Are Homophones Acoustically Distinguished in Child-Directed Speech?

Erin Conwell

Department of Psychology, North Dakota State University

ABSTRACT

Many approaches to early word learning posit that children assume a one-to-one mapping of form and meaning. However, children’s early vocabularies contain homophones, words that violate that assumption. Children might learn such words by exploiting prosodic differences between homophone meanings that are associated with lemma frequency (Gahl, 2008). Such differences have not yet been documented in children’s natural language experience and the exaggerated prosody of child-directed speech could either mask the subtle distinctions reported in adult-directed speech or enhance them. This study measured the duration, vowel characteristics, and pitch information of homophone tokens taken from a corpus of child-directed speech. The results show that homophone meanings are acoustically distinct in child-directed speech as a function of lemma frequency, particularly in utterance-final positions. Such distinctions may allow children to maintain separate phonetic representations of homophones until their cognitive and linguistic abilities are robust to violations of the one-to-one bias.

Introduction

Learning the meanings of words is a critical step in learning a language. As infants and young children acquire their first language, they must deduce from their experience how speech sounds match up with meanings. Because neither words nor their referents typically appear in perfect isolation, this process is rife with ambiguity, and yet most children successfully map at least some sounds to meanings within the first 6–9 months of life and have productive vocabularies of dozens of words by their second birthdays (Bergelson & Swingley, 2012; Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Brown, 1973). This indicates that even infants are skilled and efficient word learners, suggesting that young children have certain expectations about how word-referent pairings will work.

These expectations have been formalized in a number of theoretical accounts, which differ in their details, but have similarities in their broader points. For example, most accounts of lexical acquisition propose that children might constrain their hypotheses about the meaning of a new word by assuming that it will not refer to anything that already has a label. The flip side of that same assumption is that a familiar word will only denote its previously established referent. Under various approaches to word learning, such assumptions are called one form/one function (Slobin, 1973), the taxonomic constraint (Markman, 1990), or the mutual exclusivity hypothesis (Markman & Wachtel, 1988). The details of this word learning constraint vary depending on the particular instantiation, but they all center on the same general idea. In short, each word gets one referent and each referent gets one word.
Empirical support for the existence of such a constraint is legion. Children’s use of mutual exclusivity for fast-mapping of form to meaning has been reported for nearly 40 years (e.g., Carey, 1978; Markman, Wasow, & Hansen, 2003; Merriman, Bowman, & MacWhinney, 1989; inter alia) and is used as the basis of many paradigms in word learning research. A one-to-one relationship between linguistic forms and their meanings goes a long way toward constraining children’s hypotheses regarding possible meanings and accounts for the acceleration of word learning over early childhood. The more words a child knows, the better able she is to make informed guesses about the intended referent of a new word. From a learning perspective, there’s only one problem with assuming that each word has one meaning and each meaning has one word: it isn’t true.

Homophony violates the assumption of a one-to-one relationship between words and their referents because a homophone is a single phonological form that maps to two (or sometimes more) unrelated referents. Accounting for homophones has not been a primary concern of much literature on lexical development in spite of the fact that homophony (and its counterpart, polysemy, in which a word has multiple related meanings) is widespread in the world’s languages. If children constrain the possible meanings of a word by assuming that each word has only one meaning, homophones should be particularly problematic for learners. One would expect that children would learn only one meaning for a homophone until they are sufficiently linguistically or cognitively advanced to allow violations of their word learning constraints.

Studies of children’s acquisition of homophones or pseudohomophones (familiar words used with a novel meaning, e.g., egg used to refer to a jackhammer) in laboratory settings generally support the idea that children under the age of five adhere to a one-to-one mapping assumption and fail to learn homophones. In particular, when the contrastive meaning of a word is available, children struggle to interpret even familiar homophones and appear not to entertain the possibility that a familiar word might have a second meaning, but do not typically have trouble learning the meaning of a completely novel word (Beveridge & March, 1991; Doherty, 2004; Mazzocco, 1997; see Casenhiser, 2005; Storkel & Maekawa, 2005 for important exceptions). Such findings indicate that learning a new phonological string with a new meaning may be easier than attaching a new meaning to an existing phonological form. These studies constitute strong support for the use of a one-to-one mapping assumption as a guiding principle for word learning. In controlled settings, children struggle to learn homophones.

At some point, however, children do learn homophones. Their environments contain homophones (Conwell, 2014; Nelson, 1995), children produce noun/verb homophones (e.g., kiss) in both categories by their third birthdays (Clark & Clark, 1979; Conwell & Morgan, 2012; Nelson, 1995), and, of course, adults are fluent with homophony. Given their apparent inability to learn homophones in the lab, how do children gain adult-like abilities with ambiguous words?

One possibility is that children are equipped with cognitive structures to support multiple meanings at a very young age. Studies of polysemy find cross-linguistic regularities in polysemes that point to likely cognitive biases in how people entertain polysemous meanings (Srinivasan & Rabagliati, 2015). Children’s ability to interpret multiple related meanings appears to develop in tandem with certain conceptual abilities. For example, as children become better able to recruit world knowledge and context during language processing, they become able to interpret different classes of denominal verbs (Srinivasan & Barner, 2013). Lippeveld and Oshima-Takane (2014) found evidence that children’s experience with polysemous words contributes to their ability to interpret novel instances of such words. In their study, two-and-a-half year olds whose mothers produced both noun and verb tokens of familiar noun/verb polysemes during a brief interaction were able to interpret denominal verb uses of a novel noun, but those children whose mothers did not produce cross-category uses of noun/verb polysemes showed chance performance at the task. However, these studies consider polysemy, in which a word’s multiple meanings are semantically related in regular and cross-linguistically consistent ways (Srinivasan & Rabagliati, 2015). They do not speak to how children learn multiple unrelated meanings for homophonous words. Other research suggests that children need to acquire certain meta-cognitive abilities before they can begin to entertain the
possibility that a word might have two meanings (Garnham, Brooks, Garnham, & Ostenfeld, 2000). However, such claims do not explain how very young children come to use both meanings of homophones in their spontaneous productions.

Although conceptual and semantic development likely play significant roles in children’s ability to assign more than one meaning to a single word form, very early homophone learning may be supported by a different set of cues. In particular, growing evidence indicates that homophones may differ in their pronunciation depending on their intended meaning. On the surface, this doesn’t make much sense, as homophones are, by definition, words that have the same phonological form but distinct meanings. However, the distinctness of the meanings may result in different pronunciation over time. Nygaard, Patel, and Queen (2002) reported that speakers produce homophones with emotional valence appropriate to the intended meaning, leading to differences in duration and pitch between such words as bridal and bridle. Jurafsky, Bell, and Girand (2002) found that some function words with multiple meanings differ in duration in spontaneous speech depending on the intended meaning. Gahl (2008) further reported that, over a large corpus of spoken English, the more frequent member of a homophone pair is shorter in duration than the less frequent meaning, even when sentence position and category of use are controlled for. Such findings provide evidence that adults who are speaking to other adults pronounce homophones slightly differently depending on the intended meaning, which could facilitate processing of potentially ambiguous words and sentences.

The availability of durational and pitch differences in homophones in adult-directed speech raises the possibility that such differences might also be available to child language learners. If they are, then children might be able to use these differences to side-step the potential learning problems posed by homophones. Homophones should be difficult to learn because they do not adhere to a one-to-one relationship between form and meaning; they are a single word form that has two meanings. If, however, those two meanings are actually attached not to one word form, but to two word forms that are segmentally the same, but exhibit subtle differences in duration or other acoustic cues, homophones would no longer violate the one-to-one mapping assumption. Children might be able to associate one meaning with the version of the word that has a shorter duration and the other meaning with the longer duration. For that to be feasible, however, the durational distinctions between homophones that are present in adult-directed speech must also be available in speech to children.

Speech to young children differs from speech between adults in a number of ways, in particular in terms of its prosody. Child-directed speech is prosodically exaggerated, with a slower speaking rate, elongated vowel durations, and longer pauses than adult-directed speech (Bernstein Ratner, 1984, 1986; Fernald et al., 1989; Fisher & Tokura, 1996; Ko, 2012). Pitch contours are larger in child-directed speech (Ferguson, 1964; Fernald, 1989) and the vowel space is also enlarged in speech to children (Bernstein Ratner, 1984; Cristia & Seidl, 2014; Kuhl et al., 1997). Because child-directed speech is prosodically and phonetically distinct from adult-directed speech, the differences reported between homophone meanings in adult-directed speech may be masked by the exaggeration of prosodic features. The durational differences between homophones in adult-directed speech are small, but consistent, which means that the elongation and slowed speech rate that characterize child-directed speech have the potential to introduce variability and overall lengthening that could easily swamp those durational differences. Alternatively, the differences may be exaggerated in child-directed speech. Because prosodic contours and speaking rate are exaggerated in child-directed speech, individual words might also be subject to that exaggeration, thus enhancing prosodic differences between homophone meanings. Even if the differences between homophone meanings are not exaggerated in child-directed speech, they might be similar to those in adult-directed speech. Either of these alternatives would mean that young children are exposed to low-level acoustic information that could help them distinguish the multiple uses of homophones, provided that they are perceptually sensitive to that information.

Some evidence suggests that child-directed speech does contain at least some differences in the pronunciation of homophones, but examination of those differences has, thus far, been largely based on
the syntactic category of use. In a study of Canadian French speakers, Shi and Moisan (2008) found that mothers reading to their children will alter their pronunciation of disyllabic novel words depending on the category of use. Specifically, mothers elongated and increased the pitch of the second vowel of noun uses but not of verb uses. In a similar study of Mandarin Chinese speakers, mothers altered the relative duration of vowels in disyllabic nonsense words as a function of category of use (Li, Shi, & Hua, 2010). Conwell and Morgan (2012) reported differences in duration and vowel formants in a small number of noun and verb tokens of noun/verb homophones in the natural speech of one English-speaking mother to her child. Furthermore, they found that infants demonstrate perceptual sensitivity to those differences. In a larger-scale corpus analysis, Conwell (in press) reported that noun tokens of noun/verb homophones in child-directed speech are consistently longer than verb tokens in sentence-medial, but not sentence-final contexts. All of these studies, however, consider only the differences between cross-category uses of nouns and verbs. This is an important limitation in the previous research, as these studies do not differentiate between polysems and homophones, nor do they consider the role of meaning frequency in the production of these words. No research to date has examined the nature of homophone pronunciation in child-directed speech in general and the role of lemma frequency has yet to be considered as a factor in homophone production in speech to children.

Characterizing the presence of perceptual cues to homophone meaning in child-directed speech is important to understanding whether those cues might facilitate homophone learning in children, especially in the early stages of lexical acquisition. To resolve the issue of whether durational and other prosodic differences between homophone senses are present in child-directed speech and whether those differences are, as in adult-directed speech, associated with the relative frequency of the meaning (Gahl, 2008), the study presented here examines the prosodic characteristics of homophones in a large corpus of child-directed speech. If child-directed speech contains the same patterns of pronunciation that have been reported for homophones in adult-directed speech, such information could be used by children to distinguish homophone meanings in acquisition. If, however, the exaggerated prosody of child-directed speech masks the durational differences that have been reported in homophones in adult-directed speech, such perceptual differences are unlikely to support homophone learning in children.

Method

Corpus

This study uses the child-directed speech from the six mothers in the Providence Corpus (Demuth, Culbertson, & Alter, 2006). This corpus consists of hour-long naturalistic recordings of mothers interacting with their children in their homes. Recordings began when the children uttered their first words (11–16 months) and were taken every other week for 2–3 years. Two of the mothers had rhotic dialects characteristic of southern New England, while the other four spoke non-rhotic dialects. This corpus includes approximately 364 h of transcribed interaction and was selected because audio recordings are available for all transcribed sessions. The ages of the children and the total number of recordings per child are presented in Table 1.

Procedure

A list of 33 homophone pairs was compiled from various sources, including type frequency lists from the CHILDES database (MacWhinney, 2000) and literacy materials for elementary school students. The maternal speech in the Providence Corpus was searched for all of these homophones. To reduce effects of interpersonal variation in speaking rate and pronunciation, each mother’s use of a particular homophone pair was included only if she produced both meanings at least five times. Of the original 33 pairs, 25 appeared with both meanings at least 5 times in the speech of at least one mother. Research assistants verified that pairs consisted of the same phonetic segments by listening to the mother’s pronunciation of
A complete list of the target words that were extracted is presented in Appendix A. Using the kwal function in CLAN (MacWhinney, 2000), each use of the target words was identified in the maternal speech and categorized by hand as a noun, verb, adjective, adverb, or other use (e.g., *new* in *New York*). Isolated tokens were not included in the analysis. To coarsely assess the role of sentence prosody on production of homophones, utterance position was also hand-coded as either utterance-medial or utterance-final. Utterance-initial tokens were fairly small in number (301 total) and therefore included in the utterance-medial category, as proximity to an utterance boundary was the feature of interest. A total of 2906 tokens were included in the analysis. Of these, 627 were nouns, 1187 were verbs, 449 were adjectives, 103 were adverbs, and 540 were other uses. A total of 2318 medial and 588 final tokens were represented.

To assess how frequency of use affects the acoustic properties of homophones, each homophone meaning was classified as either higher or lower frequency. Relative frequency was based on the total number of uses in the Providence Corpus (Demuth et al., 2006). For homophones that are also heteronyms (e.g., *read*), uses were hand-coded for pronunciation prior to analysis and only those uses that matched the target pronunciation were included in the frequency counts. No homophone pair had equal numbers of uses for both meanings.

Each token was extracted from the audio recordings associated with each transcript using the Audacity program. After extraction was completed, trained research assistants used the PRAAT program (Boersma & Weenink, 2014) to place boundaries at the beginnings and ends of each token as well as the beginning and end of the vowel. For disyllabic words, only the beginning and end of the stressed vowel were marked with boundaries. Boundaries were placed using a combination of visual and auditory examination of the waveform and spectrogram. Word types with rhotic syllables were excluded from the vowel duration analysis, bringing the number of tokens in that analysis to 2195. To assess the reliability of the vowel duration measure, 10% of tokens were re-marked by a second researcher. Reliability was high ($r = .949$). On the basis of these boundaries, a PRAAT script extracted the token and vowel duration (in ms) and mean, minimum, and maximum pitch (in Hz). Vowel duration was included to determine whether any lengthening was primarily due to enhanced production of the vowel. The minimum and maximum pitch were used to compute the pitch range in semitones, which, along with mean pitch, were included as coarse measures of sentence prosody and lexical stress.

The data were analyzed using linear mixed effects models (*lme4*; Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2015). For each measure, a model was constructed with relative frequency, utterance position, vowel type and category of use as fixed effects. Speaker (mother) and homophone pair were included as random effects with by-speaker and by-pair random slopes for the effect of frequency. Statistical significance of each effect was determined by a maximum likelihood test comparing the full model to a model with that effect omitted. Statistical significance for each estimate was calculated using the Satterthwaite approximation for degrees of freedom as implemented in *lmerTest* for R (Kuznetsova, Brockhoff, & Christensen, 2015) to provide information regarding each level of the vowel type and category of use factors.

---

1The research assistants were mostly from the North Plains region of the U.S. and spoke the dialect associated with that region, so it is possible that their perception of the segments was not consistent with the dialect of the speakers. However, their categorization was verified by either the author or a member of the lab staff who is from the northeastern U.S.
Results

Token duration was significantly affected by both fixed and random factors. The data for the effects of frequency and utterance position on both durational measures are presented in Figure 1. Token duration showed a significant main effect of relative frequency of meaning ($\chi^2(2) = 14.91, p < .001$), a significant main effect of utterance position ($\chi^2(2) = 384.15, p < .001$) and a significant interaction of meaning frequency and utterance position ($\chi^2(1) = 11.47, p < .001$). Lower frequency meanings were longer than higher frequency meanings and utterance-final uses were longer than medial uses. The effect of frequency on token duration was statistically significant in utterance-final cases ($\mu_{\text{lower}} = 500\text{ ms}; \mu_{\text{higher}} = 438\text{ ms}; t(524) = 3.162; p = .002$), but not in utterance-medial position ($\mu_{\text{lower}} = 266\text{ ms}; \mu_{\text{higher}} = 258\text{ ms}; t(2273) = 1.29; p = .2$). Token duration was also significantly affected by the category in which a word was used ($\chi^2(4) = 24.98, p < .001$), vowel type ($\chi^2(8) = 30.93, p < .001$), speaker identity ($\chi^2(3) = 306.09, p < .001$), and homophone pair ($\chi^2(3) = 130.29, p < .001$). Complete results from the mixed model analysis of token duration are presented in Table 2.

Vowel duration showed a significant main effect of utterance position ($\chi^2(2) = 242.26, p < .001$), but did not show a significant main effect of frequency of meaning ($\chi^2(2) = 3.01, p = .22$), nor a significant interaction of position and frequency ($\chi^2(1) = 2.34, p = .13$). Vowel duration was also significantly affected by category of use ($\chi^2(4) = 18.84, p < .001$), speaker identity ($\chi^2(3) = 136.81, p < .001$), vowel type ($\chi^2(5) = 21.79, p < .001$), and homophone pair ($\chi^2(3) = 60.1, p < .001$). The results of the mixed model analysis of vowel duration are presented in Table 3.

The effects of relative frequency and sentence position on the two pitch measures are presented in Figure 2. Mean pitch showed no main effect of relative frequency of use, no main effect of utterance position and no interaction of utterance position and meaning frequency (all $p > .5$). Category of use also did not affect mean pitch ($\chi^2(2) = 1.154, p = .562$). As expected, vowel type had a significant effect on mean pitch ($\chi^2(8) = 16.99, p = .03$), as did homophone pair ($\chi^2(3) = 91.81, p < .001$) and speaker identity ($\chi^2(3) = 91.29, p < .001$). Pitch range was significantly affected by utterance position ($\chi^2(2) = 138, p < .001$), with utterance-final tokens showing greater pitch range than utterance medial tokens ($\mu_{\text{final}} = 11.32\text{ ST}; \mu_{\text{medial}} = 6.22\text{ ST}; t(784) = 14.003, p < .001$). However, pitch range showed no main effect of meaning frequency and no interaction (both $p > .2$). Category of use also significantly affected pitch range ($\chi^2(4) = 30.44, p < .001$), as did vowel type ($\chi^2(8) = 21.7, p = .005$), homophone pair ($\chi^2(3) = 12.5, p = .006$), and speaker identity ($\chi^2(3) = 48.37, p < .001$). Complete results of the analyses of both pitch measures are in Table 4.

Figure 1. Token durations (left) and vowel durations (right) in seconds of child-directed homophones by utterance position and relative frequency of meaning.
Table 2. Results of the linear mixed effect model for token duration. Noun was the reference category for category of use and /ɔ/ was the reference category for vowel type. Statistical significance of the fixed effects estimates was established using the Satterthwaite approximation, while statistical significance of the random effects was established by maximum likelihood estimation. (*p < .05; **p < .01; ***p < .001).

| Estimate | SE  | df    | t    |
|----------|-----|-------|------|
| **Fixed Effects** |     |       |      |
| Intercept | .041 | .005  | 76.2 | 8.21*** |
| Position: Middle | −.006 | .001  | 2820 | 12.14*** |
| Relative Frequency: Lower | .124 | .0018 | 58.2 | 3.65*** |
| Position x Relative Frequency | −.005 | .0014 | 2830 | 3.39*** |
| **Category** |     |       |      |
| Adjective | −.0038 | .0015 | 154  | 2.55*   |
| Adverb | −.0043 | .0023 | 214  | 1.88     |
| Other | −.0028 | .0016 | 69.3 | 1.69     |
| Verb | −.0094 | .0013 | 24   | 6.94*** |
| **Vowel Type** |     |       |      |
| au | .0258 | .0053 | 54.5 | 4.87*** |
| at | .0031 | .0046 | 67.4 | .67      |
| e | .0004 | .0048 | 61.2 | .44      |
| et | .0004 | .0059 | 72.6 | .07      |
| i | .0008 | .0047 | 66.7 | .17      |
| oo | .0062 | .0045 | 73.2 | 1.36      |
| U | .001 | .0047 | 65.8 | .21      |
| **Random Effects** |     |       |      |
| Homophone Pair | .0034*** | .0117 |       |         |
| Speaker | .00008*** | .0012 |       |         |
| Residual | .0195 | .0026 |       |         |

Table 3. Results of the linear mixed effect model for vowel duration. Noun was the reference category for category of use and /aɪ/ was the reference category for vowel type. Statistical significance of the fixed effects estimates was established using the Satterthwaite approximation, while statistical significance of the random effects was established by maximum likelihood estimation. (*p < .05; **p < .01; ***p < .001).

| Estimate | SE  | df  | t    |
|----------|-----|-----|------|
| **Fixed Effects** |     |     |      |
| Intercept | .337 | .0021 | 23   | 15.7*** |
| Position: Middle | −.105 | .001 | 1230 | 10.48*** |
| Relative Frequency: Lower | −.0027 | .0016 | 48.1 | 1.66 |
| Position x Relative Frequency | −.0022 | .0014 | 1630 | 1.55 |
| **Category** |     |     |      |
| Adjective | −.0036 | .0014 | 85.5 | 2.63* |
| Adverb | −.0048 | .002 | 124  | 2.42* |
| Other | −.0007 | .0016 | 64.8 | .418 |
| Verb | −.0059 | .0013 | 40.3 | 4.53*** |
| **Vowel Type** |     |     |      |
| e | −.0038 | .0021 | 10.6 | 1.79 |
| et | −.0029 | .0018 | 14.3 | 1.62 |
| i | −.0042 | .0019 | 16   | 2.17* |
| oo | −.0006 | .0014 | 12.6 | .402 |
| u | −.0039 | .0015 | 12   | 2.53* |
| **Random Effects** |     |     |      |
| Homophone Pair | .0016*** | .0098 |       |         |
| Speaker | .000002*** | .0006 |       |         |
| Residual | .0139 | .0025 |       |         |

Overall, frequency of meaning affects the duration of homophones in child-directed speech, even when speaker identity, homophone pair, lexical category, vowel type, and utterance position are considered. No effects of frequency are found for vowel duration, mean pitch, or pitch range once other factors are controlled for.
Discussion

This study finds that the frequency effects on homophone pronunciation that have been reported in adult-directed speech (e.g., Gahl, 2008; Jurafsky et al., 2002) are also present in child-directed speech. These findings indicate that the prosody of child-directed speech does not overwhelm frequency-based durational differences in homophones. The presence of such distinctions might support homophone acquisition by children by allowing them to maintain distinct word forms for homophones, one for each meaning, at least until their lexical and conceptual development are sufficiently advanced to support multiple meanings for a single form.

Table 4. Results of the linear mixed effect model for mean pitch and pitch range. Noun was the reference category for category of use and /ɔ/ was the reference category for vowel type. Statistical significance of the fixed effects estimates was established using the Satterthwaite approximation, while statistical significance of the random effects was established by maximum likelihood estimation. (*p < .05; **p < .01; ***p < .001).

| Fixed Effects                  | Estimate | SE  | df  | t       | Estimate | SE  | df  | T       |
|--------------------------------|----------|-----|-----|---------|----------|-----|-----|---------|
| Intercept                      | 201.76   | 21.91 | 155.3 | 9.21*** | 11.17   | 1.63 | 466.5 | 6.93*** |
| Position: Middle               | −1.85    | 5.71  | 2079.2 | .324    | −3.81   | .459 | 1692.2 | 8.29*** |
| Relative Frequency: Lower      | −9.12    | 11.04 | 50.4  | .827    | 1.25    | .716 | 48.5  | 1.75    |
| Position x Relative Frequency  | −1.25    | 8.06  | 2120.9 | .155    | −.645   | .649 | 1967.8 | .993    |
| Category                       |          |      |      |         |          |      |      |         |
| Adjective                      | 2.97     | 8.24  | 180.5 | .361    | −.184   | .578 | 194.6 | .318    |
| Adverb                         | −13.2    | 12.77 | 243.1 | 1.03    | −.76    | .954 | 401   | .796    |
| Other                          | 8.03     | 8.1   | 41.6  | .991    | −1.95   | .535 | 303.8 | 3.64*** |
| Verb                           | −10.86   | 7.37  | 34.4  | 1.47    | −2.43   | .503 | 84.1  | 4.84*** |
| Vowel Type                     |          |      |      |         |          |      |      |         |
| aʊ                              | 20.25    | 22.37 | 103   | .905    | 2.91    | 1.68 | 1994.1 | 1.73    |
| aɪ                              | 56.87    | 20.22 | 140.9 | 2.81**  | −.399   | 1.55 | 2508.3 | .257    |
| ɛ                               | 32.88    | 20.05 | 160.4 | 1.64    | −.225   | 1.54 | 2191.1 | .146    |
| eɪ                              | 53.18    | 20.47 | 125.2 | 2.6*    | 1.03    | 1.56 | 2382.6 | .656    |
| i                              | 21.85    | 25.63 | 171.4 | .852    | −.972   | 1.97 | 2100.8 | .493    |
| u                              | 49.7     | 20.32 | 135.5 | 2.45*   | .051    | 1.55 | 2424.1 | .033    |
| oʊ                              | 41.51    | 19.87 | 166   | 2.09*   | .194    | 1.53 | 2223.6 | .127    |
| u                               | 56.48    | 20.43 | 133.7 | 2.76**  | −.823   | 1.55 | 2206.4 | .529    |

Random Effects

| Homophone Pair  | 1206.1*** | 6.95 | 2.516*** | .317       |
| Speaker         | 109.9***   | 4.28 | .377***   | .251       |
| Residual        | 6231.4     | 1.47 | 43.83     | .123       |
In addition to supporting the hypothesis that homophone distinctions are present in child-directed speech, these data have an unexpected feature. Token duration showed an interaction between utterance position and relative frequency. That utterance position affects duration is not in itself surprising. However, the particular interaction between position and frequency does not reflect other effects of sentence position on homophone pronunciation that have been previously reported. Conwell & Barta (2015) reported durational differences in adult-directed noun/verb homophones in sentence-medial position only. Likewise, Conwell (in press) found that noun/verb homophones in child-directed speech differ in duration in utterance-medial, but not utterance-final positions. The data presented here show differences as a function of frequency only in final position.

One important difference between the data reported here and the studies by Conwell and colleagues is that the previous work was concerned only with differences that result from lexical category and did not consider the role of meaning frequency in the production of the target words. The previous studies also include only noun and verb uses, while the analysis in this article includes a wider range of category types and controls statistically for category effects. The data presented here concern a set of word types that is wholly distinct from those examined in Conwell (in press) and Conwell & Barta (2015). Those studies that find only utterance-medial distinctions examine noun/verb homophones, many of which (e.g., kick) are in fact polysemes. Homophones and polysemes may differ in the extent to which adults represent their meanings separately (e.g., Gahl, 2008), which could affect the role of meaning frequency in creating durational differences. The study described in this article shows that meaning frequency contributes to durational differences over and above the effects of lexical category and sentence position.

Because this study finds consistent durational differences in homophones in child-directed speech, we can conclude that the slower pace and exaggerating lengthening of child-directed speech does not add variability that might overwhelm the rather subtle differences in homophone duration that have been reported in adult-directed speech (Gahl, 2008). Rather, children, like adults, are exposed to differences in homophone duration even when sentence position and grammatical category of use are controlled for. This finding has important implications for how children might learn homophones. Traditional theories of word learning posit that children will assume a one-to-one form-to-meaning mapping while in the early stages of word learning (Markman, 1990; Slobin, 1973). These accounts do not typically explain how children decide to violate that assumption to learn homophones. The evidence presented here creates the possibility that homophones do not violate the one-to-one assumption for children because the durational differences reported here are perceptible by young word learners. Conwell and Morgan (2012) reported that toddlers show perceptual sensitivity to the durational differences that distinguish noun/verb homophones, which could allow them to maintain separate forms for the two meanings. Because child-directed speech contains durational differences in homophones more broadly, children could similarly maintain distinct forms of homophonous words early in acquisition, which would allow them to circumvent the one-to-one mapping assumption that could impede homophone learning. For example, a child might encounter an unusually long token of a familiar word and use that durational information to conclude that this token has a novel meaning. This process might be additionally supported by the absence of the familiar referent in the context.

Other work on word learning, however, indicates that infants do not use small changes in pronunciation to assume that a word has a new meaning (e.g., Stager & Werker, 1997), suggesting that learners may not be able to use the durational differences described in this article to support homophone learning. Word familiarity, age, and vocabulary size also affect infants’ ability to detect differences in pronunciation in word identification and learning tasks (Fennell & Werker, 2003; Werker, Fennell, Corcoran, & Stager, 2002). Children become better able to detect minor changes in pronunciation and use them in learning tasks as they age and as their vocabularies expand, although their use of changes in pronunciation to assume that the speaker intends to refer to a novel object also depends on the nature of those changes (White & Morgan, 2008). Therefore, whether durational differences in homophones are recruited by children during the word learning process remains an open question. Completely addressing this question would require experimental manipulation of duration in a word learning study.
Furthermore, because a wide range of factors, including sentence position, lexical category and speaker identity also contribute to durational differences, these findings raise the question of how children determine the various sources of differences in token duration and allocate those differences appropriately, which would be a critical step in using these differences to learn multiple meanings for homophones. One possibility is that children have expectations about how much variability is associated with frequently encountered sources of durational differences, such as sentence position, and notice cases that lie outside the usual distribution. These outlier cases might be key to detecting a new source of durational variation. However, this account (or any account of how children detect and attribute durational variation) needs extensive empirical examination.

There are, of course, other cues that children might use to learn about homophones. In particular, the plausibility of the interpretation is a highly reliable source of the intended meaning, although syntactic and contextual factors can also disambiguate homophone meanings. However, Srinivasan and Barner (2013) showed that even 4- and 5-year-old children do not seem to use plausibility to interpret denominal verbs, which indicates that early homophone learning may not be well-constrained by plausibility. Lippeveld and Oshima-Takane (2014) reported that older toddlers can use a generalized instrument-action relationship to interpret novel denominal verbs, although that ability depends on children’s experience with cross-category word use. Such findings indicate that children may not have access to top-down information like plausibility and syntax early in the lexical acquisition process (but see Dautriche, Swingley, & Christophe, 2015, for data regarding the role of syntax in infants’ learning of phonological neighbors). Low-level acoustic differences such as those reported here could serve as an early, bottom-up cue to facilitate homophone learning in children whose understanding of syntax and context is as yet insufficient to allow them to use more top-down information.

The data described in this article show that child-directed speech contains reliable acoustic distinctions between homophone meanings and that those differences are a function of the frequency of the target meaning. These findings are similar to those reported in a large corpus of adult-directed speech (Gahl, 2008). Such differences, if they are perceptible, may allow children to learn distinct meanings for homophones in spite of their general disinclination to permit a single form to have more than one meaning. This low-level cue may give children a foothold to begin acquiring homophones before they have access to the kinds of top-down information that might play a role in homophone acquisition later in development.

Acknowledgments

The author wishes to thank Brenden Melvie, Katelyn Tallas, Matthew Kramer, Felix Pichardo, Cheyenne Brady, Adrienne MacDonald, Samantha Hamernick, and Stephanie Leach for their assistance with token extraction and measurement. Thanks also to Alejandrina Cristia for sharing her PRAAT scripts and three anonymous reviewers for their helpful comments on previous versions of the manuscript.

Funding

This research was funded by Grant 1R15HD077519-01 to the author from the National Institute of Child Health and Human Development, which is part of the National Institutes of Health. The contents of this article are the sole responsibility of the author and do not necessarily represent the official views of NICHD or NIH.

References

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

Bergelson, E., & Swingley, D. (2012). At 6 to 9 months, human infants know the meanings of many common nouns. Proceedings of the National Academy of Sciences, 109, 3253–3258. doi:10.1073/pnas.1113380109

Bernstein Ratner, N. (1984). Patterns of vowel modification in mother-child speech. Journal of Child Language, 11, 557–578.
Bernstein Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Phonetics, 14*, 303–309.

Beveridge, M., & March, L. (1991). The influence of linguistic context on young children’s understanding of homophonic words. *Journal of Child Language, 18*, 459–467. doi:10.1017/S0305000900011168

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

Bortfeld, H., Morgan, J., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech stream segmentation. *Psychological Science, 16*, 298–304. doi:10.1111/j.0956-7976.2005.016.issue-4

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

Casenhiser, D. (2005). Children’s resistance to homonymy: An experimental study of pseudohomonyms. *Journal of Child Language, 32*, 319–343. doi:10.1017/S0305000904006749

Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language, 55*, 767–811. doi:10.2307/412745

Conwell, E. (2014). Parents do not reduce their use of homophones when speaking to infants. Poster presented at the 19th International Conference on Infant Studies, Berlin.

Conwell, E. (in press). Prosodic disambiguation of noun/verb homophones in child-directed speech. *Journal of Child Language*.

Conwell, E., & Barta, K. (2015). Prosodic effects in noun/verb homophone production. Manuscript in preparation.

Conwell, E., & Morgan, J. L. (2012). Is it a noun or is it a verb? Resolving the ambicategoricality problem. *Language Learning and Development, 8*, 87–112. doi:10.1080/15475441.2011.580236

R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language, 41*, 913–934. doi:10.1017/S0305000912000669

Dautriche, I., Swingley, D., & Christophe, A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition, 143*, 77–86. doi:10.1016/j.cognition.2015.06.003

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, Epenthesis and Coda Licensing in the Early Acquisition of English. *Language and Speech, 49*, 137–173. doi:10.1177/0023830904009020201

Doherty, M. J. (2004). Children’s difficulty in learning homonyms. *Journal of Child Language, 31*, 203–214. doi:10.1017/S030500090300583X

Fennell, C. T., & Werker, J. F. (2003). Early word learners’ ability to access phonetic detail in well-known words. *Language and Speech, 46*, 245–264. doi:10.1177/0023830903460020901

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist, 66*, 103–114. doi:10.1525/aa.1964.66.suppl_3.02a00060

Fernald, A. (1989). Intonation and communicative intent in mothers’ speech to infants: Is the melody the message? *Child Development, 60*, 1497–1510. doi:10.2307/1130938

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., De Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers’ speech to preverbal infants. *Journal of Child Language, 16*, 477–501. doi:10.1017/S0305000900010679

Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: Crosslinguistic evidence. *Child Development, 67*, 3192–3218. doi:10.2307/1131774

Gahl, S. (2008). *Time and thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language, 84*, 474–496. doi:10.3353/lan.0035

Garnham, W. A., Brooks, J., Garnham, A., & Ostenfeld, A. (2000). From synonyms to homonyms: Exploring the role of metarepresentation in language understanding. *Developmental Science, 3*, 428–441. doi:10.1111/1467-7687.00087

Jurafsky, D., & Martin, J. H. (2000). *Speech, language, and mind*. Upper Saddle River, NJ: Prentice Hall.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*, 684–686. doi:10.1126/science.277.5326.684

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *Package ‘lmerTest’*. Retrieved from https://cran.r-project.org/web/packages/lmerTest/index.html

Li, A., Shi, R., & Hua, W. (2010). Prosodic cues to noun and verb categories in infant-directed Mandarin speech. *Speech Prosody, 10*, 1–4.

Lippeveld, M., & Oshima-Takane, Y. (2014). The effect of input on children’s cross-categorical use of polysemous noun-verb pairs. *Language Acquisition, 22*, 209–239. doi:10.1080/10489223.2014.943902

MacWhinney, B. J. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum.
Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science, 14*, 57–77. doi:10.1207/s15516709cog1401_4

Markman, E. M., & Wachtel, G. F. (1988). Children’s use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20*, 121–157. doi:10.1016/0010-0285(88)90017-5

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology, 47*, 241–275. doi:10.1016/S0010-0285(03)00034-3

Mazzocco, M. M. M. (1997). Children’s interpretations of homonyms: A developmental study. *Journal of Child Language, 24*, 441–467.

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. doi:10.2307/1166130

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*. Mahwah, NJ: Erlbaum.

Nygaard, L. C., Patel, N., & Queen, J. S. (2002). The link between prosody and meaning in the production of emotional homophones. *The Journal of the Acoustical Society of America, 112*, 2444. doi:10.1121/1.4780051

Shi, R., & Moisan, A. (2008). Prosodic cues to noun and verb categories in infant-directed speech. In H. Chan, H. Jacob, & E. Kapia (eds.), *Proceedings of the 32nd Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson, & D. I. Slobin (eds.), *Studies of child language development*. New York, NY: Holt, Rinehart & Winston.

Srinivasan, M., & Barner, D. (2011). The Amelia Bedelia effect: World knowledge and the goal bias in language acquisition. *Cognition, 128*, 431–450. doi:10.1016/j.cognition.2013.05.005

Srinivasan, M., & Rabagliati, H. (2015). How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua, 157*, 124–152. doi:10.1016/j.lingua.2014.12.004

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*, 381–382. doi:10.1038/41102

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. doi:10.1017/S0305000905007099

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants’ ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3*, 1–30. doi:10.1207/S15327078IN0301_1

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language, 59*, 114–132. doi:10.1016/j.jml.2008.03.001

---

**Appendix A. Homophone pairs included in the analysis**

| Ate/eight | Bare/bear |
| Berry/bury | Blew/blue |
| Board/bored | Buy/bye |
| Close/clothes | Dear/deer |
| Flour/flower | Hear/here |
| Hi/high | Hole/whole |
| Knew/new | Know/no |
| Knows/nose | Meat/meet |
| Pair/pear | Plain/plane |
| Read/red | Right/write |
| Road/rode | Sea/see |
| Threw/through | Toe/tow |
| Wear/where |  

---

**Note:** The list includes a total of 16 homophone pairs that were included in the analysis.