Journal of Phonetics 100 (2023) 101267

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Phonetic variation in English infant-directed speech: A large-scale corpus analysis

Ekaterina A. Khlystova [a], Adam J. Chong [b], Megha Sundara [a,*]

[a] University of California Los Angeles, Los Angeles, 90095 CA, USA
[b] Queen Mary University of London, Mile End Road, London E14NS, United Kingdom

ARTICLE INFO

Article history:
Received 10 November 2022
Received in revised form 14 August 2023
Accepted 17 August 2023

Keywords:
Phonetic categories
Allophones
Corpus
Morphology
Acquisition
Pronunciation variants

ABSTRACT

Learning sound categories is central to language acquisition – but we know little about the extent of phonetic variability in the learner's input. In this study, we phonetically annotated coronal segments (/t/, /d/, /s/, /z/, and /n/) in a corpus of naturalistic American English infant-directed speech (IDS). We did not find evidence that IDS is consistently more canonical than adult-directed speech (ADS), challenging the notion of IDS as a learning register. While IDS is not more canonical than ADS overall, the canonical form was nonetheless the most frequent form in IDS for all segments except /t/. We also considered how infants may move beyond the task of identifying the canonical form to how they may learn to cluster allophones; for this purpose, we quantified the dissimilarity in the phonological environments of the variants in question. Lastly, we investigated a case in which the overwhelming majority of instantiations were *not* canonical – word-final *t* and *d* – and demonstrated that morphologically-conditioned suffixes were more canonical than other word final segments. This corpus is a vital step towards understanding how infants can learn to categorize sounds from their input and will be an invaluable tool for future sociolinguistic, computational and theoretical modeling of language learning.

1. Introduction

Typically developing children learn their native language(s) at a spectacular rate – most children are highly competent users of their language by the time they are 5–6 years old. The magnitude of this tremendous feat is highlighted by the fact that the linguistic input children hear is variable, contains overlapping sound categories, and frequently demonstrates semantic or syntactic ambiguity. One of the first tasks with which a child is faced is learning the sound systems of their language(s). This task is vital, as these systems serve as the foundational building blocks for the acquisition of morphology, syntax, semantics, and pragmatics. Furthermore, the acquisition of speech sound categories entails learning not only prototypical variants, but also the less typical ones. Because this mapping is language-specific, all of this must be learned solely from their linguistic input.

We know from research on adult-directed speech (ADS) that some types of phonetic variation are predictable and positionally dependent. For instance, in North American English, /t/ is aspirated in word-initial position, as in "**t**ip" [tʰɪp]. In word-medial position, /t/ is often tapped intervocalically, as in "bu**tt**er" [bʌɾə˞], while word-finally it typically occurs as a glottal or glottalized stop as in "ca**t**" [kæʔ]. In addition to these position-specific variants, /t/ can also be unreleased, as in "firs**t** time" [fɪɹst̚taim], or it can be released and unaspirated, as in "s**t**op" [stap]. Other variation, such as that arising from differences in rates of speech – including segment deletion (Bell et al., 2003; Johnson, 2004), vowel reduction (Dalby, 1986; Patterson, LoCasto, & Connine, 2003) and voicing assimilation (Ernestus, Lahey, Verhees, & Baayen, 2006; Snoeren, Hallé, & Segui, 2006) – is less predictable. While variation, especially of /t/, is extensively studied in adult speech (e.g., Pitt, Dilley, & Tat, 2011), and particularly in sociolinguistics (e.g., Bybee, 2000; Guy, 1980, 1991; Labov, Cohen, Robins, & Lewis, 1968; MacKenzie & Tamminga, 2021; Tamminga, 2016), relatively little is known about the extent of phonetic variation for other segments. In this paper, we evaluated how variable alveolar coronals, some of the most common segments of English (Carterette & Jones, 1974; Denes, 1963; Tobias, 1959), are in the everyday speech directed to infants.

[*] Corresponding author.
E-mail address: megha.sundara@humnet.ucla.edu (M. Sundara).

https://doi.org/10.1016/j.wocn.2023.101267
0095-4470/© 2023 The Author(s). Published by Elsevier Ltd.

1.1. Infant-directed speech as a learning register

Infant-directed speech, or IDS, has long been described as a learning register. In early descriptions, IDS (also known as "baby talk" or "motherese") was reported to have beneficial modifications for learning. These modifications include more canonical segments compared to ADS and hyperarticulation of corner vowels, resulting in decreased overlap between vowel categories (Bernstein Ratner, 1984; Burnham, Kitamura, & Vollmer-Conna, 2002; Ferguson, 1964; Kuhl et al., 1997). Such segmental modifications in IDS have been argued to facilitate language acquisition (Eaves, Feldman, Griffiths, & Shafto, 2016; Ferguson, 1964). Consistent with this idea, Dilley, Millett, McAuley, and Bergeson (2014) found that IDS contained more canonical forms than ADS when examining regressive place assimilation of alveolar stops in recordings of parents reading to their children. Similarly, Fritche, Shattuck-Hufnagel, and Song (2021) also report more canonical productions of /t/ when mothers were reading to their children. However, read speech has been shown to be intentionally slower and clearer than natural speech (Ludusan, Mazuka, & Dupoux, 2021).

The characterization of IDS as clear speech is consistent with documented speech rate differences between IDS and ADS. Cross-linguistically, IDS is slower than ADS (Fernald & Simon, 1984; Bernstein-Ratner, 1984; Tang & Maidment, 1996; Sjons, Hörberg, Östling, & Bjerva, 2017; among others). Because more reduction and deletion processes tend to occur at faster speech rates (Johnson, 2004; among others), IDS, with its slower speech rate, is likely to involve fewer reduction and deletion contexts, thus making it more canonical overall.

Nonetheless, the enhancements typically cited as beneficial modifications in IDS tend to be unreliable. For instance, increased formant distance between vowels has been reported for some but not all vowel pairs. Cristia and Seidl (2014) examined tense and lax vowels in IDS and ADS in English (e.g., /i-ɪ/), and found little evidence for hyperarticulation of vowels differing in tenseness, although corner vowels were more separated as has been previously reported. In other studies, ADS has been reported to be just as canonical as IDS. In an analysis of a naturalistic speech corpus, Lahey and Ernestus (2014) found that IDS contains as many non-canonical segments as ADS. However, this comparison was conducted only for 2 lexical items. Similarly, Buckler, Goy, and Johnson (2018) examined place assimilation in read speech directed to English-learning 18-month-olds, recorded in the lab like Dilley and colleagues (2014), but found as many non-canonical realizations of words in place assimilation contexts in IDS as in ADS. Finally, there are also reports that when compared to ADS, IDS is less canonical, or even hypoarticulated. This includes comparisons in English (Shockey & Bond, 1980), Japanese (Martin et al., 2015; Miyazawa, Shinya, Martin, Kikuchi, & Mazuka, 2017), as well as Danish (Englund, 2018).

In sum, it is unclear whether the everyday speech addressed to infants is more canonical than adult-directed speech. In Study 1, using a phonetically annotated corpus based on home recordings of speech addressed to infants, we evaluated whether IDS is more canonical than ADS. We did this for English /t/, a segment whose variation is well-documented in ADS. Additionally, we also evaluated the extent of variation for /d/, /n/, /s/ and /z/, some of the most frequent segments in English.

1.2. Identifying the canonical variant

Further complicating the role canonical instances play in the acquisition of sound categories, there is often no clear consensus as to which variant is canonical. This is most evident in the case of English /t/. Although it is generally accepted that canonical /t/ has a distinct stop closure and release, there are mixed views on the underlying specification with regards to aspiration. Some have argued that English voiceless stops are underlyingly [-aspirated] (Odden, 2005), while others adopt released, [+aspirated] as canonical (Vaux, 2002), and still others treat both [±aspirated] variants as canonical (Dilley, Gamache, Wang, Houston, & Bergeson, 2019).

Despite the lack of consensus about which variant is canonical, experimental research on word recognition in adult listeners has shown that some variants – the ones typically referred to as canonical – are privileged. Sumner and Samuel (2005), for example, found long distance priming effects for the canonical [t] (released, unaspirated, non-glottalized) word-final variants of /t/, but not for the glottalized or glottal stop variants that are most frequent in this position. Similarly, Ranbom and Connine (2007) examined nasal flaps in words containing word-medial /nt/ (as in "wi**nt**er", which can be produced [wɪn̦ɛɹ]) and found that words produced with the 'canonical' [nt] (canonical [n] with an aspirated [t^h]) pronunciation were significantly more likely to be identified as words than those with the nasal flap ([ɾ̦]) pronunciation, even when controlling for differences in stimulus length. Lastly, Pitt et al. (2011) found that words pronounced with the canonical [t] (released or unreleased, aspirated or unaspirated [t]) were identified with equally high accuracy as position-specific variants such as taps and glottal stops (and with higher accuracy than inappropriate variants). Accuracy was high for canonical [t] words even in contexts in which [t] is not the most frequent variant.

Taken together, this line of research shows that canonical variants have a processing advantage in adult word recognition. Unlike other variants that are favored in specific positions, canonical variants are often recognized as well as, if not better than, position-specific variants, even in contexts where they are not the most likely to be heard. In Study 2, we evaluated whether the variants described as canonical are the most frequent variants in infants' input.

1.3. Discovering positional allophony

Identifying the canonical variant is not the only learning problem infants face. At least some phonetic variation is obligatory and contextually-determined. For example, if a /t/ occurs in between two vowels in which the first is stressed and the second is unstressed, a North American English speaker will typically produce a tap. This is not the case for speakers of other dialects of English, for instance British or Indian English. Thus, in addition to identifying the canonical variant from a set of variants, in traditional accounts, infants must also uncover when and where other variants surface, while distinguishing them from the less intentional results of reduction and assimilation at high rates of speech. Because at least some contextual variation is language-specific, infants must rely on their input to discover it.

Extant developmental research challenges the traditional account where infants gradually discover contextual variants. There is evidence that infants are sensitive to contextual variation in pronunciation – that is, allophones – very early in life. English-learning 2-month-olds are able to discriminate between the /t/ in “night rate” [naɪ**ʔ** reɪt] versus “nitrate” [naɪ**t͡ʃ**reɪt] (Hohne & Jusczyk, 1994). By 10.5 months, they are even able to use this contextual, allophonic variation to segment words (Jusczyk, Houston, & Newsome, 1999). Experimental research, however, shows that the ability to group these variants into phonemes emerges later; it is in place only by the end of the first year (Seidl, Cristià, Bernard, & Onishi, 2009; Sundara, White, Kim, & Chong, 2021).

Consistent with these results, in more recent proposals, infants’ early sound categories are characterized, at best, as context-sensitive (Feldman, Griffiths, & Morgan, 2009; McMurray & Aslin, 2005; Port, 2007), which they subsequently combine to construct abstract phonemes (e.g., Peperkamp, Le Calvez, Nadal, & Dupoux, 2006; Swingley, 2009; Feldman et al., 2009; see also Martin, Peperkamp, & Dupoux, 2013). In one such proposal, infants construct phonemes bottom-up by clustering variants based on their complementary distribution in the input (Peperkamp et al., 2006; Martin et al., 2013; see also Hitczenko & Feldman, 2022). In Study 3, we quantify the extent to which distributions of variants in IDS are in complementary distribution in order to generate hypotheses about how infants might begin to construct phonemes from their input.

1.4. Effects of morphology on the occurrence of variants

In addition to phonological context, morphological structure also has an impact on the occurrence of pronunciation variants. Some of these effects are categorical, and others are more gradient. For instance, differences in morphological structure have been shown to affect the duration of specific segments, although these durational correlates are inconsistent. For example, the duration of /s/ and /z/ has been reported to be longer in monomorphemic cases (e.g., *freeze*) compared to cases in which it is a suffix (e.g., *free-s*) in distinct homophones (Plag, Homann, & Kunter, 2017; Schmitz, Baer-Henney, & Plag, 2021; Tomaschek, Plag, Ernestus, & Baayen, 2021; Zimmermann, 2016). Contradictorily, the duration of /s/ and /z/ has also been reported to be consistently longer when it is suffixed as compared to /s/ and /z/ in monomorphemic words (Seyfarth, Garellek, Gillingham, Ackerman, & Malouf, 2018). Similarly, the nasal in the English suffix *un-* or *in-* has been reported to be longer in words where it is a prefix (Ben Hedia & Plag, 2017).

Morphological structure has also been documented to influence the specific variant produced. For example, in North American English, coronal stops in word-final clusters in monomorphemic words like *mist* are deleted more often than word-final suffix [t, d] as in *missed* (e.g., Labov et al., 1968; Guy, 1980, 1991; Bybee, 2000; MacKenzie & Tamminga, 2021; but see also Seyfarth et al., 2018). Similarly, word-medial voiceless stops are more likely to be aspirated if the word is morphologically complex (e.g., with a prefix, *disclaim*) compared to when it is not (e.g., *discover*) (Baker, Smith, & Hawkins, 2007; Smith, Baker, & Hawkins, 2012).

Given the mixed findings that morphological structure can affect variant realization, in Study 4 we examined whether morphological status also affects which specific variants of /t/ and /d/ are produced in infant-directed speech. Crucially, whether morphological structure affects phonetic realization has implications for the architecture of speech production models. In traditional feed-forward accounts of speech production (Chomsky & Halle, 1968; Kiparsky, 1982; Levelt & Wheeldon, 1994; Levelt, Roelofs, & Meyer, 1999) morphological encoding is inaccessible at the point of phonetic production. In other proposals, there is more direct interaction of morphology and phonetics, either mediated by prosodic structure in relatively constrained ways (for example, Booij, 1983; Nespor & Vogel, 2007), or via more extensive interactions between morphology and phonetics (see Bybee, 2001; Gahl & Yu, 2006; Goldinger, 1998; Pierrehumbert 2001, 2002).

1.5. The present study

In the present study, we used a phonetically annotated corpus of IDS to answer four questions about the nature of variation in IDS. In Study 1, we evaluated phonetic variation in two corpora of naturalistic speech to determine whether IDS is more canonical than ADS for some of the most frequent segments in English: /t/, /d/, /s/, /z/, and /n/. Next, in Study 2, we investigated whether the variants that have been hypothesized as canonical for these segments are indeed the most frequent variant in the IDS corpus. In Study 3, we quantified the extent to which variants of /t/ and /d/ in IDS are in complementary distribution. We did this in order to generate hypotheses about how infants might cluster distributionally distinct variants to construct phonemes. Lastly, in Study 4, we asked how (if at all) morphological structure affects variation in word-final segments, focusing on /t/ and /d/.

To answer these questions, we transcribed ~6500 utterances from the Providence Corpus (Demuth, Culbertson, & Alter, 2006) to quantify the degree of variation present in alveolar coronals, some of the most frequent segments in English (Carterette & Jones, 1974; Denes, 1963; Tobias, 1959). The Providence Corpus is longitudinal and consists of home audio and video recordings of parent–child interactions with 5 monolingual English-learning children during everyday activities. For each of these parent–child dyads, we sampled recordings at two ages for phonetic transcription: 16–18 months and 22–24 months. Data from both ages are combined here.

In doing so, we have compiled one of the largest corpora of phonetically transcribed utterances in IDS to date. This is particularly important as documenting the extent of allophonic variation in naturalistic IDS is critical in order to make future theoretical and computational modeling of phonological acquisition ecologically valid. We used this corpus to determine how variant forms of the coronal segments are distributed in naturalistic IDS, the contexts in which they are observed most often, and the extent to which this positional variation is predictable. From this analysis, we can begin to chart how and what infants can learn about and from positional variation in their IDS input.

2. Study 1: Is IDS more canonical than ADS?

In Study 1, we evaluated whether IDS is more canonical compared to ADS. Ideally, a comparison of IDS and ADS would be based on speech produced in the two registers by the same individual in similar settings. The IDS in this study was based on phonetically transcribed, home recordings from the Providence Corpus (Demuth et al., 2006). Although there was some speech between adult caretakers (i.e., ADS) in the Providence corpus, it was not sufficient to allow a statistical comparison between the two (~50 ADS utterances out of the ~6500 IDS utterances included in the final analysis). For this reason, we sought an alternate ADS corpus for comparison.

Our choice of the ADS corpus was limited because although there are several phonetically annotated corpora of adult speech, few involve naturalistic speech, and none completely match the dialect of the speakers in the Providence Corpus. Out of 5 parents in the Providence Corpus, 3 are described as speaking Mainstream American English (MAE) and two parents are described as speaking a variety of the New England (NE) dialect (Song, Sundara, & Demuth, 2009). Because 3 out of the 5 mothers in the Providence Corpus spoke what was described as MAE, we chose the Buckeye Corpus to sample ADS (Dilley & Pitt, 2007).

The Buckeye Corpus consists of connected speech from 40 different adult speakers, also described as speaking MAE. This includes men and women under 30 as well as over 40 years old. We sampled the first 5 female speakers under 30 from the Buckeye corpus because they are most similar in age and gender to the mothers in our IDS corpus. While speakers in both corpora have been described as speaking MAE, there are undoubtedly differences between MAE dialects spoken in Columbus, Ohio, and Providence, RI. However, there are no documented reports of differences between coronal segments in MAE spoken in Ohio and Providence. To enable a comparison to the ADS corpus, and in keeping with the phonetic transcription in the Buckeye Corpus (Dilley et al., 2019), we treated [±aspirated] stop variants as canonical in this study.

Recall that two of the remaining speakers in the Providence Corpus were described as speaking a NE dialect. The NE dialect of American English differs from MAE in two ways: vocalization of /ɹ/ (as in /kaɹ/ → [ka] 'car') and a possible low back vowel merger that is currently in progress (Labov, Ash, & Boberg, 2008). In our corpus as well, the two NE speakers deleted and/or produced more /ɹ/s as vowels than the three MAE speakers. Although there are no documented reports of differences specific to coronal segments between MAE and NE dialects, the phonological environments in which some coronal segments occur are likely to be different due to the vocalization of /ɹ/. Because variants are likely to be governed by the phonological environment in which they occur, difference in the vocalization of /ɹ/ between MAE and NE dialects could affect our analyses here as well as in Study 3, where we characterize the distributions of variants of /t/ and /d/ (Section 5). To ascertain how much (if at all) our findings were affected by including speakers of the NE dialect, we ran all relevant analyses excluding the two parents reported to speak NE English. These analyses can be found on our OSF page. We mention these results explicitly in this paper only if excluding the two NE speakers changed the pattern of results.

2.1. Methods

2.1.1. The infant-directed speech corpus

First, we identified each mother's utterance in the selected samples from the Providence corpus containing the target segments. This was done by extracting any utterances in the orthographic transcript that were coded as the mother's utterances and contained "t", "d", "s", "z", or "n", since the orthographic symbols of these segments correspond almost exclusively to their phonemic equivalents. Utterances that contained 'th' were initially extracted, but these were not transcribed (since these correspond to either [θ] or [ä]) unless they contained one of the target coronals. The time points for each utterance were then used to extract the relevant portion of the audio recording, which was then force-aligned using the Forced Alignment and Vowel Extraction program suite (FAVE; Rosenfelder, Evanini, & Prichard, 2014). This force-aligner uses an HTK Toolkit for phonetic alignment, referencing the CMU Pronouncing Dictionary to transform orthographic transcription into phonemic notation. Altogether, this yields a set of Praat (Boersma & Weenink, 2013) TextGrids containing a time-aligned, segmented phone (phonemic) tier and a word tier. Any segments on which FAVE failed because certain words were not in the pronunciation dictionary causing large misalignment errors were excluded and not annotated (~730 utterances or 9%).

2.1.1.1. Exclusions. Because we were interested only in naturalistic IDS, tokens were excluded from the analysis for the following reasons: mechanical/acoustic noise (such as microphone static or feedback); clearly adult-directed speech; reading or singing; and child vocalizations and speech. Further, contracted expressions, such as *"wanna"* or *"gonna",* were excluded due to the ambiguity of target forms, specifically, whether targets (and therefore segments) should be analyzed compositionally or not (i.e. target for "wanna" = "want to" or "wanna"). Finally, because the files were sampled using the corresponding orthographic symbols for each segment, some number of files were sampled that did *not* contain any of the target segments – e.g., an orthographic "t" could actually correspond to [θ], leading the file to be sampled, but ultimately excluded if there were no alveolar segments.

A total of 552 utterances and 19,035 additional tokens were excluded from the original transcripts for these combined reasons. The number of utterances analyzed in the final sample was 6662, with 94 165 total segments, including all vowels and consonants, annotated. While it may initially seem alarming that so many utterances were excluded from the analysis, a significant portion of the recorded input contained parents reading to their children, which we chose to exclude due to the fact that read speech is often intentionally slower and clearer than natural speech (Ludusan et al., 2021). We also excluded speech between adult caretakers (i.e., ADS), because they only constituted ~50 utterances in our sample and we were specifically interested in IDS (but see Shneidman & Goldin-Meadow, 2012; Shneidman, Arroyo, Levine, & Goldin-Meadow, 2013; and Weisleder & Fernald, 2013 for discussions of the role of overheard ADS in language acquisition).

2.1.1.2. Annotating the IDS corpus. The interval boundaries of all the force-aligned consonants were then checked on the text grid, corrected (when needed), and crucially, phonetically annotated by one of five trained research assistants, all native speakers of American English. In the final output, each segment has a phonemic form based on the CMU pronunciation dictionary (the automatic FAVE output), the surface form (the phonetic variant that was annotated by the phonetically trained coder), word position (initial, medial, or final), and surrounding segmental context. The surface variants that were annotated for each of the coronal segments analyzed in this paper are shown in Table 1. Annotators were given acoustic landmarks to help with perceptual transcription of categorical variants. The distinction between taps and deletions hinged on (a) the presence of a clear occlusion (b) and the percept. In the absence of a clear occlusion, for example in the cases where the intervocalic stops were produced as approximants (e.g., Warner & Tucker, 2011), transcription was based on perception. Annotation criteria for each of the variants coded and representative spectrograms for each variant are available on our OSF page, along with the annotation landmarks for all other consonants coded in the corpus.

A portion of the data (650 utterances or 9.7% of the final sample) was annotated by all 5 coders to evaluate reliability of phonetic annotation. Cross-coder reliability was determined using Fleiss's Kappa for each segment: for /t/, κ = 0.712 (p < 0.001); for /d/, κ = 0.757 (p < 0.001); for /s/, κ = 0.602 (p < 0.001); for /z/, κ = 0.682 (p < 0.001), and for /n/,[1] κ = 0.383 (p = 1).

Once the IDS corpus annotation was complete, the annotated segments were extracted using a custom-written Python program. This script, along with all other scripts used in the extraction and analysis of data, are available at our OSF page. The total number of analyzed coronals was 23 446 (out of the ~ 94 000 segments annotated).

2.1.2. The adult-directed Buckeye speech corpus

ADS was sampled from the phonemic and phonetic annotations of conversational speech from 5 young adult women (s01, s04, s08, s09, s12) in the Buckeye Corpus (Pitt, Dilley, Johnson, Kiesling, Raymond, Hume, & Fosler-Lussier, 2007). These annotations were extracted using the Python program provided with the corpus as well as a custom-written script available on the project OSF page and were then filtered for the coronal segments /t/, /d/, /s/, /z/, and /n/. Deletions were hand-aligned. This yielded 19 344 coronals analyzed from the Buckeye Corpus, out of a total of 44 493 segments.

For comparison purposes, because Buckeye collapses [tʰ], [t], and [t̚], these variants were also collapsed and considered canonical for this analysis of our IDS corpus. Similarly, for /d/, both [d] and [d̚] were both counted as canonical. Lastly, for /s/, /z/, and /n/, the canonical forms were [s], [z], and [n], respectively.

Table 1
Set of surface variants of coronal consonants annotated.

Phoneme	*Surface Variants*[a]
/t/	[tʰ] (aspirated), [t] (unaspirated), [ɾ] (tap), [ʔ] (glottal stop), [t̚] (unreleased),[g, k, m, n, p, s] (assimilated), [tʃ] (affricated), [∅] (deleted)
/d/	[d] (canonical), [ɾ] (tap), [d̥] (voiceless), [d̚] (unreleased), [dʒ] (affricated),[b, ä, n, s, ʒ] (assimilated), [∅] (deleted)
/n/	[n] (canonical), [ɾ̃] (nasalized tap), [m, ŋ] (assimilated), [∅] (deleted)
/s/	[s] (canonical), [ʃ] (assimilated), [∅] (deleted)
/z/	[z] (canonical), [z̥] (devoiced), [ʃ, ʒ] (assimilated), [∅] (deleted)

[a] Categories with fewer than 50 tokens were not included; glottalized tokens were collapsed with glottal stops.

2.1.3. Analysis

A logistic mixed effects model was used to analyze the log odds of canonical pronunciation; the final model had a random intercept for speaker and fixed effects of segment (/t/, /d/, /s/, /z/ or /n/), position in word (initial, medial or final), and register (IDS or ADS), and all two- and three-way interactions.

2.2. Results & discussion

The frequency of canonical forms for /t/, /d/, /s/, /z/ and /n/ across word positions, for both IDS and ADS, are presented in Table 2. We discuss the position effects first, followed by the register effects. Cross-linguistically, syllable onsets are less variable than codas (e.g., Beckman, 1998) – by extension, word-initial consonants are expected to be more canonical than word-final consonants. Consistent with this, there was a significant main effect of position, such that compared to word-final position there were more canonical pronunciations in initial (z = 22.6, p < 0.0001) and in medial position (z = 4.3, p < 0.0001). There was also a significant main effect of register, with more canonical productions in IDS than in ADS (z = 4.5, p < 0.0001). However, the three-way interaction between segment, position and register was also significant, indicating that the effect of register varied by segment and position; thus, we probed the 3-way interaction using *emmeans* (Lenth, 2022) to determine register effects for individual segments by position. We excluded /s/ when evaluating register effects, because in both registers more than 90% of the /s/ variants were canonical.

In the initial position where the log-odds of canonical instances were the highest, effects of speech register were minimal. Compared to ADS, IDS had significantly more canonical instances only for initial /d/ (z = 3.8, p = 0.0062), that is, 1 of the 4 segments. In medial position as well, register effects were limited: IDS had significantly more canonical instances only for 2 of the 4 segments: /t/ (z = 0.46, p = 0.01) and /n/ (z = 10.7, p < 0.0001). The register effects in the final position were the most variable. Word-finally, there were significantly more canonical instances in IDS for 2 of the 4 segments: /d/ (z = 4.5, p = 0.0003), and /n/ (z = 6.3, p < 0.0001); there was no difference between the registers for /t/; and significantly more canonical instances in ADS for /z/ (z = -3.73, p = 0.007). In a more conservative comparison where we restricted the analysis to only the 3 speakers from the Providence Corpus described as speaking MAE, we found even fewer differences

[1] There was complete agreement for 97.1% of tokens of /n/ – all of these tokens were faithful. However, Fleiss's Kappa weights those tokens outside of the majority more heavily than those in line with the majority outcome; these 5 tokens happened to be the only tokens on which complete agreement wasn't observed, leading to a low κ value.

Table 2
Frequency (raw counts and percent) of canonical forms.

Segment	*Position in word and register*							
	Initial		*Medial*		*Final*		*Overall*	
	ADS	*IDS*	*ADS*	*IDS*	*ADS*	*IDS*	*ADS*	*IDS*
/t/	1039 (73.8%)	1030 (81.0%)	1096 (52.9%)	1308 (64.4%)	1283 (23.5%)	1028 (24.8%)	3418 (38.2%)	3366 (45.2%)
/d/	929 (71.6%)	1090 (82.3%)	330 (42.5%)	273 (43.3%)	1011 (34.4%)	815 (49.1%)	2270 (45.3%)	2178 (60.3%)
/z/	29 (100.0%)	43 (95.6%)	250 (87.4%)	172 (88.7%)	1995 (81.9%)	1870 (74.6%)	2247 (82.5%)	2085 (75.9%)
/s/	1636 (98.6%)	1701 (98.9%)	1261 (95.0%)	497 (96.7%)	1235 (92.8%)	2207 (96.1%)	4132 (95.7%)	4405 (97.2%)
/n/	882 (97.1%)	861 (99.2%)	3204 (69.1%)	1886 (91.1%)	1794 (82.2%)	2065 (92.3%)	5880 (76.1%)	4812 (93.0%)

Percent of total was calculated out of the total number of segments in that position (e.g., all /t/'s in initial position).

between IDS and ADS (data available on the project OSF page). In this conservative analysis, only /n/ in medial position and /d/ and /n/ in final position had more canonical instances in IDS compared to ADS. Thus, we failed to find evidence that speech in IDS is more canonical than ADS across the board.

3. Study 2: Which is the most frequent variant in infant-directed speech?

Next, we turn to the question of how children may identify the canonical form from their input. This question is of interest given previous reports of an advantage for canonical forms in adult listeners such that these forms facilitate word recognition even when they are not presented in the contexts that typically license them (Section 1.2). Identifying the canonical form is likely to be challenging given the extent of variability in IDS overall, as well as across positions and segments. In this study, we examined if the processing advantage observed for canonical forms in adult listeners may arise because canonical forms are the most frequent in a child's input.

3.1. Method

In this study, we analyzed only the phonetically annotated IDS corpus described in Section 2.1. The methods here were the same as those of the previous study with the following exception: because we were interested in a fine-grained analysis of the variants present in IDS, and were no longer comparing to the Buckeye Corpus, we did not collapse unaspirated/aspirated and unreleased/released /t/ or unreleased/released /d/. Instead, we treated them as three separate categories: unreleased, aspirated, and faithful (meaning faithful to the phonetic symbol – that is, released and unaspirated).

3.1.1. Analysis

For each segment, we used multinomial logistic regression to identify the most frequent variant in IDS. Analyses were carried out using the *mblogit* function from the *mclogit* package in R (Elff, 2022); in addition to determining the most frequent variant overall, we also evaluated the most frequent variant in initial, medial and final position in a word by including position as a fixed effect with a random intercept for speaker. For the multinomial regression, any variants that comprised less than 0.1% of the total instances of the segment were collapsed into the "Other" category. Importantly, we did not collapse the aspirated and unaspirated tokens as we were interested in the relative frequencies of each of these categories. Full coefficients, standard errors, and p values for this analysis can be found on our OSF page.

3.2. Results & Discussion

The frequency of variants in IDS for each of the segments examined is shown in Fig. 1, where *faithful* refers to the surface form that is consistent with the phonetic symbol. Note that /s/ had the fewest non-canonical variants overall as we previously reported in Table 2 (<3%), so the canonical /s/ is the most frequent variant. We present the frequencies for /s/ in Fig. 1 only for completeness.

The most frequent variant of /t/ across word positions was the glottal stop ($p < 0.001$); the faithful form – that is, unaspirated, released [t] – was slightly more frequent than the aspirated variant ($p < 0.05$), and more frequent that all others (p's < 0.001). However, as expected, the most frequent variant differed by position. Word-initially, affricated and aspirated variants were most frequent (p's < 0.001) compared to faithful, which was more frequent than all other variants (p's < 0.001). In word-medial position, all variants were significantly less frequent than the faithful variant (p's < 0.001). Lastly, word-finally, deletions, glottal stops, taps, and unreleased variants were all significantly more common than faithful variants (p's < 0.001), while all others were significantly less common (p's < 0.001).

In contrast, the faithful variant of /d/ – the released [d] – was the most frequent variant overall, as well as word-initially. In word-medial position, taps were most frequent ($p < 0.001$), whereas deleted variants were most frequent word-finally ($p < 0.001$). For /z/ and /n/ as well, like for /d/, the faithful variant was the most frequent overall, and in every position.

In summary, canonical variants were indeed the most frequent variants of /s/, /z/, and /n/ in infant-directed speech. This was also the case for /d/ overall, although taps and deletions were more likely in medial and final position respectively. The variants of /t/, however, presented the most complex picture. Faithful (unaspirated, released) /t/ was not the most frequent variant overall; neither was the aspirated variant. Instead, glottal stops were the single most likely variant of /t/ in the corpus. We highlight the challenges this poses for infants' ability to identify the canonical variants in the General Discussion.

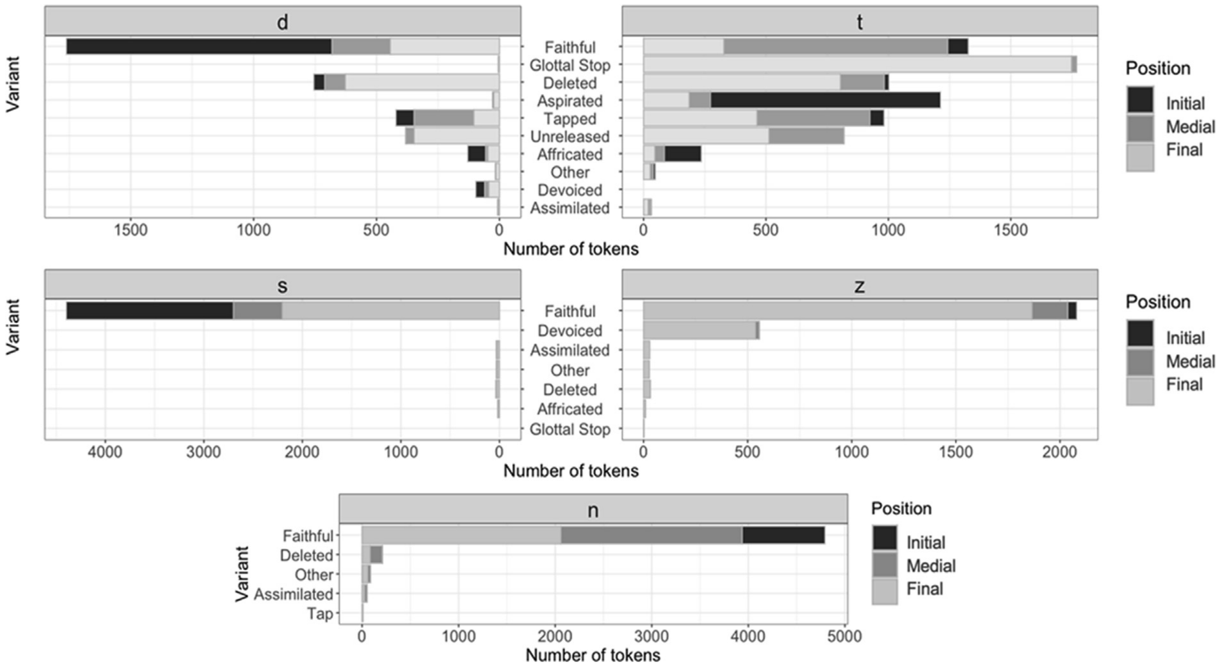

Fig. 1. Frequency of variants in IDS. Faithful refers to the surface form that is consistent with the phonetic symbol, with no reference to the underlying form.

4. Study 3: How might infants find phonemes?

Findings from Study 1 and Study 2 highlight the different variants present in IDS. The phoneme /t/ has the most variants, and it is the only coronal segment where the canonical variant is not the most frequent one. Its voiced counterpart /d/ is also variable, such that variants other than the canonical one are more frequent, at least in word medial and final position. Likely because of this variability, it is challenging for infants to relate even frequent variants like taps with the canonical, unaspirated stop variant /t/: they fail to do so even at 12-months, although they are able to relate taps with the perceptually more similar /d/ (Sundara et al., 2021). In Study 3, we investigated how infants might start to cluster variants of /t/ and /d/ to construct phonemes.

According to one proposal, infants use bottom-up distributional information about the phonological environments of variants to discover allophones, which they can then cluster to build phonemes (Peperkamp et al., 2006; Martin et al., 2013; see also Hitczenko & Feldman, 2022). In other words, variants with large differences in the contexts in which they occur are in complementary distribution and therefore likely to be allophones. In Study 3, we quantified the dissimilarity between the contexts in which each pair of variants of /t/ and /d/ occurs to explore this hypothesis.

Following Peperkamp and colleagues (Martin et al., 2013; Peperkamp et al., 2006), we quantified the dissimilarity between the probability distribution of two variants across a set of contexts using the Kullback-Leibler (KL) divergence metric (Kullback & Leibler, 1951). When the distributions of any two variants are identical, as in the case of phonemes that occur in minimal pairs, KL divergence is 0. When two distributions are dissimilar, KL divergence is greater than 0, with higher numbers for more dissimilar distributions. Therefore, when comparing the distribution of two variants, a higher KL divergence is consistent with a more complementary distribution of the variant pair.

In Study 3, we used the KL divergence measure to compare the dissimilarity of the contexts in which pairs of variants of /t/ and /d/ occur with a view to generating a developmental timeline for the discovery of the /t/ and /d/ phonemes. Under a bottom-up learning account, variants with larger KL divergence scores should be identified as allophones first because they have more complementary distributions.

4.1. Methods

Using the conventions of Pitt et al. (2011) as a starting point, we identified environments of /t/ that favor faithful [t], aspirated $[t^h]$, glottal stops [ʔ], taps [ɾ] and deletions. The phonological environments that favored [t] were those in which /t/ occurs before an unstressed vowel and is preceded by a voiceless stop consonant, voiceless fricative, or /l/. Environments expected to favor [ɾ] are those in which /t/ appears intervocalically, following a stressed vowel. Next, to identify environments that favor deletions, we followed both the conventions of Pitt et al. (2011) and the phonotactic rules from a well-known phonetics textbook (Ladefoged & Johnson, 2014): /t/ preceding an /n/, and those preceded and followed by a consonant, are deleted. Because our corpus is not annotated by syllables, we used the following proxy for environments where the aspirated form is favored (“Favors $[t^h]$”), based on Zuraw and Peperkamp (2015): word-initially, or when the preceding segment is an unstressed vowel and the following is a stressed vowel. Lastly, we use a combination of rules from Ladefoged and Johnson (2014) and Seyfarth and Garellek (2020) to identify environments that favor glottal stops [ʔ]: alveolar stops become glottal stops phrase-finally or in non-initial position when preceded by a vowel and followed by a sonorant. Variants of /t/ that occur in each of these environments were then isolated, and their distributions across environments were compared for each pair of allophones.

Environments that favor specific variants were also identified for /d/. Those in which /d/ occurs between two consonants were expected to favor deletion (Ladefoged & Johnson, 2014). As with /t/, environments expected to favor [ɾ] were those in which /d/ appears intervocalically, following a stressed vowel. All other environments were expected to favor [d].

4.1.1. Analysis

We calculated the symmetrical KL divergence for the contexts in which each pair of variants occurs to determine the extent to which their distributions are complimentary. These were calculated using the KL function in the *philentropy* package in R (Drost, 2018). In the context of our analysis, KL divergence was calculated for 10 pairs of variants, where the distribution of one variant across all possible environments (described above) was compared to the distribution of another variant in the same set of environments.

4.2. Results & Discussion

The matrices in Fig. 2 depict the observed distribution of variants within the specified environments for /t/ and /d/ across all positions. As we can see from the last column in the matrix on the left, for /t/, there were 821 environments favoring taps (the sum of the last column in the matrix), of which 605 were produced as taps, 41 were produced as glottal stops, 25 as [t], 52 were aspirated, and 98 were simply deleted.

Comparing two rows of Fig. 2 allows us to compare the distribution of a pair of variants produced by all speakers across the set of environments identified. Recall that KL divergence values index the extent of dissimilarity of these distributions. We discuss the results for /d/ first, then /t/.

The KL divergence scores were highest for deletions and taps (3.89), followed by taps and faithful /d/ (1.29), and were lowest for faithful variants of /d/ and deletion (1.11). Even when we restricted our analysis to only the 3 mothers described as speaking MAE, there was no change in the rank ordering of KL divergence values for /d/ variants. That is, this ranking is stable across MAE and NE dialects. The table containing the KL divergence for each allophone pair for just the 3 MAE speakers is available on our OSF page. If infants rely on the extent of complementary distribution across contexts, these findings suggest they should relate deleted and tap variants of /d/ as allophones (highest scores) before they relate faithful and tap variants of /d/.

The KL divergence values for variants of /t/ from all of the speakers in the corpus are presented in Table 3. From Table 3 we can see that the top 3 pairs of /t/ variants, with the highest KL divergence scores, are glottal stops and taps (6.37), followed by glottal stops and deleted variants (4.12), followed by deleted variants and taps (3.34). If infants rely on the extent of complementary distribution across contexts, they should relate glottal variants and taps as allophones before relating glottal variants and deletions, and only then relate deleted and tap variants. Again, there was no change in the rank ordering of the top 3 pairs even when we restricted our analysis to only the 3 mothers described as speaking MAE. The table containing the KL divergence for each pair for just the 3 MAE speakers can be found on our OSF page. We discuss further implications for learning, as well as empirical predictions for acquisition experiments, in the General Discussion.

5. Study 4: Does morphology affect variation?

Beyond phonological environments, there is some evidence that morphological structure may also impact which specific phonetic variant is produced. Because morphological complexity has been shown to affect both categorical and continuous acoustic properties of suffixes (Drager, 2011; Gahl, 2008; Sugahara & Turk, 2009; Zuraw, Lin, Yang, & Peperkamp, 2021), we conducted Study 4. We focused on word-final seg-

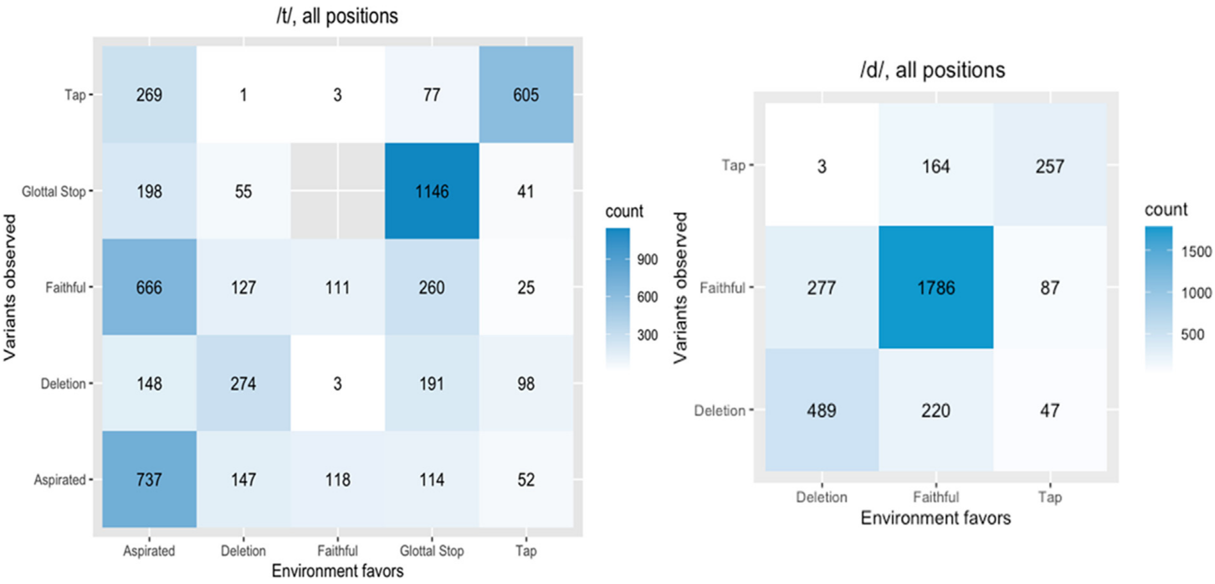

Fig. 2. Observed distributions of variants in favored environments for /t/ (left) and /d/ (right) – the x-axis denotes environments typically thought to favor a variant, and the y-axis denotes the possible variants. Note that in these matrices, “faithful” collapses both released and unreleased variants.

Table 3
KL divergence values for /t/ (ranked ordered).

Variant pair	*KL divergence*
Glottal stop - tap	6.37
Deleted - glottal stop	4.12
Deleted - tap	3.34
Aspirated - glottal stop	2.02
Aspirated - tap	1.97
Faithful - tap	1.94
Faithful - glottal stop	1.88
Aspirated - deleted	1.06
Deleted - faithful	0.84
Aspirated - faithful	0.08

ments because they were the most variable and because in English, word-final coronal segments, specifically [t, d] and [s, z], serve a morphological function. Inflectional suffixes, such as the regular past tense marker *-ed* and regular plural, third singular, or possessive *-s*, are typically instantiated as word-final [t, d] and [s, z] respectively.

In Study 4, we assessed whether there is a difference in the frequency of variants between morphologically-conditioned word-final segments and other word-final coronals. Because there is little categorical variation in the production of English [s, z] we do not discuss it here, although a future analysis of the acoustic realization of these might reveal differences based on morphological status.

We focused specifically on examining the frequency of variants of [t, d] as a function of morphological status. A detailed breakdown of the variation in the production of word-final /t/ and /d/ in our corpus is presented in Fig. 3. In 4133 instances of word-final /t/ in our corpus, glottal stops were the most likely variant (n = 1725) followed by deletions (804), unreleased variants (512), and taps (463), all of which were more frequent than the faithful variant (328). In 1657 instances of word-final /d/, it was deletions (627) that outnumbered the faithful variant (444), while all others were less frequent. Thus, the phonetic instantiation of word-final /d/ and /t/ was sufficiently variable to investigate whether some of it was conditioned by morphological structure.

5.1. Methods

To determine any differences in variant frequencies between morphologically-conditioned /t/ and /d/ and other word-final instances of /t/ and /d/*,* we compared variants in words that were suffixed with past tense *-ed* and words that were monomorphemic. One set of words included monomorphemic stems suffixed with regular past tense *-ed* like *walked* or *showed* (145 tokens; 62 types). Irregular past tense words like *kept* and *felt* were not included in this set. We also did not include words like *impressed* in this set because it is controversial whether *im-* is a prefix (e.g., Baroni, 2000; Aronoff, 2019). Next, we identified monomorphemic words ending in *-d* or *-t* like *need* or *put* (4984 tokens; 244 types). Contractions like *what's* or *what're* were not included in this set because in these cases *-t* and *-d* are not word final; to make the most conservative comparison we also did not include contractions like *aren't*. Finally, also excluded from this set were adjectives like *bored*, which are ambiguous between adjectives and past tense.

It is important to note that the established phonological analysis for this past-tense suffix is that the underlying form is /-d/ (e.g., Albright & Hayes, 2002). In our analysis, however, we treated the phonologically expected surface [-t] in instances like *walked* [wak-t] as phonemic for ease of comparison between suffixed and non-suffixed word-endings in these two segments.

5.1.1. Analysis

A multinomial regression model was used to determine how frequent variants were, relative to the faithful form for suffixed and non-suffixed forms for each category. The model was run using the *mblogit* function in the *mclogit* package (Elff, 2022).

5.2. Results & Discussion

The frequency of variants of regular past-tense *-ed* compared to word-final /t/ and /d/ in monomorphemic words are shown in Fig. 4. Our analyses indicate that morphological /t/ and /d/ are predominantly faithful, whereas non-suffixed word-final /t/ and /d/ are *not*: non-suffixed word-final /t/ is significantly more likely to surface as a glottal stop (z = 6.05, p < 0.001), tap (z = 1.71, p < 0.001), be unreleased (z = 1.78, p < 0.001), or to be deleted (z = 2.37, p < 0.001) than to be faithful, while /d/ is significantly more likely to be deleted (z = 1.57, p < 0.001). In contrast, for suffixed /t/ and /d/, all variants were significantly less likely to surface than the faithful form (z's < 0.38, p's < 0.01 for /t/ and z's < 0.57, p's < 0.01 for /d/). This suggests that the morphological structure also contributes to the nature of variation in IDS.

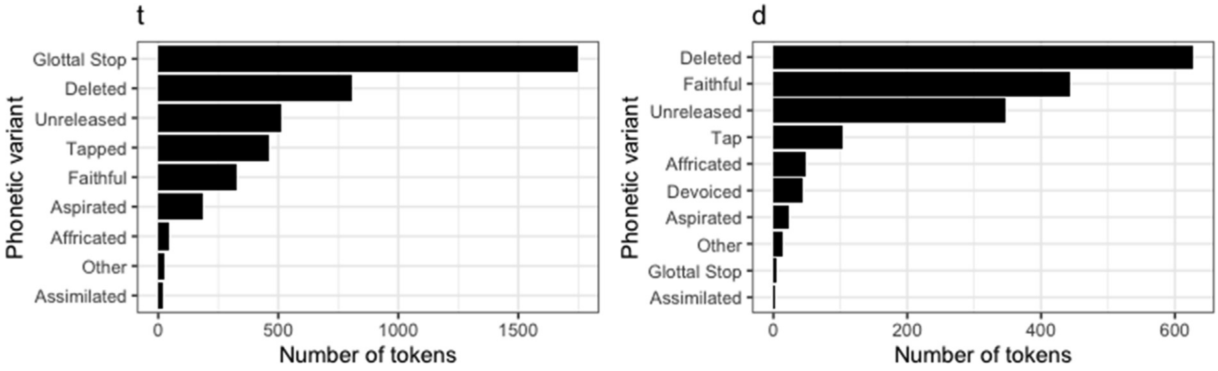

Fig. 3. Frequency of variants in word-final /t/ (left) and /d/ (right).

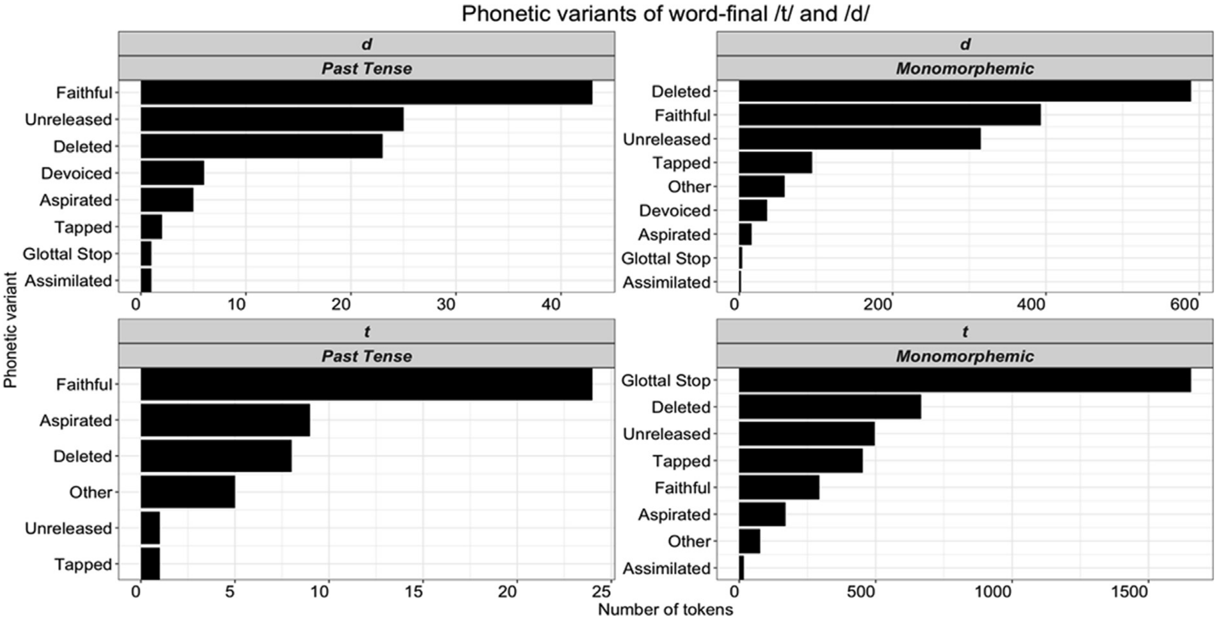

Fig. 4. Morphologically conditioned word final *-d* and *-t* (top left and bottom left) and monomorphemic word-final *-t* and *-d* (top right and bottom right).

6. General discussion

We created one of the largest phonetically annotated corpora of IDS to date in order to quantify the degree of phonetic variation present in the day-to-day input that children receive – specifically, the variation observed for alveolar coronals, some of the most frequent segments in English. Such a corpus is crucial to ensure that future theoretical and computational modeling of phonological acquisition is ecologically valid. We used this corpus to address four questions in Studies 1 through 4.

In Study 1, we did not find compelling support for the claim that IDS is more canonical than ADS. However, the canonical form was the most frequent for all segments except /t/; for /t/, the glottal stop variant outnumbered all others (Study 2). In Study 3, we quantified the extent to which the contexts where pairs of variants occur are dissimilar in order to predict which pairs of variants are likely to be identified as allophones using purely bottom-up information. Finally, in Study 4 we showed that the frequency of variants for /t/ and /d/ is governed by morphological structure, such that more faithful variants are produced when they signal inflectional suffixes.

We present two caveats before we discuss the implication of these findings for acquisition. Our IDS corpus is annotated only for categorical variation. However, we know from previous research that phonetic variability, including reduction, exists on a continuum (e.g., Barry & Andreeva, 2001; Warner & Tucker, 2011; Ernestus & Warner, 2011; Davidson, 2016; among others), which we did not annotate. Because we have multiple tiers of annotation, including phonetic, phonemic, and word, and the audio quality is sufficient to run an automatic forced aligner, our corpus can be used to investigate more gradient variation. We leave this as a promising avenue for future research. The second caveat is specific to interpreting the results from Study 1. Recall that our IDS and ADS corpora were not perfectly matched with respect to dialect. This was due to a dearth of phonetically annotated naturalistic corpora of ADS, especially with the exact dialect mix of our speakers. This dearth is unsurprising given the scale of effort needed to phonetically annotate corpora – a task that requires coding by trained human listeners.

Given the lack of a dialect-matched ADS corpus, in Study 1 we chose to compare our IDS corpus with 3 MAE speakers and 2 speakers described as speaking an NE dialect to 5 MAE speakers in the Buckeye corpus. All the speakers were young women. Additionally, we conducted two analyses, one with all speakers and one excluding the 2 NE speakers. Our results from both analyses converged, which we take to mean that the pattern of results reported in this paper, particularly Study 1, is robust. We have shared our corpus on the project OSF page to allow for future fine-grained analysis of phonetic differences between IDS and ADS, as well as the study of phonetic variation in IDS as a result of dialectal differences.

By comparing the extent to which the pronunciation of coronal segments is canonical in IDS vs ADS, in Study 1 we investigated the extent to which IDS is a learning register. However, we did not find compelling evidence that there are more canonical variants in IDS. Instead, using our phonetically annotated corpus of everyday speech directed to infants, we found that there were more canonical instances in IDS than ADS for only some segments, in some positions. Our lack of evidence in support of the claim that IDS is more canonical than ADS across the board thus challenges the proposal that IDS is a learning register (c.f., Burnham et al., 2002; Ferguson, 1964; Kuhl et al., 1997).

Instead, our findings are consistent with reports from machine learning studies that learning phonetic categories from IDS is neither easier nor more accurate. For instance, Ludusan and colleagues (2021) compared the robustness of vowel category learning from IDS, ADS, and read speech in Japanese and English using six different machine learning algorithms and two different speech representations and found more robust learning in both read speech and ADS compared to IDS. This is similar to findings with several other algorithms trained on English IDS performing worse on ADS than those trained on ADS (Kirchhoff & Schimmel, 2005; McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013). Our finding demonstrating that IDS is not more canonical than ADS may provide one explanation why IDS does not facilitate category learning across a large array of machine learning algorithms.

Instead, it is read speech that typically contains more canonical instances (e.g., Dilley et al., 2014; Fritche et al., 2021; but see also Buckler et al., 2018). Further, due to the reduced overlap in phonetic categories in read speech, machine learning algorithms are more accurate at learning vowel categories from it, consistent with expectations of a learning register. Although reading has been shown to have positive effects on vocabulary acquisition (e.g., Dickinson et al., 2019), further research is needed to investigate the specific role of read speech from caregivers in phonetic category acquisition.

The focus on the canonicality of IDS involves an assumption that variation is noise to be filtered out from the category signal. As a result, input with more canonical forms is considered beneficial for a language learner. However, our findings in Study 4 show that the relative frequency of variants can be helpful to signal the morphological structure of a sequence. Specifically, /t/ and /d/ are less likely to be produced as glottal stops or deleted when they are used as suffixes to signal past tense. The finding that morphological /t/ and /d/ are deleted less frequently in IDS is consistent with findings from adult directed speech in North American English (Bybee, 2000; Guy, 1980, 1991; Labov et al., 1968; among others). These differences in the frequency of variants in suffixed forms, as compared to monomorphemic forms, could potentially facilitate morphological decomposition over the course of development. Given recent findings that infants are sensitive to English morphological suffixes as early as 6-months (Kim & Sundara, 2021), before they learn meanings of verbs, it is likely that distributional differences in the input play a critical role in morpheme discovery.

Even when items are controlled for morphological complexity, research shows that both the category of phonetic variant (e.g., aspirated vs. unaspirated) as well as its acoustic instantiation can differ. This could be due to differences in word length (Johnson, 2004; Turnbull, 2018), phonological neighborhood density (Munson & Solomon, 2004; Wright, 2004), grammatical function (e.g., Drager, 2011) or lexical frequency (e.g., Aylett & Turk, 2004; Gahl, 2008; Turnbull, 2018). For example, Drager (2011) found that different functions of *like* (quotative, discourse particle, and lexical verb) showed systematic differences in phonetic realization. Similarly, it has been shown that the plural *-s* and 3rd singular *-s* can differ in their durations, likely because the former is more likely to be affected by lengthening in sentence-final position compared to the latter (Hsieh, Leonard, & Swanson, 1999). We also know that monomorphemic high frequency words, like *time*, are shorter than their low frequency homophones like *thyme* (Gahl, 2008). Similarly, the rate of initial stop aspiration for English prefixed stems in words like *disclaim* has also been shown to be sensitive to the frequency of the whole word and the stem (Zuraw et al., 2021). We were, however, not able to disentangle the contribution of all these different variables from the effect of morphological structure in Study 4 because of our limited dataset. Therefore, it is possible that the effect of morphological structure on variant frequency reported in Study 4 is due to factors not considered in this study (i.e., lexical frequency, Aylett & Turk, 2004; Gahl, 2008; Turnbull, 2018; grammatical function, Drager, 2011; and phonological neighborhood density, Munson & Solomon, 2004; Wright, 2004).

The effect of morphological structure on variant frequency reported here may also be due to the past tense marker *-ed* occurring only in a subset of all possible phonological environments. Further analysis will be needed to determine whether there are any significant differences in the distribution of phonological environments as a result of the morphological and syntactic patterns of English and whether morphological conditioning is evident after controlling for the environment. If phonetic evidence for morphological structure as we have demonstrated here is robust, it is incompatible with traditional feedforward models of speech production (e.g., Levelt & Wheeldon, 1994; Levelt et al., 1999) where morphological information is assumed to be inaccessible at the point of phonetic production. Instead, it is more compatible with exemplar-based models (e.g., Goldinger, 1998; Pierrehumbert 2001, 2002; Bybee, 2001; Gahl & Yu, 2006).

We also found evidence that the most frequent variant for /d, s, n, z/ in IDS is the variant that has been claimed to be the canonical variant in ADS. For /t/, however, glottal stops were significantly more frequent than [t] when word position is not taken into account, with aspirated variants only slightly less frequent than the faithful variant. Given that the faithful variant of /t/ has been shown to be privileged in the processing literature (Pitt et al., 2011; Ranbom & Connine, 2007; Sumner & Samuel, 2005), this raises questions about the roots of that processing advantage.

If infants are tracking the overall frequency of variants, then we would expect an advantage for the most frequent form which is not the faithful [t], but rather the glottal stop. Even if infants are tracking position-specific frequencies of variants, we would not expect to see any canonical advantage in early learning. Under this view, the canonical advantage observed in adult priming studies is expected to emerge later and would likely require abstraction over the set of variants encountered. Any advantage that the faithful variant of /t/ has in recognition thus has to be learned. That is, the faithful form may be privileged if and when the allophones are clustered into phonemes. Alternatively, it may be privileged by the learning of orthography (Ranbom & Connine, 2007). However, the fact that 2-year-old North American English children (from the same corpus as the one we have examined here) produced more faithful forms of coda /t/s and /d/s relative to their adult caregivers (Song, Shattuck-Hufnagel, & Demuth, 2015) suggests that any advantage for the faithful form arises before children learn orthography.

One difficulty in comparing our findings regarding the frequency of the canonical variant in our IDS corpus to the existing research showing a canonical advantage in word recognition is that in previous research, phonetically distinct categories are either combined, or only a subset of variants are considered. Ranbom et al. (2009), for example, do not discuss whether the word-final [t]s in their experiment contained instances of unreleased or aspirated stops. Based on our findings, we might expect at least some word-final [t]s to be produced as variants other than the unaspirated fully released [t] that they coded for.

In this paper, we provided a more fine-grained analysis of which variants are most frequently associated with coronal segments in IDS. Our separation of the variant types into aspirated released and unaspirated released reflects the fact that in many of the world's languages, aspiration of stops is a contrastive feature (e.g., Hindi: Ohala, 1999; Thai: Tingsabadh & Abramson, 1993). Similarly, we separated out released and unreleased variants as well, given that in many languages, like Korean (Sohn, 1994) or Thai (Smyth, 2014; Tingsabadh & Abramson, 1993), unreleased stops are variants whose occurrence is predictable from phonological context (e.g., ends of words). These cross-linguistic differences mean that whether or not these variants are functionally treated as distinctive categories has to be learned. By separating out aspirated and unreleased variants from [t], we found that the faithful [t] form was no longer the most frequent form for /t/. This is in contrast to previous claims that [t] is the most frequent, and therefore canonical, variant of /t/ (c.f. Pitt et al., 2011; Ranbom et al., 2009). Instead, glottal stops were most frequent. Our findings, therefore, reveal a more complex learning problem than is generally alluded to in the processing literature.

Lastly, beyond learning to privilege the canonical variant, infants must also learn to abstract away from phonetic categories to identify phonemes. Results from infant experiments are consistent with the idea of gradual emergence of phonemes over the first year. Infants have difficulty distinguishing between allophones by 12 months (Pegg & Werker, 1997) as do adults (Pegg & Werker, 1997; Peperkamp et al., 2003; Whalen, Best, & Irwin, 1997). Thus, infants appear to discover the equivalence between some variant forms by 12-months. Further confirming this timeline, we know that 12-month-olds can use this knowledge to successfully relate [ɾ] to /d/, but not to /t/, in morpho-phonological alternations when verb stems are suffixed with *-ing* (Sundara et al., 2021). What information could infants be relying on in their input to build these abstract categories?

It has been proposed that phonemes could be discovered using a bottom-up approach by clustering variants based only on their distributions across phonological environments (Maye, Werker, & Gerken, 2002; Peperkamp et al., 2006; Maye, Weiss, & Aslin, 2008; among others). Clustering variant pairs with the most complementary distribution is one possible way for infants to construct phonemes (e.g., Peperkamp et al., 2006). In Study 3, using KL divergence as a metric, we found that the most divergent variant pair for /t/ is glottal stop and tap, and for /d/ is deletion and tap. A purely distributional account would predict that infants learn to relate glottal stops and tap variants of /t/ and deleted and tap variants of /d/ first in acquisition. These findings thus predict an empirically testable developmental trajectory which, if supported, could lend weight to the idea that infants use complementary distributions to cluster phonetic categories. We leave this for future research.

The idea that infants can use the complementary distribution of variants to discover phonemes is not uncontroversial. Martin and colleagues (2013) show that a bottom-up approach to learning phonemes is not accurate when applied to all phonemes simultaneously, given the large number of variants and environments alongside the uncertainty about the number of phoneme categories. In this study, we have simplified the problem of learning phonemes in at least two ways. First, we restricted the environments to those that favor one or other variants. We also restricted the variants themselves – for example, we only calculated the KL divergence score for pairs of variants of /t/ or /d/, but not both. We also did not integrate any prosodic information, which additionally conditions phonetic variation (Parker & Walsh, 1982; Keating, Cho, Fougeron, & Hsu, 2004; Random et al., 2009; among others), into these calculations, as this was not annotated in the corpus. Thus, it is possible that the simplifying assumptions we made mean that the variants identified here as most different distributionally are not the same as in the child's language input.

Through computational modeling, it has been demonstrated that having additional information about sequencing restrictions (Martin et al., 2013) or access to word forms (Feldman, Myers, White, Griffiths, & Morgan, 2013) can facilitate the discovery of phonemes from bottom-up information. Further, experimental evidence shows that infants favor clustering perceptually similar variants before those that are less similar (Sundara et al., 2021) identifying another kind of bias that explains how infants start to construct abstract phoneme categories. In sum, evidence is accumulating that a purely distributional model is not sufficient to fully explain how phonemes can be learned from the speech signal (see also Bion, Borovsky, & Fernald, 2013, Antetomaso et al., 2017). What is less clear is whether a sub-optimal distributional learning model could reflect a similarly sub-optimal, earlier stage of infant acquisition.

7. Conclusion

In this paper, we evaluated the degree of phonetic variation in naturalistic infant-directed speech. This was achieved by phonetically transcribing ~6500 utterances from the Providence Corpus (Demuth et al., 2006). We focused on the degree of variation in coronals, some of the most frequent segments in English. First, we evaluated whether IDS is more canonical than ADS across the board and found that this was only true for certain segments in certain positions. We also found that the phonetically faithful form was the most frequent form for all segments except for /t/, where glottal stops were more frequent than [t] overall and aspirated [tʰ] were only slightly less frequent. One exception was in the case of regular past tense *-ed*, which were overwhelmingly faithful, more so than other word-final instances of these segments in naturalistic IDS. We also identified the variants of /t/ and /d/ that are most in complementary distribution, with the view that infants might cluster them first to discover phonemes. Our focus here was on categorical variation, although the corpus can also be used to measure more gradient phonetic patterns in American English IDS. We expect that this corpus will be a critical resource for future research on phonetic variation in English IDS.

CRediT authorship contribution statement

Ekaterina A. Khlystova: Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Adam J. Chong:** Data curation, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Megha Sundara:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Funding to support this project was provided by the National Science Foundation [grant number BCS-2028034 to MS]. This work would not have been possible without the help of our undergraduate research assistants: AJ Yoo, Dianely Mendoza, Dylan Ross, Irene Chou, Oliver Huang, Sam O'Brien, Theo Chen, and Thu Phan. We would also like to thank Isabelle Lin, who was involved in earlier stages of this project. Lastly, we thank the audiences at LabPhon 16, the Boston University Conference on Language Development (BUCLD) 2020, and the 181st Meeting of the Acoustical Society of America (ASA), as well as members of the UCLA Phonetics Lab, for their valuable feedback.

References

Albright, A., & Hayes, B. (2002). Modeling English past tense intuitions with minimal generalization. In *Proceedings of the ACL-02 workshop on Morphological and phonological learning* (pp. 58–69).

Antetomaso, S., Miyazawa, K., Feldman, N., Elsner, M., Hitczenko, K., & Mazuka, R. (2017). Modeling phonetic category learning from natural acoustic data. *Proceedings of the annual Boston University Conference on Language Development*.

Aronoff, M. (2019). Competitors and alternants in linguistic morphology. *Competition in Inflection and Word-Formation*, 39–66.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*(1), 31–56.

Baker, R., Smith, R., & Hawkins, S. (2007). Phonetic differences between mis-and dis-in English prefixed and pseudo-prefixed words. *Proceedings of ICPhS XVI, 16*, 6–10.

Baroni, M. (2000). *Distributional cues in morpheme discovery: A computational model and empirical evidence*. Los Angeles: University of California.

Barry, W., & Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association, 31*(1), 51–66.

Beckman, Jill N. (1998) *Positional Faithfulness.* [Doctoral Dissertation, University of Massachusetts Amherst]. ProQuest Dissertations Publishing.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America, 113*(2), 1001–1024.

Ben Hedia, S., & Plag, I. (2017). Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics, 62*, 34–49.

Bernstein Ratner, N. (1984). Patterns of vowel modification in mother–child speech. *Journal of Child Language, 11*(3), 557–578.

Bion, R. A., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition, 126*(1), 39–53.

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 6.2.23. *Online:* http://www.praat.org.

Booij, G. E. (1983). Principles and parameters in prosodic phonology.

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.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*(5572). 1435–1435.

Bybee, J. (2000). Alternating environments. *Papers in Laboratory Phonology V: Acquisition and the Lexicon, 5*, 250.

Bybee, J. (2001). Phonology and Language Use. Cambridge: CUP. *Cambridge Studies in Linguistics, 94*.

Carterette, E. C., & Jones, M. H. (1974). *Informal speech: Alphabetic & phonemic texts with statistical analyses and tables*. Univ of California Press.

Chomsky, N., & Halle, M. (1968). The sound pattern of English.

Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language, 41*(4), 913–934.

Dalby, J. M. (1986). Phonetic structure of fast speech in American English. *Phonetics and Phonology*. Bloomington, IN: Indiana University Linguistic.

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics, 54*, 35–50.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech, 49*(2), 137–173 [data source].

Denes, P. B. (1963). On the statistics of spoken English. *The Journal of the Acoustical Society of America, 35*(6), 892–904.

Dickinson, D. K., Collins, M. F., Nesbitt, K., Toub, T. S., Hassinger-Das, B., Hadley, E. B., & Golinkoff, R. M. (2019). Effects of teacher-delivered book reading and play on vocabulary learning and self-regulation among low-income preschool children. *Journal of Cognition and Development, 20*(2), 136–164.

Dilley, L. C., & Pitt, M. A. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *The Journal of the Acoustical Society of America, 122*(4), 2340–2353 [data source].

Dilley, L. C., Millett, A. L., McAuley, J. D., & Bergeson, T. R. (2014). Phonetic variation in consonants in infant-directed and adult-directed speech: The case of regressive place assimilation in word-final alveolar stops. *Journal of Child Language, 41*(1), 155–175.

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

Drager, K. K. (2011). Sociophonetic variation and the lemma. *Journal of Phonetics, 39*(4), 694–707.

Drost, H. G. (2018). Philentropy: Information theory and distance quantification with R. *Journal of Open Source Software, 3*(26), 765.

Eaves, B. S., Jr, Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review, 123*(6), 758.

Elff, M. (2022). mclogit: Multinomial Logit Models, with or without Random Effects or Overdispersion_. R package version 0.9.4.2, <https://CRAN.R-project.org/package=mclogit>.

Englund, K. T. (2018). Hypoarticulation in infant-directed speech. *Applied Psycholinguistics, 39*(1), 67–87. https://doi.org/10.1017/S0142716417000480.

Ernestus, M., Lahey, M., Verhees, F., & Baayen, R. H. (2006). Lexical frequency and voice assimilation. *The Journal of the Acoustical Society of America, 120*(2), 1040–1051.

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics, 39*(SI), 253–260.

Feldman, N., Griffiths, T., & Morgan, J. (2009). Learning phonetic categories by learning a lexicon. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 31, No. 31).

Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., & Morgan, J. L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition, 127*(3), 427–438.

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist, 66*(6), 103–114.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*(1), 104.

Fritche, R., Shattuck-Hufnagel, S., & Song, J. Y. (2021). Do adults produce phonetic variants of/t/less often in speech to children? *Journal of Phonetics, 87* 101056.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language, 84*(3), 474–496.

Gahl, S., & Yu, A. C. L. (2006). Introduction to the special issue on exemplar-based models in linguistics. 213–216.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review, 105*(2), 251.

Guy, G. (1980). Variation in the group and the individual: The case of final stop deletion. In *Locating Language in Time and Space* (pp. 1–36). Academic Press.

Guy, G. R. (1991). Contextual conditioning in variable lexical phonology. *Language Variation and Change, 3*(2), 223–239.

Hitczenko, K., & Feldman, N. H. (2022). Naturalistic speech supports distributional learning across contexts. *Proceedings of the National Academy of Sciences, 119*(38). e2123230119.

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

Hsieh, L., Leonard, L. B., & Swanson, L. A. (1999). Some differences between English plural noun inflections and third singular verb inflections in the input: The contributions of frequency, sentence position, and duration. *Journal of Child Language, 26*, 531–543.

Johnson, K. (2004). Acoustic and auditory phonetics. *Phonetica, 61*(1), 56–58.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*(3–4), 159–207.

Keating, P., Cho, T., Fougeron, C., & Hsu, C. S. (2004). Domain-initial articulatory strengthening in four languages. *Phonetic Interpretation: Papers in Laboratory Phonology VI*, 143–161.

Kim, Y. J., & Sundara, M. (2021). 6–month–olds are sensitive to English morphology. *Developmental Science, 24*(4), e13089.

Kiparsky, P. (1982). Word-formation and the lexicon. Mid-America Linguistics Conference.

Kirchhoff, K., & Schimmel, S. (2005). Statistical properties of infant-directed versus adult-directed speech: Insights from speech recognition. *The Journal of the Acoustical Society of America, 117*(4), 2238–2246.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., ... Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684–686.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics, 22*(1), 79–86.

Labov, W., Cohen, P., Robins, C., & Lewis, J. (1968). A study of the Nonstandard English of Negro and Puerto Rican speakers in New York City, Final Report. *US Office of Education Cooperative Research Project*, (3288).

Labov, W., Ash, S., & Boberg, C. (2008). *The atlas of North American English: Phonetics, phonology and sound change*. Walter de Gruyter.

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics*. Cengage learning. [PUBLISHING INFO].

Lahey, M., & Ernestus, M. (2014). Pronunciation variation in infant-directed speech: Phonetic reduction of two highly frequent words. *Language Learning and Development, 10*(4), 308–327.

Lenth, R. (2022)._emmeans: Estimated Marginal Means, aka Least-Squares Means_. R package version 1.7.4-1, <https://CRAN.R-project.org/package=emmeans>.

Levelt, W. J., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50*(1–3), 239–269.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38.

Ludusan, B., Mazuka, R., & Dupoux, E. (2021). Does infant-directed speech help phonetic Learning? A machine learning investigation. *Cognitive Science, 45*(5), e12946.

MacKenzie, L., & Tamminga, M. (2021). New and old puzzles in the morphological conditioning of coronal stop deletion. *Language Variation and Change, 33*, 217–244.

Martin, A., Peperkamp, S., & Dupoux, E. (2013). Learning phonemes with a proto-lexicon. *Cognitive Science, 37*(1), 103–124.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis. *Psychological Science, 26*(3), 341–347.

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

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*(1), 122–134.

McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition, 95*(2), B15–B26.

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition, 129*(2), 362–378.

Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H., & Mazuka, R. (2017). Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition, 166*, 84–93.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research, 47*, 1048–1058.

Nespor, M., & Vogel, I. (2007). *Prosodic Phonology*. Berlin, Boston: De Gruyter Mouton.

Odden, D. (2005). *Introducing phonology*. Cambridge University Press.

Ohala, M. (1999), “Hindi”, in International Phonetic Association (Ed.), Handbook of the International Phonetic Association: a Guide to the Use of the International Phonetic Alphabet, Cambridge University Press (pp. 100–103).

Parker, F., & Walsh, T. (1982). Blocking alveolar flapping: A linguistic analysis. *Journal of Phonetics, 10*(3), 301–314.

Patterson, D., LoCasto, P. C., & Connine, C. M. (2003). Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica, 60*(1), 45–69.

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

Peperkamp, S., Pettinato, M., & Dupoux, E. (2003). Allophonic variation and the acquisition of phoneme categories. Proceedings of the 27th annual Boston University conference on language development, Vol. 2, Cascadilla Press, Sommerville, MA (2003), pp. 650–661.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. *Typological Studies in Language, 45*, 137–158.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory Phonology, 7*(1), 101–140.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). Buckeye Corpus of Conversational Speech. Columbus, OH: Department of Psychology, Ohio State University (Distributor); 2007. (2007; Final release) [www.buckeyecorpus.osu.edu] [data source].

Pitt, M. A., Dilley, L., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics, 39*(3), 304–311.

Plag, I., Homann, J., & Kunter, G. (2017). Homophony and morphology: The acoustics of word-final S in English1. *Journal of Linguistics, 53*(1), 181–216.

Port, R. F. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology, 25*, 143–170.

Rosenfelder, Fruehwald, Evanini, Seyfarth Gorman, Prichard & Yuan (2014). FAVE (forced alignment and vowel extraction) program suite v1.2.2 10.5281/zenodo.22281.

Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language, 57*(2), 273–298.

Ranbom, L. J., Connine, C. M., & Yudman, E. M. (2009). Is phonological context always used to recognize variant forms in spoken word recognition? The role of variant frequency and context distribution. *Journal of experimental psychology: Human perception and performance, 35*(4), 1205.

Schmitz, D., Baer-Henney, D., & Plag, I. (2021). The duration of word-final /s/ differs across morphological categories in English: Evidence from pseudowords. *Phonetica, 78*(5–6), 571–616.

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.

Seyfarth, S., & Garellek, M. (2020). Physical and phonological causes of coda/t/glottalization in the mainstream American English of central Ohio. *Laboratory Phonology, 11*(1).

Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., & Malouf, R. (2018). Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience, 33*(1), 32–49.

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science, 15*(5), 659–673.

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language, 40*(3), 672–686.

Shockey, L., & Bond, Z. S. (1980). Phonological processes in speech addressed to children. *Phonetica, 37*(4), 267–274.

Sjons, J., Hörberg, T., Östling, R., & Bjerva, J. (2017). Articulation rate in Swedish child-directed speech increases as a function of the age of the child even when surprisal is controlled for. *arXiv preprint arXiv:1706.03216*.

Smith, R., Baker, R., & Hawkins, S. (2012). Phonetic detail that distinguishes prefixed from pseudo-prefixed words. *Journal of Phonetics, 40*(5), 689–705.

Smyth, D. (2014). Pronunciation. In *Thai* (pp. 5–10). Routledge.

Snoeren, N. D., Hallé, P. A., & Segui, J. (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics, 34*(2), 241–268.

Sohn, H. (1994). Korean. Descriptive Grammars. *Issues in Applied Linguistics, 5*(2). London and New York: Routledge.

Song, J. Y., Shattuck-Hufnagel, S., & Demuth, K. (2015). Development of phonetic variants (allophones) in 2-year-olds learning American English: A study of alveolar stop/t, d/codas. *Journal of Phonetics, 52*, 152–169.

Song, J. Y., Sundara, M., & Demuth, K. (2009). Phonological constraints on children's production of English third person singular–s. *Journal of Speech, Language, and Hearing Research, 52*(3), 623–642.

Sugahara, M., & Turk, A. (2009). Durational correlates of English sublexical constituent structure. *Phonology, 26*(3), 477–524.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final/t. *Journal of memory and language, 52*(3), 322–338.

Sundara, M., White, J., Kim, Y. J., & Chong, A. J. (2021). Stem similarity modulates infants' acquisition of phonological alternations. *Cognition, 209* 104573.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1536), 3617–3632.

Tang, J. S. Y., & Maidment, J. A. (1996). Prosodic aspects of child-directed speech in Cantonese. *University College London: Speech, Hearing and Language—Work in Progress, 9*, 257–276.

Tamminga, M. (2016). Persistence in phonological and morphological variation. *Language Variation and Change, 28*(3), 335–356.

Tobias, J. V. (1959). Relative occurrence of phonemes in American English. *The Journal of the Acoustical Society of America, 31*(5). 631–631.

Tomaschek, F., Plag, I., Ernestus, M., & Baayen, R. H. (2021). Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning. *Journal of Linguistics, 57*(1), 123–161.

Tingsabadh, M. K., & Abramson, A. S. (1993). Thai. *Journal of the International Phonetic Association, 23*(1), 24–28.

Turnbull, R. (2018). Patterns of probabilistic segment deletion/reduction in English and Japanese. *Linguistics Vanguard, 4*(s2).

Vaux, B. (2002). *Aspiration in English*. University of Wisconsin-Madison.

Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America, 130*(3), 1606–1617.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*(11), 2143–2152.

Whalen, D. H., Best, C. T., & Irwin, J. R. (1997). Lexical effects in the perception and production of American English/p/allophones. *Journal of Phonetics, 25*(4), 501–528.

Wright, R. (2004). A review of perceptual cues and cue robustness. *Phonetically Based Phonology, 34*, 57.

Zimmermann, E. (2016). The power of a single representation: Morphological tone and allomorphy. *Morphology, 26*(3), 269–294.

Zuraw, K., & Peperkamp, S. (2015). Aspiration and the gradient structure of English prefixed words. *ICPhS*.

Zuraw, K., Lin, I., Yang, M., & Peperkamp, S. (2021). Competition between whole-word and decomposed representations of English prefixed words. *Morphology, 31*(2), 201–237.