meanings suggests differentiation of homophone meanings as a function of the frequency of the intended meaning in both child-directed and adult-directed English. Such differences may support the disambiguation of homophone meaning by young word learners by allowing them to maintain separate phonetic representations for each homophone meaning, thus removing the need to violate the one-to-one mapping principle to learn homophones and providing motivation to consider that a familiar phonemic sequence might have a second meaning. In other words, if children actually have two distinct phonetic forms for *flower* and *flour*, then they may be able to associate one meaning (e.g., a blossom) with one phonetic form and another meaning (e.g., a grain product) with the other. Indeed, Gahl (2008) proposed that adults have separate phonetic forms for homophones, each associated with a distinct meaning. Those distinct forms must have been acquired at some point during development. Children may create distinct phonetic entries for homophones based on their experience and maintain them over their lifespan, which would not only facilitate acquisition of such words, but would also reduce the effects of lexical ambiguity on sentence processing through adulthood. Evidence from studies of noun/verb homophones suggests that infants can perceive the acoustic differences that exist in some homophone pairs (Conwell & Morgan, 2012) and adults show different neural responses to isolated noun tokens and verb tokens of noun/verb homophones in an event-related potential paradigm with electroencephalography (Conwell, 2015).

Although there is evidence that listeners can perceive some of the acoustic cues that distinguish homophone meanings, the issue of whether children incorporate those distinctions into their own productions of homophones remains unaddressed. Corpora of child speech for which audio recordings are available do not contain enough child-produced homophone tokens to address this question and preschool-aged children cannot participate in reading studies. Therefore, an elicitation paradigm is needed to obtain evidence on whether and how children distinguish homophones in their own speech. Elicitation also mitigates the effects of speaking rate, syntax, and referential context that might influence spontaneous productions.

The present study asks whether children produce the same distinctions in homophones that adults do. The analysis examines a range of acoustic features that might vary between productions of the same word form: duration, vowel duration, mean pitch, pitch range, and vowel quality, while many previous studies have examined only duration. Adults are expected to differentiate homophone meanings as a function of lemma frequency using duration, pitch, and vowel formant cues. If children fully incorporate those differences into their lexical representations of homophones, then they should show the same patterns of differentiation as adults. However, children's developing lexical and phonological representations may result in a pattern of homophone production that is not adult-like.

2 Method

2.1 Stimuli

The stimuli for this study comprised 22 brief stories (3–4 sentences each). Stories were presented one sentence at a time with accompanying illustrations to support production of the target word (Figure 1). Sixteen of the stories could be completed with a homophone from one of four

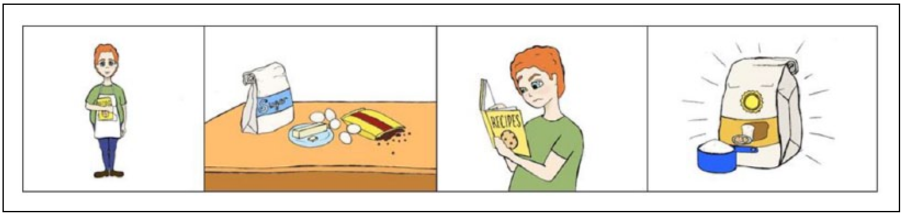

Figure 1. A sample of the pictures used for one story. "Nate plans to make cookies. He has sugar, butter, eggs and chocolate chips. He is forgetting one very important thing. There it is! There's the ________." (Target: flour).

homophone pairs (flower/flour; plane/plain; red/read; whole/hole). Each homophonous meaning could felicitously complete two of the stories. The illustrations were printed in full color on card-stock and the stories were produced by the experimenter. The other six stories were used as practice trials and filler trials. The appendix includes the full text of the stories.

For each homophone pair in this study, one meaning was more frequent in child-directed speech than the other. The relative frequency of homophone meanings was based on the total number of adult uses in the Providence Corpus, which consists of approximately 360 hours of speech between parents and children (Demuth, Culbertson & Alter, 2006). The more frequent targets were *flower, plane, red*, and *whole*, while *flour, plain, read*, and *hole* were less frequent than their counterparts. For *read*, a heteronym, the frequency of the target /ɹɛd/ (as opposed to the non-target /ɹid/) was determined by hand-counting occurrences of each pronunciation.

2.2 Participants

Child participants were 68 English-learning children ages 4 years, 11 months to 6 years, 11 months (mean age: 5 years, 9 months). One child was not included in the analysis due to documented speech delays. Eight children (five male; three female) did not produce both meanings of any of the homophone pairs; their data were also excluded from analysis. The final sample included 59 child participants (21 male; 38 female). Parental consent to participate was obtained for all child participants prior to the study. All children received a book upon completing the study to thank them for their participation.

Adult participants were 35 monolingual native speakers of American English, 12 male and 23 female. All adult participants who were included in the analysis had lived the majority of their lives in the north-central region of the United States and spoke the dialect of English associated with that region. Six additional participants completed the study, but were excluded from the analysis for having non-regional dialects that affected vowel categories. All adult participants received course credit for their participation.

2.3 Procedure

All participants were seated in front of a Blue Snowball USB microphone connected to a laptop computer. The experimenter sat across the table from the participant with the story illustrations. The stories were presented in random order across participants. The experimenter displayed one illustration at a time and read the corresponding sentence before setting the illustration aside. Participants were instructed to produce the word they thought best completed the story. Sentences were not repeated. Adult participants were not prompted to produce alternate words if they did not produce the target on their first attempt. When child participants did not produce the target word on the first try, they were prompted once to produce an alternative, using the phrase "Can you think of another way to say that?" If children did not produce the target word after one use of this prompt, the experimenter moved on to the next set of illustrations. At no point during the experimental session did the experimenter produce any of the target words. The session was recorded continuously using the Audacity program (Audacity, 2014).

2.4 Measurement

All productions of target words were extracted from the recordings and analyzed using PRAAT (Boersma & Weenink, 2014). Trained research assistants placed boundaries at the beginning and end of each word as well as at the beginning and end of the vowel. These boundaries were placed based on auditory characteristics as well as on visual examination of the spectrogram and waveform. A PRAAT script was used to extract token duration, vowel duration, pitch characteristics (mean, minimum and maximum), and the first and second formant frequencies at the midpoint of the vowel, as well as at 20% of the vowel and 80% of the vowel. Pitch range was computed in semitones using the minimum and maximum measurements. Static formant measures reported here were taken at the midpoint of the vowel.

Following Ferguson and Kewley-Port (2007), the formant measures from 20% and 80% of the vowel were converted to Barks and used to calculate two dynamic measures of vowel quality that are associated with clear speech: spectral change (λ) and spectral angle (Ω). Spectral change (λ) describes the amount of formant movement that occurs during the middle 60% of the vowel. It was calculated as $\lambda = |F1_{80} - F1_{20}| + |F2_{80} - F2_{20}|$.

Spectral angle (Ω) describes the rate of change of the vowel formants by characterizing the angle of the vowel spectra in radians. It is calculated by adding the absolute values of the angles of the first and second vowel formants, which were computed as $\theta_n = \arctan\left(\frac{Fn_{80} - Fn_{20}}{d}\right)$, where $d = \frac{time_{80} - time_{20}}{100}$.

These dynamic vowel measures provide some indication of how clearly articulated the vowels are, as larger λ and larger Ω are both associated with speech that was intended by the speaker to be very clear (Ferguson & Kewley-Port, 2002; Ferguson & Kewley-Port, 2007). Larger values for both measures additionally predict vowel intelligibility (Hillenbrand & Nearey, 1999). Therefore, the prediction is that, if speakers exaggerate lower frequency meanings of a homophone with the intent of producing clearer exemplars of those words, the lower frequency meanings should have larger λ and Ω.

In addition to measures of the tokens themselves, two variables were calculated and included in the model to control for variables known to affect the duration of words. As speaking rate has been reported to influence duration of words (Bell, et al., 2003; Cohn, et al., 2005), the rate at which the experimenter produced the story prior to the target word was calculated for every trial by measuring from the onset of the experimenter's speech to the offset of the experimenter's speech and then dividing that time by the number of syllables in that trial. Predictability, specifically bigram probability, can also affect word duration (Gahl, 2008; Jurafsky, et al., 2002). Therefore, the probability of each target word given the preceding word in the child-directed speech of the Providence Corpus (Demuth, et al., 2006) was calculated as the total number of productions of the bigram in the corpus divided by the total number of uses of the preceding word. Because the probability of sequences larger than bigrams does not affect word duration once bigram effects are controlled for (Bell,

Table 1. Mean values of each dependent measure for each homophone pair by age.

Word (frequency)	Total tokens	Mean token duration (ms)	Mean vowel duration (ms)	Mean token pitch (Hz)	Mean pitch range (ST)	Mean F1 (Hz)	Mean F2 (Hz)	Mean λ	Mean Ω (radians)
Flower (higher)	155								
Adults	63	487	193	233.3	14.81	811.4	1307	1.75	3.04
Children	92	656	259	264.9	11.97	825.2	1416	3.74	3.07
Flour (lower)	158								
Adults	64	501	195	229	15.69	812.9	1295	2.06	3.04
Children	94	707	264	274	13.76	838.2	1397	3.8	3.07
Plane (higher)	110								
Adults	45	432	216	223.8	10.97	550.9	1615	3.34	3.08
Children	65	573	263	266.2	12.93	561.6	1494	3.82	3.09
Plain (lower)	118								
Adults	61	441	217	221.5	15.61	501.2	1726	3.85	3.1
Children	57	600	266	264	14.09	538.9	1397	3.87	3.07
Red (higher)	147								
Adults	70	331	186	201.1	9.84	675.6	1635	2.55	3.09
Children	77	502	269	261	9.38	791.1	1543	4.29	3.07
Read (lower)	118								
Adults	60	347	185	209.4	10.86	656.5	1640	2.2	3.0
Children	58	482	285	245.6	8.44	772.3	1537	4.63	3.1
Whole (higher)	93								
Adults	60	392	192	216.2	14.81	532.8	865.1	1.28	3.01
Children	33	533	241	269.7	9.81	582.7	982.5	1.81	3.01
Hole (lower)	104								
Adults	58	387	196	210.4	13.45	519.6	913.6	1.36	3.03
Children	46	538	244	273	11.49	572.6	993.2	1.88	3.01

et al., 2003), these bigram probabilities alone were used as a measure of predictability for each trial.

3 Results

Children's production of the 16 target words was highly variable (mean: 11.3; range: 4-16). Adult production was more consistent (mean: 13.8; range: 11-16). To permit comparisons within a homophone pair, a participant's production of a target word was included only if the participant produced at least one token of each meaning of the homophone pair. The numbers of productions of each target word are presented by age group in Table 1.

The data were analyzed using linear mixed models (lme4; Bates, Maechler, Bolker & Walker, 2015) implemented in R (R Core Team, 2015). Frequency of the target word relative to its homophonic counterpart (higher vs. lower) and age group (child vs. adult) were included as fixed effects with an interaction. Speaking rate of the experimenter and bigram probability were also included as fixed effects. Participant and homophone pair were included as random effects. Statistical significance was tested by comparing models with and without the relevant effect or interaction using maximum likelihood estimates. As all measures had a non-normal distribution, they were

Table 2. Results of the linear mixed model analyses for the two durational measures. Statistical significance was determined using likelihood ratio tests. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

Effect	Parameter	Token duration		Vowel duration	
		Estimate	*SE*	Estimate	*SE*
Fixed Effects					
Intercept	β	0.35	0.037	0.698	0.03
Frequency of meaning	β	0.012**	0.006	0.006	0.008
Age group	β	0.137***	0.017	0.13***	0.022
Speaking rate	β	0.116**	0.036	0.056	0.417
Predictability	β	0.249	1.426	0.78	1.87
Frequency x age	β	0.005	0.008	0.001	0.012
Random effects					
Participant	σ^2	0.006***	0.076	0.009***	0.095
Homophone pair	σ^2	0.004***	0.061	0.0005***	0.024
Residual	σ^2	0.004	0.067	0.008	0.092

log-transformed before the analysis. The results of the analyses of the durational measures are presented in Table 2, the results of the analyses of the pitch measures are in Table 3 and the results from the vowel formant measures are shown in Table 4.

Token duration in homophone production showed a significant effect of the relative frequency of the target meaning, $\chi^2(2) = 12.37$, $p = 0.002$, as well as a significant effect of the age of the participant, $\chi^2(2) = 52.33$, $p < 0.001$. Lower frequency meanings had longer durations than higher frequency meanings and children produced target words with longer durations than adult participants did. However, the interaction of age and frequency of meaning was not statistically significant for token duration, $\chi^2(1) = 0.381$, $p = 0.54$, indicating that adults and children do not differ in the extent of their exaggeration of token duration for lower frequency homophone meanings. An additional fixed effect of speaking rate on token duration was also found, $\chi^2(1) = 10.22$, $p = 0.001$, but there was not an effect of predictability, $\chi^2(1) = 0.03$, $p = 0.86$. This model also showed significant random effects of participant, $\chi2(1) = 527.1$, $p < 0.001$ and homophone pair, $\chi2(1) = 452.9$, $p < 0.001$. Such effects likely reflect individual differences in typical pace of production and the different numbers of phonemes in each homophone pair. The token duration results are shown in Figure 2.

Vowel duration did not show a significant effect of meaning frequency $\chi^2(2) = 1.09$, $p = 0.58$, but did show a significant effect of age, $\chi^2(2) = 25.02$, $p < 0.001$. The interaction of age and frequency on vowel duration was not significant, $\chi^2(1) = 0.006$, $p = 0.94$. No effects of speaking rate, $\chi^2(1) = 1.3$, $p = 0.25$ or predictability, $\chi^2(1) = 0.18$, $p = 0.67$ on vowel duration were found. These results indicate that, while children elongate vowels relative to adults, vowel duration is not affected by meaning frequency, speaking rate, or predictability in any age group. Participant and pair both showed significant effects on vowel duration, $\chi2(1) = 473$, $p < 0.001$; $\chi2(1) = 29.01$, $p < 0.001$, respectively, likely indicating individual differences in production speed and effects of different vowel types on vowel duration.

Mean pitch was not affected by the frequency of target meaning, $\chi^2(2) = 0.81$, $p = 0.67$, but, as expected, age did affect mean pitch, with children showing higher average mean pitch than adults, $\chi^2(2) = 28.78$, $p < 0.001$. The interaction of frequency and age was not significant for mean pitch, $\chi^2(1) = 0.056$, $p = 0.813$ and neither speaking rate, $\chi^2(1) = 2.53$, $p = 0.111$ nor predictability, $\chi^2(1) = 0.141$, $p = 0.71$ were significant predictors of mean pitch. Participant identity significantly affected

Table 3. Results of the linear mixed model analyses for the two pitch measures. Statistical significance was determined using likelihood ratio tests. (*p < 0.05, **p < 0.01, ***p < 0.001).

Effect	Parameter	Pitch		Pitch range	
		Estimate	*SE*	Estimate	*SE*
Fixed effects					
Intercept	β	2.35	0.029	0.976	0.074
Frequency of meaning	β	0.004	0.01	0.036	0.029
Age group	β	0.1***	0.019	0.013	0.035
Speaking rate	β	0.089	0.056	0.036	0.14
Predictability	β	0.687	1.78	1.77	3.35
Frequency x age	β	0.003	0.014	0.005	0.04
Random effects					
Participant	σ^2	0.006***	0.075	0.008***	0.092
Homophone pair	σ^2	0.0001*	0.013	0.006*	0.075
Residual	σ^2	0.012	0.111	0.099	0.315

mean pitch, $\chi^2(1) = 214.1$, $p < 0.001$, reflecting individual differences in fundamental frequency. Somewhat unexpectedly, however, homophone pair also had a significant effect on mean pitch, $\chi^2(1) = 4.86$, $p = 0.027$. Pitch range showed no effects of meaning frequency, $\chi^2(2) = 2.87$, $p = 0.24$ or age, $\chi^2(2) = 0.16$, $p = 0.92$ and no interaction of the frequency and age, $\chi^2(1) = 0.015$, $p = 0.9$. Speaking rate, $\chi^2(1) = 0.065$, $p = 0.8$ and predictability, $\chi^2(1) = 0.279$, $p = 0.6$ also did not affect pitch range. Participant identity did significantly affect pitch range, $\chi^2(1) = 22.73$, $p < 0.001$, again likely due to individual variation in prosody. Pitch range was also significantly affected by homophone pair, $\chi^2(1) = 42.29$, $p < 0.001$. The effects of homophone pair on both pitch and pitch range suggest natural variability in the prosody associated with specific phonological sequences. Critically for the question at hand, no effects of meaning frequency were found for any pitch measure.

The first and second vowel formants were not significantly affected by relative frequency, age or the interaction of the two (all $p > 0.36$). Predictability also did not significantly affect the first or second vowel formants, $\chi^2(1) = 2.8$, $p = 0.09$; $\chi^2(1) = 1.07$, $p = 0.3$, respectively. Speaking rate significantly affected the first formant, $\chi^2(1) = 10.01$, $p = 0.002$, but not the second formant, $\chi^2(1) = 0.342$, $p = 0.559$. The vowel formants were both significantly affected by participant, F1: $\chi^2(1) = 126.75$, $p < 0.001$; F2: $\chi^2(1) = 18.23$, $p < 0.001$ and by homophone pair, F1: $\chi^2(1) = 381.61$, $p < 0.001$; F2: $\chi^2(1) = 398.2$, $p < 0.001$, indicating expected differences in vowel production between participants and as the result of different vowel types.

In addition to static measure of vowel quality, two measures of formant change were analyzed. The first, spectral change (λ), was significantly affected by participant age, $\chi^2(2) = 42.73$, $p < 0.001$, but not by meaning frequency, $\chi^2(2) = 1.92$, $p = 0.37$, speaking rate, $\chi^2(1) = 0.17$, $p = 0.68$, or predictability, $\chi^2(1) = 9$, $p = 0.34$. The interaction of meaning frequency and age was also not significant for spectral change, $\chi^2(1) = 0.047$, $p = 0.83$. Spectral change was significantly affected by participant, $\chi^2(1) = 16.99$, $p < 0.001$ and homophone pair, $\chi^2(1) = 159.4$, $p < 0.001$. The other dynamic measure, spectral angle (Ω), was not significantly affected by meaning frequency, age, speaking rate, predictability, or participant (all $p > 0.14$). The interaction of frequency and age was also not significant for spectral angle, $\chi^2(1) = 0.66$, $p = 0.42$. Homophone pair did significantly affect spectral angle, $\chi^2(1) = 4.49$, $p = 0.034$, likely due to the different vowels in each homophone pair.

Overall, the results show consistent use of token duration to distinguish homophone meanings as a function of meaning frequency across both age groups, suggesting that children

Table 4. Results of the linear mixed model analyses for the vowel formant measures. Statistical significance was determined using likelihood ratio tests. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Effect	Parameter	F1		F2		Λ		Ω	
		Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
Fixed effects									
Intercept	β	2.87	0.048	3.125	0.049	0.227	0.104	0.49	0.005
Frequency of meaning	β	0.014	0.01	0.009	0.01	0.034	0.03	0.002	0.002
Age group	β	0.013	0.017	0.004	0.012	.215***	0.035	0.001	0.004
Speaking rate	β	0.18**	0.023	0.029	0.049	0.06	0.144	0.011	0.01
Predictability	β	2.79	1.65	1.22	1.18	3.21	3.38	0.337	0.227
Frequency x age	β	0.012	0.014	0.012	0.014	0.009	0.042	0.003	0.003
Random effects									
Participant	σ^2	0.004***	0.066	0.001***	0.033	0.007***	0.087	<0.001	0.002
Homophone pair	σ^2	0.006***	0.08	0.008***	0.088	0.026***	0.162	<0.001*	0.002
Residual	σ^2	0.012	0.109	0.012	0.109	0.108	0.329	<0.001	0.027

incorporate the durational differences from their experience with homophones into their lexical representations of these words. Vowel duration, pitch measures, and vowel quality are not significantly affected by the frequency of the target meaning. All age-related differences may arise from vocal tract size.

4 Discussion

The data presented in this article provide evidence for acoustic differentiation of homophone meanings by both adults and children. Taken together, the results show that speakers of all ages lengthen homophonous words when producing the lower frequency meaning. These findings are consistent with an account of homophone acquisition in which children perceive the acoustic distinctions between homophone meanings in their input and incorporate those distinctions into their lexical representations to guide their learning of homophonous words.

Acoustic differences between homophone meanings may support children's early acquisition of homophones in natural contexts and account for their difficulties in learning pseudohomophones in laboratory tasks, where tokens are pre-recorded and not differentiated in this way (e.g., Doherty, 2004; Mazzocco, 1997). In the real world, children have access to information, including acoustic information, that allows them to disambiguate homophone meanings. When children encounter a lower frequency homophone such as *flour* in their natural experience, they may notice that the word is notably longer than it has been on previous encounters, when the intended meaning was really its more frequent homophonic counterpart, *flower*. Children may also notice an absence of a blossom from the context and assume that a different meaning is intended, although this cue may be less reliable and harder to detect. Children do hear words in the absence of their referents with some frequency and referential context is often ambiguous (Nichol Medina, Snedeker, Trueswell & Gleitman, 2011), but the absence of the familiar referent may, along with other semantic cues, support the assignment of a new meaning to a familiar phonological sequence under the right circumstances (Backsheider, et al., 1999; Dautriche, Chemla & Christophe, 2016). However, there is some evidence that very young children struggle to use context and plausibility to determine the intended meaning when a word that they already know is used in an unfamiliar way (Srinivasan &

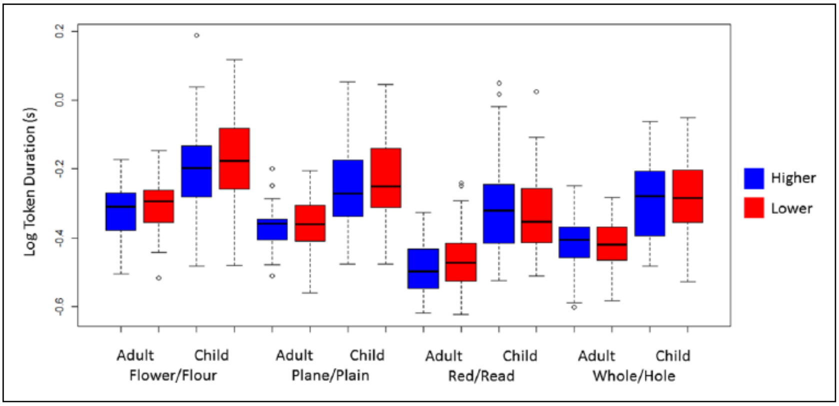

Figure 2. Data from the two durational measures (token duration and vowel duration) for each homophone pair by relative frequency of the homophone meaning and age of the participant.

Barner, 2013). The availability of acoustic information may support that assignment by providing children with bottom-up evidence that the two words are distinct.

The data from the present study are consistent with an account of homophone representation in which acoustic distinctions are encoded as part of the lexical representation and directly associated with different meanings (e.g., Gahl, 2008). Indeed, they further bolster the claim that acoustic differences in homophones are not merely byproducts of sentence-level prosodic processes, but rather part of the word itself. Appeals to production processes, such as sentence prosody and articulatory planning, as the cause of durational differences in homophone meanings (e.g., Lavoie, 2002) do not explain why such differences are found in isolated tokens in an elicitation task. If these differences were entirely the result of phrasal or sentential prosody, isolated tokens would not show them. One interpretation of these data is that these differences are encoded in the lexicon from childhood on the basis of auditory experience. Whether this arises as a result of an exemplar-based lexicon or whether some more abstract lexical representation might contain these differences as part of the phonological specification is beyond the scope of these data. Another explanation of the similarity of adult and child patterns of lengthening in homophones as a function of lemma frequency could be that articulatory planning processes that interact with lemma frequency are consistent across development and may not be a direct result of representational differences.

An alternative account of these findings is that the lower frequency targets were exaggerated because they were being used contrastively in the study. That is, participants who had already produced the more frequent *whole* may have stressed their productions of *hole* later in the task to indicate a contrast in meaning (Chafe, 1976; Cohn, et al., 2005), creating longer durations. In this case, one would expect that participants would exaggerate second uses of words relative to first uses, regardless of meaning frequency. Although stories were presented in random order, higher frequency meanings could have been presented first more often than lower frequency meanings were, resulting in exaggeration of lower frequency targets. However, examination of session records shows that lower frequency homophone meanings were presented after their higher frequency counterparts only 47.9% of the time for adult participants and 46.2% of the time for child participants, indicating that lower frequency meanings were not elongated to contrast them with their higher frequency meaning.

One important feature of these data is that the only measure affected by meaning frequency was token duration. Pitch and vowel measures (including vowel duration) were not affected by the frequency of the target meaning. These results are consistent with the findings from spontaneous child-directed speech (Conwell, 2017a). This shows that speakers are not just exaggerating all aspects of their productions of less frequent items, but rather that the two meanings differ in their phonetic realization along the single dimension of overall duration. The lack of an effect of meaning frequency on vowel articulation or on pitch shows that these effects are not the result of pitch accent, increased lexical stress, or clearer articulation of lower frequency meanings. The lack of an effect of frequency of meaning on vowel duration, in spite of the effect on token duration, is a bit surprising: what is lengthened if not the vowel? Based on these data, the durations of all segments in the less frequent meaning are likely slightly lengthened, leading to a cumulative difference that is not carried by any single segment. The isolation of the effect of frequency to a single measure further supports the claim that this effect is not the product of articulation or production processes alone, but rather is part of the lexicon.

Token duration was additionally affected by the speaking rate of the experimenter, which is potentially important because speaking rate may be affected by lexical frequency (Lavoie, 2002). One explanation of this finding is that the experimenter may have unconsciously altered speaking rate based on the homophone meaning that they knew to be the target. Although the analyses presented here statistically control for the effect of speaking rate on token duration, it is possible that the participants' productions of the target words were influenced by the speaking rate of the experimenter. Some caution may be warranted in interpreting these results as reflecting truly isolated tokens of the target words.

One potential limitation of this study is that the analyses were conducted over a small number of homophone types. The small number of types is due to the short attention spans and performance limitations of children in the age range. Many children failed to produce target words even though their caregivers indicated that the words were highly familiar to the children. Nevertheless, the small number of types raises questions regarding both the generalizability and the robustness of these effects. It is possible that these specific homophone pairs show an effect of lemma frequency on token duration, but that others would not. Prior literature does show a robust effect of lemma frequency on token duration in both adult-directed and child-directed speech (Conwell, 2017a; Gahl, 2008), however, suggesting that the pattern seen in the adult data in this study is likely to generalize to other homophone pairs. The findings from child participants in this study indicate that children do not differ from adults in terms of how lemma frequency affects their production of homophones. While conservativism is wise when considering a finding based on a small number of word types, this finding converges with other evidence on adult productions and provides the first real evidence that children's productions of homophones might also be influenced by lemma frequency.

This article asked whether children's representations of homophones include the acoustic distinctions that differentiate homophone meanings in their experience by examining whether children exhibit those distinctions in their own productions of homophonous words. Children, like adults, elongate the duration of lower frequency homophone tokens. These findings support an account of homophone acquisition in which children incorporate acoustic differences from their experience with homophone meanings into their representations of those words and could, therefore, use those differences as a bottom-up cue to the existence of two distinct lexical items with the same phonological form.

Acknowledgements

I thank Felix Pichardo, Brenden Melvie, Elisabeth Dukowitz, Samantha Hamernick, and Adrienne MacDonald for assistance with data collection and analysis and Katherine Wong for creating the drawings that accompanied the stories. Two anonymous reviewers made comments that significantly improved the manuscript.

Funding

This research was supported by the National Institute of Child Health and Human Development, part of the National Institutes of Health (grant number 1R15HD077519-01). The contents of this article are the sole responsibility of the author and do not necessarily represent the official views of either organisation.

References

Audacity (2014). Audacity [Computer program]. Version 2.0.6. Retrieved from http://audacityteam.org/

Backscheider, A. G., & Gelman, S. A. (1995). Children's understanding of homonyms. *Journal of Child Language*, *22*, 107–127.

Backscheider, A. G., Gelman, S. A., Martinez, I., & Kowieski, J. (1999). Children's use of different information types when learning homophones and nonce words. *Cognitive Development*, *14*, 515–530.

Bates, D., Maechler, M., Bolker, B.M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. ArXiv e-print. Retrieved from http://arxiv.org/abs/1406.5823

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of dysfluencies, predictability and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, *113*, 1001–1024.

Bergelson, E., & Swingley, D. (2015). Early word comprehension in infants: Replication and extension. *Language Learning and Development*, *11*, 369–380.

Bernstein Ratner, N. (1984). Patterns of vowel modification in mother-child speech. *Journal of Child Language*, *11*, 557–578.

Beveridge, M., & Marsh, L. (1991). The influence of linguistic context on young children's understanding of homophonic words. *Journal of Child Language*, *18*, 459–467.

Boersma, P., & Weenink, D. (2014). Praat: Doing Phonetics by Computer [Computer Program]. Version 5.3.67. Retrieved from http://www.praat.org

Casenhiser, D. (2005). Children's resistance to homonymy: An experimental study of pseudohomonyms. *Journal of Child Language*, *32*, 319–343.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics and points of view. In C. Li (Ed.), *Subject and Topic*. New York: Academic Press.

Cohn, A. C., Brugman, J., Crawford, C., & Joseph, A. (2005). Lexical frequency effects and the phonetic duration of English homophones: An acoustic study. *Journal of the Acoustical Society of America*, *118*, 2036.

Conwell, E. (2015). Neural responses to category ambiguous words. *Neuropsychologia*, *69*, 85–92.

Conwell, E. (2017a). Are homophones acoustically distinguished in child-directed speech? *Language Learning and Development*, *13*, 262–273.

Conwell, E. (2017b). Prosodic disambiguation of noun/verb homophones in child-directed speech. *Journal of Child Language*, *44*, 734–751.

Conwell, E., & Morgan, J. L. (2012). Is it a noun or is it a verb? Resolving the ambicategoricality problem. *Language Learning & Development*, *8*, 87–112.

Dautriche, I., Chemla, E., & Christophe, A. (2016). Word learning: Homophony and the distribution of learning exemplars. *Language Learning and Development*. Retrieved from http://dx.doi.org/10.1080/15475441.2015.1127163

Dautriche, I., Fibla, L., & Christophe, A. (2015, November). *Learning homophones: Syntactic and semantic contexts matter*. Paper presented at the 40th Annual Boston University Conference on Language Development, Boston, MA.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the acquisition of English. *Language and Speech*, *49*, 137–174.

Doherty, M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language*, *31*, 203–214.

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist*, *66*, 103–114.

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *112*, 29–271.

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language and Hearing Research*, *50*, 1241–1255.
Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Development*, *60*, 1497–1510.
Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*, 474–496.
Guion, S. G. (1995). Word frequency effects among homonyms. *Texas Linguistic Forum*, *35*, 103–116.
Hillenbrand, J., & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *The Journal of the Acoustical Society of America*, *105*, 3509–3523.
Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*, *10*, 163–169.
Jurafsky, D., Bell, A., & Girand, C. (2002). The role of the lemma in form variation. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp 1–34). Berlin: Mouten de Gruyter.
Kelly, M. H. (1988). Phonological biases in grammatical category shifts. *Journal of Memory and Language*, *27*, 343–358.
Kelly, M. H., & Bock, J. K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 389–403.
Lavoie, L. (2002). Some influences on the realization of *for* and *four* in American English. *Journal of the International Phonetic Association*, *32*, 175–202.
Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*, 57–77.
Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–157.
Mazzocco, M. M. M. (1997). Children's interpretations of homonyms: A developmental study. *Journal of Child Language*, *24*, 441–467.
Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech Hearing and Language Research*, *47*, 1048–1058.
Nelson, K. (1995). The dual category problem in the acquisition of action words. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: young children's acquisition of verbs*. New Jersey: Lawrence Erlbaum.
Nichol Medina, T., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, *108*, 9014–9019.
R Core Team (2015). R: A language and environment for statistical computing [Computer program]. *R Foundation for Statistical Computing, Vienna, Austria*. Retrieved from http://www.R-project.org/
Sereno, J. A., & Jongman, A. (1995). Acoustic correlates of grammatical class. *Language and Speech*, *38*, 57–76.
Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies in child language development*. New York: Holt, Reinhart, Winston.
Srinivasan, M. & Barner, D. (2013). The Amelia Bedelia effect: World knowledge and the goal bias in language acquisition. *Cognition*, *128*, 431–450.
Storkel, H. L., & Maekawa, J. (2005). A comparison of homonym and novel word learning: The role of phonotactic probability and word frequency. *Journal of Child Language*, *32*, 827–853.
Whalen, D. H. (1991). Infrequent words are longer in duration than frequent words. *Journal of the Acoustical Society of America*, *90*, 2311.

Appendix

Laura is going to take a trip. Her car is broken, so she can't drive. She has a great idea. Instead of driving, she could fly on a __________________ (target: plane).
Tim is trying to choose a bagel. He wanted blueberry, but they are all out. He doesn't like the other flavors. He decides to get one that is ______________ (target: plain).

Brendan has a lot of different vehicles. He loves his bike. He also has a boat. He really wishes that someday he could buy a _______________ (target: plane).
Meredith does not like her dress. One of her friends has a dress with flowers. Her other friend's dress has ruffles. Her dress has nothing on it; it's very ___________________ (target: plain).
Nate plans to make cookies. He has sugar, butter, eggs and chocolate chips. He is forgetting one very important thing. There it is! There's the ___________________ (target: flour).
Ruth loves to work in her garden. She gets very messy, but she grows beautiful things, like daffodils and tulips. The rose is her very favorite _______________________ (target: flower).
Anne has made a big mess. She was going to bake a cake. She wasn't careful when she was reaching into the cupboard. She knocked over the ____________ (target: flour).
James went for a walk with his dog yesterday. They walked past houses and trees. When they got to the park, the dog ran around and James picked a ____________________ (target: flower).
Bella's pet is a snake named Horace. Horace can't chew his food. When he has lunch, he has to swallow it ___________________ (target: whole).
Dan can't figure out why the water keeps coming out of his cup. When he drinks, the water dribbles on his chin. He looks closely and finds a big ___________________ (target: hole).
Kai is going to have a slice of pizza. His mom asks if she should cut it up. He says no thank you. Kai wants the slice to stay ________________ (target: whole).
Zoe loves to play outside. Yesterday while she was running, she tripped and fell on her knee. Now her pants have a big _____________ (target: hole).
Max really loves stories. He goes to the library all the time to get new books. Today, though, he wants a story he knows, so he picks out a book that he has already _________________ (target: read).
Jean loves all of the colors, but one of them is her very favorite. She decided to get a new shirt in that color. Her new shirt is _______________ (target: red).
Maddie thinks that story time is the best part of school. She likes to look at the pictures while the teacher reads. She thinks "The Cat in the Hat" is the best book they have ever ______________ (target: read).
Clay's room was very boring. All of the walls were white. He asked his mom if he could change that. He wanted to paint the walls _______________ (target: red).
Filler and warm-up items
Tasha likes to help her dad cook. She is careful with the knife and knows that the stove is hot. Her favorite part of cooking is peeking at the soup in the _____________ (target: pot).
Patrick is learning how to grow his own plants. He started by putting the seeds in little trays. Now that the plants are bigger, he is going to move them to a __________________ (target: pot).
Ian was not nice to his sister. He took one of her toys without asking and called her name. She was so mad that she started to _________ (target: yell).
When Betsy gets excited, she can be very loud. She saw a rainbow on the way to school. She told her dad about it in a very excited way. He said, "I like it, too, Betsy, but please don't ___________" (target: yell).
Margot wanted to make supper. She chopped up some vegetables. She put them in the pan. The pan got hot and the vegetables started to _________________ (target: cook).
Drew loves watching animals in the woods. Yesterday, he went for a walk there. He saw birds and a deer. When he got to the pond, he also saw a _____________ (target: frog).