Journal of Phonetics 87 (2021) 101056

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

Do adults produce phonetic variants of /t/ less often in speech to children?

Robin Fritche [a], Stefanie Shattuck-Hufnagel [b], Jae Yung Song [a,*]

[a] Department of Linguistics, University of Wisconsin-Milwaukee, Milwaukee, WI 53211, USA
[b] Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 8 April 2020
Received in revised form 25 March 2021
Accepted 28 March 2021

Keywords:
Phonetic variation
Canonical form
Language acquisition
Child-directed speech
Flapping
Glottalization
Aspiration

ABSTRACT

The surface phonetic details of an utterance affect how 'native' a speaker sounds. However, studies have shown that children's acquisition of context-appropriate variation (sometimes called allophones) is late. This study's goal was to understand how caregivers use phonetic variation in the production of American English /t/ in child-directed speech (CDS), compared to in adult-directed speech (ADS). We hypothesized that mothers modify their input to children in order to produce more limited variation in CDS than in ADS, to potentially assist children in the development of contrastive phonemic categories. To this end, we recorded eight mothers of children under the age of 2 years in both ADS and CDS conditions. Results reveal that CDS contains significantly more canonical cues to /t/ than ADS does, and fewer non-canonical cue patterns, including fewer unreleased tokens and fewer glottalized tokens in utterance-medial position. Also, we found larger aspiration duration differences in CDS between aspirated singleton [t^h] vs. unaspirated [t] in /st/ contexts, suggesting that mothers exaggerate this cue to the phonemic context in which the /t/ occurs. Overall, the findings suggest that CDS more clearly signals the phonemic category, which could in turn assist children learn the relationship between the underlying and surface forms.

1. Introduction

As children learn their first language, linguistic input, i.e. the language that is spoken to them and around them, plays a critical role. Understanding the characteristics of linguistic input to children, therefore, provides a window into the way children acquire a first language. Past work investigating adults' speech to children (also called child-directed speech or CDS) has found that there are key differences between CDS and adult-directed speech (or ADS). These include higher pitch (Fernald & Simon, 1984), larger pitch variability (Fernald & Simon, 1984), shorter utterances (Broen, 1972), slower speech rate (Broen, 1972), restricted vocabulary (Broen, 1972) and signals of greater positive affect (Masataka, 1992). Investigations at the level of phonemic segments have found that CDS exhibits stronger cues to phonemic contrasts, including a larger (or 'stretched') vowel space for vowels (which makes vowels more distinct from one another, often referred to as hyperarticulation) (Bernstein Ratner, 1984a; Burnham & Kitamura, 2002; Cristia & Seidl, 2014; Hartman, Bernstein Ratner, & Newman, 2017; Kuhl et al., 1997), an increased voice onset time (VOT) independent of speaking rate (Englund, 2005), less VOT overlap between voiced and voiceless stops (Malsheen, 1980), and a greater tendency to lengthen vowels preceding word-final voiced consonants (Bernstein Ratner & Luberoff, 1984). Other differences found include prosodic patterns that are more informative about communicative intent (Fernald, 1989) and some structural variations such as a higher occurrence of questions and the placement of more important words at the end of an utterance, where they are perceptually more salient (Fernald & Mazzie, 1991).

Historically, researchers in this field have focused on suprasegmentals such as intonation and speaking rate, as well as segmentals on the phonemic level, which deals with contrastive sounds that, when changed, alter the word form to a separate item that can have a different meaning (e.g., *p*ack vs. *b*ack). However, these are not the only aspects of the sound system that a child is exposed to. Below the phonemic level lies the level of phonetic variation within a phoneme. Phonetic variation refers to non-contrastive change, an example of which is the way /t/ is pronounced in these three words in American English: *t*ar, bu*tt*er, and go*t*. In the first word, /t/ is pronounced clearly with a complete closure made between

* Corresponding author.
E-mail address: songjy@uwm.edu (J.Y. Song).

https://doi.org/10.1016/j.wocn.2021.101056
0095-4470/

the tongue and the alveolar ridge behind the teeth, with pressure buildup behind that constriction released with a puff of air and turbulent noise. In the second, however, what we still think of as /t/ often sounds like a very fast /d/. In the last, /t/ is often realized without any release, and/or with a concomitant full or partial glottal constriction. These latter two variations of /t/ are still interpreted as /t/ by native speakers of English, despite the fact that they have different acoustic properties than the 'canonical' /t/ with distinct stop closure and release. These phonetic variants are often referred to as allophones, and are signaled by different sets of acoustic cues.[1]

The appropriate production of phonetic variants with different acoustic characteristics is an important marker of 'belonging' to a particular speech community, and American English is no exception to the existence of such variation. For example, in many varieties of American English, both voiced and voiceless intervocalic alveolar stops are often flapped when the second of the two vowels is unstressed, as in *city* and *writer* (Byrd, 1993; De Jong, 1998; Herd, Jongman, & Sereno, 2010; Zue & Laferriere, 1979). More explicitly, the word-internal flap can be defined as "a momentary apicoalveolar single trill necessarily preceded and followed immediately by vowels. The first (vowel) is either stressed or unstressed, while the second is always unstressed" (Malécot & Lloyd, 1968, p. 264). When followed by a syllabic nasal, such as in the words *kitten* and *button,* the voiceless alveolar stop is often glottalized (Byrd, 1993). Word-final stops are sometimes unreleased (such as the /t/ in *late* and *what)* and sometimes glottalized (as in *hot dog*), and voiceless stops occurring directly after syllable-initial /s/ (e.g., *stool*) lack the clear aspiration of the canonical prevocalic voiceless alveolar stop /t/ (e.g., *tool*) (Byrd, 1993)[2].

While there have been some studies comparing CDS with ADS at this level of context-driven phonetic variation in recent years, reviewed below, coverage has been far from comprehensive. Furthermore, very few CDS vs. ADS comparisons have specifically addressed phonetic variation of one specific phoneme using the same speakers for both conditions (as was done in Dilley, Gamache, Wang, Houston, & Bergeson, 2019, discussed more in detail below).

To fill this gap, the present study aims to compare the frequency of occurrence of phonetic variation in productions of the voiceless alveolar stop /t/ in both CDS and ADS by the same speakers. The /t/ phoneme was chosen because of its unique richness of phonetic variants in American English, where it is subject to flapping, glottalization, lack of release, and lack of aspiration (or extremely short VOT), depending upon the surrounding phonemes and structural position (Ladefoged & Johnson, 2006).

In the rest of this Introduction, we will review studies looking at the production of phonetic variants in CDS. Although there is only limited work on the level of context-appropriate phonetic variation, the studies that are available suggest that mothers do change their speech in a particular way when addressing their children, producing clearer, more canonical speech with fewer phonetic variants. For example, studying seven mothers of children with Mean Length of Utterances (MLUs) between 1.0 and 2.5, Bernstein Ratner (1984b) focused on four processes that occur at word boundaries – palatalization (*did you* → [ˈdɪdʒju]), dental deletion (*want it* → ['wɑnə]), [ð] deletion – (*get them* → [ˈgɛɾəm]), and /ts/ → [s] conversion (*that's good* → [ðæs] *good*) – and showed that while palatalization was more common in CDS than ADS, other processes such as dental deletion, /ð/ deletion, and /ts/ → [s] conversion were significantly less common in CDS than in ADS.

In another study of phonetic variation across word boundaries, Dilley, Millett, McAuley, and Bergeson (2014) investigated regressive place assimilation in word-final alveolar stops in both ADS and CDS. Forty-eight American mothers of infants aged approximately 0;3, 0;9, 1;1, or 1;8 (year; month) who were asked to read a specifically prepared storybook, to both their children and to an adult researcher, exhibited "a reliable tendency for canonical variants to occur in infant directed speech more often than adult directed speech" (p. 154), meaning that mothers produced canonical forms of stops more often in CDS than in ADS. Dilley et al. (2014) also hypothesized age effects on CDS, predicting consonant hypoarticulation (more phonetic variation) when mothers spoke to infants at the lower and upper end of the study's range and hyperarticulation (less phonetic variation) around the age that infants first show definite signs of comprehension (1;0, also the approximate age at which infants produce first words). While results did hint at such a trend, they were not statistically significant. Buchan and Jones (2014) also investigated age effects on CDS in a longitudinal study focusing on what they term deletion of word-initial /h/ and word-final /v/ in Australian English. They recorded four mother–child dyads when the children were aged 1;6, 2;0, and 2;6. Though their hypothesis was that occurrence of those processes would increase over time, they found changes to be non-linear, with deletion increasing between 1;6 and 2;0 and decreasing between 2;0 and 2;6. These authors suggested that children's linguistic development could be a factor in determining how and when mothers vary the phonetic shape of their words in CDS.

Foulkes, Docherty, and Watt (2005) investigated phonetic variants of /t/ in CDS in conjunction with sociolinguistic cues in a speech community in Tyneside, England and compared results with ADS data from a previous study in that same area. They focused on word-medial and word-final pre-vocalic variations of /t/ and reported that production of variants in CDS differs from that in ADS in that there is a higher incidence of [t] and a lower incidence of [ɹ] (a phonetic variant of /t/ present in the language variety spoken in that community) and glottal

[1] One point we wish to make clear here is that phonetic variants or phonetic variation need not be thought of as separate symbolic categories, but rather can be construed as different sets of acoustic cues to the features of a given contrastive phonemic category. For convenience in analysis, in this paper we use binary terms like released/unreleased, flapped/unflapped, glottalized/unglottalized, and aspirated/unaspirated to designate differences among sets of cues that occur together with high frequency in certain contexts, to signal a phonological category. But this notational decision does not represent a commitment to the claim that these high-frequency cue sets correspond to symbolic categories at the phonetic level.

[2] The stop consonant /t/ is canonically produced with a closure of the tongue tip against the alveolar ridge, followed by pressure buildup within the mouth as air continues to flow into the oral tract, followed by an abrupt onset of a noise burst due to air turbulence as the constriction is released. There seems to be no question that the distinct stop closure and release are critical components of the canonical (or underlying) form of /t/. In contrast, there seem to be mixed views on whether English voiceless stops are underlyingly specified as being [+aspirated].Sometimes they are specified as being [−aspirated] (Odden, 2005), sometimes [+aspirated] (or [spread glottis]) (Vaux, 2002), and sometimes not specified (Dilley et al., 2019). In the present study, we consider the variant that occurs in stressed, syllable-initial position as the canonical form of /t/. Therefore, the assumption here is that the underlying form of /t/ is [+aspirated].

variants, and that these rates of variance are influenced by the gender and age of the child.

The very recent study by Dilley et al. (2019) is especially relevant to the present study, as they looked at the distribution of alveolar stop variants in CDS and ADS in the same speakers, in American English. They hypothesized that mothers produce more instances of canonical variants in CDS than in ADS. They recorded 53 mothers of typically-developing children with an average age of 1;0, both during play sessions with their child and during an interview with an adult researcher. They examined three stop consonants, /t/, /d/, and /n/, in word-final position, and the target sounds were judged to be canonical, assimilated, deleted, or glottalized on the basis of acoustic displays and auditory perception. They found a clear statistical predominance of canonical /t/ in CDS, though not for canonical /d/ or /n/.

In contrast, other studies have indicated that there is more phonetic variation (regarding at least some processes) in CDS than in ADS. Shockey and Bond (1980), studying the four processes (palatalization, dental deletion, [ð] deletion, and /ts/ → [s] conversion) later addressed by Bernstein Ratner (1984b, mentioned above), collected data from eight British mothers speaking both to their two- to four-year-old children and to an adult. Although there were insufficient tokens in the data to judge palatalization, results showed that the latter three processes were actually more likely to occur in CDS than in ADS. Bernstein Ratner and Luberoff (1984) also found evidence of more phonetic variation in CDS. They studied nine mothers of children aged 0;9 to 2;3 (with MLUs of zero to 2.5) and found that there was a greater tendency for mothers to delete or glottalize word-final consonants in CDS than in ADS.

Such a disparity in results creates conflict, though it has been suggested that size, dialect, procedures, and even age of the child may factor into this issue (Bernstein Ratner, 1984b). For example, mothers may have used a more formal register when interacting with researchers than when interacting with their children, and CDS may contain more canonical cues only during a limited stage of the child's development.

Although the studies mentioned above have begun to address the problem of how CDS differs from ADS on the level of phonetic variation, many questions remain, and the present study aims to contribute to resolving them. Specifically, this study examines mothers' production of four different types of context-appropriate phonetic variants of the alveolar stop /t/: flapped (as in *wa<u>t</u>er*), unreleased (as in *ha<u>t</u>*), glottalized (as in *bu<u>tt</u>on*), and unaspirated (as in *s<u>t</u>op*). Based upon the studies cited above that have indicated that CDS is in general more clearly articulated, and in particular that CDS is less subject to phonetic variation (Bernstein Ratner, 1984b; Dilley et al., 2014, 2019; Malsheen, 1980), we hypothesized that mothers would produce the canonical variant of /t/ (with a clear closure, release burst, and aspiration) more often in CDS than in ADS.

For three of the variants, flapped, unreleased, and glottalized stops, we hypothesized that mothers produce the canonical variant of /t/ more often when speaking to their young children than when speaking to adults. In other words, we predicted that the use of phonetic variants would be more restrained in CDS. Thus, for these three variants, the frequency of occurrence was measured. For the fourth variant, unaspirated stops, we measured the duration of aspiration noise and also compared it with that of the canonical variant of /t/ (e.g., *s<u>t</u>ool* vs. *<u>t</u>ool*). We hypothesized that more aspiration (i.e., longer aspirated duration) would be present in both unaspirated and aspirated/-canonical variants in CDS than in ADS. If the duration of aspiration increases in both variants to a similar degree, the difference between the two should remain constant. However, considering Malsheen (1980) finding of less VOT overlap between voiced and voiceless consonants, it is possible that the aspiration of canonical [tʰ][3] could increase to a greater extent than that of an unaspirated variant. In this way, there will be a larger difference between aspirated and unaspirated variants in CDS than in ADS, signaling the different contexts of occurrence more robustly. We shall consider both possibilities.

The findings from the present study will expand our understanding of linguistic input to a language-acquiring child in at least two ways. First, this is one of the first systematic studies looking at the phonetic variants of /t/ in CDS and as mentioned above, this may be particularly useful since /t/ is especially rich in phonetic variants in American English. To our knowledge, there are only two other studies that systematically examined the phonetic variants of /t/ in CDS versus ADS in spontaneous speech: Foulkes et al. (2005) and Dilley et al. (2019). Foulkes et al. (2005) differs from the present study in that they investigated a dialect of British English with a slightly different set of phonetic variants and the participants were not the same for the ADS condition as they were for the CDS condition; the ADS data was taken from a corpus that was previously recorded in the same community. Therefore, comparisons could not be made between the ADS and CDS of individual speakers. Although Dilley et al. (2019) investigated some of the same variants as the present study, they did so in only utterance-medial position, focusing upon 'assimilable contexts' (contexts in which the following segment is expected to trigger phonetic variation) for word-final /t/, /d/, and /n/ (e.g., *gree<u>n b</u>all*). The tasks completed in ADS and CDS conditions were also different. In ADS, mothers were asked interview questions by a researcher, and in CDS, mothers played with their children on the floor using toys, and they were instructed to speak to their children as they normally would. Although Dilley et al. (2019) had more participants, their sessions were shorter, and may not have included many of the same target words across both conditions, due to the different tasks. In the present study, the same materials were used in both conditions, in an attempt to elicit the same set of target words for comparison.

A second way in which this study may expand our understanding more directly concerns the relationship between what the child hears and what the child produces. In a previous study, researchers found that children begin producing canonical variants of phonemes before they are able to produce systematic context-governed phonetic variation appropriately (Song, Shattuck-Hufnagel, & Demuth, 2015). Examining mothers'

[3] The International Phonetic Alphabet (IPA) was developed with the constraint that each symbol that it uses to designate a separate phoneme must correspond to contrastive phonological category in some known language, but the urge to capture more fine-grained differences has resulted in the development of additional symbols for narrower phonetic transcription. As more and more evidence accumulates that speakers and listeners attend to systematic fine-grained context-governed differences in the numeric values of the acoustic cues, this move to capturing more detail in the phonetic transcription increasingly seems wise, but should not be taken as evidence for finer and finer grained symbolic categories.

production of phonetic variants in the input will help us better understand why children start out by producing a fully articulated canonical variant of a phoneme, rather than context-appropriate phonetic variants. If the use of canonical forms is more prevalent in CDS than in ADS, this might suggest that mothers' speech more clearly signals the phonemic category. This could help children build a firm phonemic representation of the contrastive sound categories of their language early in development. However, if there is no difference in mothers' use of phonetic variants between CDS and ADS, this might suggest that children do not simply store and reproduce the context-driven variability of phonetic patterning in the adult language, but rather, they 'reanalyze' what they hear as they go through the learning of their phonological system, so that their productions reflect the canonical acoustic cues for a phoneme rather than the variable patterns they have heard. To test these hypotheses, we designed a study in which mothers of young children would be recorded speaking both to their children and to an adult researcher. The next section describes the method of this study.

2. Materials and methods

2.1. Participants

The participants in this study were eight mothers of young children under two years of age. The average age of the children was one year, four months. This age group encompasses the age range described by Ingram (1989) as the phonological stage of the first 50 words, when mothers may be more apt to speak to their children in a manner aimed to facilitate acquisition. Half of the children were boys and half were girls (See Table 1 for the age and gender of each participant's child). All participants were middle class, white, native speakers of American English. All but one participant grew up in Wisconsin (one grew up in Nebraska). All were the primary caregivers of their children. Three mothers spoke a second language at an advanced level, though none of the participants indicated they were teaching their children a language other than English. All participants were high school graduates. In addition, one participant had two associate's degrees, one had a bachelor's degree, three had a master's degree, and one had a doctorate. Participants were recruited through social media and personal invitation and were offered monetary compensation for their time.

Mothers were also specifically chosen over other adults and fathers because, traditionally, in the field of first language acquisition, researchers have studied CDS primarily in mothers (Buchan & Jones, 2014; Dilley et al., 2014; Foulkes et al., 2005; Kuhl et al., 1997). Also, Byrd (1994) found that the gender of the speaker affects a speaker's production of phonetic variants, with men producing many variants more often than women. The exclusion of one gender was an attempt at reducing variables, and also, in this culture, primary-caregiving mothers are generally more plentiful than primary-caregiving fathers, increasing the likelihood of finding qualifying participants.

Table 1
Ages and genders of participants' children, along with recording locations.

Participant	Child age	Child gender	Recording location
1	0;10	F	Home
2	0;11	M	Home
3	0;11	F	Home
4	1;1	M	Lab
5	1;2	F	Lab
6	1;4	M	Home
7	1;9	F	Home
8	1;11	M	Lab
Average	1;4		

2.2. Procedure

Testing took place in a quiet room either at University of Wisconsin-Milwaukee or at the location of participant's choosing (i.e., participant's home). The environments were comparable, being a quiet room in which the participant could interact, one-on-one, with their interlocutors. We acknowledge that it is possible that one recording location may have felt more 'formal' to participants than the other, possibly affecting the results, so this will be addressed further, in the Discussion. Each participant was outfitted with a Nady DKW-8U UHF wireless lavalier microphone, which was connected to a computer running the Audacity recording software. The lavalier microphone was chosen to allow the mothers to move freely while interacting with their children. Each mother participated in two recording sessions on the same day, a CDS session and an ADS session. To control for any effect that session order might have on production, the sequence of sessions alternated with each participant. Each session was approximately 20 minutes long.

Two storybooks containing target words with corresponding illustrations were created for this study. During both sessions, participants were asked to read the storybooks in their own words (in a paraphrased manner) to either their child (the CDS session) or an adult researcher (the ADS session). The participants were asked to paraphrase in order to elicit target words in more natural-sounding, running speech while still providing a guideline from the text to encourage the production of target words.

2.3. Stimuli

A total of 93 different target words were included in the storybooks to help elicit the production of qualifying words, which were generally words with a high frequency of occurrence (For a complete list of words analyzed, see Appendix). Four different phonetic variants were focused upon, corresponding to common phonetic variants of /t/: flapped, unreleased, glottalized, and unaspirated (See Table 2). *Flapped* corresponds to the variant of /t/ in which /t/ is flapped between two vowels if the second vowel is unstressed, *unreleased* corresponds to the variant of /t/ in which a word-final /t/ is unreleased, *glottalized* corresponds to the variant of /t/ in which /t/ is expressed as a glottal stop between a vowel and a syllabic nasal, and *unaspirated* corresponds to the characteristic lack of, or reduced, aspiration of /t/ following /s/. Words typically containing the canonical variant of /t/ (with clear closure, release burst and aspiration) were also included for comparison. Additionally, three different flap environments were targeted, because researchers have found that in child language acquisition, there is a difference in children's production rate of flaps in these environments: /t/ followed by *–er*, /t/ followed by *–y* or *–ie*, and /t/ followed by *–ing* (Klein & Altman, 2002). Because

Table 2
Phonetic variants and example target words.

Phonetic variant		Examples
Flapped [ɾ]	*-er*	water, better
	-y or *-ie*	kitty, naughty
	-ing	sitting, petting
Unreleased [t̚]		night, hot
Glottalized [ʔ]		button, mitten
Unaspirated [t]		stop, star
Canonical [tʰ]		toy, take

children's acquisition is so influenced by the input from caregivers, it makes sense to investigate whether mothers' rate of production of flaps in these environments differs, and if so, whether the results are in line with the findings from child language data. Examples of target words in each phonetic variant's environment are included in Table 2.

At least 8–10 different words corresponding to each variant's set of target word criteria were present in the storybooks. Although most target words were intended to come from the story, any words uttered by the mothers that fit the criteria set for any of the target variants were included in the analysis. The criteria for target words were thus: two-syllable words where /t/ occurs between two vowels, the second of which is unstressed, to elicit flapping; one-syllable words where /t/ occurs word-finally, preceded by a vowel, to elicit an unreleased variant; two-syllable words in which /t/ is preceded by a vowel and followed by an unstressed syllabic nasal, to elicit glottalization; one-syllable words starting with /st/, followed by a vowel, to elicit an unaspirated variant; and one-syllable words starting with /t/ followed by a vowel to elicit the canonical variant of /t/ (See Table 2, above, for an example of each).

Tokens were labeled for either utterance-medial (a word occurring within the utterance) or utterance-final (being the final word of the utterance) status. These two utterance positions were separated due to the findings of previous studies that found lengthening (Klatt, 1976; Oller, 1973) and strengthening effects (Fougeron & Keating, 1997) in utterance-final or phrase-final position. By separating utterance-final tokens from utterance-medial tokens, we have lessened the chance that we will attribute an effect to speaking condition (ADS vs. CDS) when it is simply an effect of being in utterance-final position. In the present study, utterance-medial tokens were defined as target words uttered in the middle of an intonational phrase that were not followed by any discernible pause before the next segment was uttered. Initially, tokens were analyzed for several other types of utterance-medial positions (depending on prosodic position and length of following pause), but because not enough tokens were produced in most of these positions to allow differential analysis, only true utterance-medial tokens, those not followed by any discernible pause, were included in the analysis.

Utterance-final tokens, in this study, were defined by two main criteria, both of which had to be met. The first criterion was a perceptual judgement of prosodic signals indicating the termination of a 'thought-unit.' The second criterion was acoustic measurement of the duration of the following pause, and although the lower limit was set to be at least 70 ms, the majority of tokens were followed by pauses of more than 200 ms. Our review of the literature suggests that there is no consensus on the length of pause that define separate utterances, probably because utterances can be defined differently depending on the listener, the type of speaking task, or even the theoretical stance of the investigator. Many of the studies relevant to the present study either did not focus on more than one utterance position (Dilley et al, 2014, 2019), or else simply did not separate the data by utterance position at all (Foulkes et al., 2005). Studies that have defined utterances have used different criteria; some didn't specify in terms of pauses or specific pause length (Henning, Striano, & Lieven, 2005; Kemper, Herman, & Nartowicz, 2005; Murray, Fees, Crowe, Murphy, & Henriksen, 2006; Schaffer & Crook, 1979; Toivola, Lennes, & Aho, 2009), some used 200 ms (Lowit, Marchetti, Corson, & Kuschmann, 2018), some used 300 ms (Fernald & Simon, 1984; Kitamura, Thanavishuth, Burnham, & Luksaneeyanawin, 2001), some even used 10 ms (Kemper, Herman, & Lian, 2003). Although the majority of the utterances in the current dataset were separated by pauses greater than 200 ms, there were a few utterances that required us to set the lower limit to 70 ms.

Table 3 shows the total number of tokens analyzed for each participant, separated by both utterance position and speaking condition. These numbers vary due to the spontaneous nature of the sessions. Because a few words occurred with high frequency in the stories, a limit was set to no more than 10 tokens per target word in each of the four possible conditions (utterance-medial ADS, utterance-final ADS, utterance-medial CDS, utterance-final CDS) to avoid the results being too influenced by a few individual words being produced over and over. The first 10 tokens were analyzed, as encountered chronologically in each recording.

2.4. Coding

After the recording sessions were completed, the resulting data were mined for tokens of the target words, which were analyzed using Praat (Boersma & Weenink, 2019). The criteria for each variant are clarified in the following sections, complete with example waveforms and spectrograms of target words uttered by participants, both with canonical [tʰ] and with the variant in question. Judgement of each token was made primarily based on acoustical analysis, but also backed by perception.

2.4.1. Flapping

Fig. 1 illustrates the two main possibilities for production in an environment conducive to flapping. Fig. 1(a) illustrates a word with an intervocalic /t/ (*pretty*) that is usually flapped, pronounced here as a canonical [tʰ]. A closure is clearly visible in the middle of the word (indicated with an arrow), followed by a release burst and aspiration. Fig. 1(b) displays the word *kitty*, with a flapped /t/. Compared to the example in Fig. 1(a), notice the extremely brief reduction in signal that indicates a fast closure and release (area of interest indicated with an arrow) and the dark bar across the bottom of the spectrogram that indicates continuation of voicing, which is not a usual acoustic characteristic of /t/, since it is a voiceless consonant. In addition to our interpretation of the acoustic signal, we made a perceptual judgement as to whether we also heard a flap-like variant instead of a canonical /t/.

Table 3
Token count for each participant in both speaking conditions and utterance positions.

Participant	Utterance-medial		Utterance-final		Total
	ADS	CDS	ADS	CDS	
1	185	208	65	70	528
2	199	200	59	85	543
3	174	171	85	95	525
4	162	129	62	59	412
5	190	169	57	83	499
6	198	166	56	66	486
7	240	214	90	108	652
8	196	216	54	98	564
Total	1544	1473	528	664	4209

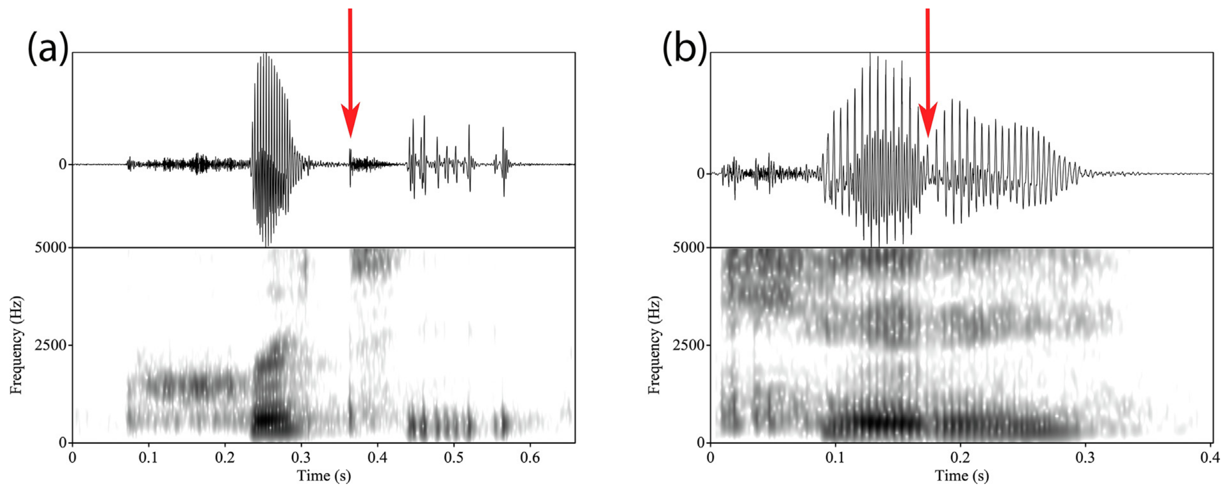

Fig. 1. (a) Shows the waveform and spectrogram of *pretty* with a fully released, canonical [t^h], and (b) shows those of *kitty*, produced with a flapped [ɾ]. Both tokens were produced in CDS.

2.4.2. Unreleased

For the unreleased variant of /t/, the target of the analysis was to determine whether or not a word-final post-vocalic /t/ was realized with a release burst. An example of a clearly released /t/ is displayed in Fig. 2(a), again showing a sharp release spike (indicated by an arrow). Fig. 2(b) shows an example of a word in which the signal does not include any acoustic evidence for the articulatory release of the /t/ closure. The word ends and there is no release burst visible (point of interest is indicated by an arrow). In both examples, you may notice that irregular pitch periods are visible in the signal, indicating that the /t/ was also glottalized, which is one of the many possibilities for a word-final /t/. For this variant, the corresponding perceptual judgement was whether or not an aspirated release burst was heard.

2.4.3. Glottalization

Fig. 3(a) illustrates a token containing the word-medial glottalization environment targeted in this study, uttered with a canonical [t^h], as shown by release bursts (area of interest indicated with an arrow). In Fig. 3(b), the canonical cues to a /t/ are replaced by irregular pitch periods reflecting a glottal constriction (area of interest indicated with an arrow). Perceptual judgment involved auditory detection of the sound of a glottal constriction instead of the silent or low-amplitude closure, release burst and aspiration of a canonical /t/.

2.4.4. Non-aspiration

Because the characteristics of the unaspirated variant of /t/ include a much-reduced (or no) duration of aspiration after release, the duration of aspiration was the target of this measure. Aspiration is difficult to measure separately from a release burst of a stop, so the VOT (a brief release burst followed by more continuous aspiration noise) of each of these tokens was actually what was measured. However, in keeping with the theme of what this variant represents (the presence or absence of aspiration), we will be referring to these VOT measures as *aspiration* throughout this paper. The aspiration duration was measured from the release of the stop to the beginning of a clear, regular pattern of pulses that signals the beginning of the vowel, even if there was still a little noise left in the periodic waveform. Aspiration was measured for each token recorded for both the word-initial [t^h] (where a canonical variant is expected) and [t] in /st/ (where an unaspirated variant is expected).

In Fig. 4(a), below, the aspiration of canonical [t^h] is illustrated. Notice the sharp spike signaling a clear release burst (indicated by the first arrow), first followed by irregular aspiration noise and then the regularization of the waveform that begins the vowel (indicated by the second arrow). Fig. 4(b) illustrates the reduced aspiration in the unaspirated variant of /t/. Again, a clear release burst is visible (relevant part is indicated by an arrow), but the duration is much shorter between the burst and regularization of the signal.

2.5. Reliability check

All data were coded by the first author. To assess intra-coder reliability, a randomly chosen 25% of the total data (2

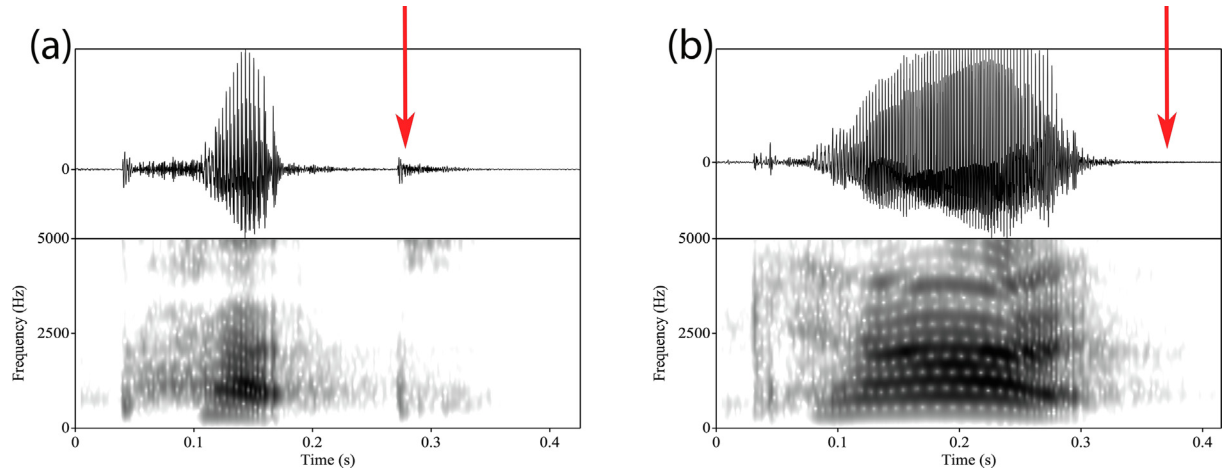

Fig. 2. (a) Shows the waveform and spectrogram of *pet* with a fully released, canonical [t^h], in CDS, while (b) shows those of *cat*, with an unreleased [t̚], in ADS.

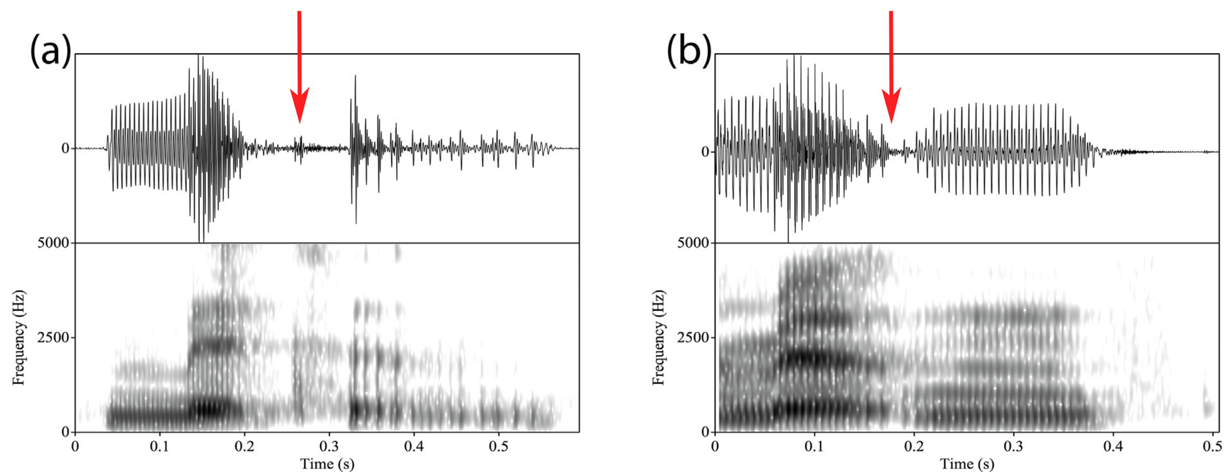

Fig. 3. Waveform and spectrogram of *mitten,* with a fully-released, canonical [t^h] in (a), and glottalized /t/ in (b), both in ADS.

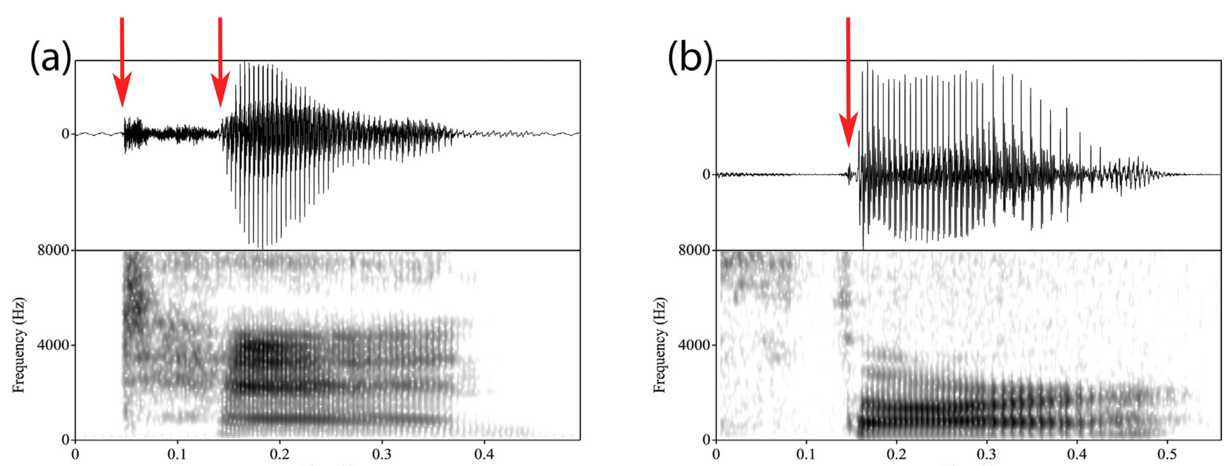

Fig. 4. (a) Shows the waveform and spectrogram of *ten* with an aspirated [t^h], in CDS, and (b) shows those of *star*, with an unaspirated [t], in ADS.

out of 8 mothers) were recoded by the same researcher, approximately 1.5 years after the original coding. The comparison between original and recoded data suggested high intra-coder reliability. For the three phonetic variants that involved binary decisions (flapped, released, and glottalized), only one out of 677 recoded tokens differed from the original coding; a token that was coded as 'unreleased' in the original coding was coded as 'released' in recoding. For the duration of aspiration noise, the average durations for aspirated [t^h] in the original coded data and recoded data were 67.1 ms (*SD* = 27.9) and 66.5 ms (*SD* = 28.3), respectively, based on 182 recoded tokens. The Pearson *r* correlation coefficient calculated between the two was 0.99 ($p < .001$). The average durations for unaspirated [t] in the original coded data and recoded data were 17 ms (*SD* = 6.2) and 16.6 ms (*SD* = 6.1), respectively, based on 135 recoded tokens. The Pearson *r* correlation coefficient calculated between the two was 0.98 ($p < 0.001$).

3. Results

To ascertain whether or not any differences between ADS and CDS were significant, mixed-effects regression models were utilized via the R statistical computing software (R Core Team, 2017). Mixed-regression models were particularly appropriate for our spontaneous speech data, which, by nature, are uneven across participants due to their individual speech choices. The speaking condition/listener (ADS or CDS) and utterance position (utterance-medial or utterance-final) were entered as fixed effects and participants and words were included as random effects. The dependent variable in each model was production (or non-production) of the target variant of /t/. Flapped, released, and glottalized variants involved only a binary decision, depending on whether or not the target variant was present. Therefore, for these variants, we used the *glmer* command in the lme4 package (Bates, Maechler, & Bolker, 2015), which is suited to analyze this type of binary data.[4] On the other hand, for continuous measurement data, such as aspiration duration, we needed to employ a different command, *lmer*, in the lme4 package (Bates et al., 2015).[5] For average aspiration duration difference between aspirated and unaspirated /t/, because we were working from averaged data for each participant (instead of raw), we did not include a random effect for word.[6] The data files that we used for the statistical analyses are available in the Supplementary Materials.

Now that we have explained our statistical models, we will start with a discussion of flapped [ɾ] results in comparison with canonical [tʰ], and then continue with discussions of results for unreleased [t̚], glottalized [ʔ], and unaspirated [t].

3.1. Flapping

The results for the three word-internal flap environments (-*er*, -*y*/-*ie*, -*ing*) have been collapsed into one flap category and presented in Fig. 5, because nearly all tokens collected for the flap environments were judged as flapped for all participants, in both utterance positions and speaking conditions, with most results close to 100%. As there was no variation, results were not significant for speaking condition ($\beta = 0.143$, $SE = 1.352$, $z = 0.105$, $p = .916$), utterance position ($\beta = 28.472$, $SE = 193.52$, $z = 0.147$, $p = .883$), or the interaction between the two ($\beta = -26.171$, $SE = 193.519$, $z = -0.135$, $p = .892$). This suggested that the mothers' production of flapped /t/ in the flap environment did not vary as a function of the listener or the utterance position.

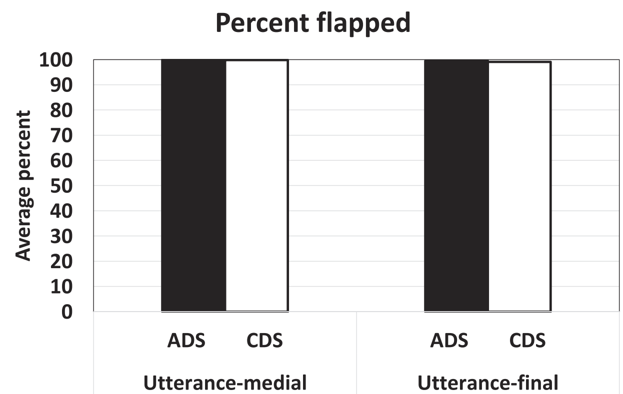

Fig. 5. Percentages of tokens of flapped /t/ (e.g., *water*), in both speaking contexts and utterance positions.

Table 4 shows individual results for each participant. Although there is some degree of overlap between what is presented on Fig. 5 and Table 4, it is useful to look at both individual data and group averages, because of the relatively small sample size. Notice in Table 4 that only three participants (Participants 2, 3, and 7) uttered the four unflapped tokens in this category, with Participant 2 uttering two of them.

Because the analysis based on a binary decision (flapped or not flapped) revealed so little variation in this area (almost every opportunity for a flapped variant did result in a flapped variant), we decided to look into the production of flaps more closely using acoustic measurements. In particular, we focused on the reduction (or non-reduction) of the flap gesture. An unreduced flap presents with a clear closure and loss of amplitude in the acoustic signal, as well as the cessation of vowel formants during the closure. When a flap is reduced, however, it resembles an approximant in the acoustic signal; the closure is not complete, and the vowel formants carry through, at least faintly, to the next vowel (Tucker, 2011). Of the 1,103 total tokens targeting flaps that were indeed flapped (1,099), 790 were judged to be reduced flaps (not showing a clear closure or cessation of formants in the acoustic signal), 257 were unreduced, and 52 were not possible to be measured precisely due to noise or vocalization by the listener (generally, this happened in the sessions with the children). As can be seen in Table 5, below, participants varied widely in the percentage of flap tokens that were unreduced. While participants, on average, did reduce flaps more often in CDS, the difference was small and not significant ($\beta = 0.404$, $SE = 0.266$, $z = 1.518$, $p = .129$). However, it was found that participants were less likely to reduce word-medial flaps (as in *water*) when words occurred in utterance-final position than when they occurred in utterance-medial position ($\beta = 1.115$, $SE = 0.245$, $z = 4.554$, $p < .001$). The interaction between speaking condition and utterance position was not significant ($\beta = -0.45$, $SE = 0.334$, $z = -1.347$, $p = .178$).

Unreduced flaps were further measured for duration, but reduced flaps were not, as there was no precise way to mark the exact beginning and end of such flaps. In his 2011 study, Tucker had devised a way to measure reduced flaps. He stated that duration of reduced flaps "was conservatively deter-

[4] An example of the code used for this model is: glmer(variant ~ listener*utteranceposition + (1|participant) + (1|word), family=binomial). 'Variant' stands for whichever variant chosen for the analysis and 'listener' indicates ADS or CDS – the type of person, adult or child, who was 'listening.'

[5] For the aspiration of aspirated and unaspirated /t/ and other continuous variables, we used this code: lmer(variant ~ listener*utteranceposition + (1|participant) + (1|word)).

[6] Here is the code used for average aspiration duration difference: lmer(averagedifference ~ listener*utteranceposition + (1|participant)).

Table 4
Number of tokens flapped/number of tokens targeting flapping (percentage of flapped tokens) for each participant.

Participant	Utterance-medial				Utterance-final				Total	
	ADS		CDS		ADS		CDS			
1	48/48	100%	54/54	100%	16/16	100%	22/22	100%	140/140	100%
2	53/53	100%	56/56	100%	12/12	100%	24/26	92.3%	145/147	98.6%
3	44/44	100%	46/46	100%	24/25	96%	28/28	100%	142/143	99.3%
4	46/46	100%	41/41	100%	17/17	100%	17/17	100%	121/121	100%
5	45/45	100%	33/33	100%	14/14	100%	23/23	100%	115/115	100%
6	51/51	100%	46/46	100%	13/13	100%	19/19	100%	129/129	100%
7	64/64	100%	56/57	98.3%	21/21	100%	33/33	100%	174/175	99.4%
8	42/42	100%	49/49	100%	10/10	100%	32/32	100%	133/133	100%
Total Avg	393/393	100%	381/382	99.8%	127/128	99.5%	198/200	99%	1,099/1,103	99.6%

Table 5
Number of flap tokens unreduced/number of flapped tokens, excluding 52 unmeasured tokens (percentage of flap tokens unreduced) for each participant.

Participant	Utterance-medial				Utterance-final				Total	
	ADS		CDS		ADS		CDS			
1	4/48	8%	5/53	9%	3/16	19%	3/20	15%	15/137	11%
2	3/53	6%	10/54	19%	7/12	58%	8/21	35%	28/140	20%
3	7/44	16%	10/46	22%	5/24	20%	6/28	21%	28/142	20%
4	15/39	39%	9/36	25%	9/16	56%	9/15	60%	42/106	40%
5	28/45	62%	21/30	70%	13/14	93%	14/19	74%	76/108	70%
6	4/51	8%	2/37	5%	0/13	0%	5/17	29%	11/118	9%
7	16/63	25%	11/55	20%	12/21	57%	7/33	21%	46/172	27%
8	3/42	7%	2/44	5%	2/10	2%	4/28	14%	11/124	9%
Total Avg	80/385	21%	70/355	20%	51/126	40%	56/181	31%	257/1,047	25%

mined by a combination of 'best-fit' judgment and selecting the distance between the half-way points along the intensity curve between peak and minimum intensity" (p. 314) and admits that "the duration measure may overestimate the duration of the consonant by including the approach and retraction of the tongue during the reduced productions" (p. 314). Because of this and because comparing flap duration was not the original aim of this study, we opted not to measure reduced flaps. Regarding the duration of unreduced flaps, the results showed that there was no effect of speaking condition (β = 0.264, *SE* = 2.16, *t* = 0.122, *p* = .903) or utterance position (β = $-$3.086, *SE* = 2.086, *t* = $-$1.48, *p* = .14), nor was there a significant interaction (β = 0.235, *SE* = 2.823, *t* = 0.083, *p* = .934). Thus, the acoustic measurements reinforced the result based on a binary decision that flap productions did not vary as a function of these factors. Individual data are presented in Table 6.

3.2. Unreleased

Although the variant of interest here is the unreleased variant [t̚], we decided to present the unreleased variant results in terms of percentage released, rather than percentage unreleased, as whether or not a token ended with a clear closure and a release burst was the most consistent way to measure a variant that is concerning the absence of release. Therefore, a larger percentage released means that fewer tokens were unreleased.

As is visible in Fig. 6, in utterance-medial position, participants released about 2% of tokens in both speaking conditions. This means that an unreleased variant was used about 98% of the time. In utterance-final position, as expected, tokens were released more often overall, though they were released much more often in CDS. In ADS, 29% of tokens in this category were released, whereas, in CDS, 48% of tokens were released. Results were found to be highly significant for speaking condition (β = 1.12, *SE* = 0.273, *z* = 4.096, $p < .001$), as well as utterance position (β = $-$3.773, *SE* = 0.541, *z* = $-$6.98, $p < .001$). The interaction, however, was not significant (β = $-$1.035, *SE* = 0.568, *z* = $-$1.822, *p* = .069). This means that mothers were less likely to employ an unreleased variant in CDS than they were in ADS, and also less in utterance-final position than in utterance-medial position.

Table 6
Average duration (in ms) of flapped tokens judged unreduced for each participant. Number of tokens measured precedes each duration.

Participant	Utterance-medial				Utterance-final				Total	
	ADS		CDS		ADS		CDS			
1	4	33.1	5	39.7	3	42.1	3	29.3	15	36.3
2	3	28.9	10	29.7	7	40.2	8	49.3	28	37.8
3	7	34.8	10	38.6	5	37.4	6	34.5	28	36.1
4	15	30.9	9	31.8	9	35.4	9	32.1	42	32.3
5	28	30.2	21	30.9	13	34.2	14	33.9	76	31.8
6	4	28.8	2	39.3	0	N/A	5	30.3	11	31.4
7	16	20.2	11	21.9	12	20.7	7	18.5	46	20.5
8	3	32.3	2	24.4	2	32.8	4	46.7	11	36.2
Total Avg	80	29.1	70	30.7	51	32.5	56	32.4	257	31.4

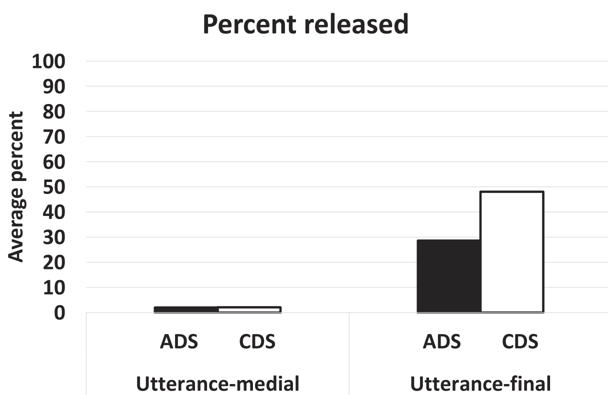

Fig. 6. Percentages of tokens of released /t/ in word-final context (e.g., *night*) in both speaking contexts and utterance positions.

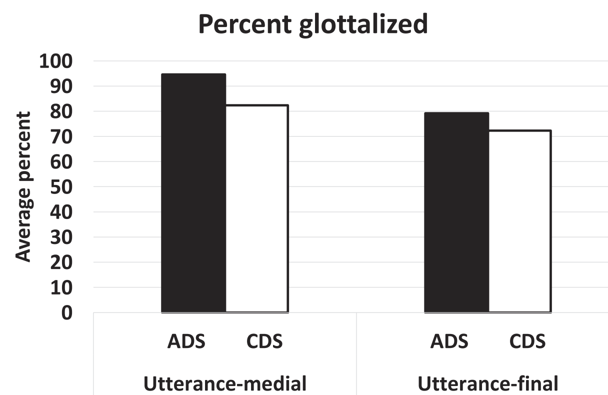

Fig. 7. Percentages of tokens of glottalized /t/ (e.g., *button*) in both speaking contexts and utterance positions.

Table 7 shows individual percentages of released tokens. In utterance-final position, all participants followed the pattern of more release in CDS than in ADS. However, in utterance-medial position, only Participants 5, 6, and 8 released more often in CDS than in ADS, though it is worth noting that the amount of release in this position was very low across the board.

3.3. Glottalization

Fig. 7, below, displays the average percent glottalized across participants for the glottalized [ʔ] variant. In utterance-medial position, participants glottalized relevant tokens 95% of the time in ADS, as opposed to 82% in CDS. Utterance-finally, fewer tokens were glottalized overall, though in ADS, participants glottalized 79% of tokens, while in CDS, they glottalized only 72%. Although percent glottalized did not vary significantly as a function of speaking condition ($\beta = -0.508$, $SE = 0.44$, $z = -1.156$, $p = .248$), there was a significant interaction between speaking condition and utterance position ($\beta = -1.352$, $SE = 0.629$, $z = -2.149$, $p = .032$). The effect of utterance position was also significant ($\beta = 2.143$, $SE = 0.516$, $z = 4.157$, $p < .001$).

In order to understand the nature of the significant interaction between speaking condition and utterance position, we compared ADS to CDS in utterance-medial and utterance-final positions, separately. Utterance-medially, the effect of the listener was highly significant; participants glottalized tokens significantly more in ADS than in CDS ($\beta = -1.901$, $SE = 0.479$, $z = -3.972$, $p < .001$). Utterance-finally, although participants also tended to glottalize tokens more often in ADS, the difference was not significant ($\beta = -0.455$, $SE = 0.423$, $z = -1.075$, $p = .282$).

For this variant, there was greater individual variability than for any other variant in this study. About half of the participants (Participants 1, 4, 5, 6, 7), almost always produced glottalization, while the rest glottalized far less often, and showed large differences in amount of glottalization between ADS and CDS. Most notably, in utterance-medial CDS, Participant 8 glottalized tokens only 24% of the time (vs. 96% in ADS), and in utterance-final CDS, both Participant 2 and Participant 8 produced very few glottalized tokens, 20% and 10%, respectively (vs. 50% in ADS, for both Participants) (Table 8).

3.4. Non-aspiration

Recall that we considered two possibilities for results in this category. The first possibility is that the average duration of aspiration would increase similarly for both aspirated and unaspirated /t/ in CDS (as opposed to shorter durations expected in ADS). The second possibility is that the difference between them would be larger in CDS than in ADS. To test these possibilities, it was deemed useful to first compare the measurements of unaspirated [t] with aspirated [tʰ], to see how much each participant differentiates the two variants by duration of aspiration.

In Fig. 8, the vertical axis shows the average duration of aspiration for /t/ in the word-initial position, where aspirated (or canonical) [tʰ] is expected (as in *top*), and for /t/ after /s/, where unaspirated [t] is expected (as in *stop*). Regarding unaspirated [t] (i.e., in the post-[s] environment; white bars),

Table 7
Number of tokens released/number of tokens targeting release (percentage of released tokens) for each participant.

Participant	Utterance-medial				Utterance-final				Total	
	ADS		CDS		ADS		CDS			
1	0/59	0%	0/63	0%	3/16	25%	5/16	31%	8/153	5%
2	0/57	0%	0/61	0%	5/17	29%	9/19	47%	14/154	9%
3	4/51	8%	2/52	4%	9/21	43%	14/22	64%	29/146	20%
4	0/49	0%	0/42	0%	0/12	0%	3/11	27%	3/114	3%
5	2/61	3%	4/62	7%	1/18	6%	7/22	32%	14/163	9%
6	0/63	0%	1/50	2%	10/17	59%	15/20	75%	26/150	17%
7	4/81	5%	2/67	3%	10/25	40%	21/34	62%	37/207	18%
8	0/66	0%	1/72	1%	6/22	27%	13/28	46%	20/188	11%
Total Avg	10/487	2%	10/469	2%	44/147	30%	87/172	51%	151/1,275	12%

Table 8
Number of tokens glottalized/number of tokens targeting word-medial glottalization (percentage of glottalized tokens) for each participant.

Participant	Utterance-medial				Utterance-final				Total	
	ADS		CDS		ADS		CDS			
1	21/21	100%	23/23	100%	8/8	100%	12/12	100%	64/64	100%
2	19/22	86%	9/18	50%	5/10	50%	3/15	20%	36/65	55%
3	16/20	80%	17/20	85%	8/14	57%	11/17	65%	52/71	73%
4	18/18	100%	14/14	100%	13/13	100%	14/15	93%	59/60	98%
5	21/21	100%	18/18	100%	6/6	100%	14/14	100%	59/59	100%
6	22/22	100%	16/16	100%	8/9	89%	9/10	90%	55/57	97%
7	19/20	95%	21/21	100%	14/16	88%	13/13	100%	67/70	96%
8	22/23	96%	5/21	24%	4/8	50%	1/10	10%	32/62	52%
Total Avg	158/167	95%	123/151	82%	66/84	79%	77/106	73%	424/508	84%

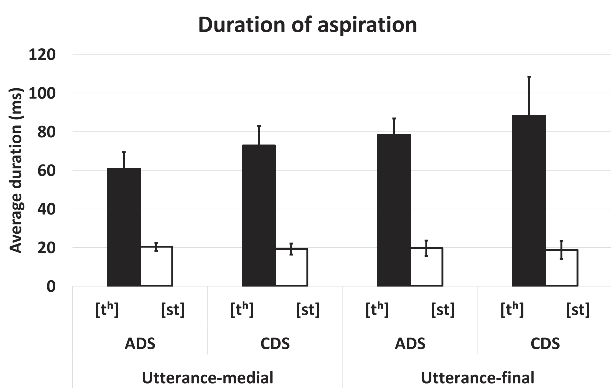

Fig. 8. Average duration of aspiration (in ms) in each context. Error bars indicate standard deviation.

notice that the average duration is comparable across utterance positions and speaking conditions. However, note that results for aspirated [tʰ] (black bars) were more varied, with longer average durations overall in utterance-final position. Crucially, we found longer average durations for aspirated [tʰ] in CDS than in ADS. The mixed-effects regression analyses showed that the differences were significant for both speaking condition and utterance position. Participants had significantly longer average aspiration durations in CDS than in ADS overall (β = 8.663, *SE* = 3.859, *t* = 2.245, *p* = .025), and they also had longer average aspiration durations in utterance-final position than in utterance-medial position (β = −11.537, *SE* = 4.039, *t* = −2.857, *p* = .004). Specifically, in utterance-medial position, the average duration of aspiration for aspirated [tʰ] was 61 ms (*SD* = 9) in ADS and 73 ms (*SD* = 10) in CDS. In utterance-final position; the average durations for aspirated [tʰ] were 78 ms (*SD* = 8) and 88 ms (*SD* = 20), in ADS and CDS, respectively. The interaction between speaking condition and utterance position was not significant (β = 3.037, *SE* = 4.436, *t* = 0.685, *p* = .494).

In summary, the majority of participants followed the pattern of significantly longer average duration of aspiration for aspirated [tʰ] for both the CDS speaking condition and utterance-final position, though there was some individual variation.[7] For unaspirated [t] (/t/ following /s/), average duration of aspiration results are fairly similar to one another and there appears to be no clear pattern. Indeed, these data did not show a significant difference in speaking condition (β = −1.464, *SE* = 1.221, *t* = −1.199, *p* = .231), utterance position (β = 0.48, *SE* = 1.096, *t* = 0.438, *p* = .662), or the interaction between the two (β = 0.455, *SE* = 1.457, *t* = 0.312, *p* = .755).

Fig. 9 shows the differences in aspiration duration between canonical [tʰ] and unaspirated [t] for ADS and CDS in each utterance position. Because the average duration of aspiration did not increase similarly for both aspirated and unaspirated /t/ in CDS, we can rule out the first probability we considered (that both would increase similarly). Rather, the latter possibility is what we found when we examined the size of differences between canonical [tʰ] and unaspirated (or post-[s]) [t], primarily due to the increased average duration of aspirated [tʰ] in CDS compared to ADS. Indeed, we found greater differences in CDS than in ADS (β = 10.732, *SE* = 4.874, *t* = 2.202, *p* = .036). We also found greater differences in utterance-final position than in utterance-medial position (β = −18.372, *SE* = 4.874, *t* = −3.769, *p* < .001). The interaction between speaking condition and utterance position was not significant (β = 2.613, *SE* = 6.893, *t* = 0.379, *p* = .708). In short, the duration of aspiration difference measures displayed in Fig. 9 show that the contrast between canonical [tʰ] and unaspirated [t] was greater in CDS independent of utterance positions.

Table 9 shows the average aspiration duration differences for each individual. All participants, with the exception of Participant 6, show a greater average aspiration duration difference in both CDS and utterance-final position. Participant 6 did have a greater average duration difference between ADS and CDS in utterance-final position, but in utterance-medial position, she had roughly the same amount of difference in both speaking conditions, and not a greater amount than she has in ADS.

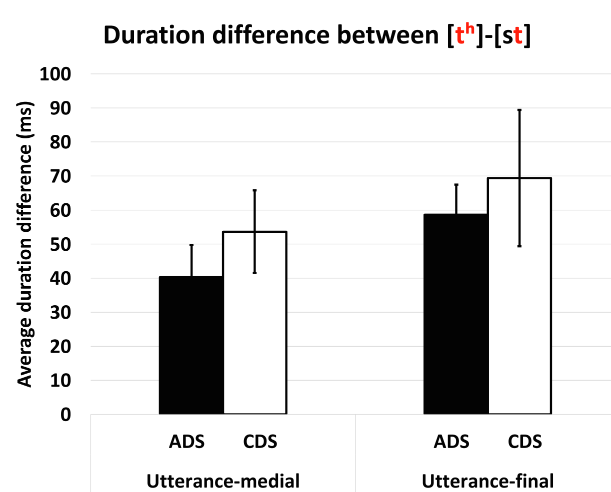

Fig. 9. The differences between canonical [tʰ] and unaspirated [t] values for average duration of aspiration. Error bars indicate standard deviation.

Table 9
Average aspiration duration difference (in ms) between aspirated and unaspirated /t/ for each participant.

Participant	Utterance-medial		Utterance-final	
	ADS	CDS	ADS	CDS
1	59.9	68.8	55.5	72.8
2	31.2	59.2	64.7	97.5
3	39.7	59.7	54.9	63.2
4	45.3	62.1	75.3	90
5	38	57.5	64.2	63.9
6	32.8	32.4	49.4	33
7	42.9	49	55.9	77
8	32.3	40.5	49.4	57.9
Average	40.3	53.6	58.7	69.4

So far, we have compared mothers' production of four types of phonetic variants of alveolar stop /t/ between ADS and CDS. While some of these results appear to confirm our hypotheses that mothers produce fewer phonetic variants of /t/ in CDS than in ADS, and that aspiration duration difference may increase in CDS in comparison with ADS, it might be argued that these differences have merely arisen due to a change in speaking rate, as a slower speaking rate is a noted feature of CDS (Bernstein Ratner, 1985; Fernald & Simon, 1984). However, it does not immediately follow that a slower speaking rate is responsible for our results. Note that the flapping, glottalization, and unaspirated variants did not differ significantly between ADS and CDS (though for glottalization, there *was* a significant interaction between speaking condition and utterance position). If speaking rate were the main reason behind any difference in results, we would expect these differences to apply to all target variants uniformly, in both utterance positions. Because that is not the case, something else may be at work here. To delve into this question more deeply, we chose to compare speaking rates, in particular articulation rates, between ADS and CDS. The results are given in the next section.

3.5. Relationship between production of phonetic variants and articulation rate

The utterances in our spontaneous recordings sometimes included pauses interrupted by utterances from the participants' interlocutors (e.g., child). A version of each recording was altered to remove any speech by the interlocutor (or any overlapping speech) to ensure that the correct speaker was analyzed. For this reason, conventional speech rate (number of syllables divided by total duration including pauses, as defined by De Jong & Wempe, 2009) was not deemed to be a reliable measure for this study, because pauses between some utterances had been shortened or removed. However, it was possible to calculate articulation rate, which is defined as the total number of syllables uttered by the participant, divided by their total phonation time (total speaking time, excluding pauses). The articulation rates for each participant in both ADS and CDS were extracted via Praat script (De Jong & Wempe, 2009), and are given in Table 4. In addition, average syllable duration was calculated, but as the results were very highly correlated with articulation rate ($r = -0.998$, $p < .001$, in both ADS and CDS), these analyses will not be reported separately here.

As can be seen in Table 10, nearly all participants had a faster articulation rate in ADS (all except Participant 5, who had basically the same rate in both speaking conditions). The participants were on average 0.26 syllables per second slower in CDS than in ADS.

Pearson correlation analyses were subsequently conducted using R (R Core Team, 2017) to ascertain whether articulation rate significantly affected our data, though we acknowledge that the present study does not meet the minimum sample size for correlation analysis (according to Bonett & Wright, 2000), and therefore the results should be taken with a degree of caution. First, participants' articulation rates in ADS were compared to their rates in CDS, and a significant positive correlation was found between them ($r = 0.758$, $p = .029$). This indicates that if a participant's articulation rate in one speaking condition was relatively faster or slower than that of others, her articulation rate in the other speaking condition was also relatively faster or slower, in the same direction. This is what was expected, and it confirms that the participants were all performing in a similar way in both conditions. Knowing this allows for further analysis into how articulation rate correlated with participants' production of the target variants. Table 11 gives the correlation coefficients (*r* value) between articulation rate and all four target variants, in both ADS and CDS, and their corresponding significance levels. Results for the two utterance positions were collapsed, as there was no statistical difference between them. Note that only glottalization in CDS was significantly correlated with articulation rate ($r = 0.761$, $p = .028$), suggesting that those who had a faster articulation rate in CDS also produced glottalization more frequently, but that other variant forms were not related to articulation rate in this way.

In addition to examining the correlation between articulation rate and aspiration duration differences (as shown in Table 11), we also examined correlations between articulation rate and the duration of the aspirated (canonical) and unaspirated variants respectively, to make sure that the raw durations were not correlated with articulation rate. These correlations were also not significant, which may be expected, due to the non-significance of correlations between articulation rate and the difference in aspiration duration.

Because the correlation between articulation rate and glottalization in CDS was the only significant correlation found, that relationship warranted a closer look. Fig. 10(a) and (b) shows the correlation between articulation rate and glottalization in ADS and CDS, respectively. It is clear in Fig. 10(a) that glottalization is more prevalent in ADS than in CDS, and there was no significant correlation between articulation rate and production of glottalization. The lack of significant correlation in ADS may simply be due to lack of individual differences in production of glottalization. That is, the majority of the tokens were glottalized in ADS, and once the ceiling is reached, there is

Table 10
Articulation rate for each participant (number of syllables/phonation time).

Participant	ADS	CDS	Difference (ADS-CDS)
1	4.3	4.32	−0.02
2	4.43	4.12	0.31
3	4.54	4.42	0.12
4	4.59	4.35	0.24
5	4.73	4.45	0.28
6	4.98	4.67	0.31
7	5.03	4.43	0.6
8	4.4	4.14	0.26
Average	4.63	4.36	0.26

Table 11
Articulation rate correlation data for the four target variants.

	Flapping		*Release*		*Glottalization*		*Aspiration duration difference*	
	ADS	CDS	ADS	CDS	ADS	CDS	ADS	CDS
r value	0.128	0.521	0.415	0.513	0.348	0.761	−0.439	−0.579
p value	0.763	0.186	0.306	0.193	0.399	0.028*	0.276	0.133

* Indicates that correlation was significant at the 0.05 level.

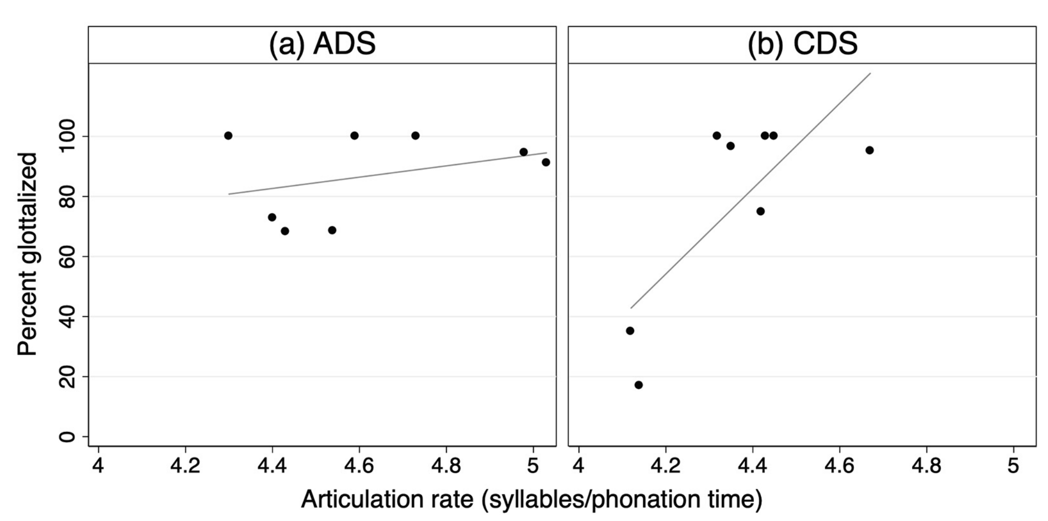

Fig. 10. Scatterplots of the relationship between articulation rate and glottalization in (a) ADS and (b) CDS.

nowhere higher to go. As shown in Fig. 10(b), participants' production of glottalization was significantly correlated with articulation rate in CDS. The findings suggest that participants with a slower articulation rate produced less glottalization, and as the articulation rates of participants increased, their individual percentages of glottalization production increased as well. In Fig. 10(b), it is also noteworthy that for many of the participants, as an articulation rate of about 4.3 is reached, the percentage of glottalized tokens increases closer to ceiling.

Slower articulation rates in CDS are what we would expect, because a slower speaking rate is an attested feature of CDS, and since a slower speaking rate is so integral to CDS, it would be impossible to tease it apart from other aspects of CDS without creating an unnatural speaking situation. However, our findings suggest that mothers' propensity to produce canonical [tʰ] over variants is not necessarily a result of slower articulation rate. If it were, we would expect similar correlations between production of variants and articulation rate across the board. However, there was significant correlation only for glottalization, not for the flapping environment, nor for release. We would also have expected the duration of aspiration to lengthen for both canonical [tʰ] and unaspirated [t] in CDS, but it lengthened selectively only for canonical [tʰ]. In the following section, we discuss the implications of our findings for the acquisition of phonemic categories and phonetic variants.

4. Discussion

Although previous work in the field of first language acquisition has shown CDS to be clearer and more carefully articulated than ADS, the bulk of this work has been on the phonemic (contrastive) level, as opposed to the phonetic (non-contrastive). This has resulted in a gap in the understanding of how caregivers speaking to children produce context-appropriate phonetic variants that make a native speaker sound 'native.' To address this gap, the present study compared mothers' production of four types of phonetic variants of alveolar stop /t/: flapped (as in *butter*), unreleased (as in *got*), glottalized (as in *button*), and unaspirated (as in *stop*). Due to previous findings of relative clarity in CDS versus ADS, it was hypothesized that mothers would produce fewer variants of /t/ in CDS than in ADS, i.e., more canonical tokens. To test this hypothesis, eight native American English-speaking mothers of children under the age of two years were recorded while they told stories out of picture books to their child and to an adult researcher. Results showed significantly less use of phonetic variants of /t/ in CDS, i.e., fewer unreleased tokens, fewer glottalized tokens in utterance-medial position, and also a significantly greater difference in aspiration duration between canonical [tʰ] and unaspirated [t].

Results obtained for the unreleased and glottalized variants show that mothers do indeed produce fewer phonetic variants in CDS than they do in ADS. Also, while the results for the unaspirated variant do not show that mothers are producing longer VOTs for /t/ following /s/ (i.e., there is no evidence indicating that mothers produce canonical long-aspiration variants in this context), it is interesting that mothers did increase the VOT for canonical (aspirated) [tʰ], resulting in a larger difference between these two variants (i.e., singleton onset /t/ with longer aspiration intervals, cluster-onset /st/ with shorter aspiration intervals). This might allow children to more easily distinguish between them. Overall, these results show that while mothers do produce a range of phonetic variants of /t/ in CDS, their productions vary less in CDS than in ADS.

Our findings reinforce previous studies that found less-frequent production of variants due to phonetic processes in CDS. The recent study by Dilley et al. (2019) similarly found that mothers produced fewer instances of phonetic variants of /t/ in CDS than in ADS. Although direct comparison of the results of the two studies is not possible, due to little overlap between the target phonemes and variants, it is possible to compare the results for the category targeting word-final unreleased variants of /t/ (e.g., *ca*t̚) in the present study with the results in the Dilley et al. study. Our results for utterance-medial tokens were that mothers produced canonical /t/ only 2% of the time in this context, in both ADS and CDS. Conversely, Dilley et al. (2019) found that mothers produced canonical /t/ 27% of the time in ADS and 41% in CDS, which is quite different. However, there are methodological differences between this study and theirs which could be possible reasons for the difference in outcomes. In Dilley et al. (2019), it was possible that utterance-medial tokens were followed by pauses less than 100 milliseconds. In contrast, in the present study, utterance-medial tokens were almost immediately followed by another segment, without any discernible pause, because utterance-medial tokens that are followed by different lengths of pauses were omitted, to control for the possible effects of being at the boundaries of varying sizes of constituents. This is a likely explanation of why the alveolar stop was observed to be less often released in utterance-medial position in the present study.

Another reason why direct comparison between this study and the Dilley et al. (2019) study is difficult is that in Dilley et al. (2019), the tasks in ADS and CDS were not the same, and the recording sessions lasted 2–5 minutes only. In contrast, in the present study, the tasks in ADS and CDS were identical, and lasted roughly 20 minutes each. While the number of participants differed quite a bit (53 for Dilley et al. (2019) and 8 for the present study), the token counts for /t/ in this position (word-final, utterance-medial) were comparable for the two studies: Dilley at al. obtained a total of 948 (538 in ADS and 410 in CDS) and the current study obtained 956 (487 in ADS and 469 in CDS). On the one hand, having a larger pool of participants may give a good representation of the population we are aiming to describe (i.e., mothers), but on the other, longer recording sessions may give participants time to become more comfortable in their task and for any effect of the novelty of the situation to be reduced. That is, even when participants are asked to speak as they normally would, the simple situation of being in a research study can have an effect, especially at the beginning of the session, when the situation is new, so a longer session may be beneficial in providing more time for the participants to relax and speak more normally. Aside from session length, the ADS tasks in these two studies also differed from one another in that Dilley et al. (2019) used an interview format, which could be interpreted as a more formal situation than telling a researcher a story, which was the method used in the present study. This may have actually inhibited the production of phonetic variants in ADS. Conversely, the CDS format used by Dilley et al. (2019) was a play session with the child, which is possibly an even more informal speaking situation than the task employed in the current study, which was telling a story. Currently we do not have a clear understanding of the effects of different elicitation methods on the occurrence of phonetic variants. As more researchers investigate this issue, it will be interesting to see how their results compare.

Because we have recorded participants in two different locations, and have participants interacting with children of two different genders, one may also wonder if these variables had any effect upon our results. At this point we have no strong empirical evidence to offer on the issue, because the variables were not systematically manipulated in our experiment due to the fact that they were not part of our research question. Thus, our purpose here is simply to sketch out our observations and speculations. We will discuss the possible effect of location first. Recall that three of our participants (4, 5, and 8) were recorded in the lab and five (1, 2, 3, 6, and 7) were recorded at home. Participants recorded in the lab and those recorded in their homes were grouped separately, and the average production of each phonetic variant was calculated for each group (See Table B3 in the Appendix). One notable pattern was that the lab-recording group on average produced fewer releases (and therefore more of the unreleased variant) than the home-recording group, in both ADS and CDS, which might indicate the lab was a less 'formal'-feeling location. This pattern was unexpected as we would have expected that mothers to use a more formal register in the lab.

Likewise, to explore whether the gender of child interlocutor might have been an influence, the average production of each phonetic variant was computed with the participants grouped according the genders of the children (See Table B4 in the Appendix). Recall that the half of our participants (1, 3, 5, and 7) interacted with their daughter and the other half (2, 4, 6, and 8) interacted with their son during the recording. Overall, we found that the two groups performed similarly regarding ADS vs. CDS. One observation we would like to make is on the production of glottalization. In both utterance positions, we found that the difference between ADS and CDS in percent glottalization was bigger for male-child group than the female-child group. The female-child group produced more glottalization overall than the male-child group, regardless of the speech register. Studies have widely documented the gender difference in the production of glottalization in adults, with some reporting that females produce more glottalization than males (Byrd, 1994). However, we do not know if the gender of the listener will have similar effect on the speakers' production of glottalization, so this is a question for future studies. More crucially, we acknowledge that the small sample size is a limitation of the present study.

Even though mothers in this study produced some phonetic variants of /t/, the results indicate that the input young children receive still contains more instances of the canonical variant than the input an adult would receive. Mothers could be providing input both in canonical form and with context-appropriate phonetic variation, as a way of providing scaffolding to help children make a connection between underlying forms and their variants. If the child hears both *kitten* [ˈkɪtʰɪn] and *kitten* [ˈkɪʔn] from her mother, while referring to the same thing, this could give her entry into the understanding of both the phonological categories and how they may be realized in different contexts. The canonical form provides the strongest signal of the underlying category, which is what ties all of these variants together, and that form is also what children seem to produce first. Studies have shown that children start by producing the canonical, or underlying, form of a phoneme before they acquire its phonetic variants (Song et al., 2015). The results of this study could help to explain why children are learning underlying forms before surface forms, and possibly how mothers tailor their speech to aid their children's acquisition of phonological categories.

At this point, one may wonder why, if mothers do tailor their speech to aid their children's language acquisition, we do not find more use of the canonical form across the board, as our results show flapped forms and utterance-medial unreleased forms to be as prevalent in CDS as they are in ADS. Regarding utterance-medial tokens, the work done by Dilley et al. (2019) allows us to predict that the difference in percent unreleased between ADS and CDS would have been significant, had our criteria for utterance-medial tokens allowed for short pauses. Flapping, however, presents a different case, because flapping (unlike lack of release) occurred word-internally in this study. Recall that mothers in this study produced word-medial flaps nearly 100% of the time, even in CDS. This is particularly interesting considering the previous findings showing that young children consistently produce the canonical form in this context before they begin to produce flapped variants (Song et al., 2015). This suggests that children produce flaps less often than adults, despite receiving plenty of input containing flapped variants. A possible reason why flaps are so prevalent in both CDS and ADS produced by mothers is that word-medial flaps are simply obligatory in American English, once adult-like production of the language is acquired. Some researchers even characterize flaps as intrinsic phonetic variants, as they arise from a process that is triggered automatically in certain contexts, due to the movements of the articulators in the vocal tract from one segment to the next (Ladefoged & Johnson, 2006). That is, when producing flaps, the tongue often does not make full contact with the place of articulation, and this might be triggered automatically as the tongue travels from the position of one vowel to that of another. An example of an intrinsic variant would be how the alveolar nasal /n/ becomes dentalized when preceding the voiceless interdental fricative /θ/, such as in the word *tenth* - this variation is thought to be directly related to the tongue preparing to travel forward to the teeth to articulate the interdental fricative. The movement happens early, and so the placement of the closure in the oral cavity for the nasal consonant moves to a more forward location. Conversely, word-medial glottalization can be considered extrinsic, arising from a process that is not related to the movement of the articulators between adjacent sounds (Ladefoged & Johnson, 2006), and may therefore be more optional. Extrinsic variants are not necessarily similar to the phoneme from which they arise, and they cannot be explained by the physical movement from and to adjacent segments. This distinction aligns with the difference between the glottalization and flapping distributions in this study.

We note that variants of /t/ other than the specific variants investigated here are possible. Take, for example, the target word *kitten*, which often elicits word-medial glottalization of /t/ preceding a syllabic nasal (as in the pronunciation [ˈkɪʔn]). One alternative pronunciation of /t/ in this context might be a canonical /t/, [ˈkɪtʰɪn], but this /t/ could also be flapped, [ˈkɪɾɪn], if the vowel is not deleted between /t/ and /n/. Although this might be a possible way of producing the word in causal conversation, this variant was not produced in the present study. In principle, we chose to focus on the most common variants for these environments in American English, and whether canonical [tʰ] was produced instead. The issue of multiple possible variants is most relevant to the unreleased variant in this study, as there were several different unreleased variants employed (such as flapped or glottalized, anything other than a canonical [tʰ] with a full closure and audible release). For these tokens, in addition to noting whether or not the /t/ was released, the specific nature of the unreleased variant was also recorded, as well as the type of following segment (vowel, stop, fricative, etc.). However, further analysis of these results was deemed unnecessary due to the finding of no significant difference between ADS and CDS in the utterance-medial position, as that was the position in which the following segment would vary. In utterance-medial position, participants produced the canonical form of the stop with a clear release only in 2% of tokens in both speaking conditions. Again, the main question was whether or not a fully released, canonical [tʰ] was produced rather than a variant, so any other variant was counted as unreleased. Utterance-finally, every token was followed by a pause, and therefore the following environment for all of those tokens was identical.

Articulation rate was another factor taken into consideration as a possible influence on the results. However, correlational analyses revealed only one significant relationship, between articulation rate and glottalization in CDS. The articulatory mechanism behind these results is not certain – why, out of the four variants examined, did only glottalized /t/ correlate significantly with articulation rate? Although glottalized /t/ involves irregular pitch periods towards the end of the preceding vowel, it also often lacks the discernable stop closure and release burst associated with the canonical /t/, which could together take almost 100 ms (Byrd, 1993). Therefore, glottalized /t/, when it is produced without identifiable stop closure and release burst, could be expectedly shorter compared to the canonical /t/, which has the two components. When lacking a full closure and release in the oral tract, it is possible that glottalized /t/ is produced more rapidly than the canonical /t/. This may explain why glottalized stops are positively correlated with articulation rate. To our knowledge, little work has been done on the interaction between articulation rate and glottalization (although cf. Markó, Deme, Bartók, Gráczi, & Csapó, 2017), so this may warrant investigation in the future. In any case, critically, articulation rate does not appear to be the reason for all the differences in variant production found between ADS and CDS in this study, which raises the question: Why do these differences occur? In line with the literature supporting the role of the linguistic (pedagogical) function of CDS (Fernald et al., 1989, among many others), we argue that mothers may be producing language in a way that could facilitate the language acquisition of their children.

Previous work by Cho, Lee, and Kim (2014) into prosodic effects on /s/-stop clusters revealed that the effect of prosodic strengthening on an already short lag VOT (a voiceless stop consonant after /s/) was to shorten it further, which was interpreted as an enhancement of the context-appropriate phonetic feature (unaspirated) of a voiceless stop following /s/. One of the possibilities we had considered, regarding our target unaspirated variant of /t/ ([t] following /s/), was that there would be longer durations of aspiration in CDS than in ADS, which would be viewed as a move more toward a canonical /t/. We found, however, that the average duration of aspiration for unaspirated /t/ remained nearly constant across both speaking condition and utterance position, and we argued that the larger difference between aspirated and unaspirated /t/ in CDS, created by the increase in aspiration durations for aspirated /t/, may make the difference more salient to children. An investigation including tokens in positions of differing prominence in ADS and CDS could uncover whether this type of strengthening is enhanced further in speech addressed to a child rather than to an adult. Therefore, a promising area for future research would be to investigate how prosody interacts with speaking condition regarding context-appropriate phonetic variation in general, to ask how prominence and boundary markings could possibly strengthen the effects we have found.

Another point we wish to raise concerns how CDS compares with clear speech produced to aid an adult listener. It has been well documented in the literature that CDS and clear ADS share many acoustic properties in common, such as slow speaking rate and hyperarticulated vowels (Lindblom, 1990; Uther, Knoll, & Burnham, 2007). However, the literature on CDS and the majority of work on clear ADS have been on the phonemic level, as opposed to the phonetic. Therefore, this is an area for further research which might provide a more complete understanding of clear speech and its role in both speech communication and language learning.

5. Conclusions

This study set out to fill a gap in the understanding of how caregivers produce different phonetic variants in CDS, compared to in ADS, for the phoneme /t/ in American English. Results showed that while there is little difference between ADS and CDS in the use of the flap variant of /t/, there were significant differences for other variants. As expected, mothers released word-final /t/ (therefore not producing the 'unreleased' variant) more often in CDS than in ADS. Concerning word-medial glottalized /t/, when the significant interaction between utterance position and speaking condition was broken down, mothers produced significantly more glottalization in ADS than in CDS in utterance-medial position. Finally, the difference between the VOT of aspirated and unaspirated variants was found to be significantly larger in CDS than in ADS. Overall, these results suggest that mothers are more likely to produce canonical variants in CDS than in ADS, although a portion of their CDS still contains non-canonical variants as well. Mothers could be providing input in both canonical form and with some degree of context-appropriate phonetic variation, as a way of providing scaffolding to help children make a connection between underlying forms and their variants. These results could help explain why children produce canonical forms that are more expressive of underlying forms (phonemes) before they produce surface forms that correspond to context-appropriate phonetic variants.

CRediT authorship contribution statement

Robin Fritche: Conceptualization, Investigation, Methodology, Resources, Formal analysis, Validation, Writing - original draft, Writing - review & editing. **Stefanie Shattuck-Hufnagel:** Conceptualization, Writing - original draft, Writing - review & editing. **Jae Yung Song:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing.

Acknowledgements

The authors would like to thank Aimee Roekle for file preparation and research assistance, Tessa Culleton for script writing assistance, and Anne Pycha and Hanyong Park, as well as our anonymous reviewers, for thoughtful critique and suggestions.

Appendix A. Target words

*Target words not present in the stories but included in the analysis.

Flap [-ɾɚ]	eight	**Aspiration**
better	feetfeet	take
brighter	foot*	tale*
butter	fruit*	talk
heater	get	tall
later	got	taste
latter*	great	taught
letter*	hate	tea
litter	heat	tears
matter	height	tell
sweater	hot	ten
water	let	Tess
	light	tie
Flap [-ɾi]	lit	time
beauty	lot	tired*
Betty	might	toast
bitty*	night	told
catty	not	tone*
city	out	too
itty*	pet	took
kitty	put	touch
Mattie	right	turn
naughty	sat	two
pretty	shut*	
sweetie*	sit*	**Non-aspiration**
	taught	stalk
Flap [-ɾɪŋ]	thought	Stan
eating	tie	stand
fighting	wait	star
floating	white	stare
getting		start
petting	**Glottalization**	stay
putting	bitten	steal
sitting	button	step
writing	cotton	still
	eaten	stop
Unreleased	gotten	store
bet	kitten	stuck
bit	mitten	stuff*
bright	rotten	style*
cat	smitten	
cute*	sweeten	
eat	written	

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wocn.2021.101056.

References

Bates, D., Maechler, 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>.

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

Bernstein Ratner, N. (1984b). Phonological rule usage in mother-child speech. *Journal of Phonetics, 12*(3), 245–254.

Bernstein Ratner, N. (1985). Dissociations between vowel durations and formant frequency characteristics. *Journal of Speech and Hearing Research, 28*(2), 255–264.

Bernstein Ratner, N., & Luberoff, A. (1984). Cues to post-vocalic voicing in mother-child speech. *Journal of Phonetics, 12*(3), 285–289.

Broen, P. A. (1972). The verbal environment of the language-learning child. ASHA Monography, 17. Washington, D.C: American Speech and Hearing Association.

Boersma, P. & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.09, retrieved 20 January 2019 from http://www.praat.org/.

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika, 65*(1), 23–28.

Buchan, H., & Jones, C. (2014). Phonological reduction in maternal speech in northern Australian English: Change over time. *Journal of Child Language, 41*(04), 725–755.

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

Byrd, D. (1993). 54,000 American stops. *UCLA Working Papers in Phonetics, 83*, 97–116.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication, 15* (1–2), 39–54.

Cho, T., Lee, Y., & Kim, S. (2014). Prosodic strengthening on the /s/-stop cluster and the phonetic implementation of an allophonic rule in English. *Journal of Phonetics, 46*, 128–214.

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

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

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods, 41*(2), 385–390.

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.

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*(01), 155–175.

Englund, K. T. (2005). Voice onset time in infant directed speech over the first six months. *First Language, 25*(2), 219–234.

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

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*(2), 209.

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

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*, 477–501.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America, 101*(6), 3728–3740.

Foulkes, P., Docherty, G. J., & Watt, D. (2005). Phonological variation in child-directed speech. *Language, 81*(1), 177–206.

Hartman, K. M., Bernstein Ratner, N., & Newman, R. S. (2017). Infant-directed speech (IDS) vowel clarity and child language outcomes. *Journal of Child Language, 44*(5), 1140–1162.

Henning, A., Striano, T., & Lieven, E. V. (2005). Maternal speech to infants at 1 and 3 months of age. *Infant Behavior and Development, 28*(4), 519–536.

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.

Ingram, D. (1989). *Phonological disability in children*. London Jersey City: Whurr Publisher.

Kemper, S., Herman, R. E., & Lian, C. H. (2003). The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech of noise. *Psychology and Aging, 18*(2), 181.

Kemper, S., Herman, R. E., & Nartowicz, J. (2005). Different effects of dual task demands on the speech of young and older adults. *Aging, Neuropsychology, and Cognition, 12*(4), 340–358.

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development, 24*(4), 372–392.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America, 59*(5), 1208–1221.

Klein, H. B., & Altman, E. K. (2002). The acquisition of medial /t, d/ allophones in bisyllabic contexts. *Clinical Linguistics & Phonetics, 16*(3), 215–232.

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

Ladefoged, P., & Johnson, K. (2006). *A course in phonetics (5th)*. Thomson Wadsworth.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Netherlands: Kluwer.

Lowit, A., Marchetti, A., Corson, S., & Kuschmann, A. (2018). Rhythmic performance in hypokinetic dysarthria: Relationship between reading, spontaneous speech and diadochokinetic tasks. *Journal of Communication Disorders, 72*, 26.

Malécot, A., & Lloyd, P. (1968). The t:D distinction in American alveolar flaps. *Lingua, 19* (3–4), 264–272.

Malsheen, B. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. *Child Phonology, 2*, 173–184.

Markó, A., Deme, A., Bartók, M., Gráczi, T. E., & Csapó, T. G. (2017). Word-initial irregular phonation as a function of speech rate and vowel quality in Hungarian. *International Seminar on Speech Production (134–145)*. Cham: Springer.

Masataka, N. (1992). Motherese in a signed language. *Infant Behavior and Development, 15*(4), 453–460.

Murray, A. D., Fees, B. S., Crowe, L. K., Murphy, M. E., & Henriksen, A. L. (2006). The language environment of toddlers in center-based care versus home settings. *Early Childhood Education Journal, 34*(3), 233–239.

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

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America, 54*(5), 1235–1247.

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Schaffer, H. R., & Crook, C. K. (1979). Maternal control techniques in a directed play situation. *Child Development*, 989–996.

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

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.

Toivola, M., Lennes, M., & Aho, E. (2009). Speech rate and pauses in non-native Finnish. In Tenth Annual Conference of the International Speech Communication Association. Proc. 10th Interspeech 2009, Brighton, 1707–1710.

Tucker, B. V. (2011). The effect of reduction on the processing of flaps and/g/in isolated words. *Journal of Phonetics, 39*(3), 312–318.

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner-and infant-directed speech. *Speech Communication, 49*(1), 2–7.

Vaux, B. (2002). Aspiration in English. Ms., Harvard University. Accessed on November 27, 2020.

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