Journal of Phonetics 52 (2015) 152–169

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

Development of phonetic variants (allophones) in 2-year-olds learning American English: A study of alveolar stop /t, d/ codas

Jae Yung Song [a,*], Stefanie Shattuck-Hufnagel [b], Katherine Demuth [c,d]

[a] Department of Linguistics, University of Wisconsin-Milwaukee, WI 53211, USA
[b] Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[c] Department of Linguistics, Centre for Language Sciences (CLaS), ARC Centre of Excellence in Cognition and its Disorders (CCD), Macquarie University, Sydney, New South Wales 2109, Australia
[d] The Santa Fe Institute, Santa Fe, NM 87501, USA

ARTICLE INFO

Article history:
Received 26 June 2014
Received in revised form
28 May 2015
Accepted 11 June 2015

Keywords:
Language development
Allophones
Alveolar stops

ABSTRACT

This study examined the emergence of the phonetic variants (often called allophones) of alveolar phonemes in the speech production of 2-year-olds. Our specific question was: Does the child start by producing a "canonical" form of a phoneme (e.g., /t/ with a clear closure and a release burst), only later learning to produce its other phonetic variants (e.g., unreleased stop, flap, and glottal stop)? Or, does the child start by producing the appropriate phonetic variants in the appropriate contexts and only later learn that they are phonetic variants of the same phoneme? In order to address this question, we investigated the production of three phonetic variants (unreleased stop, flap, and glottal stop) of the alveolar stop codas /t, d/ in the spontaneous speech of 6 American-English-speaking mother–child dyads, using both acoustic and perceptual coding. The results showed that 2-year-old children produced all three variants significantly less often than their mothers, and produced acoustic cues to canonical /t, d/ more often. This supports the view that young children start out by producing a fully articulated canonical variant of a phoneme in contexts where an adult would produce non-canonical forms. The implications of these findings for early phonological representations are discussed.

1. Introduction

One of the central questions in child language development is what early phonological representations look like and how these develop over time. Because sounds are the building blocks of words, this question is crucial to understanding the emergence and development of words in children. The purpose of the present study was to shed light on the nature and the organization of early phonological representations by examining 2-year-olds' production of the phonetic variants (sometimes called allophones) of phonemes.[1]

It is generally agreed by adult native speakers of English that there is a certain relationship between the underlined sounds in words like *hat*, *writer*, and *fountain*, i.e., that they are all variants of the same sound, represented here by the same orthographic letter. In linguistic terms, they are all phonetic implementations of the phonological category of voiceless alveolar stop consonant /t/, which can be released or unreleased, flapped or not flapped, or glottalized or not glottalized. For children, however, this relationship might not be obvious in the early stages of language learning; instead it might be something that is acquired later during the course of language development. Furthermore, although some phonetic variants can be in free variation, the choice of variant is often dependent on its environment. Therefore, in order to sound natural while speaking, an individual must learn the correct use of variants in the appropriate phonological contexts. These observations pose an important question: During development, when are the phonetic variants of a phoneme acquired and used appropriately?

There is only limited information about the development of phonetic variants in children, as most studies to date have focused on children's production and perception at the phonemic level, rather than at the phonetic level. Although there is a substantial body of

* Corresponding author. Tel.: +1 414 229 6166; fax: +1 414 229 5041.
E-mail address: songjy@uwm.edu (J.Y. Song).

[1] The terms 'phonetic variants' and 'allophones' are often used interchangeably, including in this paper. However, we believe that phonetic variants are not categorical in terms of articulation and acoustics, unlike what is implied by the term 'allophones'. We expand this idea below.

0095-4470/$ - see front matter © 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2015.06.003

literature examining the characteristics of allophones in adults' speech production, there are only a handful of studies exploring their use in young children. Therefore, in the present study we aimed to examine the use of phonetic variants in children around 2 years of age, focusing on the production of the alveolar stop codas /t, d/. In general, the realization of coda stops tends to vary much more compared to onset stops. We chose to examine alveolar stop codas because these are especially rich in phonetic variants in English. The allophones that we investigated are: the unreleased stop [t̚], [d̚], the flap [ɾ], and the glottal stop [ʔ]. In the present study, we focused on the following specific questions: When do children first start producing these three phonetic variants of alveolar stop codas? Do children begin by producing a “canonical” variant of a phoneme (e.g., /t/ with a clear closure and a release) and only later learn to produce the various phonetic variants in their appropriate contexts? Or do they start by producing various other non-canonical phonetic variants and only later learn how these are related to a particular phoneme? To address these questions, we examined whether children's production of phonetic variants differs from that of adults, i.e., How often are the different variants used and in what way does the acoustic shape of children's productions differ from that of adults? To give a preview, the 2-year-olds in the present study produced all three phonetic variants of alveolar stop codas less frequently than adults. In the discussion, we explore the implications of these findings for the nature of early phonological representations, especially considering children's developing articulatory skills and the role of speech input to children.

The rest of this introduction consists of two parts. In the first part, we review the literature on the characteristics of phonetic variants of alveolar stops in adult speech. Once we know what children are targeting, we will be better equipped to understand the path of phonetic development. In the second part of the introduction, we review the existing studies on the development of phonetic variants in children.

However, before we proceed, we would like to clarify one point. Although we will be using terms like released/unreleased, flapped/unflapped, and glottalized/unglottalized throughout the paper, we believe that phonetic variants are not binary or contrastive in terms of articulation and acoustics. Rather, we believe that these should be understood in terms of degree or amount of release, flap and glottalization, although they could be perceived categorically by the listener. For example, even when the listener judges that an alveolar stop /t/ was produced as unreleased [t̚], it could be that /t/ is not fully unreleased articulatorily. Therefore, we suggest that phonetic variants should not necessarily be thought of as belonging to one category vs. another category in terms of articulation and acoustics, but rather as at some point along a continuum of acoustic cues to the features of that phonemic category. To reflect this, we employed acoustic measurements in our analysis when determining the presence of a phonetic variant, conducting perceptual judgments of phonetic variants only when necessary.

1.1. *The characteristics of phonetic variants of alveolar stops in adult speech*

In this section, we review the literature on the articulatory, acoustic, and perceptual characteristics of different variants of the alveolar stops /t, d/ in adult speech. We primarily discuss three variants: unreleased stops, flaps, and glottalized stops, as they are the focus of our study. According to Zue and Laferriere (1979), the articulatory gestures used in the production of a flap involve the tongue tip making brief contact with the alveolar ridge, followed by an immediate release. The tongue tip can either make an up-and-down movement or a front-and-back movement to touch the alveolar ridge, and the closure can be complete or partial. These articulatory movements produce a wide variety of acoustic realizations. For example, in the case of partial closure, sometimes noise will be generated at the constriction, resulting in a voiced fricative. Flaps are generally quite short in duration, often with little or no sign of a release burst. Also, flaps are more likely to occur within a word if the vowel following the phonological stop is reduced or unstressed, as in *bu<u>tt</u>er* (although flapping can occur across a word boundary in other stress contexts, as in *kni<u>t</u> a lot*; see Fukaya & Byrd, 2005).

Byrd (1993) provides a broad picture of the characteristics of American English stops using data from the TIMIT database, which includes sentences read by 630 American speakers from a range of geographical locations. The data set included 54,384 oral and nasal stops, affricates, oral and nasal flaps, and glottal stops, which Byrd described and analyzed for frequency of occurrence, mean segmental durations, voice onset times, and certain effects of voicing, place, word position, and speaker gender. Concerning the characteristics of oral stops, 24,414 closures and 21,847 releases were found in the TIMIT transcription. In sentence-final position, where the environment (i.e., a following silence) was more controlled than in sentence-medial position, bilabial stops were released 49.5% of the time, alveolar stops 57% of the time, and velar stops 83.11% of the time. Also transcribed in the TIMIT database were flaps, which were classified as oral or nasal. A total of 4980 flaps occurred: 3649 oral flaps (e.g., *wa<u>t</u>er*) and 1331 nasal flaps (e.g., *sui<u>t</u> in*). Mean duration of flaps was 29 ms, and there was no difference in mean duration between oral and nasal flaps. As for glottal stops, there were 4834 in the TIMIT database. Word-initial glottal stops made up 49% of the total, word-medial glottal stops constituted 6%, word-final glottal stops constituted 16%, and unaffiliated glottal stops (glottal stops occurring between two vowels at a word boundary) were 29% of the total.

De Jong (1998) provides an investigation into the nature of flapping of /t, d/ in American English. In particular, he evaluated two models of flapping: the traditional model in which flapping is considered as a categorical switch from stop to flap in a specified linguistic environment, and another model in which flapping is produced as a by-product of articulatory changes related to the prosodic structure. To address the issue, the data were collected from three speakers using the X-ray microbeam systems, which tracked the location and motion of radio-opaque pellets attached to the articulators during speech. De Jong (1998) showed several acoustic correlates that distinguished the perceptually transcribed flaps from the tokens transcribed as stops. In general, flaps had shorter occlusion durations, shorter voice onset times and a greater percentage of voicing during closure. In addition, flaps were similar to /d/ in that both had longer preceding vowel durations than /t/, and also similar to /t/ in that both had smaller changes in F2 than /d/. Articulatory measures showed that the tongue body was more retracted during flaps than during stops. In terms of the location and movement of the tongue tip/blade, flaps were more similar to /d/.

De Jong's (1998) results showed that all four transcribers were quite consistent in transcribing the presence/absence of flaps. This suggested that flapping across a word boundary might be a perceptually categorical phenomenon. However, for speakers, production of flaps appeared to be optional or inconsistent across a word boundary. These findings suggest a model in which a gradient articulatory change results in quantized acoustic outputs, giving rise to the consistent transcriptions of either flaps or stops. This challenges both the categorical rule account and the prosodic by-product account. If this is the case, from a speaker's point of view, there is no need for a rule that demands a specific production of a flap before an unstressed vowel. Rather, what is important for the speaker is to understand the segmental and prosodic conventions of the language sufficiently to know when a salient consonant release is needed.

Further acoustic and perceptual analyses of flaps were carried out by Herd, Jongman, and Sereno (2010), who examined whether /t/ and /d/ are neutralized in flapping environments. When the authors examined the distribution of /t/ or /d/ duration, which included both closure and release burst, they found a clear separation of durations between flapped and fully articulated stops for all speakers; consistent with previous studies, flaps were shorter than fully articulated stops. Therefore, they were able to determine within speaker cut-off values for flaps; tokens below the cut-off value were considered flapped and those above were considered unflapped. The average flap cut-off value across 20 speakers was 56 ms, with a range of 43–69 ms. The average flap duration was 32 ms, with a range of 24 to 41 ms. When this method was applied, word-medial /t/ and /d/ were found to be flapped 76% and 99% of the time, respectively.

One of the three phonetic variants we examined was glottalization. We review some of the studies of the acoustic and perceptual characteristics of glottalization below. Acoustically, glottalization is often characterized by irregularly spaced pitch periods and characteristics such as low amplitude and low F0 and breathiness (Pierrehumbert & Talkin, 1992). The frequency of occurrence of glottalization is known to be affected by various factors including the position of the word within the utterance and intonational phrase (Pierrehumbert & Talkin, 1992; Dilley, Shattuck-Hufnagel, & Ostendorf, 1996), gender (Byrd, 1994), and dialect (Docherty & Foulkes, 1995).

To understand the factors and articulatory mechanisms involved in glottalization, Redi and Shattuck-Hufnagel (2001) examined glottalizations at phrase boundaries in both medial and final positions within an utterance. Of the two corpora they examined, the first, called the Labnews, consisted of read speech produced by 6 professional radio announcers (from the BU FM news corpus) and 4 non-professionals, all native speakers of American English. The second corpus, called the ABC corpus, included read speech produced by four non-professional speakers. In both corpora, the frequency of occurrence of glottalization differed greatly from individual to individual. Also, not all speakers produced the same acoustic characteristics, and the results were inconclusive about whether males or females produced more glottalizations. On the other hand, the frequency of occurrence of glottalization was significantly affected by position within the utterance; speakers produced glottalizations more frequently in utterance-final positions compared to utterance-medial positions.

As the above review indicates, the alveolar stop codas in English are rich in phonetic variants, each with its own characteristic acoustic pattern. At the same time, the review of the literature suggests that, despite numerous studies on flap production, it is challenging to fully characterize flaps (for relevant discussion, see Herd et al., 2010). In the next section, we review the literature on the development of phonetic variants in young children, where much less is known.

1.2. Development of phonetic variants

To our knowledge, there are only a few studies that have examined children's speech production at the level of phonetic variants. Klein, Altman, and Tate (1998) examined how closure duration in young children's speech affects the adult listener's perception of flap. To this end, 34 adults listened to audio recordings of two children, a male and female, ages between 30 and 48 months. The target words, which contained medial /t, d/ in flap contexts (e.g., *water*, *feeding*), were taken from a series of monthly recordings made for a longitudinal study on the acquisition of /t, d/ allophones. The tokens were chosen based on closure duration, and were placed in one of five successive closure-duration categories increasing by 20 ms at each step; tokens in these categories had closures of 3–17 ms, 23–37 ms, 43–57 ms, 63–77 ms, and 83–97 ms, respectively.

The results showed that tokens were less likely to be judged to be flaps by adult listeners as closure duration increased. Flap duration of 3–17 ms produced the greatest percentage of flap judgments (at 75%); 23–37 ms ranked second (at 61%); 43–57 ms and 63–77 ms were perceived as flaps 50% of the time, and 83–97 ms had the lowest percentage of flap judgments (at 32%). In sum, children's medial /t, d/ productions were most often judged as flaps when their closure durations were at or under 37 ms. This is consistent with the range of closure duration for adult flaps from Zue and Laferriere (1979), 10–40 ms, and suggests that a listener is most likely to identify an alveolar stop as a flap when its closure duration is in this range, regardless of the age of the speaker. However, it is interesting to note that not all tokens in the shortest duration category were perceived as flaps; likewise, not all tokens in the longest duration category were judged as being stop-like /t, d/s. Therefore, there might be other acoustic parameters playing a role in addition to closure duration.

Although Klein et al. (1998) showed the relationship between the production and perception of children's flaps, they did not address at what age children develop flaps. In a subsequent speech production study, Klein and Altman (2002) carried out a longitudinal study to examine the acquisition of medial /t, d/ allophones in bisyllabic words in 4 typically-developing children. For analysis purpose, the longitudinal data were consolidated into three sessions: (1) under 36 months, (2) from 36 to 47 months, (3) from 48 to 60 months. The allophones of /t, d/ examined included flaps (as in *ladder*), laterally-released forms (as in *bottle*), and nasally-released forms (as in *button*). Because the authors wanted to examine how the various phonetic and prosodic contexts affected the production of flaps, the flap contexts were further divided into three groups: before ‘y’ (as in *kitty*), before ‘er’ (as in *butter*) and before ‘ing’ (as in *eating*).

The phonetic transcriptions of the productions by the parents of the children revealed that the parents consistently produced the flap, and used both lateral release and nasal release. On the other hand, children's speech productions seldom exhibited lateral and nasal release, but importantly, flaps increased by about 15% with each additional session. By the last session (between 48 and 60 months), the children produced flaps in the adult flapping context about 50% of the time. In the flapping environment in children's productions, a '-ty' ending facilitated flap production most strongly, while '-ding' was found to be the least facilitative. Overall, there was much variability in the 2–5-year-olds' use of variants for /t, d/, suggesting that there are many factors that determine the rate and progress toward adult-like context-appropriate medial /t, d/ in young children.

Rimac and Smith (1984) compared word productions from 8 children (around 8 years of age) and 8 adults with the aim of examining the relationship between children's and adults' speech segment durations. Consistent with previous developmental studies of speech timing, they found that the 8-year-olds' segment durations were longer than those of adults, including both stressed vowels and flaps. Overall, Rimac and Smith (1984) found that acquisition of the flap is a gradual process. The authors suggest that this delay could be due to the fast movement of the tongue involved in the production of a flap, since it is known that children generally speak more slowly than adults. However, as they point out, it could also be attributed to non-motoric factors; for example, children may be unaware of the phonetic contexts in which flaps are typically produced. In sum, both Rimac and Smith (1984) and Klein and Altman (2002) suggest that children around 5–8 years of age are still learning to produce flaps in an adult-like manner.

In a study of acoustic cues to coda contrasts, Song, Demuth, and Shattuck-Hufnagel (2012) showed that English-learning children as young as 1;6 produced many adult-like cues to voicing (voiced vs. voiceless) and place (alveolar vs. velar) contrasts in coda stops. At the same time, there were two aspects of the children's production of these cues that were generally different from adults. Overall, 1;6-year-olds had more frequent releases of stops than adults. In addition, these children did not produce glottalization at the end of the vowel before the coda stop as often as adults, especially in sentence-medial position. These findings suggest that a canonical variant of the coda stop (e.g., /t/ with a clear closure and a release) may be more common than non-canonical variants in young children's speech production.

In addition to these studies of the acquisition of phonetic variants in speech production, evidence from some infant studies shows that awareness of phonetic variants in perception might be acquired early in life, well before their production is mastered. For example, Hohne and Jusczyk (1994) showed that two-month-olds were able to discriminate pairs like *nitrate* and *night rate*, which differ in phonetic realizations of /t/ and /ɹ/. This suggests that infants at this age are able to discriminate phonetic differences that could provide information about word boundaries. In a subsequent study, Jusczyk, Hohne and Bauman (1999) further asked when infants develop sensitivity to the way that variants are systematically distributed within words, and when they begin to use these cues to segment words from fluent speech. The results showed that by 10.5 months of age, infants were able to take advantage of what the authors called allophonic cues to word boundaries in recognizing *nitrates* vs. *night rates* in fluent speech. The authors argue that, compared to other types of cues to segmentation that are utilized earlier in development (Newsome & Jusczyk, 1995; Saffran, Aslin, & Newport, 1996), the acquisition of allophonic cues might be later because the infant learner needs to be exposed to a sufficient number of instances of words before being able to learn the mapping between context-governed variants and the contexts in which they appear. More recently, there has been increasing interest in infants' ability to learn about allophonic variation on the basis of distributional information (Peperkamp & Dupoux, 2002; Peperkamp, Pettinato, & Dupoux, 2003; White, Peperkamp, Kirk, & Morgan, 2008). These studies suggest that by the end of their first year of life, infants are sensitive to complementary distribution where phonological alternations occur, and can learn patterns of phonological alternations on the basis of the distributional cues in the input. This ability might be a prerequisite for producing context-appropriate phonetic variants.

Although these studies on the development of speech perception suggest that infants below one year of age might be sensitive to the distribution of phonetic variants in the input, to date only limited information is available on the development of phonetic variants in young children's speech production. Therefore, in the current study, we aimed to expand the existing literature in several directions. First, in contrast to previous speech production studies in children, which have primarily focused on the production of flaps, we investigated the production of three different variants for /t, d/: unreleased stops, flaps, and glottalized stops. By examining more than one kind of variant, we hoped to be able to draw more general conclusions about the path of development of phonetic variants in young children. Second, we examined the use of phonetic variants in children between 1;6-2;6 years of age, an age group that is younger than that found in most early speech production studies. Children produce their first words around one year of age (MacNeilage & Davis, 1990), and around 1;6 years of age, many children gain speed in word learning, a shift in rate of vocabulary development which is often cited as the 'vocabulary spurt' (Goldfield & Reznick, 1990). Therefore, examining children in the age range 1;6–2;6 could provide valuable information on the initial processes of phonological development. Finally, we combined transcriptional methods and acoustic measures to compare the use of phonetic variants in children and adults, rather than relying on phonetic transcriptions alone.

Existing literature suggests that production of some variants of /t, d/, such as flaps, is articulatorily challenging for children; moreover, acquisition of adult skill with these variants also involves learning the phonetic and prosodic environments in which they occur. Therefore, we predicted that children would exhibit a late acquisition of the three phonetic variants, showing non-adult-like acoustic patterns at the beginning of the development process.

2. Method

2.1. Database

The data examined in this study came from the Providence Corpus (Demuth, Culbertson, & Alter, 2006), a collection of spontaneous speech interactions between six mother–child pairs from the New England area (for further information and access to the corpus, see the Child Language Data Exchange System [CHILDES: http://childes.psy.cmu.edu/]). All six children (three boys, three girls) were typically developing, monolingual speakers of American English. Digital audio/video recordings were collected in the children's homes, approximately one hour every two weeks for two years. Two of the children had denser corpora for part of this time, with weekly recordings for approximately a year during the recording period. Recording started between the ages of 0;11 and 1;4, depending on when the children started producing their first words. During the recording, both mother and child wore a wireless Azden WLT/PRO VHF lavalier microphone pinned to their collar as they engaged in everyday activities. The recordings were made using a Panasonic PV-DV601D-K mini digital video recorder. The audio from the video was later extracted and digitized at a sampling rate of 44.1 kHz.

Both the mothers' and children's speech in the Providence Corpus were orthographically transcribed using Codes for the Human Analysis of Transcripts (CHAT) conventions (MacWhinney, 2000). The children's sentences were also transcribed by trained coders using International Phonetic Alphabet (IPA) transcription, providing the phonetic representations of words and the location of stressed syllables. Ten percent of the child data from each recording session were re-transcribed by a second transcriber. Transcription reliability of overall coded segments ranged from 80–97% across files in terms of presence/absence of segments and place/manner of articulation. The feature voicing is difficult to reliably transcribe in young children's speech (Stoel-Gammon & Buder, 1999), and was therefore not assessed in these reliability measures.

2.2. Data

The data we used for this study were part of a larger project looking at the acoustic characteristics of the coda in children's speech production. As described in Song et al. (2012), to select the data we first identified the most frequently produced monosyllabic CVC (consonant-vowel-consonant) content words ending in /t, d/ in the Providence Corpus. Words starting with glides or liquids (e.g., *red*) were excluded, due to the difficulty of identifying the beginning of vowels in such words. The final set of target words included six words ending in phonologically voiceless /t/ codas (*bat*, *boat*, *cat*, *feet*, *hat*, *hot*) and five words ending in voiced /d/ codas (*bed*, *food*, *good*, *head*, *side*). We examined both child and adult productions of these words when the children were between 1;6 and 2;6 years of age. Initially, the data were sampled at 1;6, 2;0, and 2;6, plus one month before and after these ages. However, due to the lack of the use of phonetic variants in children at these ages, data were collapsed and are presented in a single age group in the Results section.

After identifying the target words, we extracted all the audio files of the sentences containing these target words. For individual mothers and children at each age, we coded the first ten acoustically clean tokens of a target word in sentence-final position (e.g., *It's good.*). In the current data, syntactic sentences and prosodic utterances usually ended at the same point. In sentence-medial position, we coded the first five acoustically clean tokens in each of four contexts: before non-glide consonant-initial words (e.g., *good boy*), before glide-initial words (e.g., *good wagon*), before words beginning with a stressed vowel (e.g., *good apple*), and before words beginning with an unstressed vowel (e.g., *good as*). To clarify, acoustically clean tokens were those without any interruption in acoustic quality (which was sometimes compromised by overlap with other speaker's vocalizations or background noise). In addition, it is worth noting that the sentence-medial contexts were determined based on the actual pronunciation of the following word, rather than the target phonological shape of the following word. For example, if a child stressed a vowel-initial word that often does not receive a stress in adult speech (e.g., *head on*), the sentence-medial context for the target word was classified as “before a word beginning with a stressed vowel”, rather than “before a word beginning with an unstressed vowel”. We identified stressed vowels based on the overall perception of variations in the three accepted perceptual correlates of phase-level stress: Increased length, loudness, and a pitch marker (e.g., Beckman & Edwards, 1994; Bolinger, 1961; Lieberman, 1960).

These selection criteria often resulted in fewer than five tokens in each of the four sentence-medial contexts. However, setting this as an upper bound provided some controlled variability for the phonetic contexts in which these medial target words appeared across mothers and children. For both mothers and children, more than half of the sentence-medial tokens were followed by non-glide consonant-initial words (Mothers: 52.7%, Children: 57.1%). About a quarter of the sentence-medial words were followed by words beginning with an unstressed vowel (Mothers: 28.8%, Children: 25.4%). Target words followed by glide-initial words (Mothers: 13.5%, Children: 7.1%) and by words beginning with a stressed vowel (Mothers: 5%, Children: 10.3%) were the least common sentence-medial contexts.

The initial number of tokens contributed by each subject in the sentence-medial and sentence-final positions is listed under the column titled *Initial dataset* in Table A1 in Appendix A. The data are presented separately for /t/ and /d/ target words. Note that Mothers 3 and 4, and their children, had a greater number of tokens compared to other mother–child pairs; this is due to the fact that they had denser corpora, with weekly recordings during the time of investigation. As expected, mothers produced a greater number of tokens than children. Furthermore, in sentence-medial position, it turned out that mothers had a greater number of target words ending in /d/ (332) than /t/ (112). This appears to be due to the high frequency of the word *good*; in fact, the word *good* accounted for the 50% (223/444) of the mothers' sentence-medial /t, d/ tokens. The breakdown of token numbers by word type is shown in Table A2 in Appendix A. When we discuss the results in a later section, we will also present the breakdown of the results by word type in Tables B1–B5 in Appendix B.

2.3. Acoustic and perceptual coding

In this section we discuss the coding of unreleased stops, flaps, and glottalized stops. Our goal in coding was to determine whether children produce fully occluded stops with a release burst indicating pressure buildup during occlusion (i.e., canonical stop cues), or the context-appropriate unreleased, flapped, or glottalized versions, as adults do. Coding was carried out in two different ways: acoustically and perceptually. Although acoustic coding was the preferred method, perceptual coding was particularly useful in two cases. First, perceptual coding was used to distinguish between those tokens produced with or without cues (to an adult listener at least) that indicated the presence of a coda consonant. This was important because children in this age group often produce tokens which seem to an adult listener to be missing the coda. Perceptual coding allowed us to determine which tokens were heard as including a coda, even if the acoustic correlates of that coda were not apparent. Within the set of tokens determined by a listener to have been produced with a coda, perceptual coding was also used to identify those produced as a flap. This was necessary because the acoustic definition of a flap has not yet been clearly determined in the literature.

2.3.1. Coding of unreleased stops

In order to determine whether the stop coda of a word token was released or not, first we looked for acoustic evidence of the release of an oral occlusion, in the form of a noise burst. The release of the oral occlusion was evidenced by at least one strong vertical spike in the waveform and the spectrogram, signaling the transient at the abrupt release of pressure built up during the closure of a stop. The codas that had such acoustic evidence of the release of the oral occlusion were labeled as *released*.

If a coda was determined to lack acoustic evidence for the release of such pressure buildup behind a constriction, there were two possibilities: i) There was an occlusion for the coda stop, whether oral or glottal (or both), but the articulatory release did not produce a burst of frication, possibly due to lack of pressure buildup behind the occlusion, or ii) there was no release because the coda stop was omitted in the production. The tokens corresponding to first possibility were labeled as *unreleased*, whereas the tokens corresponding to the second possibility were labeled as *no coda*. To determine whether a token was unreleased as opposed to lacking a coda, we looked for acoustic evidence of an occlusion. However, there were some tokens that did not show acoustic evidence of an occlusion but still perceptually sounded as if they were produced with a coda. Among the tokens without acoustic evidence of an occlusion (or release), to differentiate the tokens produced with a coda from the tokens produced without a coda, we employed perceptual judgment. Therefore, there were two types of tokens labeled as unreleased; those based on acoustic evidence of occlusion (without a release burst) and those based on perceptual judgment (even when lacking acoustic evidence of occlusion).

To test the first possibility, we looked for acoustic evidence that suggested an oral or glottal occlusion. One signal characteristic that provides evidence for the presence of an oral occlusion is a voice bar, i.e., the signal produced when there is both an oral and velopharyngeal occlusion for the stop closure, and there is still a sound source in the form of vocal fold vibration. A voice bar is characterized by a low frequency, low amplitude signal with a simpler sinusoidal waveform, reflecting continued vocal fold vibration during the stop closure without radiation of higher-frequency components of the source (see Fig. 1). The example shown in Fig. 1 exhibits two cues to the existence of a closure: the voice bar during closure, and the noise burst at release, which indicates pressure build up behind the closure. In contrast, a signal characteristic that indicates the presence of a gesture toward glottal occlusion is the presence of irregular, creaky-sounding pitch periods at the end of the vowel, suggesting glottalization (see Fig. 2). Although there is no

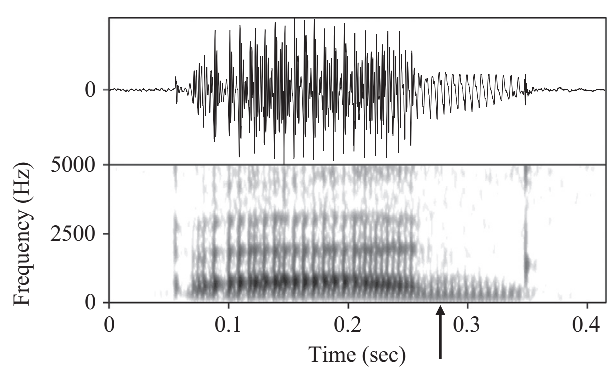


Fig. 1. Representative waveform and spectrogram for the word *bed* produced by a mother. The voice bar during the closure for /d/ is marked with an arrow. Note that the closure is fully voiced, i.e., the voice bar continues until the release burst; this was not the case for all tokens.

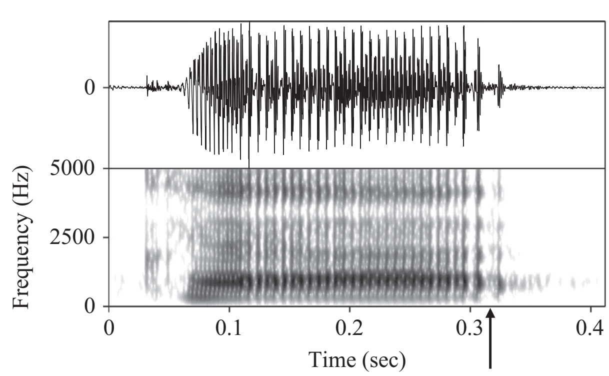


Fig. 2. Representative waveform and spectrogram for the word *cat* produced by a mother. The glottalization at the end of the vowel is marked with an arrow.

acoustic evidence of oral closure and release in the token shown in Fig. 2, it exhibits evidence of glottal occlusion. This particular token also shows a change in the vocal fold vibration pattern at around 0.1 s, well before the vowel-final irregularity begins at around 0.3 s; only tokens with irregularity that continued to the end of the vowel were counted as glottalized at the end of the vowel. Tokens that showed acoustic evidence of either oral or glottal occlusion, but lacked acoustic evidence of release, were labeled as *unreleased*.

As mentioned above, there were some tokens that did not show acoustic evidence of either an oral or a glottal occlusion, but nonetheless sounded as if they were produced with a coda stop, presumably because of other cues such as formant transitions at the end of the vowel or duration characteristics. Those were also labeled as *unreleased*. Of the 588 tokens that were labeled as unreleased, 458 tokens (78%) exhibited either a voice bar or glottalization, or both. The other 130 tokens (22%) were perceptually categorized as unreleased because they sounded to the listeners as if there was an occlusion, although they lacked acoustic evidence of either an oral or a glottal occlusion. Finally, tokens without any acoustic/perceptual evidence of occlusion were labeled as having *no coda*.

2.3.2. Coding of flaps

Because the focus of the present study lies in the possible difference in flapping behavior between mothers and children, we first had to determine which tokens were flapped. However, it is challenging to determine the occurrence of flaps acoustically, because the acoustic characteristics of flaps have not been fully agreed upon in the literature. Therefore, we used perceptual coding to determine the presence/absence of flaps. The codas that were perceived to be flapped were labeled as *flapped*, whereas the codas that were perceived to be non-flapped were labeled as *unflapped*. Fig. 3 shows two examples of /d/ tokens perceived as flaps; although both were perceived as flaps, notice that they differ markedly in their acoustic shape. The token shown on the left lacks a clear complete closure (note presence of higher formants) and release, whereas the token shown on the right exhibits acoustic evidence of (voiced) oral closure and release burst (most clearly visible in the spectrogram).

However, as pointed out by Rimac and Smith (1984) and Klein et al. (1998), the reliability of flap analysis based solely on perception is somewhat limited. Therefore, our perceptual coding was supplemented by acoustic analysis of two characteristics, to further determine possible acoustic differences in flapping behavior of mothers vs. children. The two acoustic cues that we examined were: a) percent of tokens produced with a release burst (for all tokens), and b) percent of the closure interval produced with voicing (calculated only for the tokens with an oral occlusion and a release). De Jong (1998) suggests that the lack of a salient release burst and the presence of voicing in the closure are the clearest acoustic correlates of a perceived flap. A final point to mention about the coding of flaps is that flaps were coded only for the sentence-medial tokens, in which we expected flaps to occur. Sentence-final tokens were not included in this part of the analysis.

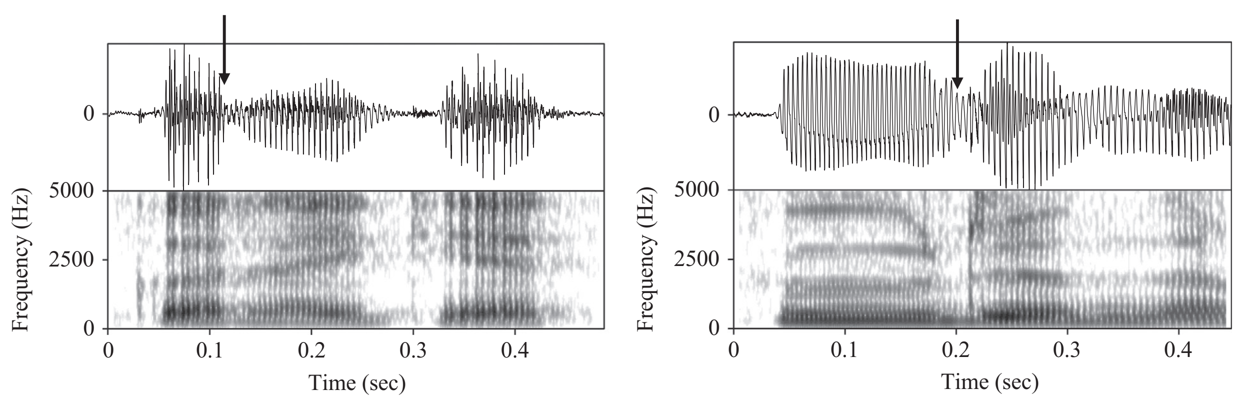


Fig. 3. Representative waveform and spectrogram for the words *good idea* (left) and *food in the* (right) produced by two different mothers. The regions depicting codas (/d/ in these cases) are marked with arrows. Both of the codas were perceived as flaps.

2.3.3. Coding of glottalized stops

To determine whether the stop coda of a word token was produced with glottal occlusion or not, we looked for acoustic evidence of glottal occlusion toward the end of the vowel. The presence of glottal occlusion (or at least a glottal gesture in the direction of closure) was evidenced by irregular, creaky-sounding pitch periods at the end of the vowel (see Fig. 2). The codas that exhibited irregular pitch periods at the end of the vowel were labeled as *glottalized*, whereas the codas lacking these were labeled as *unglottalized*.

2.3.4. Reliability of coding

The coding of unreleased stops, flaps, and glottalized stops was carried out by eight phonetically trained research assistants. To check inter-coder reliability, 10% of the tokens (127 out of 1270 total tokens) were randomly chosen proportionally to the size of group (mothers, children), age (1;6, 2;0, 2;6), target (/t/, /d/), and position within the sentence (medial, final), and coded again by the first author. Average agreement for the presence or absence of each of the three phonetic variants is as follows for mothers and children: (1) unreleased: mothers (84%), children (94%), (2) flapped: mothers (91%), children (85%), (3) glottalized: mothers (95%), children (94%). For all six groups of data (three variants × mothers, children), correlations between the measurements of the original and recoded data were significant at the level of $p<0.01$, suggesting high reliability of coding.

3. Results

In the following three sets of analyses, we compared the frequency of occurrence of three non-canonical phonetic variants of /t, d/ (i.e., unreleased stop, flap, glottalized stop) between 2-year-olds and adults (mothers). Before we examine the results, we will first discuss the tokens labeled as *no coda*. Because these tokens were determined to have no acoustic or perceptual evidence of a realized coda, they were left out in the main analyses where we examined the use of variants in children's speech and mothers' speech. Tables A1 and A2 in Appendix A show the proportion of /t/ and /d/ target words coded as *no coda* that were excluded in the main analyses (under the column *No coda*). Distribution of the remaining tokens that were included in the main analyses is shown under the column titled *Final dataset* in the same tables. Overall, mothers omitted codas in only 1% of the tokens, whereas the children omitted codas in 13% of the tokens. The prevalence of the omission of the coda consonant in children's speech as part of early phonological processes is consistent with the developmental literature (Ingram, 1986).

3.1. Mother–child contrast for unreleased stops

In the first analysis, we explored children's production of unreleased stops compared to their mothers. Fig. 4 shows the number and percentage of tokens *unreleased* for the /t/ and /d/ target words out of all tokens analyzed. The numbers are shown separately for mothers and children in sentence-medial and sentence-final position. The numbers for individual participants and word items can be found in Tables B1 (/t/ targets) and B2 (/d/ targets) in Appendix B.

Two observations can be made: First, mothers produced unreleased codas more often than children, and second, codas were unreleased more often sentence-medially than finally. To confirm the significance of these differences, we used mixed-effects logistic regression models, which incorporated both fixed and random effects. Mixed-effects regression models were particularly appropriate for our spontaneous speech corpus data, because of the unbalanced size of data in individual speakers (for further information on mixed-effects regression models, see Baayen, 2008; Janda, Nesset, & Baayen, 2010). Both individual speakers and individual words were incorporated into the models as random-effect factors. Fixed-effect factors were age (mothers vs. children), position within the sentence (sentence-medial vs. sentence-final), and the interactions between the two factors (age × position). As our dependent variable was binary (unreleased or not), we used logistic mixed-effects models. Statistical calculations were carried out using the R statistical computing software (R Development Core Team, 2011), and in particular the lme4 package (Bates, Maechler, & Bolker, 2012). We analyzed voiceless /t/ and voiced /d/ target words separately.

As shown in Table 1, the effect of age was statistically significant for /t/ target words, suggesting that mothers produced unreleased /t/ codas more often than children. Although /t/ codas were overall unreleased more often in sentence-medial position than in sentence-final position, the effect of sentence position was not significant. However, there was a significant age × position interaction; as shown in Fig. 4, the medial-final difference was greater in mothers. Also, the difference between mothers and children was greater in sentence-medial position. For the /d/ target words, although mothers overall produced unreleased codas more often than children, the difference was not significant. Curiously, compared to /t/ targets, mothers produced unreleased /d/s less often, while children produced them more often, narrowing the gap between the two, especially in sentence-final position (see Fig. 4). The effect of sentence position was significant, with more frequent unreleased codas in sentence-medial position than in sentence-final position. There was also a significant interaction between age × position. Similar to /t/ targets, the difference between mothers and children was greater sentence-medially than finally.

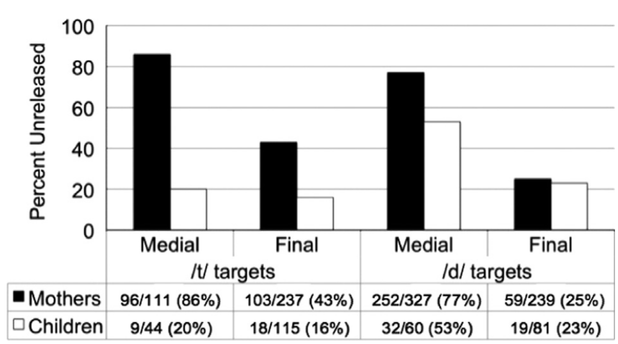


Fig. 4. The number (percentage) of tokens categorized as *unreleased*.

Table 1
The estimates of the coefficients, standard errors, and their *z*-statistics for the logistic mixed-effects models with respect to unreleased codas.

	/t/ targets				/d/ targets			
	Estimate	Std. error	*z*-Value	*p*-Value	Estimate	Std. error	*z*-Value	*p*-Value
(Intercept)	−1.682	0.289	−5.820	$p<0.001$	−1.095	0.352	−3.112	$p<0.01$
Age	1.428	0.293	4.870	$p<0.001$	−0.061	0.315	−0.192	0.848
Position	0.325	0.461	0.704	0.481	1.225	0.379	3.230	$p<0.01$
Interaction	1.795	0.553	3.246	$p<0.01$	1.122	0.430	2.609	$p<0.01$

3.2. Mother–child contrast for flaps

In the second analysis, we examined children's production of flaps compared to their mothers. As noted earlier, the production of flaps was examined only in sentence-medial position, and codas that were perceived to be flapped were labeled as *flapped*, whereas codas that were perceived not to be flapped were labeled as *unflapped*. Fig. 5 shows the number and percentage of tokens for codas perceived as flaps for the /t/ and /d/ target words. Of the 4 sentence-medial contexts for the target words (i.e., before non-glide consonants, before glides, before stressed vowels, and before unstressed vowels), the two contexts before consonants were expected to be inappropriate segmental contexts for flapping; indeed, none of the mothers or children produced flaps in these two contexts. Therefore, the results shown in Fig. 5 are from the two vowel-initial contexts where flapping was possible. The breakdown of the numbers by participant and item can be found in Table B3 in Appendix B.

As shown in Fig. 5, alveolar stop codas were realized as flaps more often in mothers' speech production than in children's speech production. Consistent with the results of unreleased stops, this suggests that the children used non-canonical phonetic variants of alveolar stops less often, producing more canonical versions. In fact, all 6 instances of flaps in children's speech occurred toward the end of our sample period, at 2;6. Furthermore, only 3 of the 6 children produced flaps by this age, suggesting the possibility that children are able to produce flaps in an adult-like manner only by 2;6 or later. To compare, the youngest group in Klein and Altman (2002), children under 36 months, produced flaps about 25% of the time in flap contexts. Klein et al. (1998) also commented that the first flaps were produced between 2 and 4 years of age in these children.

We used logistic mixed-effects models to examine whether there was a difference in the frequency of occurrence of flaps between mothers and children. Again, we performed analyses separately for the voiceless /t/ target words and voiced /d/ target words. The dependent variable was whether a token was perceived as produced with a flap or not. As we were primarily interested in the overall mother–child difference, the two vowel-initial contexts were pooled together to provide increased statistical power. As shown in Table 2, the results showed that mothers produced flaps significantly more often than children for both /t/ and /d/ target words.

In addition to comparing the frequency of occurrence of flapped codas between mothers and children, we also examined the acoustic characteristics the flapped and unflapped codas to determine possible acoustic differences in the flaps that mothers and children produced. In particular, we focused on two aspects: (a) the percentages of flapped and unflapped codas that were acoustically released or unreleased, and (b) the portion of stop closure duration that was voiced (as indicated by periodicity in the waveform and vertical striations in the spectrogram) for flapped vs. unflapped codas. However, because children produced very few flapped codas, the data were often sparse, which prevented us from doing statistical analysis and drawing any general conclusions. Although we do not present the acoustic results here, preliminary descriptions of the data are given in Appendix C for those who are interested in the acoustic aspects of these productions.

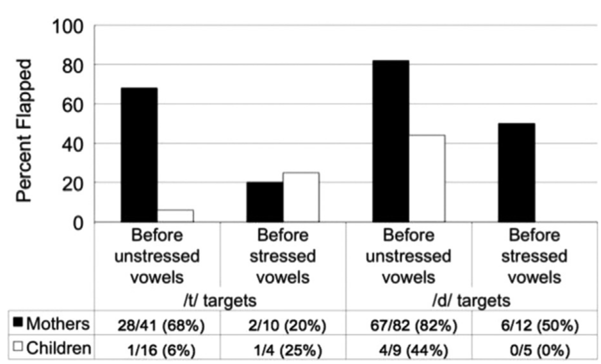


Fig. 5. The number (percentage) of medial /t/ and /d/ targets perceived as a flap.

Table 2
The estimates of the coefficients, standard errors, and their *z*-statistics for the logistic mixed-effects models with respect to flapped codas.

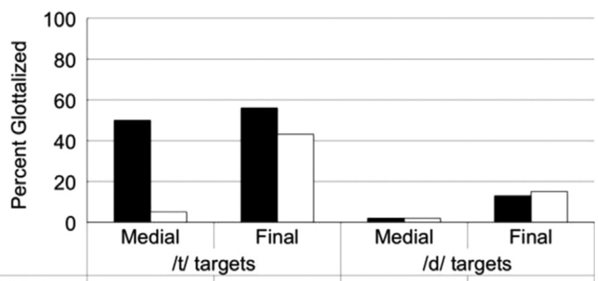
	/t/ targets				*/d/ targets*			
	Estimate	Std. error	*z*-Value	*p*-Value	Estimate	Std. error	*z*-Value	*p*-Value
(Intercept)	−2.964	0.807	−3.673	$p<0.001$	−2.883	0.602	−4.787	$p<0.001$
Age	1.961	0.771	2.543	$p<0.05$	1.548	0.547	2.831	$p<0.01$

3.3. Mother–child contrast for glottalized stops

In the third analysis, we investigated children's production of glottalized stops compared to their mothers. Recall that glottalized stops have irregular, creaky-sounding pitch periods at the end of the vowel. Fig. 6 shows the number and percentage of tokens coded as glottalized. Again, the numbers are separated by age (mothers vs. children) and sentence position (medial vs. final). The breakdown of the numbers of participant and item is listed in Tables B4 (/t/ targets) and B5 (/d/ targets) in Appendix B.

The data reveal that mothers glottalized 50% of tokens with /t/ codas, but glottalized coda /d/ much less often. Also, mothers glottalized both /t/ and /d/ more often in sentence-final position than in sentence-medial position. Consistent with the results of unreleased stops and flaps, children produced glottalization less often than mothers, but this was primarily true for sentence-medial /t/. In sentence-final position, children produced glottalizations for both /t/ and /d/ codas almost as often as their mothers.

To confirm these observations, we conducted logistic mixed-effects analyses separately for the voiceless /t/ and voiced /d/ target words. The dependent variable was whether a token was glottalized or not. For /t/ target words, the effect of age was significant, suggesting that mothers produced glottalized codas more often than children (see Table 3). The effect of sentence position was also significant, with a higher proportion of glottalized codas in sentence-final position than in sentence-medial position. The age × position interaction was also significant. As shown in Fig. 6, the mother–child difference was particularly large sentence-medially. For /d/ target words, neither mothers nor children produced glottalized codas very often, and the two sets of speakers did not significantly differ. There was a significant effect of sentence position, with more frequent production of glottalized codas in sentence-final position compared to sentence-medial position. Unlike for /t/ target words, there was no significant interaction between age × position.



	Medial /t/	Final /t/	Medial /d/	Final /d/
■ Mothers	56/111 (50%)	133/237 (56%)	8/327 (2%)	30/239 (13%)
□ Children	2/44 (5%)	48/115 (42%)	1/60 (2%)	11/81 (14%)

Fig. 6. The number (percentage) of tokens showing glottalization at the end of the vowel.

Table 3
The estimates of the coefficients, standard errors, and their *z*-statistics for the logistic mixed-effects models with respect to glottalized codas.

	/t/ targets				/d/ targets			
	Estimate	Std. error	*z*-Value	*p*-Value	Estimate	Std. error	*z*-Value	*p*-Value
(Intercept)	−0.455	0.296	−1.534	0.125	−1.910	0.412	−4.632	$p<0.001$
Age	0.618	0.244	2.533	$p<0.05$	−0.074	0.387	−0.191	0.849
Position	−2.830	0.760	−3.725	$p<0.001$	−2.182	1.055	−2.068	$p<0.05$
Interaction	2.510	0.795	3.160	$p<0.01$	0.385	1.132	0.340	0.734

3.4. Summary of results

Our analyses demonstrated that 2-year-olds produced all three phonetic variants of /t, d/ (i.e., unreleased stops, flaps, glottalized stops) less often than their mothers. Further analysis also showed that, for mothers, 67% of the total codas (/t/: 74.7%, /d/: 62.5%) were unreleased, flapped, glottalized, or produced as different combinations of these three; only 33% of the /t, d/ codas were produced in canonical form, with a clear stop closure and a release burst. In contrast, for children, 59% of the codas were produced with a clear stop closure and a release, and only 41% of the codas (/t/: 41.5%, /d/: 41.1%) contained cues to one or more of the other phonetic variants of /t, d/. The differences between mothers and children were highly significant for both /t/ and /d/ codas (see Table 4). Overall, the results suggest that, in the early stages of language development, children's alveolar stop codas are more likely to be articulated in their canonical forms (i.e., with clear acoustic evidence for closure and release) than those of their adult caretakers.

Table 4
The estimates of the coefficients, standard errors, and their z-statistics for the logistic mixed-effects models with respect to the combined three phonetic variants.

	/t/ targets				/d/ targets			
	Estimate	Std. error	*z*-Value	*p*-Value	Estimate	Std. error	*z*-Value	*p*-Value
(Intercept)	−0.416	0.208	−2.004	$p<0.05$	−0.356	0.268	−1.327	0.184
Age	1.443	0.210	6.886	$p<0.001$	0.723	0.204	3.548	$p<0.001$

4. Discussion

The goal of this study was to examine 2-year-olds' production of three phonetic variants of the alveolar coda stops /t, d/ (i.e., unreleased stop, flap, and glottalized stop). In particular, we were interested to know whether children would start by producing the more canonical form of the phoneme (i.e., /t, d/ with a clear closure and a release), or were able to produce its other phonetic variants from early on. The results from this study collectively suggest that young children produce the three phonetic variants of the alveolar coda less often than adults. Although the children were exposed to non-canonical variants in the input, they did not produce these as often as their mothers did. Thus, the question arises as to why these phonetic variants are acquired late. That is, why does the child start out by producing a canonical form of the alveolar stop and only later produce more contextually appropriate phonetic variants? The late development of phonetic variants is puzzling, especially because the choice of many variants is known to be related to the phonetic contexts in which they appear. For example, alveolar stops occurring before another homorganic stop consonant (e.g., *hot dog*) are typically unreleased in conversational speech, and it is often assumed that this is because it is more economical for speakers to close once and release once for the pair of consecutive homorganic stops, rather than releasing twice. Since this is what children typically hear in everyday speech, one might in fact predict that they would produce such forms from the onset of connected speech.

We think there are at least two possible explanations of why adult-like mastery of context-appropriate phonetic variants is acquired late. The first possibility is that young children may have different phonological representations from adults. The second possibility is that the later acquisition of allophonic variants may be due to a lack of the articulatory skills needed to produce these non-canonical phonetic variants in an adult-like fashion. In the latter case, children might have the same phonological representation as adults, but have difficulty producing the context-appropriate phonetic variants.

Expanding upon the first possibility, our results suggest that children's representations of sounds might be at the level of the phoneme, rather than at the level of the phonetic variants. For adults, the knowledge of the sound pattern of a language encompasses both the phonemes and their phonetic variants. Adults understand the relationships between the two, so that they can choose to use phonetic variants in their appropriate contexts. Although children might have a clear understanding of the canonical form of a stop, they may still be developing their understanding of the phonetic variants and their relationship to the canonical form. For example, it is widely acknowledged that many children have a late-developing phonemic representation of /ɹ/. This is often attributed to the variable production of /ɹ/ across different contexts in the speech input they hear from the speech of the adult community around them (Klein, Byun, Davidson, & Grigos, 2013; Olive, Greenwood, & Coleman, 1993). Likewise, given the variable nature of phonetic variants of /t/ and /d/ in the speech input, children might have difficulty classifying them.

Studies have long shown that children follow the patterns of universal markedness, starting out with simpler and more unmarked structures and gradually learning more marked structures later in development. For instance, an open (CV) syllable is generally considered more unmarked, and these forms are acquired earlier than closed syllables (Demuth, 1995; Fikkert, 1994). If we extend this view to the phonetic domain, we might hypothesize that a coda stop with a closure and a release is acquired earlier because it is a less marked structure. The variant with a closure and a release might be called the canonical variant because it contains more of the cues to the features that define a stop consonant. Our evidence suggests that, at least for alveolar stop codas, the child first learns the canonical variant and only later masters the context-governed changes in this form that characterize other phonetic variants in fluent adult speech.

If children's phonemic representations are indeed different from an adult's, it will be important to determine how this different representation arises. One possibility is that, in their speech to children, adults model speech that suppresses the use of non-canonical phonetic variants. The best way to address this possibility is to compare the adults' use of such phonetic variants in adult-directed and child-directed speech. If these phonetic variants are less frequently used in child-directed speech than in adult-directed speech, then this would help children build a firm phonemic representation of the contrastive sound categories of their language early in development.

A number of studies have shown that sounds are in general more emphatic and clearly produced in child-directed speech than in adult-directed speech. This appears to be particularly true for vowels (Burnham, Kitamura, & Vollmer-Conna, 2002; Kuhl et al., 1997). For consonants, findings are mixed. Some researchers have found clearer cues to consonantal contrasts in child-directed speech (Bernstein Ratner, 1984; Marsheen, 1980), whereas others found no such effect (Baran, Laufer, & Daniloff, 1977; Sundberg & Lacerda, 1999). Just as the majority of work on children's speech production has relied on analyses at the phonemic level, many studies on child-directed speech have focused on the properties of phonemes. Therefore, phonetic variants and variant suppression have received relatively little attention in the child-directed speech literature. An exception is a recent study by Dilley, Millett, McAuley, and Bergeson (2014) that examined phonetic variation in alveolar stop coda consonants in environments where regressive place assimilation was possible (e.g., *green key*). Their results showed that mothers produced more canonical variants of alveolar stop codas (with less assimilation, glottalization, and deletion) in infant-directed speech than in adult-directed speech; this suggests that infant-directed speech delivers more careful, canonical pronunciations. However, importantly, their results also showed that infant-directed speech contained a considerable proportion of non-canonical consonantal variants as well, suggesting that infants are not exposed only to canonical variants. On the basis of these previous findings, we think it is likely that the use of canonical forms is more prevalent in child-directed speech. For example, flapping is affected by speech rate. The intervocalic /t/ in strong-weak contexts (e.g., *water*), which is generally flapped in adult-directed American English speech, is often aspirated instead in slow speech (Nespor & Vogel, 1986). Given that child-directed speech is generally slower (Fernald & Simon, 1984), it is possible that children hear a higher proportion of intervocalic /t/s produced with canonical cues indicating a stop closure and an aspirated release, rather than being realized as flapped.

Our results suggest that mothers use the three phonetic variants more often than children do. However, given that the present adult data were collected from child-directed speech, we suspect that mothers may have been suppressing their use of phonetic variants compared to their adult-directed speech. Furthermore, studies have suggested that females have a tendency to produce clearer speech compared to males, further restraining the use of phonetic variants (Zue & Laferriere, 1979). Therefore, we assume that the mothers' speech in the current study would have been quite 'careful'. Comparing mothers' use of phonetic variants in adult-directed and child-directed speech was not a goal of this study, but we note that addressing this issue in the future will enhance our understanding of children's developing phonological representations. Although we did not examine this issue directly here, we did investigate a related issue in an additional analysis; we asked whether there was a correlation between the frequency of phonetic variant use in mothers and their children. Our prediction was that if a mother uses phonetic variants more frequently, her child would be more advanced in the use of phonetic variants. The results revealed, however, that only one of the 12 correlations (/t, d/ $\times$ 3 phonetic variants $\times$ 2 sentence positions) is significant: a significant positive correlation between the percent of tokens of coda /d/ produced with glottalization in sentence-final position for mothers and children, $r(4)=0.953$, $p<0.01$. However, we interpret this lack of correlation with caution, because there were only six mother–child dyads in our sample.

In addition to individual caregivers' use of phonetic variants, another potential factor in the input that might play a role in the emergence of phonetic variants is word frequency. Some of the words in the speech input to children might contain phonetic variants more often than other words, as they appear in certain phonetic environments that trigger the production of phonetic variants. Caregivers' use of these words, in turn, will facilitate children's production of context-appropriate phonetic variants. Little is known about the source of phonetic habits that children acquire, and whether this might be lexically conditioned. A systematic examination of the relationship in phonetic patterns between individual caregivers and their children for specific target words differing in frequency of use would provide valuable information on the development of phonetic variants, an obvious question for future research.

The second possibility is that variants may be acquired late because the young child lacks the articulatory skills to produce some variants in an adult-like manner. In this case, children might have adult-like phonological representations, but have difficulty producing non-canonical phonetic variants appropriately. Because their articulatory motor control is still developing, children are known to lack articulatory efficiency (Kent, 1976). For example, although children's segment durations generally decrease with age, they are not adult-like until about 11 or 12 years of age (Kent & Forner, 1980; Lee, Potamianos, & Narayanan, 1999). This is illustrated by the development of flaps, whose production involves a quick movement of the tongue to briefly touch (or simply move toward) the alveolar ridge. As Rimac and Smith (1984) and Klein and Altman (2002) point out, the articulatory mechanism in children may lack the speed and precision required to produce the motion of the flap. Also, production of appropriate variants may be further inhibited because children are not producing the adult-like phonetic contexts in which the variants occur. For example, in adult-speech, flaps often occur intervocalically before unstressed syllables. However, children's rate of speech is generally slower and they sometimes stress syllables which adults would leave unstressed (Demuth & McCullough, 2009), eliminating the context for a flap to occur. Likewise, because speech rate is slower for children, there might be more time for the occlusion for a coda stop to be formed and released, resulting in more frequent release of coda stops and fewer unreleased variants in children's speech. In fact, young children not only release coda stops more often than adults, but they are also known to produce multiple release bursts as well as longer post-release (aspiration) noise compared to adults (Imbrie, 2005; Song et al., 2012). This has been often attributed to greater intraoral and subglottal air pressure in children, partly due to their smaller vocal tracts (Bernthal & Beukelman, 1978; Stathopoulus & Sapienza, 1993).

The case of glottalized stops appears more complicated than the other two variants. In the current study, children glottalized /t/s as often as mothers in sentence-final position, but significantly less often in sentence-medial position. Glottalization at the ends of utterances could be triggered by some physiological changes such as low subglottal pressure (Slifka, 2006) and thus, it may not be necessary to explicitly plan glottalization when it occurs at the end of utterances. That is, children's production of glottalization was comparable to that of adults in the case where it may be physiologically driven. This suggests that we would need to consider more than one factor to fully understand the lack of glottalization in the sentence-medial position and the late development of phonetic variants in general.

Although these two possibilities (i.e., phonological competence vs. articulatory limitations) may seem contrasting, the late development of phonetic variants is complex and dependent upon many simultaneously-acting variables; therefore, both possibilities may play a role in explaining the late development of phonetic variants. Here we are taking the view that there are many layers or aspects of language development. That is, the development of phonetic variants would involve not only acquiring adult-like phonemic representations, but also learning to articulate phonetic variants in an adult-like manner, and determining which contexts are appropriate for the production of these different cue patterns.

One reason why we think that the gap between these children's phonological representations of /t, d/ and that of adults is not substantial is that children's productions of variants are variable, suggesting that they have at least some understanding of where each variant can occur, as well as of how to produce it. For example, in Klein and Altman (2002), children often vocalized final /l/ in the lateral release context (e.g., *bottle*), creating a flap environment for the medial alveolar stop. The children produced flaps in these words, suggesting that they are aware of the phonetic conditioning of this phonetic variant. In other words, some of the variants that the child produce are context-appropriate, suggesting that the child has begun to master this aspect of the phonetics of their language even though performance is not always adult-like. Therefore, although the development of flap was overall gradual, there was some indication that children know when and how to produce flaps.

The current study indicates that children do not start by producing different phonetic variants of /t, d/ immediately. Rather, the initial shape of the /t, d/ coda more closely resembles the canonical form of a stop, with a clear closure and release. It takes time for children to gain the degree of accuracy and proficiency of articulatory motor control required to produce allophones in an adult-like manner. This developmental process includes learning to coordinate articulatory gestures for successive sounds and syllables to create the appropriate phonetic contexts for variants or to simply exploit an existing phonetic context. During the maturation process, children also seem to go through stages of learning and experimentation in which they master the relationship between the phoneme and its phonetic variants. We believe that the acquisition of phonological competence occurs in tandem with the maturation of articulatory control necessary to produce phonetic variants of a phoneme. Our findings suggest that the development of variants comes at a cost and depends upon many simultaneously-acting factors. The current study examined phonetic variants in spontaneous speech samples from a relatively small, homogeneous group of children and mothers. Further study of the development of phonetic variants would require elicitation of productions from a larger number of children and adults, systematically manipulating phonetic contexts and ages, to determine when these variants emerge and become more consistent in child speech.

Acknowledgments

We thank the children and their mothers for their participation. We also thank the members of the Child Language Lab at Brown University and the Phonetics Lab at the University of Wisconsin-Milwaukee (especially Elizabeth Alberswerth, Alana Dust, and Kacy Kreger) for research assistance. This work was supported by NIH Grant R01HD057606 to Demuth and Shattuck-Hufnagel.

Appendix A. Number of tokens broken up by participant and word type

Table A1
Number of tokens contributed by each participant for /t/ and /d/ targets. The number of tokens initially extracted from the corpus (initial dataset), excluded as there was no produced coda, and included for the final analyses are presented in three separate columns. The initials of the child participants 1–6 in this table and in Appendix B tables are as follows: A, E, L, N, V, and W.

(a) /t/ targets

	Sentence-medial position			*Sentence-final position*		
	Initial dataset	*No coda (Excluded)*	*Final dataset*	*Initial dataset*	*No coda (Excluded)*	*Final dataset*
Mother 1	11	0	11	29	2	27
Mother 2	22	0	22	48	0	48
Mother 3	21	0	21	31	1	30
Mother 4	45	1	44	81	1	80
Mother 5	8	0	8	24	0	24
Mother 6	5	0	5	28	0	28
Total	112	1	111	241	4	237
Child 1	11	2	9	31	3	28
Child 2	7	5	2	30	9	21
Child 3	6	2	4	15	0	15
Child 4	25	1	24	41	2	39
Child 5	3	0	3	5	1	4
Child 6	4	2	2	9	1	8
Total	56	12	44	131	16	115

(b) /d/ targets

	Sentence-medial position			*Sentence-final position*		
	Initial data	No coda (Excluded)	*Final dataset*	*Initial data*	No coda (Excluded)	*Final dataset*
Mother 1	30	0	30	38	2	36
Mother 2	57	0	57	46	0	46
Mother 3	62	1	61	34	0	34
Mother 4	110	3	107	86	0	86
Mother 5	29	1	28	11	0	11
Mother 6	44	0	44	26	0	26
Total	332	5	327	241	2	239
Child 1	4	3	1	11	2	9
Child 2	6	1	5	11	1	10
Child 3	11	1	12	15	0	15
Child 4	36	3	33	46	2	44
Child 5	3	1	2	1	0	1
Child 6	10	1	9	3	1	2
Total	70	10	60	87	6	81

Table A2
Number of word tokens by word type for mothers and children. The number of tokens initially extracted from the corpus (initial dataset), excluded as there was no produced coda, and included for the final analyses are presented in three separate columns.

(a) Mothers

		Sentence-medial position			*Sentence-final position*		
		Initial data	No coda (Excluded)	*Final dataset*	*Initial data*	No coda (Excluded)	*Final dataset*
/t/ targets	bat	2	0	2	12	1	11
	boat	9	0	9	21	0	21
	cat	44	1	43	63	2	61
	feet	21	0	21	54	0	54
	hat	30	0	30	64	0	64
	hot	6	0	6	27	1	26
	Total	112	1	111	241	4	237
/d/ targets	bed	33	1	32	51	0	51
	food	38	1	37	17	0	17
	good	223	3	220	114	0	114
	head	32	0	32	45	0	45
	side	6	0	6	14	2	12
	Total	332	5	327	241	2	239

(b) Children

		Sentence-medial position			*Sentence-final position*		
		Initial data	No coda (Excluded)	*Final dataset*	*Initial data*	No coda (Excluded)	*Final dataset*
/t/ targets	bat	3	0	3	29	8	21
	boat	12	1	11	18	2	16
	cat	22	5	17	29	2	27
	feet	5	0	5	8	1	7
	hat	5	1	4	37	3	34
	hot	9	5	4	10	0	10
	Total	56	12	44	131	16	115
/d/ targets	bed	16	1	15	23	0	23
	food	7	1	6	16	0	16
	good	29	3	26	14	3	11
	head	9	1	8	23	2	21
	side	9	4	5	11	1	10
	Total	70	10	60	87	6	81

Appendix B. Results broken up participant and word type

Table B1
The number (percentage) of tokens for *unreleased* for /t/ targets, broken up by participant (up) and word type (bottom).

Participant	*Sentence-medial*		*Sentence-final*	
	Mothers	Children	Mothers	Children
1	10/11 (91%)	4/9 (44%)	5/27 (19%)	6/28 (21%)
2	17/22 (77%)	1/2 (50%)	22/48 (46%)	6/21 (29%)
3	18/21 (86%)	0/4 (0%)	18/30 (60%)	3/15 (20%)
4	41/44 (93%)	3/24 (13%)	30/80 (38%)	0/39 (0%)
5	7/8 (88%)	0/3 (0%)	15/24 (63%)	0/4 (0%)
6	3/5 (60%)	1/2 (50%)	13/28 (46%)	3/8 (38%)
Total	96/111 (86%)	9/44 (20%)	103/237 (43%)	18/115 (16%)

Word	*Sentence-medial*		*Sentence-final*	
	Mothers	Children	Mothers	Children
bat	1/2 (50%)	1/3 (33%)	3/11 (27%)	3/21 (14%)
boat	8/9 (89%)	3/11 (27%)	8/21 (38%)	4/16 (25%)
cat	38/43 (88%)	3/17 (18%)	35/61 (57%)	6/27 (22%)
feet	18/21 (86%)	1/5 (20%)	19/54 (35%)	1/7 (14%)
hat	25/30 (83%)	0/4 (0%)	30/64 (47%)	4/34 (12%)
hot	6/6 (100%)	1/4 (25%)	8/26 (31%)	0/10 (0%)
Total	96/111 (86%)	9/44 (20%)	103/237 (43%)	18/115 (16%)

Table B2
The number (percentage) of tokens for *unreleased* for /d/ targets, broken up by participant (up) and word type (bottom).

Participant	*Sentence-medial*		*Sentence-final*	
	Mothers	Children	Mothers	Children
1	19/30 (63%)	1/1 (100%)	17/36 (47%)	2/9 (22%)
2	35/57 (61%)	3/5 (60%)	6/46 (13%)	1/10 (10%)
3	57/61 (93%)	6/10 (60%)	3/34 (9%)	5/15 (33%)
4	86/107 (80%)	14/33 (42%)	16/86 (19%)	10/44 (23%)
5	25/28 (89%)	1/2 (50%)	6/11 (55%)	0/1 (0%)
6	30/44 (68%)	7/9 (78%)	11/26 (42%)	1/2 (50%)
Total	252/327 (77%)	32/60 (53%)	59/239 (25%)	19/81 (23%)

Word	*Sentence-medial*		*Sentence-final*	
	Mothers	Children	Mothers	Children
bed	28/32 (88%)	4/15 (27%)	7/51 (14%)	4/23 (17%)
food	26/37 (70%)	1/6 (17%)	2/17 (12%)	1/16 (6%)
good	170/220 (77%)	19/26 (73%)	40/114 (35%)	3/11 (27%)
head	26/32 (81%)	3/8 (38%)	6/45 (13%)	6/21 (29%)
side	2/6 (33%)	5/5 (100%)	4/12 (33%)	5/10 (50%)
Total	252/327 (77%)	32/60 (53%)	59/239 (25%)	19/81 (23%)

Table B3
The number (percentage) of tokens for *flapped*, broken up by participant (up) and word type (bottom).

Participant	*/t/ targets*		*/d/ targets*	
	Mothers	Children	Mothers	Children
1	2/5 (40%)	0/3 (0%)	2/2 (100%)	0/0 (n/a)
2	8/12 (67%)	0/1 (0%)	16/17 (94%)	0/0 (n/a)
3	10/12 (83%)	0/3 (0%)	16/17 (94%)	0/3 (0%)
4	8/17 (47%)	0/10 (0%)	28/45 (62%)	3/9 (33%)
5	1/3 (33%)	1/2 (50%)	1/1 (100%)	0/1 (0%)
6	1/2 (50%)	1/1 (100%)	10/12 (83%)	1/1 (100%)
Total	30/51 (59%)	2/20 (10%)	73/94 (78%)	4/14 (29%)

Word	*/t/ targets*		*/d/ targets*		
	Mothers	Children	Word	Mothers	Children
bat	1/1 (100%)	0/0 (n/a)	bed	13/14 (93%)	0/2 (0%)
boat	3/4 (75%)	0/2 (0%)	food	8/15 (53%)	0/2 (0%)
cat	7/15 (47%)	1/12 (8%)	good	40/50 (80%)	3/5 (60%)
feet	11/12 (92%)	1/3 (33%)	head	11/11 (100%)	0/4 (0%)
hat	8/17 (47%)	0/2 (0%)	side	1/4 (25%)	1/1 (100%)
hot	0/2 (0%)	0/1 (0%)			
Total	30/51 (59%)	2/20 (10%)		73/94 (78%)	4/14 (29%)

Table B4
The number (percentage) of tokens for *glottalized* for /t/ targets, broken up by participant (up) and word type (bottom).

Participant	*Sentence-medial*		*Sentence-final*	
	Mothers	Children	Mothers	Children
1	8/11 (73%)	0/9 (0%)	11/27 (41%)	5/28 (18%)
2	10/22 (45%)	0/2 (0%)	29/48 (60%)	9/21 (43%)
3	4/21 (19%)	0/4 (0%)	16/30 (53%)	11/15 (73%)
4	27/44 (61%)	2/24 (8%)	50/80 (63%)	19/39 (49%)
5	6/8 (75%)	0/3 (0%)	11/24 (46%)	1/4 (25%)
6	1/5 (20%)	0/2 (0%)	16/28 (57%)	3/8 (38%)
Total	56/111 (50%)	2/44 (5%)	133/237 (56%)	48/115 (42%)

	Sentence-medial		*Sentence-final*	
Word	Mothers	Children	Mothers	Children
bat	1/2 (50%)	0/3 (0%)	5/11 (45%)	6/21 (29%)
boat	2/9 (22%)	0/11 (0%)	10/21 (48%)	3/16 (19%)
cat	25/43 (58%)	1/17 (6%)	44/61 (72%)	14/27 (52%)
feet	6/21 (29%)	0/5 (0%)	21/54 (39%)	2/7 (29%)
hat	16/30 (53%)	0/4 (0%)	35/64 (55%)	19/34 (56%)
hot	6/6 (100%)	1/4 (25%)	18/26 (69%)	4/10 (40%)
Total	56/111 (50%)	2/44 (5%)	133/237 (56%)	48/115 (42%)

Table B5
The number (percentage) of tokens for *glottalized* for /d/ targets, broken up by participant (up) and word type (bottom).

Participant	*Sentence-medial*		*Sentence-final*	
	Mothers	Children	Mothers	Children
1	0/30 (0%)	0/1 (0%)	6/36 (17%)	1/9 (11%)
2	6/57 (11%)	0/5 (0%)	3/46 (7%)	0/10 (0%)
3	1/61 (2%)	1/10 (10%)	9/34 (26%)	6/15 (40%)
4	0/107 (0%)	0/33 (0%)	10/86 (12%)	4/44 (9%)
5	1/28 (4%)	0/2 (0%)	1/11 (9%)	0/1 (0%)
6	0/44 (0%)	0/9 (0%)	1/26 (4%)	0/2 (0%)
Total	8/327 (2%)	1/60 (2%)	30/239 (13%)	11/81 (14%)

Word	*Sentence-medial*		*Sentence-final*	
	Mothers	Children	Mothers	Children
bed	0/32 (0%)	0/15 (0%)	11/51 (22%)	4/23 (17%)
food	1/37 (3%)	0/6 (0%)	1/17 (6%)	1/16 (6%)
good	4/220 (2%)	0/26 (0%)	15/114 (13%)	2/11 (18%)
head	2/32 (6%)	1/8 (13%)	3/45 (7%)	3/21 (14%)
side	1/6 (17%)	0/5 (0%)	0/12 (0%)	1/10 (10%)
Total	8/327 (2%)	1/60 (2%)	30/239 (13%)	11/81 (14%)

Appendix C. Acoustic descriptions of flapped codas

C.1. Percent released for flapped and unflapped codas

It was expected that flapped codas would be more often unreleased than unflapped codas (De Jong, 1998). Several observations can be made from the data shown in Fig. C1. For mothers, flapped codas were not necessarily more often unreleased than unflapped codas; rather, both were mostly unreleased. For children, our analysis showed that 1 of the 2 flapped /t/ codas was not released (not shown in Fig. C1) and similarly for 3 of the 4 flapped /d/ codas (shown in Fig. C1). Therefore, for children, there was a tendency to release unflapped codas more often than flapped codas. In addition, the results presented in Fig. 6 are consistent with the results discussed in the previous section that children generally release codas more often than mothers (also see related discussion in Song et al., 2012).

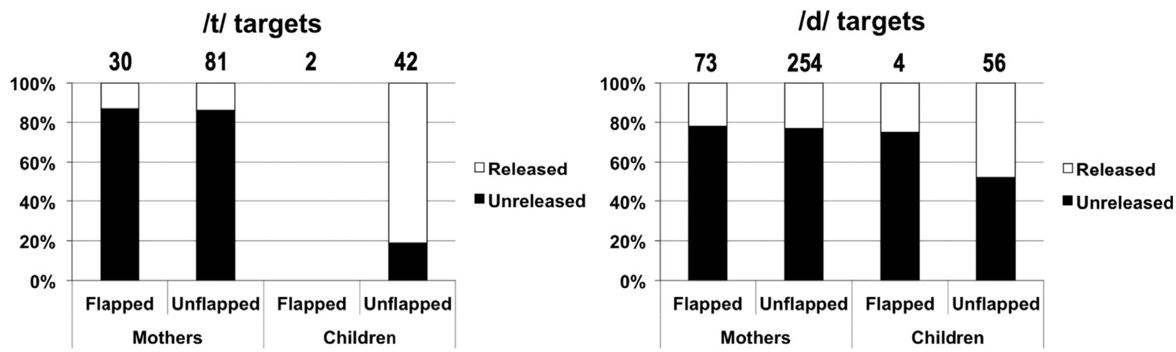


Fig. C1. Percent released and unreleased for flapped and unflapped /t/ codas (left) and /d/ codas (right). The total number of tokens used to calculate the percentages are shown above each bar. *Note*: Of the 2 instances of flapped /t/ codas, one was released and the other was not released. We decided not to present percent released for these on the figure, because we felt that it would be misleading to present the percentages of such small number of tokens.

C.2. Percent of closure duration that was voiced for flapped vs. unflapped codas

On this measure, 0% indicates that there was no vocal fold vibration during closure for the coda stop. Alternatively, 100% indicates that there was vocal fold vibration throughout closure. Following the literature (De Jong, 1998), we expected a higher percentage of voicing for flapped codas. This analysis was possible only for the tokens that had acoustic evidence for both a closure and a release, so that the duration of closure could be measured. Because the tokens coded as *unreleased* could not be included in this analysis, the data are sparse in some cells. In particular, because children had only a handful of flapped tokens to begin with, this left only one flapped token for the target /t/ and one flapped token for the target /d/ (see Table C1). Therefore, again, as in the first measure, we decided to present our observation of these tokens verbally below, rather than presenting the percentages of them in Table C1. These results need to be interpreted with caution. As expected, /d/ codas were overall more voiced than /t/ codas, and flaps were more voiced than unflapped items for both mothers and children. In particular, all 16 flapped /d/ codas in mothers' speech were fully voiced. Although not presented in Table C1, the two instances of children's flapped codas, which were produced by two different children, turned out to be also fully voiced. Of the two tokens, the average voicing and closure duration of the flapped /t/ were 62 ms, and those of the flapped /d/ were 3 ms.

Table C1
Mean percent of closure duration that was voiced for flapped and unflapped /t/ and /d/ codas in mothers' and children's speech production.

(a) /t/ targets	*Mothers*		*Children*	
	Flapped	*Unflapped*	*Flapped*	*Unflapped*
Number of tokens analyzed	4	11	1	34
Average voicing duration in ms (SD)	12 (14)	9 (15)	–	12 (19)
Average closure duration in ms (SD)	17 (12)	79 (56)	–	110 (43)
Average % voiced during closure (SD)	75 (50)	14 (31)	–	16 (30)

(b) /d/ targets	*Mothers*		*Children*	
	Flapped	*Unflapped*	*Flapped*	*Unflapped*
Number of tokens analyzed	16	59	1	27
Average voicing duration in ms (SD)	22 (14)	37 (19)	–	62 (38)
Average closure duration in ms (SD)	22 (14)	51 (31)	–	93 (39)
Average % voiced during closure (SD)	100 (0)	76 (31)	–	65 (33)

References

Baran, J. A., Laufer, M. Z., & Daniloff, R. (1977). Phonological contrastivity in conversation: A comparative study of voice onset time. *Journal of Phonetics*, *5*, 339–350.
Bates, D. M., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999999-0.
Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.
Bernstein Ratner, N. (1984). Phonological rule usage in mother–child speech. *Journal of Phonetics*, *12*, 245–254.
Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Papers in laboratory phonology III: Phonological structure and phonetic form* (pp. 7–33). Cambridge: Cambridge University Press.
Bernthal, J. E., & Beukelman, D. R. (1978). Intraoral air pressure during the production of /p/ and /b/ by children, youths, and adults. *Journal of Speech and Hearing Research*, *21*, 361–371.
Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new pussycat? On talking to animals and babies. *Science*, *296*, 1435.

Bolinger, D. (1961). Contrastive accent and contrastive stress. *Language*, *37*, 83–96.
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*, 39–54.
De Jong, K. (1998). Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics*, *26*(3), 283–310.
Demuth, K. (1995). Markedness and the development of prosodic structure. In J. Beckman (Ed.), *Proceedings of the North East Linguistic Society*, Vol. 25 (pp. 13–25). Amherst, MA: GLSA, University of Massachusetts.
Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, *49*(2), 137–173.
Demuth, K., & McCullough, E. (2009). The prosodic (re)organization of children's early English articles. *Journal of Child Language*, *36*, 173–200.
Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, *24*, 423–444.
Dilley, L. C., Millett, A. L., McAuley, J. D., & Bergeson, T. R. (2014). Acoustic-phonetic variation in word-final alveolar consonants in infant-directed speech over the first two years. *Journal of Child Language*, *41*(1), 155–175.
Docherty, G. J., & Foulkes, P. (1995). Acoustic profiling of glottal and glottalised variants of English stops. In *Proceedings of the XIIIth international congress of phonetic sciences* (pp. 350–353). Stockholm.
Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*, 104–113.
Fikkert, P. (1994). *On the acquisition of prosodic structure* (Ph.D. thesis), Holland Institute of Generative Linguistics (HIL), Leiden University. The Hague: Holland Academic Graphics.
Fukaya, T., & Byrd, D. (2005). An articulatory examination of word-final flapping at phrase edges and interiors. *Journal of the International Phonetic Association*, *35*(01), 45–58.
Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language*, *17*, 171–183.
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.
Hohne, E. A., & Jusczyk, P. W. (1994). Two-month-old infants' sensitivity to allophonic differences. *Attention, Perception, & Psychophysics*, *56*(6), 613–623.
Imbrie, A. K. K. (2005). *Acoustical study of the development of stop consonants in children* (Ph.D. thesis). Cambridge, MA: Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology.
Ingram, D. (1986). Phonological development: Production. In P. Fletcher, & M. Garmin (Eds.), *Language acquisition: Studies in first language development* (pp. 223–239). Cambridge: Cambridge University Press.
Janda, L. A., Nesset, T., & Baayen, R. H. (2010). Capturing correlational structure in Russian paradigms: A case study in logistic mixed-effects modeling. *Corpus Linguistics and Linguistic Theory*, *6*(1), 29–48.
Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Attention, Perception, & Psychophysics*, *61*(8), 1465–1476.
Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from Acoustic Studies. *Journal of Speech and Hearing Research*, *19*(3), 421–447.
Kent, R. D., & Forner, L. L. (1980). Speech segment duration in sentence recitations by children and adults. *Journal of Phonetics*, *8*, 157–168.
Klein, H. B., Altman, E. K., & Tate, J. (1998). Relationship between perceptual and acoustic judgments of children's flap productions. *Contemporary Issues in Communication Sciences and Disorders*, *25*, 25–31.
Klein, H. B., & Altman, E. K. (2002). The acquisition of medial /t, d/ allophones in bisyllabic contexts. *Clinical Linguistics & Phonetics*, *16*(3), 215–232.
Klein, H. B., Byun, T. M., Davidson, L., & Grigos, M. I. (2013). A multidimensional investigation of children's /r/ productions: Perceptual, ultrasound, and acoustic measures. *American Journal of Speech-Language Pathology*, *22*(3), 540–553.
Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*, 684–686.
Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, *105*, 1455–1468.
Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, *32*, 419–425.
MacNeilage, P. F., & Davis, B. L. (1990). Acquisition of speech production: The achievement of segmental independence. In J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 55–68). Netherlands: Springer.
MacWhinney, B. (2000). *The CHILDES project*. Mahwah, NJ: Erlbaum.
Marsheen, B. J. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology*, Vol. 2 (pp. 173–184). New York: Academic Press.
Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Fortis.
Newsome, M., & Jusczyk, P. W. (1995). Do infants use stress as a cue for segmenting fluent speech. In Proceedings of the 19th annual Boston university conference on language development (Vol. 2, pp. 415–426). Boston University, MA.
Olive, J. P., Greenwood, A., & Coleman, J. (1993). *Acoustics of American English speech*. New York: Springer.
Peperkamp, S., & Dupoux, E. (2002). Coping with phonological variation in early lexical acquisition. In I. Lasser (Ed.), *The process of language acquisition* (pp. 359–385). Frankfurt: Peter Lang.
Peperkamp, S., Pettinato, M., & Dupoux, E. (2003). Allophonic variation and the acquisition of phoneme categories. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th annual Boston University conference on language development.*, Vol. 2 (pp. 650–661). Sommerville, MA: Cascadilla Press.
Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 90–117). Cambridge: Cambridge University Press.
R Development Core Team (2011). *A language and environment for statistical computing, R foundation for statistical computing*. Vienna, Austria (Computer program). Available at 〈http://www.R-project.org〉.
Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, *29*(4), 407–429.
Rimac, R., & Smith, B. L. (1984). Acoustic characteristics of flap productions by American English-speaking children and adults: Implications concerning the development of speech motor control. *Journal of Phonetics*, *12*(4), 387–396.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.
Slifka, J. (2006). Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice*, *20*(2), 171–186.
Song, J. Y., Demuth, K., & Shattuck-Hufnagel, S. (2012). The development of acoustic cues to coda contrasts in young children learning American English. *Journal of the Acoustical Society of America*, *131*(4), 3036–3050.
Stathopoulus, E. T., & Sapienza, C. (1993). Respiratory and laryngeal measures of children during vocal intensity variation. *Journal of the Acoustical Society of America*, *94*(5), 2531–2541.
Stoel-Gammon, C., & Buder, E. (1999). Vowel length, post-vocalic voicing and VOT in the speech of two-year olds. In: Proceedings of the XIIIth international conference of phonetic sciences 3, 2485–2488.
Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica*, *56*(3–4), 186–199.
White, K. S., Peperkamp, S., Kirk, C., & Morgan, J. L. (2008). Rapid acquisition of phonological alternations by infants. *Cognition*, *107*, *238–265*.
Zue, V. W., & Laferriere, M. (1979). Acoustic study of medial /t, d/ in American English. *Journal of the Acoustical Society of America*, *66*, 1039.