Stem similarity modulates infants' acquisition of phonological alternations

Megha Sundara[a], James White[b], Yun Jung Kim[c], and Adam J. Chong[d]

[a] Department of Linguistics, University of California, Los Angeles

[b] Department of Linguistics, University College London

[c] Program in Linguistics, Emory University

[d] Department of Linguistics, Queen Mary University of London

Address for correspondence

Megha Sundara

UCLA Department of Linguistics

3125 Campbell Hall

Los Angeles, CA 90095-1543

Email: megha.sundara@humnet.ucla.edu

Revised Manuscript (Clean version)

Abstract

Phonemes have variant pronunciations depending on context. For instance, in American English, the [t] in *pat* [pæt] and the [d] in *pad* [pæd] are both realized with a tap [ɾ] when the *–ing* suffix is attached, [pæɾɪŋ]. We show that despite greater distributional and acoustic support for the [t]-tap alternation, 12-month-olds successfully relate taps to stems with a perceptually-similar final [d], not the dissimilar final-[t]. Thus, distributional learning of phonological alternations is constrained by infants' preference for the alternation of perceptually-similar segments. Further, the ability to relate variant surface forms emerges between 8- and 12-months. Our findings of biased learning provide further empirical support for a role for perceptual similarity in the acquisition of linguistically-relevant categories. We discuss the implications of our findings for phonological theory, language acquisition and models of the mental lexicon.

Keywords: acquisition, phoneme, morphology, allophones, variants

1. Introduction

During acquisition, infants must learn that morphemes sometimes have variant pronunciations in different phonological contexts. For example, in American English, the [t] in *pat* [pæt] and the [d] in *pad* [pæd] are realized with a tap [ɾ] when the *–ing* morpheme is attached in *patting* or *padding*, both [pæɾɪŋ]. Adult speakers of American English not only know that [pæɾɪŋ] is a suffixed instance of *pat* [pæt] or *pad* [pæd], but they have acquired knowledge of the [t ~ ɾ] and [d ~ ɾ] alternations and can apply them to novel cases such as made-up words (Braver, 2014). How might infants learn such phonological alternations?

Unsurprisingly, computational and experimental research indicates that learners can exploit input statistics to learn phonological alternations (Peperkamp & Dupoux, 2002; Peperkamp, Le Calvez, Nadal, & Dupoux, 2006; White, Peperkamp, Kirk & Morgan, 2008). Indeed, infants have been reported to use such a domain-general statistical learning mechanism to accomplish a variety of linguistic (e.g., Maye, Werker, & Gerken, 2002; Anderson, Morgan, & White, 2003; Chambers, Onishi, & Fisher, 2003; Saffran, Aslin, & Newport, 1996) and non-linguistic tasks (e.g., Saffran, Johnson, Aslin & Newport, 1999; Fisher & Aslin, 2002).

However, previous work suggests that the learning of phonological alternations is biased by the similarity of the sounds involved. For instance, adults in an artificial language experiment find it easier to learn novel alternations between phonetically similar sounds (e.g., [p] and [t]) than alternations between phonetically dissimilar sounds ([p] and [z]; Skoruppa, Lambrechts, & Peperkamp, 2011). Moreover, adults generalize such alternations in an asymmetric way: alternations between dissimilar sounds (e.g. [p] and [v]) are generalized to more similar sounds ([b] and [v]), but not *vice versa* (White, 2014; see also Wilson, 2006). The same asymmetry in generalization is observed in artificial language experiments with 12-month-old infants (White & Sundara, 2014). These findings suggest that although learners use distributional learning to discover phonological alternations, this learning is biased such that alternations between phonetically similar sounds are preferred. Further support for such a bias has come from studies of linguistic typology (Hayes & White, 2015) and from computational learning models (Peperkamp et al., 2006; Wilson, 2006; White, 2017; Calamaro & Jarosz, 2015).

The idea that categories—whether linguistic, auditory, visual or conceptual—are rooted in similarity, particularly perceptual similarity, has deep roots. For example, in classic theories of categorization in cognitive psychology, an object is categorized based on the extent of its similarity to the respective prototypes (e.g. Posner & Keele, 1968; Rosch & Mervis, 1975). In exemplar theories, category membership is determined by similarity to other items in that category (Medin & Schafer, 1978; Nosofsky, 1986). In linguistic theory as well, learners prefer morphologically-related forms to be as similar to each other as possible, a core principle referred to as paradigm uniformity. This bias forms the basis of some theories of language change (Kiparsky 1982), and accounts for the cross-linguistic tendency for differences between morphologically-related words to be small (Steriade 2001/2008).

In two experiments, we investigated whether the perceptual distance between alternating forms affects infants' first language acquisition. These experiments were designed to test whether English-learning 12-month-olds exhibit biased learning of the English [t ~ ɾ] and [d ~ ɾ] alternations in the *–ing* context. Comparing infants' knowledge of the [t ~ ɾ] and [d ~ ɾ] alternations (e.g. their ability to relate the [pæɾ] in *patting*/*padding* to either *pat* [pæt] or *pad* [pæd]) is an excellent case for testing the role of perceptual distance during phonological acquisition. Because [ɾ] alternates with both [t] and [d] in English, infants could learn the [t ~ ɾ] alternation first, the [d ~ ɾ] alternation first, or both at the same time. As we show in an analysis of the Brent corpus (Brent & Siskind, 2001), distributional evidence from infants' linguistic input favors learning the [t ~ ɾ] alternation. A different prediction arises when we consider the similarity of taps and [t] and [d]. Although both [t] and [d] alternate with taps, taps are perceptually more similar to [d] than [t] (de Jong, 1998; Herd, Jongman & Sereno, 2010; Malécot & Lloyd, 1968). Taps are voiced like [d]. Furthermore, X-ray microbeam data shows that the tongue position for taps is more similar to that of [d] than [t] providing another basis for the perceived similarity of taps and [d] (de Jong, 1998). So, if distributional learning is biased by the perceptual similarity of the segments involved, then infants should favor the [d ~ ɾ] alternation.

2. Corpus Analysis

Before investigating infants' burgeoning morpho-phonological abilities, we conducted a corpus study and acoustic analysis to characterize English-learning infants' speech input. We focused primarily on tapping before the *–ing* suffix because we subsequently used this context in the infant experiments.

English *–ing* is an inflectional suffix that signals grammatical information and occurs only as part of a word, typically a verb. We chose this context because *–ing* is one of the most frequent suffixes in infant-directed speech (Brown, 1973), and it is reported to be the first suffix produced by English-learning infants (e.g., Brown, 1973; de Villiers and de Villiers, 1973). For completeness, in the corpus counts we report on other tapping contexts as well.

Tapping has been studied extensively in adult-directed speech, but we know little about this process in child-directed speech. In American English, taps occur as variants of both the phonemes /t/ and /d/ in several different contexts. Word-medially, particularly between a stressed and an unstressed vowel like in the word *water*, a tap is almost obligatory (Oshika, Zue, Weeks, Neu & Aurbach, 1975; Zue & Laferriere, 1979; Kahn, 1980; Turk, 1992). Additionally, in adult-directed corpora, the probability of tapping in word-medial contexts is mediated by lexical frequency and morphological complexity: taps appear more often in words that are frequent and morphologically simpler. The morphological complexity effect is evident even when words are controlled for lexical frequency (Patterson & Connine, 2001).

Despite the variation due to frequency and morphological complexity, taps are the most frequent variant of /t/ and /d/ in word-medial contexts where they are licensed. Corpus studies of adult-directed speech as well as laboratory investigations of speech production show that, depending on the study between 76% and 99% of word-medial /t/ and /d/ are produced as taps (Patterson & Connine, 2001; Zue & Laferriere, 1979; Herd, Jongman & Sereno, 2010).

In our analysis of child-directed speech, we focused specifically on two questions. First, if infants were to rely solely on distributional information, would their input favor learning either the [t ~ ɾ] or [d ~ ɾ] alternation? In other words, do the input statistics make it more likely that [ɾ] is a variant of /t/ or /d/? To answer this question, we looked at the frequency of /t/ and /d/ in tapping contexts and elsewhere. Second, are there acoustic cues that infants could use to determine whether a given [ɾ] is derived from /t/ or /d/? We answered this question by performing an acoustic analysis to see if the distinction between /t/ and /d/ is completely neutralized when these sounds are tapped in child-directed speech.

For these analyses, we extracted tokens from the Brent corpus (Brent & Siskind, 2001) in the CHILDES database (MacWhinney, 2000). This corpus includes transcripts and audio recordings of 18 infant-mother dyads recorded at home. We selected 8 of the mothers for our analysis, all of whom were university educated; we did so because they are a good match to the demographics of the parents providing the speech corpus to the infants tested in the subsequent perception experiments. Infants were between the ages of 0;9 and 1;3 (i.e. around the age of our 12-month-old participants in the experiments) at the time of the recording.

2.1 Corpus counts

As a first step, we extracted all words ending in orthographic *–ting* or *–ding* from the 300,000-word Brent sub-corpus. One token contained a morpheme-internal /d/ (*pudding*), and this token was excluded. Overall, infants heard far more *–ting* words (41 types, 618 tokens) than *–ding* words (14 types, 130 tokens). This disparity is not limited to our sub-corpus. In the full CHILDES corpus (approximately 2 million words) as well, there are roughly twice as many words ending in *–ting* than ending in *–ding*.

Tapping, however, most commonly occurs when /t/ or /d/ appear between two vowels (the second of which is unstressed), so we next measured the frequencies of only those *–ting* and *–ding* words in which the /t/ or /d/ was preceded by a vowel. Of these, there were once again far more *–ting* words (34 types, 607 tokens) than *–ding* words (12 types, 96 tokens). In addition, the four most frequent words in the subset of words ending in *–ting*/*–ding* (where /t/ or /d/ was preceded by a vowel) were all *–ting* words: *getting* (239 tokens), *eating* (123 tokens), *sitting* (71 tokens), and *putting* (64 tokens). For comparison, the most frequent *–ding* word was *recording*, with only 33 tokens. Accordingly, MacArthur CDI lexical norms (Fenson, et al., 1994) show that verbs ending in *–ting* outnumber verbs ending in *–ding* in 12-month-olds' comprehension (6 *–ting* words vs. 2 *–ding* words) and production vocabularies (2 vs. 0).

Does this disparity in the frequency between /t/ and /d/ hold in other contexts where infants encounter taps? To address this question, we further extracted all words ending in /t/ or /d/ plus one of the endings *–er, –al,* or *–able* from our sub-corpus. These endings all correspond to unstressed suffixes that trigger tapping in American English. In addition, we extracted all words ending in *–t, –d* (combined with orthographic *–te,* or *–de*) because tapping also occurs word-finally when the following word begins with a vowel, albeit at a lower rate. We excluded the following: morpheme-internal cases like *water*; words where *–t* or *–d* were preceded by obstruent consonants, like *–s* and *–p*, where tapping is not allowed; and words where *–t* or *–d* were preceded by *–n* and *–l*, where tapping is sometimes possible, but qualitatively and quantitatively different. The frequency counts for all words with each ending as well as for only those words in which /t/ or /d/ is preceded by a vowel sound are shown in Table 1.

Table 1. Type and token frequencies from our sub-corpus, broken down by word ending. Frequencies are given for all words with a certain ending, as well as for words in which the ending is preceded by a vowel sound.

	Frequency (all words)		Frequency (preceded by vowel only)	
	Types	Tokens	Types	Tokens

-ting	41	618	34	607
-ding	14	130	12	96
-t / *-d* ratio	2.3	4.8	2.8	6.3
-ter	30	174	26	163
-der	9	129	7	101
-t / *-d* ratio	3.3	1.4	3.7	1.6
-tal	0	0	0	0
-dal	0	0	0	0
-t / *-d* ratio	--	--	--	--
-table	2	10	2	10
-dable	0	0	0	0
-t / *-d* ratio	--	--	--	--
-t or *-te* final	281	33101	256	32175
-d or *-de* final	255	6786	186	6076

-t / *-d* ratio	1.1	4.8	1.4	5.3

The corpus counts show that infants hear far more instances of /t/ than /d/ in tapping contexts. This is true for every context that we extracted, both in type and token frequency. Though these numbers do not account for every context in which infants hear taps (e.g., in morpheme-medial environments like *water*), the results strongly suggest that the disparity between /t/ and /d/ is robust across tapping contexts. These corpus counts indicate that taps encountered by infants are more frequently derived from /t/ than /d/.

2.2 Acoustic analysis

Our acoustic analysis focused on words ending in *–ting* and *–ding* where the /t/ or /d/ was preceded by a vowel. A native speaker of North American English with phonetic training segmented each token from the sub-corpus using Praat (Boersma & Weenink, 2013). To determine whether the contrast between /t/ and /d/ was fully neutralized in tapped words, we analyzed two measures: closure duration of the (potential) tap and duration of the preceding vowel. There is some evidence that acoustic differences between sounds are exaggerated in minimal pairs, where the two words are only distinguished by the sounds in question (Baese-Berk & Goldrick, 2009); however, because of limited items — only two /d/ words and 9 /t/ words in our dataset were part of a minimal pair — we did not include this as a variable in our analysis.

The boundary between the consonant and the surrounding vowels was marked at the point where there was no longer a clear second or third formant (or where there was a clear change in the darkness of these formants). In many cases, the closure was fully reduced (i.e., the formants continued without a change throughout the token) or the consonant was glottalized such that it was impossible to mark a clear beginning or ending. These cases were excluded from the acoustic analyses (n = 344, 49% of the total). Virtually all of the glottalized tokens were cases of the casual variant containing a pre-glottalized /t/ followed by a syllabic [n] with the vowel deleted (e.g., *eating* [ˈiɾɪŋ] pronounced as *eatin'* [ˈiˀtn̩]). This realization was very common for the *–ing* words. The large difference in frequency between /t/ and /d/ still holds when the glottalized or fully lenited tokens are excluded (*–ting:* 35 types, 251 tokens; *–ding:* 13 types, 111 tokens).

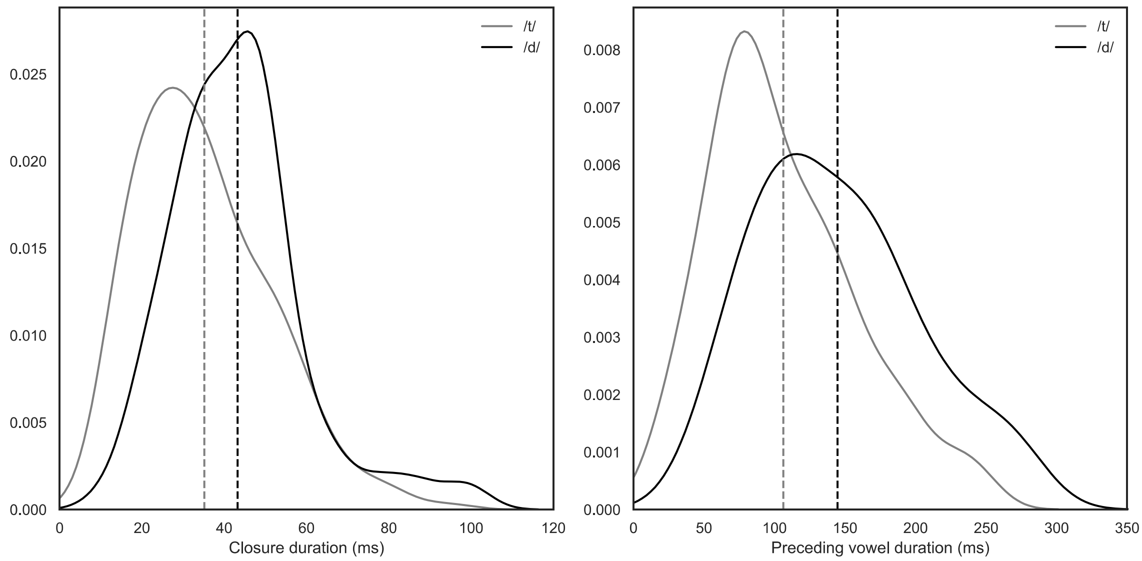
The preceding vowel was also segmented by looking for a clear second and third formant. Once segmented, each token was coded according to whether the tap was derived from a /t/ (e.g., in *eating*) or from a /d/ (e.g., in *hiding*). The closure duration and preceding vowel duration of each token were then calculated automatically using a PRAAT script.

We excluded tokens with closure durations that were more than two standard deviations (SDs) above the overall mean (i.e., tokens with closure durations greater than 110.76ms) as outliers. This resulted in the exclusion of 13 tokens (3.6% of 362 tokens). With such long durations, these tokens were most likely full stops rather than taps. We likewise excluded tokens for which the preceding vowel duration was more than two SDs above the overall mean (those greater than 289.47ms), resulting in the exclusion of a further 8 tokens (2.3% of 349 tokens). The remaining 341 tokens were included in the acoustic analysis.

Figure 1 shows kernel density estimations (KDEs) for closure duration and preceding vowel duration, separated according to whether the taps were derived from /t/ or /d/. The closure duration of taps derived from /d/ (mean = 43.26ms, SD = 16.32) was greater on average than the closure duration of taps derived from /t/ (mean = 35.17ms, SD = 16.17). The preceding vowel duration of types derived from /d/ (mean = 144.61ms, SD = 58.34) was also greater on average than the preceding vowel duration of taps derived from /t/ (mean = 106.21, SD = 52.03).

To evaluate the significance of these differences, we implemented a logistic regression model in R (R Core Development Team, 2013) predicting the log odds that a token was derived from /d/ based on closure duration (ms), preceding vowel duration (ms), and their interaction. Closure duration ($\beta = 0.052$, $z = 2.718$, $p = .007$) and preceding vowel duration ($\beta = 0.020$, $z = 3.266$, $p = .001$) were both found to be significant predictors of the underlying form of the tap. The interaction was non-significant ($p > .10$).

Figure 1. Kernel density estimations (KDEs) of closure durations (left panel) and preceding vowel durations (right panel) for taps derived from /t/ versus /d/. Means are indicated by the dashed lines.



Next, we compared the rate of tapping for /t/ versus /d/. Instead of relying on transcriptions, we used the data-driven method described by Herd et al. (2010) for this purpose. First, for each speaker we plotted a histogram of the distribution of closure duration for all /t/ and /d/ tokens. Then, we examined the histogram to identify any discontinuity between the two modes for stop and tap durations; the rate of tapping for /t/ and /d/ was calculated based on this cut-off. Overall, in our corpus there was no difference in tapping rate between /t/ (225/251, 89.6%) and /d/ (100/111, 90.1%), although there was considerable individual variation, likely because of the smaller number of tokens analyzed per speaker. Two speakers tapped both /t/ and /d/ 100% of the time, four others tapped /d/ more often than /t/, and two tapped /t/ more often than /d/.

2.3 Discussion

Our corpus study had two main findings with respect to tapping in American English child-directed speech. First, the corpus statistics show that the /t/ is more frequent than /d/ in tapping contexts. This disparity between /t/ and /d/ is not only true for the *–ing* context that we used in our experiments, but robust across a variety of contexts in which infants encounter taps. Based on frequency alone, taps encountered by infants are most likely derived from /t/. Thus, if input frequencies were leading infants to acquire one of the tapping alternations (either [t ~ ɾ] or [d ~ ɾ]) before the other, the [t ~ ɾ] alternation should be the one learned first.

Second, our acoustic analyses indicate that /t/ and /d/ are not fully neutralized when they undergo tapping in child-directed speech; taps derived from /d/ tend to have longer closure durations and longer preceding vowels than those derived from /t/. Logistic regression indicated that these acoustic parameters were significant predictors of whether a tap was derived from /t/ and /d/. In other words, the distributions of taps derived from /t/ and those derived from /d/ differ significantly from one another acoustically. From the perspective of an infant faced with learning these alternations, this suggests that closure duration and preceding vowel duration could, in principle, serve as cues to the underlying form of the taps they hear. This, in turn, could enable infants to learn both alternations if they were able to notice the differences and exploit them during acquisition.

To our knowledge, our study is the first to analyze the acoustics of taps in child-directed speech. How do these findings compare to acoustic analyses of tapping in adult-directed speech? To date, studies of adult-directed speech have been equivocal. Some have supported complete neutralization (Charles-Luce, 1997; Joos, 1942; Port, 1976), whereas others have shown differences in length of the preceding vowel, and less often, differences in closure duration or in the dip in intensity during the closure of the tap associated with /t/ versus /d/ (Braver, 2014; Fisher & Hirsh, 1976; Fox & Terbeek, 1977; Huff, 1980; Sharf, 1962; Patterson & Connine, 2001; Zue & Laferriere, 1979). The closure duration difference reported in this study (8ms) is more robust compared to previous reports on adult-directed speech (~ 1ms e.g., Zue & Laferriere, 1979; Herd et al., 2010). The vowel duration difference reported in this study (38ms) is also larger than that reported for adult-directed speech (7-16ms based on Herd et al., 2010). The larger durational differences in our study are likely due to the documented slower speaking rate in child-directed speech compared to adult-directed speech (e.g., Fernald et al., 1989).

Our finding of comparable tapping rates between /t/ and /d/ is also different from previous reports; Herd et al. (2010) report a lower rate of tapping for /t/ (76%) than /d/ (99%). We think this discrepancy stems primarily from the difference in exclusion criteria between our study and Herd et al.'s, rather than a true difference between adult- and child-directed speech. Recall that we excluded all tokens with a pre-glottalized /t/ followed by a syllabic [n] (with the vowel deleted) because tapping is not possible in the variant with the syllabic [n]; Herd et al. did not exclude them. By definition, pre-glottalized /t/ tokens are not tapped; thus, including them in the analysis would lower the tapping rate for /t/. We also had far fewer tokens from each speaker (range 23:80) than Herd et al (average ~300+ per speaker), and they analyzed data from an elicited speech task instead of a spontaneous speech corpus as analyzed here. Further research is needed to evaluate the extent to which tapping rate differs across speakers and elicited versus spontaneous speech.

In sum, infants have ample experience with tapping in spontaneous speech addressed to them. Additionally, infants hear more taps derived from /t/ than /d/ and have access to acoustic information, in the form of closure duration and preceding vowel duration, to potentially distinguish taps derived from /t/ and /d/. Thus, based on the distributional evidence alone, infants are predicted to learn the [t ~ ɾ] alternation first, before [d ~ ɾ]. We test this prediction in the next section (Experiment 1) with 12-month-old English-learning infants.

3. Experiment 1: 12-month-olds

Even before they start producing bound morphemes in the second year of life, English-learning 15-month-olds demonstrate some knowledge of the combinatorial properties of *–ing* (Mintz, 2013; Santelmann & Jusczyk, 1998; Hirsh-Pasek & Golinkoff, 1996; Golinkoff, Hirsh-Pasek, & Schweisguth, 2001; Soderstrom, Wexler, & Jusczyk, 2002). Most recently, Willits et al. (Willits, Seidenberg & Saffran, 2014) have demonstrated that when familiarized with sequences like *kissing*, where a highly familiar verb is suffixed with *–ing*, 7.5-month-olds listen longer to familiar stems like *kiss*. We have since demonstrated that 8-, but not 6-month-old English-learning infants successfully relate even nonce verbs like *babbing* and *kelling* with their nonce stems (i.e., *bab* and *kell*; Kim & Sundara, under review). Thus, 8-month-olds already demonstrate some knowledge of morphology, at least for frequent suffixes like *-ing,* when suffixed forms do not involve any change in the stem.

In Experiment 1 we used the Headturn Preference Procedure (HPP) to test whether English-learning infants familiarized with passages containing [ɾ] + *-ing* forms successfully relate the tapped suffixed forms to the corresponding /t/- and /d/-final stems produced in isolation. We targeted 12-month-olds because infants at this age have been shown to be able to learn novel phonological alternations after brief exposure to an artificial language (White et al., 2008; White & Sundara, 2014). Infants were tested using a modified version of the word segmentation paradigm in Jusczyk and Aslin (1995). Like in Jusczyk and Aslin, testing was done in two phases: familiarization and test. During the familiarization phase, infants were familiarized to two of four passages, each containing a target word ending in [ɾ] + *-ing* (either *cu[ɾ]ing* and *mee[ɾ]ing*, or *pa[ɾ]ing* and *shoo[ɾ]ing*, counterbalanced). Infants never heard the target words pronounced without *–ing* during familiarization.

In the test phase, infants were presented with all four target words in isolation. Words were presented without *–ing* and with the final /t/ or /d/ produced as a full stop consonant. That is, during the test phase, half of the infants (those tested with /t/-final words) heard *cut, meet, pat* and *shoot* whereas the other half (tested on the /d/-final words) heard *cud, mead, pad* and *shoo-ed*. All of these CVCs are phonotactically legal in English. All bare forms whether ending in [t] or [d] are also real words of English (though not all are verbs). However, based on the lexical norms of the MacArthur-Bates Communicative Development Inventories (CDI; Fenson et al., 1993), English-learning 12-month-olds neither comprehend nor produce any of the /t/-final or /d/-final words. Thus, all of the target words should be treated as nonce words by 12-month-olds. Finally, although 12-month-olds neither understand nor produce these words, they are more likely to have heard the /t/-final words given the extremely low frequency of the /d/-final words (except possibly *pad*).

Because each infant was only familiarized to two of the four passages, two target words presented in the test trials were entirely novel. The other two target words were potentially familiar; however, infants could only recognize these words as familiar if they could (a) decompose the inflected forms into stem + *–ing*, and (b) relate the stem with the altered segment, e.g., the *pa[ɾ]* of *pa[ɾ]ing,* to either the isolated form *pat* or *pad*. If infants look longer to potentially familiar words than to novel words at test, then we can conclude that they have succeeded in relating the suffixed words (with the tapping alternation applied) to the stems.

Crucially, all tapped *-ing* forms presented in the familiarization phase were derived from natural recordings of /t/-final verbs (*cut, meet, pat* and *shoot*). Based on the analysis presented in the previous section, we know that child-directed speech has acoustic cues to distinguish taps derived from /t/ and those derived from /d/. If infants can exploit these acoustic cues — either tap duration or the duration of the vowel that precedes it — they should relate the tapped *–ing* forms to potentially familiar stems ending in /t/, not /d/. Infants may also relate the tapped *–ing* forms to potentially familiar stems ending in /t/, not /d/, based on the statistics of their input. As we have shown in the corpus analysis, words ending in /t/ are more frequent than words ending in /d/ in child-directed speech, including in contexts where these sounds are tapped. However, if the learning of phonological alternations is biased by the perceptual similarity of the sounds involved, then infants should instead relate tapped *–ing* forms to potentially familiar stems ending in /d/, not /t/.

3.1. Method

3.1.1. Participants

Forty-eight full-term, monolingual English-learning 12-month-olds participated in the experiment. Half were tested on /t/-final isolated word lists (mean age = 370 days; range 349:404; 8 girls), and half were tested on /d/-final isolated word lists (mean age = 378 days; range 359:407; 9 girls). None of the infants had a history of speech, language, or hearing difficulties according to a parental report. They were all in good health and did not have a cold or ear infection on the day of testing. The parents of the infants completed a detailed language questionnaire (Bosch & Sebastián-Gallés, 2001; Sundara & Scutellaro, 2011), and only data from infants who had at least 90% of their language input in English were included. Twentyseven additional 12-month-olds were tested but their data were discarded for the following reasons: they did not complete the experiment due to fussiness (n = 21); their parents interfered during testing (n = 2); or they had a cold or ear infection on the day of testing (n = 4).

3.1.2. Stimuli

The four /t/-final CVC target words were *cut*, *meet*, *pat* and *shoot*. These four words were chosen because they could be suffixed with the *–ing* morpheme and have different vowel qualities (two front vowels, [i] and [æ], and two back vowels, [ʌ] and [u]). We used *cud, mead, pad,* and *shoo-ed* as the corresponding /d/-final CVC target words.

Each word list used in the test phase consisted of 15 repetitions of one target word concatenated with an inter-stimulus interval of 600ms. In all repetitions, the final stop (/t/ or /d/) was fully released and audible. For the familiarization phase, we recorded four six-sentence passages, each containing one of the /t/-final target words with the *–ing* morpheme (i.e. *cutting, meeting, patting,* or *shooting*). These passages are listed in Table 2. The target words appeared in sentence-initial position three times and in sentence-medial position three times, for a total of six instances per passage. All infants heard the same familiarization stimuli; since the familiarization recordings contained taps derived from /t/ (based on the written script provided to the speaker who recorded the stimuli), any acoustic cues that could potentially be used to distinguish between taps derived from /t/ and those derived from /d/ would favor the /t/ interpretation.

The stimuli were recorded by a 25-year-old female native English speaker from Southern California who was not familiar with the purpose of the study. She was instructed to read the words and passages in an animated voice as if talking to a preverbal infant. The stimuli were recorded in a soundproof booth using a Shure SM10A head-mounted microphone. All stimuli were digitized at a sampling frequency of 22050Hz and 16-bit quantization.

Table 2. Passages used in Experiment 1 and 2

Cut/cud

Cutting papers with scissors is a lot of fun. **Cutting** cakes evenly is very difficult. **Cutting** coupons is something he enjoys daily. Unlike me, my sister loves **cutting** carrots. Mommy is really good at **cutting** tofu. Daddy always takes care of **cutting** wood.

Meet/mead

Meeting people in a new environment is stressful. **Meeting** friends in a playground is always fun. **Meeting** him on the street like that was a surprise. My mom doesn't like **meeting** strangers. I was overwhelmed with **meeting** my grandparents. I get nervous before **meeting** professors.

Pat/pad

Patting dogs that you don't know can be dangerous. **Patting** sheep all day long is the best thing. **Patting** animals always relaxes me. My dog gets angry when he sees me **patting** cats. Please wash your hands before **patting** the baby. Whenever I see cats, I cannot help **patting** them.

Shoot/shoo-ed

Shooting a jump shot is very difficult. **Shooting** an arrow is hard when it's windy. **Shooting** a movie is my favorite hobby. I had fun at the carnival **shooting** balloons. He needs to get better at **shooting** free throws. I practiced a lot by **shooting** targets.

The acoustic properties of the target words, from the familiarization passages and the test phase, are reported in Table 3. All the acoustic measurements were made using Praat (Boersma & Weenink, 2013). The average duration of the full passages was 18s (SD = 0.03) and the average pitch was 235Hz (SD = 3). The average duration of the taps in the passages was 18.03 ms (SD = 7.8). The average duration of the lists was also 18s in both the /t/-final condition and the /d/-final condition (SD = 0.05 for /t/-final, 0.2 for /d/-final). All stimuli were played back at an average of 73dB.

Table 3. Acoustic measures for target words in passages and lists.

Measures	Familiarization passages	Test /t/-final	Test /d/-final
Average duration (ms)	404	646	572
Duration range (Min:Max)	331:487	530:792	482:637
Average pitch (Hz)	250	228	207
Pitch range (Min:Max)	236:264	219:233	189:221

3.1.3. Procedure

The Headturn Preference Procedure (HPP) was used to test infants (Kemler-Nelson, Jusczyk, Mandel, Myers, Turk, & Gerken, 1995; Juczyk & Aslin, 1995). Infants sat on their caregiver's lap in the center of a three-sided booth. On each side panel, a red light was located at eye level. A green light was mounted on the center panel at eye level and a camera was mounted above the green light behind this panel. The experimenter observed the infant through a monitor connected to the camera. The experimenter recorded the direction and duration of the infants' head turns, which in turn determined the presentation of the speech stimuli. Both the caregiver and the experimenter wore 3M-Peltor noise cancelling headphones that delivered masking music so they could not influence the infants' behavior. Infant looking time to the flashing lights was used as a proxy for listening time.

Each trial began when the green light on the center panel flashed. Once the infant oriented towards the center panel, one of the red lights on the side panels began to flash. When the infant turned her head towards that light, auditory stimuli began to play. Stimulus presentation continued until the infant looked away from the flashing light for more than two consecutive seconds or until the end of the trial (max = 18s). In the familiarization phase, presentation of the two passages alternated until infants accumulated 45 seconds of listening time to each passage. The two passages continued to alternate until the criterion was met for both passages individually. In the test phase, lists of each of the four words were presented in two blocks for a total of eight test trials. We used fewer test trials than Jusczyk and Aslin (eight rather than twelve) as the infants tested in this experiment were 12-month-olds, older than is typical for word segmentation studies (see Nazzi et al., 2005 for a similar design with 12-month-olds). The order of the test trials within each block was randomized for each infant.

3.1.4 Analysis

Listening time data were analyzed using linear mixed effects models in R (R Core Development Team, 2013) using *lmerTest* (Kuznetsova, Brockhoff & Christensen, 2017). Because listening times are not normally distributed, we log-transformed them (Csibra, Hernik, Mascaro, Tatone and Lengyel, 2016). The main findings remain the same even without this transformation. The fixed effects included Final Consonant ([t] or [d]) and Condition (*cu[ɾ]ing* and *mee[ɾ]ing*, or *pa[ɾ]ing* and *shoo[ɾ]ing*) as between-subjects variables, and Block (1 or 2) and Trial-Type (potentially familiar vs. novel) as within-subjects variables, and all their interactions. Additionally, the model included random intercepts for Subjects, to allow for differences in baseline listening times. We also included two random slopes, one for Subject by Trial-Type and another for Subject by Block. We report results from the highest-level random effect structure that converged (Barr, Levy, Scheepers & Tily, 2013). Fixed effects were evaluated against the full model using the *anova()* function; the output of this function provides Type III Analysis of Variance tables with Satterthwaite approximation, including *p*-values. Testing significance of fixed effects using the lmerTest package has been shown to give more acceptable Type 1 error over a range of sample sizes, including small ones, compared to Likelihood Ratio tests (Luke, 2017). We further confirmed the role of the significant variable using the step() function which uses backward stepwise selection of significant predictors in the model. Finally, planned comparisons, if warranted, were done using the *emmeans* package (Lenth, Singmann, Love, Buerkner and Herve, 2020) in *R*. Because log-transformed listening times are harder to interpret, and for ease of comparison to previously published research, we present raw listening times in figures throughout the paper.

Finally, we report power for the critical comparisons calculated using the *simr* package (Green & MacLeod, 2016). Power was calculated in three steps. First, new values for the response variable were simulated using the model provided. Next, the model was refit to the new data, in our case, 100 times. Last, a likelihood ratio test was applied to the simulated model fit. The power was estimated based on the number of successes and failures at the last step; we report 95% confidence intervals. Full analysis and results of the model are presented in the supplementary material.

3.2. Results & Discussion

Figure 2 shows the mean listening time to potentially familiar and novel stems in the /t/-final and /d/-final test groups. The final model with a maximal random effects structure that converged had only a random intercept for Subjects. The main effects of Block [$F(1, 324) = 29.6$, $p < 0.001$] and Final consonant [$F(1, 44) = 4.4$, $p = 0.04$] were significant. Crucially, there was a significant interaction of Final Consonant and Trial-Type [$F(1, 324) = 7.3$, $p = 0.007$; 95% CI for power = 96:100]. Planned comparisons using *emmeans* confirmed that there was a significant effect of Trial-Type, with infants listening longer to potentially familiar stems over novel ones in the /d/-final group [$t(324) = 2.9$, $p = 0.004$], but not the /t/-final group [$t(324) = -0.9$, $p = 0.3$].

These results show that English-learning 12-month-olds familiarized with *pa[ɾ]ing* succeeded in recognizing *pad* as a familiar word, but they failed to recognize *pat* as a familiar word (and likewise for the other words). That is, at 12-months, infants can detect morphological relatedness in the presence of small, but not large, perceptual mismatches between suffixed forms and their respective stems.

An alternative we can rule out is that infants looked longer to *pad* not because they have decomposed *pa[ɾ]ing* and learned the [d ~ ɾ] alternation, but rather because of the raw phonetic similarity of *pad* and the first chunk of *pa[ɾ]ing*. This scenario is untenable for a couple of reasons. First, previous research has shown that English-learning infants fail to map part-words (e.g. *dock*) to whole words (e.g. *doctor*), even when the part-word is a stressed syllable (Jusczyk, Houston & Newsome, 1999). Second, the *pa[ɾ]* in *pa[ɾ]ing* is acoustically more similar to *pat* than to *pad*. This can be seen in Table 4, which shows the acoustic characteristics of the target words. In terms of duration, vowels in /t/-final words are shorter than vowels in /d/-final words; the vowels followed by [ɾ] are even shorter than those in the /t/-final words. In terms of formants, the comparison is inconsistent across items, but if anything, vowels in the familiarization words are again more similar to those in the /t/-final words than to those in the /d/-final words overall. Thus, if infants were responding just based on the global phonetic similarity of *pa[ɾ]* and *pad*/*pat*, they should have mapped *pa[ɾ]ing* to *pat,* in direct contrast to the actual results.

Figure 2. Box plots showing the distribution of listening times (+/- SE) to the test passages containing the potentially familiar and novel stems sorted by final consonant, /t/-final or /d/final for Experiment 1 and 2.

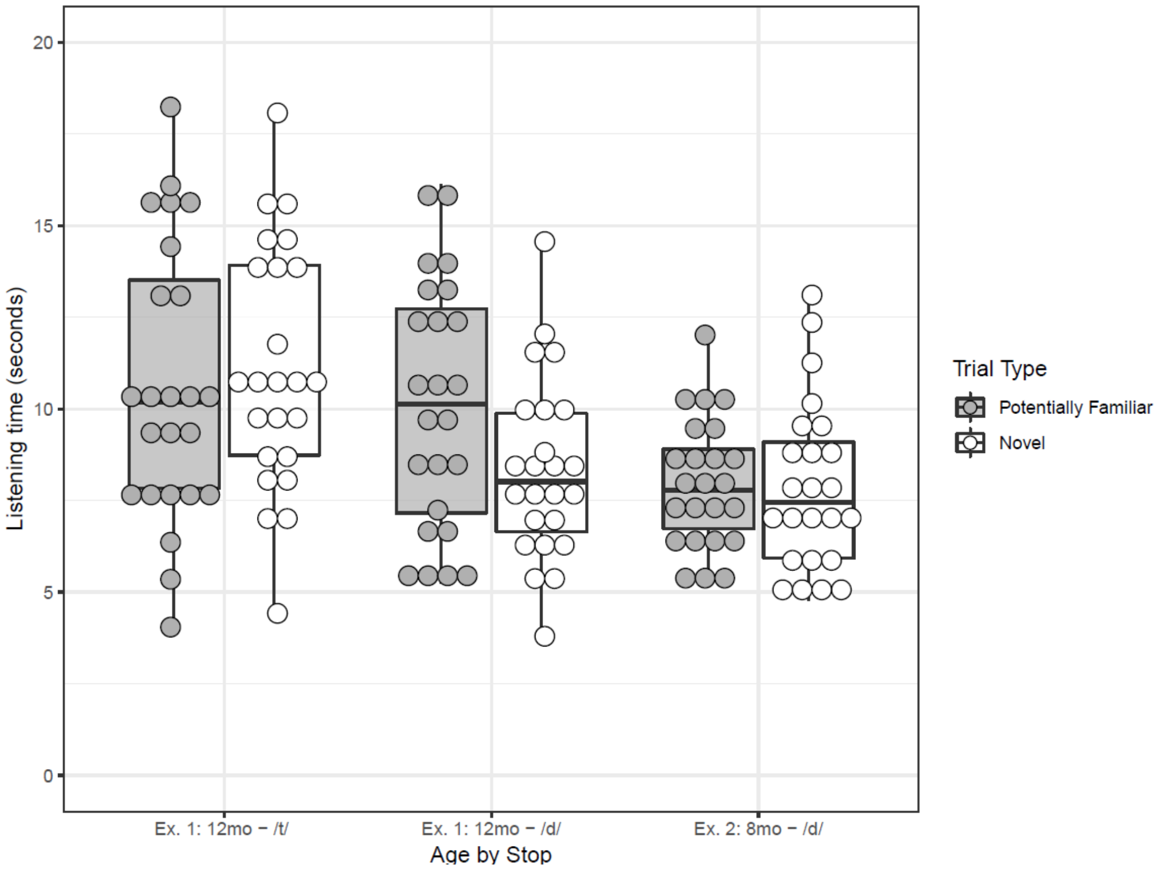


Table 4. Mean acoustic measures for target word vowels when followed by [ɾ] (in familiarization) and when followed by [t] or [d] (in test lists).

	Vowel duration (ms)	F1 (Hz)	F2 (Hz)
cu[ɾ]	69	782	1876
cut	185	849	1833
cud	334	751	1763
mee[ɾ]	92	417	2437

meet	237	430	2685
meed	420	440	2621
pa[ɾ]	87	990	1760
pat	201	1010	1756
pad	366	896	1745
shoo[ɾ]	94	446	2046
shoot	288	450	1850
shood	331	437	1912

4. Experiment 2: 8-month-olds

In Experiment 1 we showed that at 12-months infants begin to relate tapped variants in the *-ing* context to [d]-final words, not [t]-final words. To establish the earliest age at which infants begin to relate tapped variants with [d]-final words, we familiarized English learning 8-month-olds with words suffixed with *-ing* which had tap variants, but tested them on isolated words that ended in the stop consonants, as we did with 12-month-olds in Experiment 1. In word segmentation experiments, 8-month-olds are typically presented with three blocks of trials unlike older infants who are only presented with two (Jusczyk et al., 1999; Nazzi et al., 2005). To allow for comparison of our results to previously published experiments with 8-month-olds, we also presented 8-month-olds with three blocks of test trials. Recall that in previous work we have shown that infants at this age are able to relate suffixed nonce words like *babbing* and *kelling* with unchanging stems like *bab* and *kell* (Kim & Sundara, under review). If changes introduced into the stem because of alternations are challenging for infants, we expected 8-month-olds to fail to relate tapped variants produced in the context of *-ing* with stems produced with a stop.

4.1. Method

4.1.1. Participants

Twenty-four full-term, monolingual English-learning 8-month-olds participated in the experiment. All infants were familiarized with tap variants produced in the *-ing* context and tested on isolated [d]-final words produced with a full stop (mean age = 244 days; range 229:260; 12 girls). Subject inclusion criteria were exactly the same as in Experiment 1. Four additional infants were tested but their data were discarded because they did not complete the experiment due to fussiness.

4.1.2 Stimuli

The same stimuli were used as in Experiment 1

4.1.3 Procedure

The same as in Experiment 1. All the words in the familiarization passages were produced with taps, the test words ended in [d] and were produced with stops.

4.1.4 Analysis

The linear mixed effects model used to analyze the data was similar to that in Experiment 1. Log listening time was the dependent variable. The fixed effects included Condition (*cu[ɾ]ing* and *mee[ɾ]ing*, or *pa[ɾ]ing* and *shoo[ɾ]ing*) as a between-subjects variable, and Block (1, 2 or 3) and Trial-Type (potentially familiar vs. novel) as within-subjects variables, and all their interactions.

4.1.5 Results & Discussion

The mean listening times to potentially familiar and novel stems for 8-month-olds are also presented in Figure 2. The final model had a random intercept for Subject, and a random slope for Subject by Block. There was a trend towards the main effect of Block [$F(2, 22)= 2.8$, $p = 0.08$] such that listening times reduced over time, though none of the independent variables made a significant contribution to the model. That is, there is no evidence that English-learning 8-month-olds relate tapped variants in the *-ing* context to [d]-final stems. Thus, the presence of an alternation is a barrier to detecting morphological relatedness early in acquisition.

To further confirm that infants were able to detect morphological relatedness despite the presence of an alternation at 12- but not 8-months, we analyzed listening time data from Experiment 1 and 2 together. The final model included a random intercept for Subject. However, recall that 8-month-olds were presented three blocks of test trials, whereas 12-month-olds were presented with just two blocks. The interaction between Age and Trial-Type was significant whether we restricted analysis to just the first two blocks [$F(1, 330)= 6.0$, $p = 0.02$; 95% CI for power = 96:100] or included the third block for 8-month-olds [$F(1, 400.9)= 6.1$, $p = 0.02$]. In sum, these results show that the ability to relate variant forms emerges between 8- and 12-months.

5. General Discussion

We used corpus counts, acoustic analysis and behavioral experiments to determine if infants' learning of phonological alternations was constrained by the phonetic similarity of the segments involved. We showed that 8-month-olds familiarized with suffixed words with tap variants did not listen longer to isolated /d/-final stems produced with a stop release. Given infants previously reported success at relating suffixed words with unchanging stems (Kim & Sundara, under review; Willits et al., 2014), these results show that at 8 months, the presence of a phonological alternation is a barrier to detecting morphological relatedness.

By 12-months, when familiarized with suffixed words with tap variants, infants listened longer to potentially familiar stems produced with stop releases for /d/-final, but not /t/-final words. As shown in the corpus analysis, although the tapping rate for /d/ and /t/-final words is comparable, infants hear many more taps derived from /t/ than /d/; they also hear many more /t/-final words than /d/-final words. Additionally, there are at least two phonetic cues available to infants to distinguish taps derived from /d/ and /t/ — closure duration and preceding vowel duration. In fact, the taps in our familiarization passages were obtained from words derived from /t/; therefore, any potential bottom-up acoustic cues stemming from the incomplete neutralization of /t/ and /d/ as a result of tapping should have favored the /t/ interpretation. Thus, the distributional and phonetic evidence favored infants learning the tap-[t] alternation. Nonetheless, at 12-months, infants related tap and [d], not tap and [t]. The fact that English-learning infants failed to relate tap and [t] but successfully related tap and [d], despite greater distributional and (potentially) acoustic-phonetic support for the tap-[t] mapping, provides evidence that infants' learning of phonological alternations is biased. More specifically, infants' success relating tap-[d], but not tap-[t], supports a bias driven by the perceptual similarity of tap and [d].

Learning biases, like the one above, that make specific reference to phonetic information are often referred to as "substantive biases" (Wilson, 2006; see also Moreton & Pater, 2012 for a detailed discussion of such biases). One type of substantive bias holds that learners are biased against phonological processes requiring large perceptual changes (Steriade, 2001/2008). Evidence in support of such biases has been steadily accumulating in artificial language studies with adult (Skoruppa et al., 2011; White, 2014) and infant learners (White & Sundara, 2014), as well as in studies involving computational modeling (Peperkamp et al., 2006; White, 2017). The current study provides new evidence for such a bias in infant acquisition, and in particular, shows that infants are biased to prefer alternations between perceptually-similar segments while learning their first language. These results are consistent with computational models of phonological learning that assign greater prior likelihoods to alternations between perceptually similar sounds (e.g., White, 2017). More generally, this kind of a substantive bias exemplifies learners' domain-general propensity to cluster perceptually similar sounds into one category, in this case, one phoneme category. Note that not all proposed substantive biases have received the same level of empirical support from studies using a range of experimental methodologies with adults as well as infants (e.g., Jusczyk, Smolensky, & Allocco, 2002; Seidl & Buckley, 2005; for a review see Moreton & Pater, 2012).

These results also raise further questions about the phonological acquisition of the tapping alternation in English-learning infants. Given a perceptual bias against learning the alternation of phonetically distant segments like [t] and tap, how might infants eventually learn the alternation between word-final /t/ and tap in words that end with *–ing*? In other words, how do infants eventually learn that tap is not only derived from /d/, but can also be derived from /t/? One possibility is that infants merely need more input. It is possible that we tested infants at an age where they have just discovered the [d]-tap alternation, but have not yet recognized the [t]-tap alternation due to the similarity bias. If this is true, we might expect infants at a slightly older age to succeed at mapping tap to both /d/ and /t/ even for novel words, similar to an adult speaker. Another possibility is that the [t]-tap alternation can only be learned with lexical support, on an item-by-item basis. That is, once infants recognize that *ea[t]* and *ea[ɾ]ing* both refer to the same action, they might discover that *ea[t]* and *ea[ɾ]* are alternating variants of the same stem. If this is the case, infants should generalize the [t]-tap alternation to nonce words only after learning to relate multiple specific words ending in [t] and their suffixed counterparts, based on meaning. That is, we might expect infants to demonstrate their knowledge of the [t]-tap alternation in words that they know, before they are able to do so in the context of nonce words. Finally, it is possible that infants' learning of the [t]-tap mapping in English is complicated by the fact that /t/ in conversational speech can surface as a glottalized variant, including a full glottal stop, or it can even be entirely deleted, in addition to its tapped realization (Dilley et al 2019; Buckler, Goy & Johnson, 2018). This added variability might further contribute to the difficulty of learning the [t]-tap mapping. Future studies will investigate these possibilities for how development of the [t]-tap alternation unfolds.

Besides providing evidence for the constrained learning of phonological alternations, infants' success at mapping taps to [d], but not [t], also shows that English-learning infants are sensitive to voicing differences in coda consonants at 12 months of age. This findings adds to the scarce literature on infants' sensitivity to segmental contrasts in coda position (Archer, Zamuner, Engels, Fais & Curtin, 2016; Eilers, Wilson & Moore 1977; Jusczyk, Goodman & Baumann, 1999; Soderstrom, 2002; Swingley 2005; Zamuner, 2006).

In addition, our results show that infants map physically non-identical [d] and tap onto the same sound category by treating *pa[d]* and *pa[ɾ]*, for example, as alternating variants of the same stem. This finding provides evidence that infants develop abstract, phonemic representations by the end of the first year of life. Although phonemes are an intrinsic component of some models of developmental speech perception (e.g., PRIMIR: Werker & Curtin, 2005) and spoken word recognition (e.g., TRACE: McClelland & Elman, 1986) evidence that infants have abstract representations of speech sound categories in the first year of life is sparse.

How might infants arrive at an abstract representation of the speech signal? Based on their distribution alone, infants might link together variants that are conditioned by phonological context. For instance, English vowels are nasalized before nasal, but not oral, consonants. Thus, based on their complementary distribution across contexts, English-learning infants could infer that oral and nasal vowels are phonological variants (i.e., allophones) of the same abstract representation, typically called a phoneme (e.g., see Peperkamp et al., 2006; Seidl, Cristià, Bernard & Onishi, 2009). In one study that tests this hypothesis, Seidl et al. report that French- but not English-learning 11-month-olds are able to learn novel phonotactic dependencies between an oral or a nasal vowel and the following consonant. Their results are consistent with English-learning infants treating oral and nasal vowel as allophones of the same underlying phoneme (see also Seidl & Cristia, 2012). However, a recent meta-analysis (Cristia, 2018) shows that the effect size in Seidl et al (2009) and Seidl and Cristia (2012), as in other experiments where infants are taught phonotactic restrictions in an artificial grammar learning experiment, is close to 0. Thus, Seidl et al's results do not provide compelling evidence for the representation of phonemes. Our findings that infants can map perceptually distinct variants to one another, specifically those that are conditioned by morpho-phonological context, present strong, direct evidence for the representation of phonemes by 12-months of age.

Our results also contribute to the debate about how morphologically complex words are represented in the mental lexicon. Recall that although 12-month-olds failed to map [ɾ]+*ing* words to /t/-final words, they succeeded in mapping [ɾ]+*ing* words to /d/-final words. These results are not consistent with whole word models of the mental lexicon where infants learn to relate words based on the acoustic overlap of their beginnings (Rumelhart & McClelland, 1986; Plunkett & Marchman, 1993; McClelland & Patterson, 2002; most recently Baayen, Shaol, Willits, & Ramscar, 2015). Recall that 12-month-olds failed to relate *pa[ɾ]ing* to *pat* despite the greater acoustic similarity in our stimuli between the beginnings of the two sequences. Further, in whole word models, infants learn morphology as a by-product of learning overlapping forms and meaning. There is no extant evidence that infants between the ages of 8- and 12-month-olds have access to the meaning of words (typically verbs) suffixed with *-ing*.

Instead, we argue that infants' success at mapping taps to /d/- but not /t/- final words, coupled with their failure to map *dock* to *doctor,* indicates that they can segment the frequently occurring functional morpheme *–ing* (see also Willits et al., 2014). An examination of the 300,000-word Brent sub-corpus provides corroborating evidence that morphologically complex words with the -*ing* morpheme appear much more frequently (12,500 times) than even the most frequent content words like *mommy* (3,871 times). Such highly frequent sequences of sounds, whether they span word boundaries (Ngon et al., 2013), or are word internal (as shown here) are likely to be good candidates for the infants' mental lexicon (see also Marquis & Shi, 2012). Such an account is more consistent with morpheme-based models of the mental lexicon (e.g., Pinker & Ullman, 2002).

Regardless of how morphologically complex words are represented in the mental lexicon, infants' success at mapping [ɾ]+*ing* words to /d/-final words shows that the roots of morphological and alternation learning are in place at the end of the first year. In sum, we have shown that perceptual similarity constrains infants' distributional learning of phonological alternations in their first language. Based on these results, we can make a more general prediction about the developmental time course of phonological acquisition in languages with (near-)neutralizing patterns similar to tapping in American English. In any language, if two segments both alternate with a third, similar to the way that /t/ and /d/ each alternate with [ɾ] in American English, infants are predicted to learn the alternation that is more perceptually-similar first. Future studies with infants learning different languages are needed to test this prediction.

Acknowledgements

We would like to thank Robyn Orfitelli, Anya Mancillas, Victoria Mateu and Rosie Mejia for help with recruiting and testing infants and Jeffrey Lamontagne for help with the corpus acoustics. This research was funded by a UCLA COR Faculty Research Grant and NSF BCS-0951639 to MS.

References

Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech, 46*(2-3), 155–182.

Archer, S. L., Zamuner, T., Engel, K., Fais, L. & Curtin, S. (2016). Infants' discrimination of consonants: Interplay between word position and acoustic saliency. *Language Learning and Development, 12*(1), 60-78.

Baayen, H., Shaoul, C., Willits, J., & Ramscar, M. (2015) Comprehension without segmentation: A proof of concept with naive discrimination learning. *Language, Cognition, and Neuroscience 31*(1), 106-128.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language & Cognitive Processes, 24*, 527–554.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory & Language, 68*(3), 255-278. 10.1016/j.jml.2012.11.001.

Bates, D., Mächler, M., Bolker, B., & Walker S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, *67*(1), 1-48. doi: 10.18637/jss.v067.i01.

Boersma, P., & Weenink, D. (2013). *Praat: doing phonetics by computer* [Computer program]. Version 5.3.77, retrieved 18 May 2014 from http://www.praat.org/

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy, 2*(1), 29-49.

Braver, A. (2014). Imperceptible incomplete neutralization: Production, non-identifiability, and non-discriminability in American English flapping. *Lingua*, *152*, 24–44.

Brent, M., & Siskind, J. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81*(2), 31-44.

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

Calamaro, S., & Jarosz, G. (2015). Learning General Phonological Rules from Distributional Information: A Computational Model. *Cognitive Science, 39*(3), 647-666. doi: 10.1111/cogs.12167.

Buckler, H., Goy, H., & Johnson, E. K. (2018). What infant-directed speech tells us about the development of compensation for assimilation? *Journal of Phonetics*, *66*, 45-62

Chambers, K., Onishi, K., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition, 87*(2)*,* B69–B77.

Charles-Luce, J. (1997). Cognitive factors involved in preserving a phonemic contrast. *Language and Speech*, *40*(3)*,* 229-248.

Cristia. A. (2018). Can infants learn phonology in the lab? A meta-analytic answer. *Cognition, 170*, 312-327.

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology, 52*(4), 521-536. http://dx.doi.org/10.1037/dev0000083

de Jong, K. (1998). Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics, 26*(3), 283-310.

de Villiers, J. G., & de Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research, 2*(3), 267-278.

Dilley, L., Gamache, J., Wang, Y., Houston, D. & Bergeson, T. (2019). Statistical distribution of consonant variants in infant-directed speech: Evidence that /t/ may be exceptional. *Journal of Phonetics, 75,* 73-87.

Eilers, R. E., Wilson, W. R., & Moore, J. R. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech & Hearing Research, 20,* 766–780.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J., Pethick, S. J., & Reilly, J.(1993). *MacAurhur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing.

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), 1-185.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477-501.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, 99*(24), 15822-15826.

Fisher, W. M., & Hirsh, I. J. (1976). Intervocalic flapping in English. *Chicago Linguistic Society*. In *Papers from the Twelfth Regional Meeting of the Chicago Linguistic Society*, 183–198.

Fox, R. A., & Terbeek, D. (1977). Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics, 5*, 27-34.

Golinkoff, R. M., Hirsh-Pasek, K., & Schweisguth, M. A. (2001). A reappraisal of young children's knowledge of grammatical morphemes. In J. Weissenborn & B. Höhle (Eds.), *Approaches to Bootstrapping: Phonological, Syntactic, and Neurophysiological Aspects* *of Early Language Acquisition* (pp. 167-188). Amsterdam & Philadelphia: John Benjamins.

Green, P. and MacLeod, C.J. (2016), SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*: 493-498. https://doi.org/10.1111/2041-210X.12504

Hayes, B., & White, J. (2015). Saltation and the P-map. *Phonology, 32*(2), 1–36.

Herd, W., Jongman, A., & Sereno, J. (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics, 38*(4)*,* 504-516.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm reveals emergent language comprehension. In D. McDaniel & C. McKee (Eds.), *Methods for assessing children's' syntax*. Cambridge, MA: MIT Press.

Hohne, E. A., & Jusczyk, P.W. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception & Psychophysics*, *56*(6), 613-623.

Huff, C. T. (1980). Voicing and flap neutralization in New York City English. *Research in Phonetics*, *1*, 233-256.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound pattern of words in fluent speech. *Cognitive Psychology, 23*, 1-23.

Jusczyk, P. W., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*(3), 159-207.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*(8), 1465-1476.

Jusczyk, P.W., Smolensky P., & Allocco, T. (2002). How English-learning infants respond to markedness and faithfulness constraints. *Language Acquisition, 10*(1), 31-73.

Joos, M. (1942). A phonological dilemma in Canadian English. *Language, 18,* 141–144.

Kahn, D. (1980). *Syllable-based generalizations in English phonology*. New York: Garland Press.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior & Development*, *18*(1), 111-116.

Kim, Y., & Sundara, M. (under review). 6-month-olds are sensitive to English morphology. Under review.

Kiparsky, P. (1982). *Explanation in phonology*. Dordrecht: Foris. Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). *Estimated Marginal Means, aka Least-Squares Means*. R package version 1.4.6. https://cran.r-project.org/web/packages/emmeans/index.html

Kuznetsova A., Brockhoff P.B., Christensen R.H.B. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*, **82**(13), 1–26. doi: 10.18637/jss.v082.i13. Luke, S.G. (2017) Evaluating significance in linear mixed-effects models in R. *Behavioral Research Methods, 49***,** 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk*. 3rd Edition. Vol. 2: The database. Mahwah, NJ: Lawrence Erlbaum Associates.

Malécot, A., & Lloyd, P.M. (1968). The /t/-/d/ distinction in American alveolar flaps. *Lingua, 19*, 264-272.

Marquis, A., & Shi, R. (2012). Initial morphological learning in preverbal infants. *Cognition*, *122*(1), 61-66.

Mattys, S. L., & Jusczyk, P. W. (2001). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance, 27*(3), 644-655.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101–B111.

McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86.

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, *6*(11), 465-472.

Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.

Mintz, T. H. (2013). The segmentation of sub-lexical morphemes in English-learning 15-month-olds. *Frontiers in Psychology*, *4*(24), 1-12.

Moreton, E., & Pater, J. (2012). Structure and substance in artificial- phonology learning, part II: Substance. *Language and Linguistics Compass*, *6*(11), 702-718.

Nazzi, T., Dilley, L. C., Jusczyk, A. M., Shattuck-Hufnagel, S., & Jusczyk, P. W. (2005). English-learning infants' segmentation of verbs from fluent speech. *Language and Speech*, *48*(3), 279-298.

Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., & Peperkamp, S. (2013). (Non)words, (non)words, (non)words: Evidence for a proto-lexicon during the first year of life. *Developmental Science, 16*, 24-34.

Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-57.

Oshika, B. T., Zue, V. W., Weeks, R. V., Neu, H., & Aurbach, J. (1975). The role of phonological rules in speech understanding research. *IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-23*, 104-112.

Patterson, D., & Connine, C. M. (2001). Variant frequency in flap production. *Phonetica*, *58*(4), 254-275.

Pegg, J. E., & Werker, J. F. (1997). Adult and infant perception of two English phones. *The Journal of the Acoustical Society of America*, *102*(6), 3742-3753.

Peperkamp, S., & Dupoux, E. (2002). Coping with phonological variation in early lexical acquisition, in: I. Lasser (ed.) *The Process of Language Acquisition*. Berlin: Peter Lang Verlag. 359-385.

Peperkamp, S., Pettinato, M., & Dupoux, E. (2003). Allophonic variation and the acquisition of phoneme categories. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development* (Vol. Volume 2, pp. 650 – 661). Sommerville, MA: Cascadilla Press.

Peperkamp, S., Le Calvez, R., Nadal, J.-P., & Dupoux, E. (2006). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition, 101*(3)*,* B31–B41.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*(11), 456-463.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, *48*(1), 21-69.

Port, R. (1976). *The Influence of Speaking Tempo on the Duration of Stressed Vowel and Medial Stop in English Trochee Words*. PhD thesis, University of Connecticut.

R Development Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Posner, M.I., & Keele, S.W. (1968). On the genesis of abstract idea. *Journal of Experimental Psychology*, *77*, 353-363.

Rumelhart, D. E., & McClelland, J.L. (1986). On learning the past tense of English verbs. In McClelland and Rumelhart, Eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27-52.

Rosch, E., & Mervis, C.B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294)*,* 1926–1928.

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition, 69*(2), 105-134.

Sharf, D. (1962). Duration of post-stress intervocalic stops and preceding vowels. *Language and Speech, 5*, 26-30.

Seidl, A., & Buckley, E. (2005). On the learning of arbitrary phonological rules. *Language Learning and Development*, *1*(3-4), 289-316.

Seidl, A., Cristià, A., Bernard, A., & Onishi, K. H. (2009). Allophonic and phonemic contrasts in infants' learning of sound patterns. *Language Learning and Development*, *5*(3), 191-202.

Seidl, A., & Cristia, A. (2012). Infants' learning of phonological status. *Frontiers in Psychology*, *3(*448), 1-10.

Skoruppa, K., Lambrechts, A., & Peperkamp, S. (2011). The role of phonetic distance in the acquisition of phonological alternations. In S. Lima, K. Mullin, & B. Smith (Eds.), *Proceedings of the 39th North Eastern Linguistics Conference* (pp. 717–729). Somerville, MA: Cascadilla Press

Soderstrom, M., Wexler, K., & Jusczyk, P. W. (2002). English-learning toddlers' sensitivity to agreement morphology in receptive grammar. In B. Skarabela, S. Fish & A. H.-J. Do (Eds.), *Proceedings of the 26th annual Boston University Conference on Language Development* (pp. 643-652). Sommerville: Cascadilla Press.

Steriade, D. (2001/2008). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In S. Inkelas & K. Hanson (Eds.), *The Nature of the Word: Studies in Honor of Paul Kiparsky* (pp. 151–180). Cambridge: MIT Press. [Published in 2008. Originally circulated as ms. in 2001.]

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, *8*, 432–443. doi:10.1111/j.1467-7687.2005.00432.x

Sundara, M., & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics, 39*(4), 505-513.

Turk, A. (1992). The American English flapping rule and effect of stress on stop consonant duration. *Working Papers of the Cornell Phonetics Laboratory, 7,* 103-133.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development, 1*(2), 197-234.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49-63.

White, K. S., Peperkamp, S., Kirk, C., & Morgan, J. L. (2008). Rapid acquisition of phonological alternations by infants. *Cognition*, *107*(1), 238-265.

White, J. (2017). Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Language, 93*(1), 1–36.

White, J. (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition*, *130*(1), 96-115.

White, J., & Sundara, M. (2014). Biased generalization of newly learned phonological alternations by 12-month-old infants. *Cognition, 133(1),* 85-90.

Willits, J., Seidenberg, M., & Saffran, J. R. (2014). Distributional structure in language: Contributions to noun-verb difficulty differences in infant word recognition. *Cognition, 132*, 429-436.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*(5), 945-982.

Zamuner, T.S. (2006). Sensitivity to word- final phonotactics in 9- to 16- month- old infants. *Infancy*, *10*(1), 77-95.

Zue, V. W., & Laferriere, M. (1979). Acoustic study of medial /t, d/ in American English. *Journal of Acoustical Society of America, 66*(4), 1039-1050.