



## Research Article

# Statistical distributions of consonant variants in infant-directed speech: Evidence that /t/ may be exceptional

Laura Dilley<sup>a,\*</sup>, Jessica Gamache<sup>b</sup>, Yuanyuan Wang<sup>c</sup>, Derek M. Houston<sup>c</sup>, Tonya R. Bergeson<sup>d,e</sup>

<sup>a</sup> Department of Communicative Sciences and Disorders, Michigan State University, United States

<sup>b</sup> Department of Linguistics and Germanic, Slavic, Asian and African Languages, Michigan State University, United States

<sup>c</sup> Department of Otolaryngology, The Ohio State University, United States

<sup>d</sup> Dept. of Otolaryngology – Head & Neck Surgery, Indiana University School of Medicine, United States

<sup>e</sup> Department of Communication Sciences and Disorders, Butler University, United States

## ARTICLE INFO

## Article history:

Received 15 January 2018

Received in revised form 12 May 2019

Accepted 13 May 2019

Available online 24 May 2019

## Keywords:

Infant-directed speech  
phonetic variants

## ABSTRACT

Statistical distributions of phonetic variants in spoken language influence speech perception for both language learners and mature users. We theorized that patterns of phonetic variant processing of consonants demonstrated by adults might stem in part from patterns of early exposure to statistics of phonetic variants in infant-directed (ID) speech. In particular, we hypothesized that ID speech might involve greater proportions of canonical /t/ pronunciations compared to adult-directed (AD) speech in at least some phonological contexts. This possibility was tested using a corpus of spontaneous speech of mothers speaking to other adults, or to their typically-developing infant. Tokens of word-final alveolar stops – including /t/, /d/, and the nasal stop /n/ – were examined in assimilable contexts (i.e., those followed by a word-initial labial and/or velar); these were classified as canonical, assimilated, deleted, or glottalized. Results confirmed that there were significantly more canonical pronunciations in assimilable contexts in ID compared with AD speech, an effect which was driven by the phoneme /t/. These findings suggest that at least in phonological contexts involving possible assimilation, children are exposed to more canonical /t/ variant pronunciations than adults are. This raises the possibility that perceptual processing of canonical /t/ may be partly attributable to exposure to canonical /t/ variants in ID speech. Results support the need for further research into how statistics of variant pronunciations in early language input may shape speech processing across the lifespan.

© 2019 Elsevier Ltd. All rights reserved.

## 1. Introduction

Early exposure to spoken language profoundly influences the organization of the linguistic system (Kuhl, 1991; Werker & Tees, 1984; Werker, Yeung, & Yoshida, 2012). An important challenge is to understand how children, through exposure to ambient language, learn phonemes and other linguistic structures (e.g., Cristia, Dupoux, Gurven, & Stieglitz, 2017). A consensus view is that language development hinges on cognitive representations of statistical properties of the input (Gomez, 2007; Kleinschmidt & Jaeger, 2015; Norris & McQueen, 2008; Pierrehumbert, 2003; Saffran, 2003), including phonetic properties (Maye, Werker, & Gerken, 2002; Werker & Curtin,

2005; Werker et al., 2007). Therefore, characterizing statistical distributions of variant speech sounds in early input is important for explaining how children develop the ability to understand words (Ernestus & Baayen, 2007; Mitterer & Ernestus, 2006; Pitt, 2009b; Pitt, Dilley, & Tat, 2011; Sumner & Samuel, 2005, 2009; Warner & Tucker, 2011).

There are numerous sources of acoustic variability in speech sounds. One important source of variability is coarticulation, a situation in which the properties of one sound is influenced by an adjacent sound (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Stevens, 2000, 2002). Coarticulation concerns how articulation of a consonantal constriction at a given location in the vocal tract influences that of an adjacent consonant. For example, an alveolar consonant – that is, one involving a constriction at the alveolar ridge (e.g., /n/ in *green*) may take the place of articulation of an adjacent labial sound (e.g., /b/ in *ball*), resulting in a phrase like *green ball* being spoken as *greem ball*. Another source of variability is allophonic

\* Corresponding author at: Michigan State University, Department of Communicative Sciences and Disorders, 1026 Red Cedar Road, Rm. 102, Oyer Speech & Hearing Building, East Lansing, MI 48824-1220, United States. Fax: +1 517 353 3176.

E-mail address: [ldilley@msu.edu](mailto:ldilley@msu.edu) (L. Dilley).

variation. For example, the phoneme /t/ in English, which is one of the most common sounds in the language, takes many variant forms, including a *canonical* variant of /t/ (e.g., *tea* /t<sup>hi</sup>/ or *star* /star/) – that is, a variant with an overt stop closure and release (which may be aspirated or unaspirated) – a flap /ɾ/ (e.g., *city* /sɪɾ/), or a glottal variant /ʔ/ (e.g., *cat* /k<sup>h</sup>æʔ/) (Patterson & Connine, 2001; Warner & Tucker, 2011; Warner, Fountain, & Tucker, 2009). A /t/ may also be deleted, especially word-finally (Deelman & Connine, 2001; Dilley & Pitt, 2007; Dilley, Millett, McAuley, & Bergeson, 2014; Mitterer & Ernestus, 2006; Mitterer & McQueen, 2009; Pitt, 2009a, 2009b; Staum Casasanto, 2008). Coarticulatory contexts also differ in sonority (cf. obstruent vs. sonorant). Obstruent sounds, including most consonants, are produced with a significant constriction along the length of the vocal tract, while sonorant sounds are produced without such a constriction. Glottal variants have been found to be more frequent in sonorant than obstruent contexts (Huffman, 2005; Pierrehumbert, 1994).

Speech perception studies suggest that listeners draw on phonetic variant frequency statistics to process spoken language (Connine, Ranbom, & Patterson, 2008; Johnson, 2006; Pierrehumbert, 2003; Pitt et al., 2011; Pitt, 2009a, 2009b; Reinisch & Mitterer, 2016; Seyfarth, 2014). There is growing evidence that phonetic variants are important for spoken word recognition (McQueen, Cutler, & Norris, 2006; Mitterer, Chen, & Zhou, 2011; Mitterer, Cho, & Kim, 2016; Mitterer, Reinisch, & McQueen, 2018; Mitterer, Scharenborg, & McQueen, 2013; Reinisch, Wozny, Mitterer, & Holt, 2014; Sjerps & McQueen, 2010). Considerable attention has been given to processing of /t/ variants, with this literature showing that the most frequently-occurring variant pronunciation also tends to generate the most robust lexical representations (Connine et al., 2008; Connine, 2004; Patterson, LoCasto, & Connine, 2003; Pitt et al., 2011). For example, Pitt et al. (2011) manipulated allophone type and phonological environment in a lexical decision task. Listeners showed enhanced lexical representations for non-canonical variants, i.e., /ɾ/, /ʔ/, and the deleted form, in phonological contexts where these occurred frequently, respectively. However, while corpus statistics have shown that canonical /t/ is a *less* frequently occurring variant, relative to other variants of /t/ (Connine, 2004; Pitt et al., 2011), this variant type nevertheless generated clear word percepts regardless of frequency, and tended to be processed the fastest. This pattern – whereby the apparently less-frequent canonical variant /t/ is associated with more robust lexical processing – has been replicated multiple times (Ernestus & Baayen, 2007; McLennan, Luce, & Charles-Luce, 2003, 2005; Pitt et al., 2011; Pitt, 2009a, 2009b; Tucker & Warner, 2007). One explanation had been that learning to read influences the way spoken language is processed (Pitt et al., 2011; Ranbom & Connine, 2007). While some recent research has failed to support that learning to read affects processing of speech (e.g., Mitterer & Reinisch, 2015), other research has supported that learning to read can influence lexical processing in various ways (e.g., Bowers, Kazanina, & Andermane, 2016). This uncertainty over reasons for divergent findings about variant processing leads to the need for further examination of distributional facts about canonical /t/.

This paper sought to test whether early language input to infants might sometimes involve more canonical /t/ variants

than in typical conversational patterns shown by adults, which might potentially be related to the processing patterns described above. Talkers are known to adjust their pronunciations according to the communicative needs of listeners (Buz, Tanenhaus, & Jaeger, 2016; Maniwa, Jongman, & Wade, 2009; Oviatt, MacEachern, & Levow, 1998; Picheny, Durlach, & Braida, 1986; Schertz, 2013; Schertz, Cho, Lotto, & Warner, 2015; Stent, Huffman, & Brennan, 2008; Uther, Knoll, & Burnham, 2007). When talking to young children, caregivers use a speech style referred to here as *infant-directed (ID) speech*. ID speech differs from adult-directed (AD) speech in a number of ways, and some of these differences have been proposed by researchers to result from a tendency toward greater clarity in ID compared with AD speech. For example, a number of studies have found that ID speech has vowels with a more expanded area in first (F1) and second (F2) formant space than AD speech (Andruski, Kuhl, & Hayashi, 1999; Burnham et al., 2015; Burnham, Kitamura, & Vollmer-Conna, 2002; Cristia & Seidl, 2013; Kuhl et al., 1997; Liu, Tsao, & Kuhl, 2009; Miyazawa, Shinya, Martin, Kikuchi, & Mazuka, 2017; Wassink, Wright, & Franklin, 2007; Wieland, Burnham, Kondaurava, Bergeson, & Dilley, 2015). Other studies have not found this pattern but instead showed that the vowel space was reduced, shifted, or maintained in ID speech, as compared to AD speech (Benders, 2013; Burnham et al., 2015; Cristia & Seidl, 2013; Englund & Behne, 2006; Martin et al., 2015; McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013). For instance, McMurray et al. (2013) found that corner vowels /i, a, u/, but not interior vowels, showed expansion of the F1-F2 vowel space in ID compared with AD speech. ID speech also shows prosodic modifications, such as increased fundamental frequency (F0) range and variability, a slower rate, and longer pauses compared to AD speech (Bergeson & Trehub, 2002; Bergeson, Miller, & McCune, 2006; Fernald & Simon, 1984; Fernald, 1992; Papoušek, Papoušek, & Bornstein, 1985; Snow, 1977). Infants not only detect this pronunciation variation, but also prefer listening to speech presented in this style (Cooper & Aslin, 1990; Wang, Bergeson, & Houston, 2017). There is also evidence that the modification in ID speech benefits infant speech processing and language development (Hartman, Ratner, & Newman, 2017; Thiessen, Hill, & Saffran, 2005; Weisleder & Fernald, 2013), possibly because caregivers may unconsciously structure the input to their children in a way that is informative about underlying categories (Eaves, Feldman, Griffiths, & Shafto, 2016). Consistent with this, some research has shown that ID speech enhances infants' word segmentation skills compared with AD speech (Thiessen et al., 2005), although these effects have not been tied to segmental realizations *per se*. Moreover, the extent of vowel expansion in ID speech relative to AD speech predicts infants' speech perception skills (Liu, Kuhl, & Tsao, 2003).

Studies on phonetic variation at the segmental level in ID and AD speech are important because they inform understanding of how children use statistical information to form phonetic categories (Maye et al., 2002; Werker et al., 2007). Despite the findings regarding phonetic modifications of vowels in ID speech, there remain considerable questions about how the phonetic properties of *consonants* in ID speech differ statistically from AD speech. Existing studies of consonant

modification in ID speech have yielded inconsistent results (Baran, Laufer, & Daniloff, 1977; Bernstein Ratner, 1984; Englund, 2005; Fish, Garcia-Sierra, Ramirez-Esparza, & Kuhl, 2017; Lahey & Ernestus, 2013; Malsheen, 1980; McMurray et al., 2013; Shockey & Bond, 1980; Sundberg & Lacerda, 1999; Sundberg, 2001). For example, while Bernstein Ratner (1984) found that phonological reduction in consonants (e.g., *want it* → *wannit*) happened less often in ID speech than AD speech, Shockey and Bond (1980) found the opposite pattern of results: phonological reduction occurred more often in ID speech than AD speech. Further, McMurray et al. (2013) found that voice onset times for voiced vs. voiceless stop consonants were not statistically more distinctive in ID compared with AD speech. (See also Cristia & Seidl, 2013; Miyazawa et al., 2017 for similar findings.)

Looking across these studies, factors that may be implicated in explaining the inconsistent findings include the type of speech materials (read vs. spontaneous), the specific types of phonemes studied, the type of phonological context, the ages and genders of the children, and differences among words themselves (e.g., frequency, typical ages of acquisition) (Dilley et al., 2014; Foulkes, Docherty, & Watt, 2005; Lahey & Ernestus, 2013; McMurray et al., 2013; Sundberg & Lacerda, 1999; Zellou & Scarborough, 2015). Still, there may be important cases where ID provides clearer input than AD, though these may relate to a subset of segment types and/or contexts. Although study design variables may partly account for inconsistent findings about consonant clarity in ID speech, there is growing consensus that developing an understanding of statistics of consonant variant pronunciations in ID speech will require appropriately sophisticated and/or advanced statistical techniques (Barth & Kapatsinski, 2018; Th. Gries, 2015). Further, such studies will be aided by separate examinations of phonological reduction processes while carefully controlling for phonological environment (Buckler, Goy, & Johnson, 2018; Wang & Seidl, 2015; Wang, Seidl, & Cristia, 2015).

One type of consonant variation that has been subject to considerable theorizing and interest is regressive place assimilation of word-final stop consonants. Previous work has shown that when a token seems assimilated, traces of the original place of articulation are often present acoustically (e.g., in F2) (Dilley & Pitt, 2007; Gow, 2001, 2002, 2003; Holst & Nolan, 1995). Adult listeners can use this information to recover the intended phoneme when regressive place assimilation has occurred (Dilley & Pitt, 2007; Ellis & Hardcastle, 2002; Gaskell & Marslen-Wilson, 1998; Lahiri & Marslen-Wilson, 1991), although the specific patterns of recovery depend on listeners' native language (Cho & McQueen, 2008; Mitterer, Kim, & Cho, 2013). However, phonological accounts of this variation have tended to focus on categorical changes from a canonical to assimilated place of articulation without consideration for occurrences of other kinds of variants. Nevertheless, the process of perceptual recovery of an intended word-final segment in adult speech perception is more complex than envisioned by many accounts focusing just on Canonical vs. Assimilated distinctions (e.g., Gaskell & Marslen-Wilson, 1998; Lahiri & Marslen-Wilson, 1991).

Dilley and Pitt (2007) investigated the distributional statistics of phonetic variants in a corpus of spontaneous (AD) speech from 40 talkers in Ohio (Pitt, Dilley, Johnson, Kiesling,

Raymond, Hume, & Fosler-Lussier, 2007). They specifically investigated variant usage in “assimilable” environments, where regressive place assimilation of word-final alveolar stop consonants /t/, /d/, and the nasal stop /n/ was permitted (e.g., green#ball). Each token in an assimilable environment was classified as Canonical, Assimilated, Glottalized, or Deleted based on spectrographic and auditory criteria. Results showed that deleted and glottalized variants were common in this restricted phonological environment, contra to accounts focused on Canonical vs. Assimilated alternations. Moreover, canonical variants were the most frequent for /n/ across segments and the least frequent for /t/, consistent with other corpus studies showing low frequency for canonical /t/ (Patterson & Connine, 2001; Warner & Tucker, 2011). Acoustic evidence (e.g., in F2) suggested that assimilation was gradient rather than categorical, reflecting graded changes in place of articulation, so that tokens classified as “Assimilated” likely just evidenced a particularly severe form of pronunciation variation.

Moreover, Dilley et al. (2014) examined whether ID speech contained more unambiguous, canonical forms than AD speech in assimilable environments. Using a corpus of read storybook speech by 48 mothers, Dilley et al. examined mothers' rates of canonical variant usage in AD speech relative to ID speech. There was a weak statistical tendency for mothers to use canonical variants more often in ID than in AD speech, i.e., significance was found in a one-tailed *t*-test (but not a two-tailed test) at  $p < 0.05$ . However, variant pronunciation was not separately examined as a function of alveolar stop type – /t/, /d/ or /n/. Further, as they noted, read storybook speech may not be generalizable to naturalistic, spontaneous speech (Ernestus & Warner, 2011; Warner & Tucker, 2011). In addition, there was a null effect of infant age, where this finding might have been due to the lower power of the Chi-square statistics used to examine possible group effects. Further, there was a significant difference in rates of canonical usage between mothers of female vs. male infants, for the youngest group of infants: mothers who had female infants used canonical variants more frequently than those who had male infants. Different variant usage to children as a function of gender has been reported elsewhere (Foulkes et al., 2005).

Buckler et al. (2018) revisited the issue of consonant variation in ID speech by examining corpus data to examine statistics of variation in assimilable environments. Using the classification method of Dilley and colleagues (e.g., Dilley & Pitt, 2007), Buckler et al. (2018) classified consonant pronunciations in assimilable environments in both read and spontaneous speech. For read storybook speech, they found a slightly higher, but non-significant, preponderance of tokens with canonical pronunciations in ID and AD speech (46% vs. 43%, respectively). For their corpus of spontaneous speech produced by a small number of talkers (i.e., mothers of 6 children – 3 boys and 3 girls), they focused only on speech that the researchers judged to be ID. For this corpus they showed that canonical pronunciations did not predominate in ID speech (Canonical: 24%, Deleted: 33%, Assimilated: 22%, Glottalized: 21%); they concluded that ID speech does not have a statistical predominance of canonical pronunciations.

However, both Buckler et al. (2018) and Dilley et al. (2014) had limitations. First, and perhaps most importantly, neither study reported separate distributions of variants for /t/, /d/,

and /n/. Second, Dilley et al. examined read storybook speech, which tends to be pronounced more carefully and might not reflect patterns in spontaneous speech. While Buckler et al. (2018) examined spontaneous speech, the samples came from home audio recordings; therefore, the researchers might have made an inaccurate judgment of whether an utterance was directed to an infant or an adult, which may have led to sampling bias. Third, a very small sample of talkers was examined in Buckler et al. (2018) – mothers of just six children – where one of the children was later diagnosed with Asperger's syndrome. This small sample limited not only the external generalizability of Buckler et al. (2018), but precluded meaningful analysis of effects of infant age and gender. Collectively, these design factors limited the generalizability of both Buckler et al. (2018) and Dilley et al. (2014) in complementary ways.

In the present study we undertook a relatively compact, focused re-examination of statistical distributions of word-final consonant variants in spontaneous ID speech in assimilable environments, a type of variation with considerable theoretical significance (Cho & McQueen, 2008; Gaskell & Marslen-Wilson, 1998; Lahiri & Marslen-Wilson, 1991). A strength of our corpus study of spontaneous speech was the use of mixed-effect statistical modeling, a mathematical approach which has increasingly been recognized as a powerful tool in the arsenal of analysis of linguistic patterns in natural language corpus data (Th. Gries, 2015, 2016). Mixed-effect statistical modeling enables estimating statistical significance of independent variables while simultaneously “parceling out” shared variance due to random factors, such as covariance associated with individual subjects or with individual items (e.g., words). Mixed-effect models are increasingly recognized for their value in ensuring robustness of statistical significance of independent variables under conditions of imbalanced data sets that are typical of natural language corpora, including variable numbers of observations per subject or item (Th. Gries, 2015, 2016).

The present study offered multiple advantages over prior studies of a theoretically significant type of context for variation – i.e., contexts where assimilation can occur – and critically overcame several limitations. First, unlike Buckler et al. (2018), we examined both ID and AD spontaneous speech, using clearly delineated addressee conditions. Moreover, we investigated variation in a larger sample of mothers and tokens than used previously. Third, we examined consonant variants at the segmental level, looking separately at /t/, /d/, and /n/. We used the now well-established method of categorization of phonetic variants (cf. Dilley & Pitt, 2007). We hypothesized that enhanced canonicity in ID speech might reveal itself in a segment-specific manner. Findings from this research will not only allow us to characterize phonetic properties of the input infants receive, but also inform language acquisition theories.

## 2. Methods

### 2.1. Participants

Mothers of 53 typically-developing infants (23 F, 30 M) were recruited from the local Indianapolis community. These dyads constituted a pool of control subjects for a study on differences in speech to children as a function of hearing status; all the

mothers and their infants had normal hearing. The children ranged in age at the time of the visits from about 0;2–2;6, with an average age of 1;0. As part of their visits, mothers completed a demographic survey to which they provided information about their language background and their infant's home language environment. All mothers were monolingual American English speakers who lived in the Mid-Western United States, and none of the infants were in bilingual learning environments. They were given \$10 per visit for their time. This research and the recruitment of human subjects were approved by the Indiana University Institutional Review Board.

### 2.2. Materials

Mothers participated in two lab visits separated in time by about 3 months. Given that prior studies have not found differences in ID speech phonetic properties produced by mothers across such a short time frame (e.g., Dilley et al., 2014; Burnham et al., 2015), we pooled production data across both visits to increase power of analyses. The experimental procedure was identical at each visit. In the ID speech condition, mothers were asked to sit with their child on a blanket on the floor or in a chair. Mothers were instructed to speak to their child as they normally would do at home while playing with quiet toys. In the AD speech condition, an adult experimenter conducted a semi-structured short interview with each mother. The interview included questions about the typical household daily routine, the infant's favorite play activities, and the infant's communicative behaviors. Each ID and AD session lasted approximately 2–5 min.

Speech samples were recorded in a sound-attenuated booth. Mothers wore a wireless microphone (Shure) during the recording sessions. Technical equipment for recording for the first phase of the project consisted of a hypercardioid microphone (Audio-Technica ES933/H) at a fixed location in the sound booth powered by a phantom power source and linked to an amplifier (DSC 240) and digital audio tape recorder (Sony DTC-690). The equipment was updated part way through the longitudinal project to an SLX Wireless Microphone System (Shure). This system included an SLX1 Body-pack transmitter affixed to the child with a special vest with a built-in microphone and a wireless receiver SLX4 connected to a Canon 3CCD Digital Video Camcorder GL2, NTSC, which recorded the speech samples directly onto a Mac computer (Apple, Inc. OSX Version 10.4.10) via Hack TV (Version 1.11) software.

### 2.3. Phonetic analysis

For each ID and AD recording, instances of target assimilable environments were identified. Tokens analyzed were limited to word-final alveolar consonants (/t/, /d/, /n/) followed by a word-initial consonant where place assimilation could potentially occur, i.e., those consonants with a labial and/or velar place of articulation, including /b/, /p/, /m/, /g/, and /k/, as well as /w/, which has both labial and velar constrictions (Ladefoged & Maddieson, 1996). A total of 1797 tokens met these criteria. Tokens were subsequently excluded if the word-final alveolar was separated from the following consonant by a pause of greater than 100 msec ( $n = 24$ ), or if noise

prevented sufficient acoustic clarity for classification ( $n = 12$ ). After applying these criteria, a total of 1761 tokens remained.

The frequency of each token of a lexical item with a word-final /t/, /d/, or /n/ in assimilable contexts was then determined using the News on the Web (NOW) corpus (Davies, 2017), which contained approximately 5.3 billion words at the time of analysis. Relative frequency of each lexical item in the corpus was converted to a word frequency estimate by taking  $\log_{10}$  of the number of expected/imputed instances of the lexical item per one million words (e.g., Kittredge, Dell, Verkuilen, & Schwartz, 2008).

Two undergraduate assistants trained in phonetic variant classification categorized each token as belonging to one of the four phonological categories (Canonical, Assimilated, Deleted, or Glottalized) on the basis of spectrographic and auditory perceptual evidence. In instances where there was a disagreement in the label, a third trained labeler classified that token and a classification was selected corresponding to the majority opinion. If there was still no agreement among the three labelers, the token was discarded ( $n = 127$ ). The same group of three analysts classified AD and ID conditions for each mother. Kappa reliability was found to be 0.67 across the three groups, which indicates substantial agreement (Breen, Dilley, Kraemer, & Gibson, 2012; Landis & Koch, 1977; Rietveld & van Hout, 1993). (See Dilley et al., 2014 for more details.)

#### 2.4. Statistical analysis

A series of generalized linear mixed effects models (GLMMs) were conducted to assess statistical significance (Agresti, 2012; Jaeger, 2008). For analyses involving binary categories, binomial logistic regression with a logit linking function was implemented in R using the *glmer* function within the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015), and one of the two categories is designated a referent category. Multinomial logistic regression involves generalization of principles of binomial logistic regression to more than two categories (Agresti, 2012); in this case, the referent category is compared with each of the other categories. In analyses involving more than two response variable categories, we implemented multinomial logistic regression.

All GLMMs included both fixed and random effects consisting of random intercepts by items and by subjects; models that involved the factor of Condition (ID, AD) included random slopes by subjects as well. This was the maximal random effects structure justified by the study design (Barr, Levy, Scheepers, & Tily, 2013; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). Random effects in GLMMs statistically parcel out random shared covariance due to e.g., individual subjects and/or items. Such models are useful for statistically controlling for e.g., variable numbers of observations per subject or item and are a powerful way of determining statistical significance of independent variables (i.e., fixed effects) both for balanced experimental data and imbalanced corpus data (Barth & Kapatsinski, 2018; Th. Gries, 2015, 2016). For this and other corpus data, we used treatment coding, in which one level of each independent variable (i.e., fixed factor) is set as the reference level against which the other levels are compared. The *drop1* command combined with a

Chi-square test in R was used in cases of iterative model reduction to identify non-significant interaction terms that contributed the least to model fitness.

Because multinomial logistic regression with both random subject and item effects has not yet been implemented within the *lme4* package in R, we used the *MCMCglmm* package (Hadfield, 2010a) for this purpose. *MCMCglmm* utilizes Bayesian Markov chain Monte Carlo (hence, *MCMC*) methods to implement model fits, given that closed-form solutions to such mixed models cannot usually be found (Fong, Rue, & Wakefield, 2010; Hadfield, 2010a). Given this Bayesian statistical implementation, a model requires specification of prior probabilities and outputs estimates of posterior probabilities. (See Bolstad & Curran, 2016, for an introduction to Bayesian statistics.) Following recommendations for categorical distributions (Hadfield, 2010a), fixed effect priors were set at 0.5 for all of the diagonal terms and 0.25 for all of the off-diagonal terms (covariance). Further, a burn-in period of 50,000 with 500,000 iterations and a thinning interval of 250 were used. (See Hadfield, 2010b, for more information.) Random effect priors for the by-item and by-subjects intercept terms were inverse-Wishart distributed (Hadfield, 2010b). Posterior probability estimates output by the Bayesian *MCMCglmm* statistical implementation include a posterior mean given in log-odds (analogous to a  $\beta$  estimate) and a 95% credibility interval (analogous to standard error). Type I error rates are estimated using Markov chain Monte Carlo methods and quantified as *pMCMC* (Baayen, 2008) and are analogous to *p*-values. Significance was assessed at  $\alpha = 0.05$ . (See Appendix for supplemental information on statistical models.)

### 3. Results

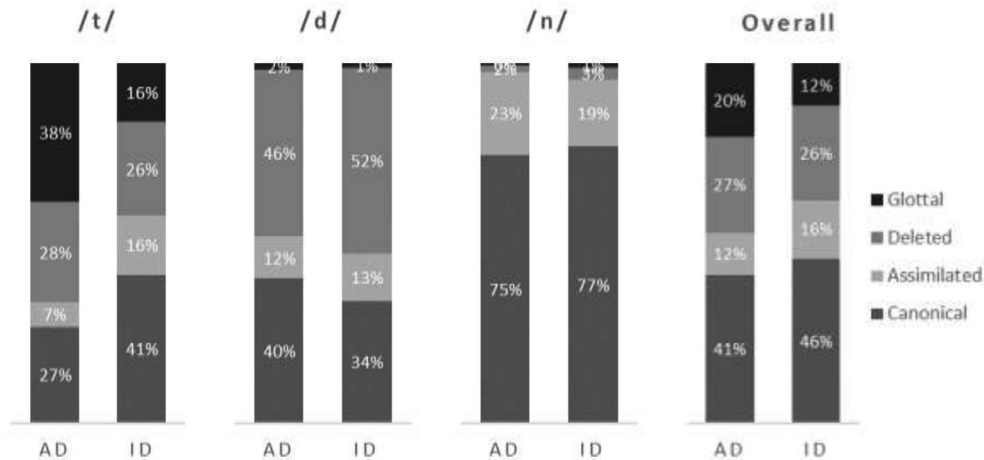
The final data set consisted of 1634 tokens. This included 584 tokens of ID speech and 1050 tokens of AD speech. The number of tokens by Segment (/t/, /d/, /n/) and by Condition (ID, AD) is shown in Table 1. Token counts are further broken down by content and function words in AD and ID conditions which ended in each Segment type. Each mother contributed an average of  $N = 31$  tokens ( $M_{ID} = 11$  tokens, range = 0–40;  $M_{AD} = 20$  tokens, range = 1–62). There was a slightly higher proportion of content words in assimilable contexts in AD speech, compared with ID speech (35% vs. 28%),  $\chi^2(1) = 8.67$ ,  $p < 0.01$ . The high overall number of function words in ID speech is consistent with high frequency of contextualized elements (e.g., pronouns like *he*) over decontextualized lexical items, as well as simpler structure in this register.

Fig. 1 shows the percentage of tokens classified as each variant type as a function of Segment and Condition. Notably, of the three segments the difference in proportion of canonical tokens between ID and AD speech was greatest for /t/.

We constructed a series of GLMMs as a means of assessing statistical significance of fixed factors on a response variable of Canonical classifications (vs. Non-canonical, i.e., Assimilated, Deleted, or Glottal). First, we constructed a model with fixed effects of Condition (ID, AD) and Segment (/t/, /d/, /n/) on canonical classification as a binomial categorical dependent variable (Agresti, 2012). Treatment (i.e., dummy) coding was used with Condition = AD and Segment = /t/ set

**Table 1**  
Counts of tokens by Segment (/t/, /d/, /n/) and by Condition (ID, AD) and Word Type (content and function).

Word-final segment	AD		ID		Total
	Content	Function	Content	Function	
/t/	155	383	96	314	948
/d/	124	148	33	50	355
/n/	85	155	32	59	331
Total	364	686	161	423	1634



**Fig. 1.** Percentage of tokens classified as Canonical, Assimilated, Deleted, or Glottal as a function of Condition (ID, AD) and Segment (/t/, /d/, /n/, overall).

as the reference levels. Further, the model included random factors with intercept and slope terms by-subject and an intercept term by-item, which was the maximal random effects structure justified by the design (Barr et al., 2013; Matuschek et al., 2017). The Wald z test was used to assess significance, assuming convergence of z and t distributions. The full model converged with the Boby optimizer (Bates et al., 2015; Powell, 2009).

Modeling showed that the ID condition significantly increased the likelihood of a canonical variant of /t/, relative to the AD condition (Table 2). Further, the odds of being canonical in ID compared to AD for /t/ was about 1.8:1 (cf.  $\exp(0.61)$ ), after controlling for subject- and item-related variation. Moreover, significant effects of Segment for both /d/ and /n/ showed that both segments had different overall likelihoods of a canonical variants, compared with /t/, for the AD Condition. After controlling for random subject- and item-related variation, the odds of being canonical variants in the AD condition was about 3.7 times or 11.5 times higher for /d/ and /n/, respectively, compared to /t/.

Moreover, there was a significant interaction between Condition and Segment for Segment = /d/, implying a different distribution for /d/ of canonicity for ID vs. AD, relative to the distribution for /t/. Notably, being a /d/ decreased the likelihood

of being canonical in ID relative to AD, compared with the effect for /t/; the ratio of [odds of canonical for (ID:AD) for /d/] to [odds of canonical for (ID:AD) for /t/] was about 1–3.7 (cf.  $\exp(-1.32) \cong 0.27$ ). Being an /n/ did not significantly affect the distribution of canonicity in ID compared with AD speech, relative to /t/.

To further assess how Condition (ID, AD) and Segment (/t/, /d/, or /n/) separately affected rates of canonical variants, we ran separate GLMMs on all tokens each of /t/, /d/, or /n/ with a fixed factor of Condition with random factors that included by-subject slopes and intercepts and by-item intercepts. For each GLMM, treatment coding was used with Condition = AD set as the reference level. These separate models are shown in Table 3. Importantly, this modeling confirmed the statistically greater likelihood of canonical variants for ID compared with AD conditions observed earlier. Specifically, the analysis underscores a significant difference for /t/ for the ID condition compared with the AD condition: /t/ was reliably more likely to be canonical in ID speech than AD speech. Further, and also importantly, this modeling showed that /d/ was not reliably less canonical in ID speech. Likewise, the analysis confirmed that there was no significant difference between ID and AD speech on degree of canonicity for /n/.

**Table 2**  
GLMM with fixed factors of Condition and Segment for a canonical dependent variable of Canonical (vs. Non-canonical). Levels of significance are \*\* $p < 0.01$  and \*\*\* $p < 0.001$ .

	$\beta$	SE	Z	p
Intercept	-0.72	0.20	-3.62	<0.001***
Condition = ID	0.61	0.19	3.21	<0.01**
Segment = /d/	1.31	0.30	4.43	<0.001***
Segment = /n/	2.44	0.32	7.62	<0.001***
Condition = ID: Segment = /d/	-1.32	0.45	-2.95	<0.01**
Condition = ID: Segment = /n/	-0.49	0.41	-1.20	0.23

**Table 3**

Separate GLMMs for word-final tokens of /t/, /d/, and /n/ for a categorical dependent variable of Canonical (vs. Non-canonical). The level of significance was \*\*\* $p < 0.001$  or \*\* $p < 0.01$ .

	/t/				/d/				/n/			
	$\beta$	SE	z	p	$\beta$	SE	z	p	$\beta$	SE	z	p
Intercept	-0.79	0.18	-4.36	<0.001***	0.77	0.37	2.11	0.04*	1.46	0.28	5.23	<0.001***
Condition = ID	0.60	0.20	3.03	<0.01**	-1.24	0.70	-1.78	0.07	0.20	0.43	0.47	0.64

Next, we investigated effects of Word Type (function vs. content word) by constructing a GLMM with fixed effects of Condition (ID, AD), Segment (/t/, /d/, /n/), and Word type (Content, Function), plus all interactions, along with random effects consisting of by-subject intercept and slope terms and a by-item intercept term. Treatment coding was used with Condition = AD, Segment = /t/, and Word Type = Function as reference levels. The full model converged with the Boby optimizer (Bates et al., 2015; Powell, 2009). This model was reduced through iterative elimination of non-significant interaction terms, starting with the three-way interaction until a likelihood ratio test revealed that the next simpler model was a significantly worse fit.

The final model is shown in Table 4. As expected, content words showed a significantly higher likelihood of a canonical variant than function words; the odds ratio of a canonical variant for content vs. function words was about 2:1 (cf. exp (-0.80)). Critically, no significant interactions between Word Type and other fixed effects were found, indicating that the word type distinction had statistically equal effects across all other conditions.

Given that levels of Word type were significantly correlated with word frequency – since function words are much more frequent than content words – it was not desirable to construct a model that included both multicollinear terms. Instead, our next models focused on effects of word frequency – a continuous variable that offered more statistical specificity than Word Type – alongside main effects and interactions with Condition or Segment; we then followed this up by examining effects of Word Frequency, Condition, and Segment separately at each level of Word Type. First, we constructed a GLMM which included all words (i.e., collapsing across both levels of Word Type) where our model had fixed factors of  $\log_{10}$  Frequency, Condition, and Segment, with random effects consisting of by-subject slope and intercept terms and a by-item intercept term. Treatment coding was used with Condition = AD and Segment = /t/ set as the reference levels. The model converged with the Boby optimizer (Powell, 2009). The full model was then iteratively reduced through elimination of non-significant interaction terms. Subsequent models lacking an interaction term were compared to the model that included it

until a likelihood ratio test revealed that the simpler model was a significantly worse fit. The final model is shown in Table 5a. Results show that  $\log_{10}$  Frequency was a significant predictor of likelihood of canonicity. Importantly, there were no significant interactions between  $\log_{10}$  Frequency and any other factor – neither Segment nor Condition. This finding reveals that word frequency does not significantly modulate, and thus cannot explain, differences in canonicity across other conditions. Specifically, word frequency cannot account for our finding that /t/ in assimilable contexts is more canonical in ID speech than AD speech.

Next, we constructed separate GLMMs for Content words and Function words. Each model had fixed and random effect factor structures which were identical to the above model. As before, each full model was iteratively reduced through elimination of non-significant interaction terms until the next simpler model resulted in a significantly worse fit. The final models for Content and Function words are shown in Table 5b and c, respectively.

In the Content word model (Table 5b), the ID condition again resulted in a significantly higher proportion of canonical variants for /t/ in assimilable contexts, compared with the AD condition. Further, /d/ and /n/ each significantly raised the likelihood of a canonical variant, relative to /t/ in AD. Condition interacted with Segment, as revealed by the fact that the ID condition significantly raised the likelihood of /t/ being canonical compared with AD but had a significant, yet opposite, effect for both /d/ and /n/. Moreover, higher word frequency resulted in lower likelihoods of a canonical variant equally across all segment types in AD. Frequency did not significantly interact with any other factors.

In the Function word model (Table 5c), the effect of ID on increasing the proportion of /t/ relative to AD was weaker and missed significance ( $p = 0.08$ ). As with Content words, /d/ and /n/ each significantly raise the likelihood of a canonical variant, relative to /t/ in AD. What is notable here is the effect of Word Frequency. Whereas there was no significant effect of frequency on realization of Function words ending in /t/, there was a pronounced effect of Word Frequency on Function words ending in /d/ and /n/. In particular, increasing Word Frequency led to decreased likelihood of a canonical variant for

**Table 4**

Final GLMM for Canonical (vs. Non-canonical) with fixed factors of Condition (AD vs. ID), Segment (/t, d, n/), and Word Type (Function vs. Content). Levels of significance were \*\* $p < 0.01$  and \*\*\* $p < 0.001$ .

	$\beta$	SE	Z	p
Intercept	-0.38	0.21	-1.81	0.07
Condition = ID	0.63	0.19	3.37	<0.001***
Segment = /d/	1.14	0.28	4.04	<0.001***
Segment = /n/	2.41	0.30	7.94	<0.001***
Word Type = Content	-0.80	0.22	-3.67	<0.001***
Condition = ID: Segment = /d/	-1.24	0.43	-2.87	<0.01**
Condition = ID: Segment = /n/	-0.48	0.40	-1.21	0.23

**Table 5**  
Final GLMMs for (a) all words, (b) Content words only, and (c) Function words only. All models included fixed factors of Condition (AD vs. ID), Segment (/t,d,n/), and log<sub>10</sub> Frequency for the categorical dependent variable of Canonical (vs. Non-canonical). Levels of significance were \**p* < 0.05, \*\**p* < 0.01 and \*\*\**p* < 0.001.

	$\beta$	SE	Z	<i>p</i>
<i>(a) All words</i>				
Intercept	-1.09	0.19	-5.60	<0.001***
Condition = ID	0.58	0.11	3.17	<0.01**
Segment = /d/	1.25	0.28	4.56	<0.001***
Segment = /n/	2.39	0.30	7.91	<0.001***
log <sub>10</sub> Frequency	-0.58	0.11	-5.12	<0.001***
Condition = ID: Segment = /d/	-1.11	0.43	-2.59	<0.01**
Condition = ID: Segment = /n/	-0.42	0.40	-1.05	0.29
<i>(b) Content words only</i>				
Intercept	-1.24	0.34	-3.67	<0.001***
Condition = ID	0.87	0.37	2.38	<0.05*
Segment = /d/	1.30	0.36	3.65	<0.001***
Segment = /n/	2.48	0.49	5.05	<0.001***
log <sub>10</sub> Frequency	-0.68	0.21	-3.29	<0.01**
Condition = ID: Segment = /d/	-1.32	0.64	-2.06	<0.05*
Condition = ID: Segment = /n/	-0.94	0.88	-1.08	0.28
<i>(c) Function words only</i>				
Intercept	-1.03	0.22	-4.65	<0.001***
Condition = ID	0.34	0.20	1.74	0.08
Segment = /d/	1.06	0.44	2.44	<0.05*
Segment = /n/	2.41	0.39	6.19	<0.001***
log <sub>10</sub> Frequency	-0.10	0.22	-0.48	0.63
log <sub>10</sub> Frequency: Segment = /d/	-1.29	0.57	-2.26	<0.05*
log <sub>10</sub> Frequency: Segment = /n/	-1.10	0.56	-1.97	<0.05*

either /d/ or /n/. Collectively, the models in Tables 5b and 5c further shed light on the distinctiveness of /t/, as compared with /d/ and /n/, for Content as opposed to Function words.

In addition, given the finding of Dilley et al. (2014) that infants' gender and age interacted with canonicity in storybook speech, we investigated whether mothers tuned their pronunciations to the infant's age and/or gender specifically when talking to their infant (i.e., in the ID condition). Note that we focused solely on tokens in the ID condition, since we had no reason to believe that the gender or age of an infant would influence how a mother produced her speech to other adults (i.e., in the AD condition). Therefore, for ID condition tokens only, we constructed a mixed-effect logistic regression model with fixed factors of Segment, Infant Age, and Infant Gender and random effects consisting of by-subject and by-item intercepts. (Note that no by-subject random slope was included in the model, due to the fact that this analysis did not include a within-subject fixed factor.) The reference level for Infant Gender was female. The model converged with a Boby optimizer (Powell, 2009); the full model was iteratively reduced by eliminating non-significant interaction terms, starting with the three-way interaction until the simpler model resulted in a significantly worse fit. The final model is shown in Table 6. No interaction terms survived model reduction. There was no significant effect of either Infant Gender or Infant Age. Regarding segment-specific effects, modeling of ID speech tokens showed that, relative to /t/, /n/ was significantly more likely to be canonical, with odds of approximately 6:1.

Finally, we considered whether Word Frequency might moderate the effects of Segment, Infant Age, and/or Infant Gender. We first constructed a GLMM with fixed factors of Segment, Infant Age, Infant Gender, and Word Frequency and all interaction terms, along with random effects consisting of by-subject and by-item intercepts. The reference level for Infant Gender was female. The model converged with a Boby optimizer (Powell, 2009). The full model was iteratively

reduced by successively eliminating the one non-significant interaction term that contributed the least to model fit. The final model is shown in Table 7. No interaction terms survived model reduction, suggesting that Word Frequency does not significantly moderate the effects of any other factor examined.

Lastly, we conducted multinomial logistic regression using *MCMCglmm* (Hadfield, 2010b) to determine how the independent variables of Segment (/t,d,n/) and Condition (ID, AD) affected distributions across all four variants (Canonical, Assimilated, Glottal, Deleted). Treatment coding was used with Condition = AD and Segment = /t/ set as the reference levels. For this analysis, the reference group was set to Canonical, while /t/ and AD were baseline conditions for fixed factors of Segment and Condition, respectively. Random effects were also included consisting of intercepts for by-item effects and intercepts and slopes for by-subject effects.

Results of the analysis are shown in Table 8. After statistically controlling for random effects, Glottal variants had comparable predicted frequency compared with the reference category (i.e., Canonical) at baseline (/t/ for the AD condition), as shown by the fact that for the Glottal condition the intercept log-odds estimate did not differ significantly from zero. Moreover, both Assimilated and Deleted variants were predicted to be significantly less common than Canonical (referent category) at baseline (/t/ for the AD condition), as shown by the significant negative mean logit values. Importantly, the analysis showed that for /t/, being in the ID condition significantly *decreased* the likelihood of Glottal variants relative to Canonical variants, compared with the reference likelihood (rate of Glottal-to-Canonical variants in the AD condition). By contrast, the relative likelihoods of Assimilated variants or Deleted variants for /t/, compared with Canonical variants, were not significantly different for the ID condition, compared with the AD condition. These findings suggest that the increase in canonicity of /t/ in ID compared with AD comes at the expense of statistically "trading off" with Glottal variants in ID: Canonical



**Table 6**

For ID speech condition tokens only, final GLMM with fixed factors of Segment, Infant Age, and Infant Gender for a categorical dependent variable of Canonical (vs. Non-canonical). The level of significance was \*\*\* $p < 0.001$ .

	$\beta$	SE	z	p
Intercept	−0.07	0.25	−0.30	0.77
Infant Age	0.08	0.11	0.77	0.44
Infant Gender = Male	−0.24	0.23	−1.02	0.31
Segment = /d/	−0.09	0.45	−0.19	0.85
Segment = /n/	1.86	0.47	3.93	<0.001***

**Table 7**

For ID speech condition tokens only, final GLMM with fixed factors of Segment, Infant Age, Infant Gender, and Frequency for a categorical dependent variable of Canonical (vs. Non-canonical). The level of significance was \*\*\* $p < 0.001$ .

	$\beta$	SE	Z	p
Intercept	−0.35	0.24	−1.47	0.14
Infant Age	0.11	0.11	1.01	0.31
Infant Gender = Male	−0.24	0.23	−1.04	0.30
Segment = /d/	0.10	0.41	0.24	0.81
Segment = /n/	1.95	0.44	4.33	<0.001***
log <sub>10</sub> Frequency	−0.55	0.17	−3.29	<0.001***

variants become more prevalent in ID, but Glottal variants become less prevalent in ID, compared with their corresponding rates in AD speech.

The type of word-final segment (/d/ or /n/) significantly altered relative rates of some variant types, but not others, relative to /t/. For /d/, there was a significant change (i.e., reduction) in the relative rate of Glottal variants in AD speech, compared with the reference rate (i.e., rate of Glottal to Canonical variants for /t/ in AD speech), but not in relative rates of any other variant type. For /n/, the relative rates of both Deleted and Glottal variants in AD were significantly lower than the reference condition rates (i.e., the rates of Deleted and Glottal variants, respectively, to Canonical variants for /t/ in AD speech), while the relative rate of Assimilated variants for /n/

was statistically comparable to that for /t/. Finally, there were no significant interactions between Segment type (/d/ or /n/) and Condition on relative rates of any variant type (Assimilated, Deleted, or Glottal), compared with rates in reference conditions. That is, being in ID condition did not significantly alter variant rate distributions compared with AD condition reference rates of variants for either /d/ or /n/. Since there was no significant interaction between ID condition and segment (/d/ or /n/) for any variant type, this implies that rates of Assimilated, Deleted, and Glottal variants for /d/ and /n/ in ID were comparable to those in AD. This analysis further shows that /t/ alone was associated with significantly higher rates of canonical variants in ID relative to AD speech, where this increased rate of canonical variants for /t/ in ID was associated with a decreased rate of glottal variants in ID relative to AD, further underscoring that /t/ is exceptional.

Given that glottal variants are more common in contexts of a following sonorant phoneme (Huffman, 2005; Pierrehumbert, 1994), we also checked whether the distribution of sonorant (vs. obstruent) following sounds could account for the significantly greater frequency of glottal variants in AD compared with ID speech. To do so, we ran an additional mixed effect multinomial logistic regression using *MCMCglmm* on the response variable of variant type distribution with fixed factors of Condition (ID vs. AD) and Sonority (Obstruent vs. Sonorant) and random effects with by-subjects random intercepts and slopes and by-items intercepts. Treatment coding was used with Condition = AD and Sonority = Obstruent set as the reference levels. Not surprisingly, the analysis (shown in Table 9) confirmed that glottal variants were significantly less likely in ID speech than AD speech, and that tokens of /t/ in AD speech which were followed by sonorants were significantly more likely to be realized as glottal variants than tokens followed by obstruents ( $p < 0.05$ ). Most importantly, there was no significant interaction between Sonority and Condition on the proportion of glottal variants ( $p = 0.28$ ). This finding indicates that there is no support for the notion that a different

**Table 8**

GLMM for the multinomial distribution of Canonical, Assimilated, Deleted, or Glottal variants; Canonical was the reference level. Fixed factors are Condition (AD vs. ID) and Segment (/t/, /d/, /n/), with AD and /t/ set as the baseline levels of Condition and Segment. Type I error rates are pMCMC values (Baayen, 2008) with \*\*\* $p < 0.001$ ; see Methods.

Comparison to baseline (Canonical for /t/ in AD condition)	Variant	Posterior mean (log-odds)	95% CI (log-odds)	p
Intercept	Assimilated	−1.87	(−2.39, −1.32)	<0.001***
Intercept	Deleted	−2.15	(−3.37, −1.16)	<0.001***
Intercept	Glottal	−0.05	(−0.55, 0.45)	0.85
Condition = ID	Assimilated	0.43	(−0.09, 1.02)	0.11
Condition = ID	Deleted	0.38	(−0.22, 1.02)	0.23
Condition = ID	Glottal	−1.56	(−2.03, −1.11)	<0.001***
Segment = /d/	Assimilated	−0.03	(−0.83, 0.70)	0.92
Segment = /d/	Deleted	−0.03	(−1.27, 1.46)	0.98
Segment = /d/	Glottal	−4.07	(−5.18, −2.97)	<0.001***
Segment = /n/	Assimilated	−0.10	(−0.90, 0.62)	0.81
Segment = /n/	Deleted	−5.66	(−8.39, −3.20)	<0.001***
Segment = /n/	Glottal	−6.24	(−8.32, −4.32)	<0.001***
Condition = ID:Segment = /d/	Assimilated	0.00	(−1.26, 1.09)	0.98
Condition = ID:Segment = /d/	Deleted	0.45	(−0.82, 1.89)	0.52
Condition = ID:Segment = /d/	Glottal	1.05	(−1.84, 3.50)	0.43
Condition = ID:Segment = /n/	Assimilated	−0.77	(−1.70, 0.34)	0.14
Condition = ID:Segment = /n/	Deleted	0.51	(−1.60, 2.76)	0.62
Condition = ID:Segment = /n/	Glottal	2.53	(−0.59, 5.61)	0.11

**Table 9**  
GLMM for the multinomial distribution of Canonical, Assimilated, Deleted, or Glottal variants; Canonical was the reference level of the response variable. Fixed factors are Condition (AD vs. ID) and Sonority (Obstruent, Sonorant), with AD and Obstruent set as the baseline levels of Condition and Sonority. Type I error rates are pMCMC values (Baayen, 2008) with \* $p < 0.05$ , \*\* $p < 0.01$  and \*\*\* $p < 0.001$ ; see Methods.

Comparison to baseline (Canonical for /t/ in AD condition)	Variant	Posterior mean (log-odds)	95% CI (log-odds)	$p$
Intercept	Assimilated	-2.28	(-3.15, -1.46)	<0.001***
Intercept	Deleted	-1.56	(-2.38, -0.58)	<0.001***
Intercept	Glottal	-0.48	(-1.04, 0.14)	0.12
Condition = ID	Assimilated	0.61	(-0.20, 1.33)	0.13
Condition = ID	Deleted	-0.36	(-1.15, 0.38)	0.37
Condition = ID	Glottal	-1.71	(-2.32, -1.02)	<0.001***
Sonority = Sonorant	Assimilated	-0.06	(-0.90, 0.83)	0.93
Sonority = Sonorant	Deleted	-0.40	(-1.14, 0.34)	0.31
Sonority = Sonorant	Glottal	0.63	(0.02, 1.18)	<0.05*
Condition = ID:Sonority = Sonorant	Assimilated	-0.70	(-1.96, 0.51)	0.25
Condition = ID:Sonority = Sonorant	Deleted	1.71	(0.66, 2.73)	<0.01**
Condition = ID:Sonority = Sonorant	Glottal	0.52	(-0.40, 1.42)	0.28

distribution of sonority across ID and AD conditions could be responsible for the observed distinct distribution of glottal variants in ID compared with AD.

#### 4. Discussion

This study sought to clarify statistical patterns of consonant pronunciations in theoretically significant phonological contexts – i.e., those where assimilation may occur – in a large corpus of American English spontaneous speech directed to infants or adults, separately examining rates of /t/, /d/, and /n/ phonetic variants. Application of mixed effect statistical models allowed robust statistical examination of the role of speech register (ID vs. AD) on variant usage while controlling for clustered variation by items and speakers, including different numbers of tokens per talker and word. These data revealed a clear statistical predominance of canonical /t/ in assimilable contexts relative to other variant forms in ID speech, a pattern which is distinct from /t/ in AD speech. This pattern was found in contexts where assimilation of place of articulation was possible (i.e., following labial and velar contexts). This relatively greater statistical prevalence of canonical variant pronunciations for /t/ in ID speech compared with AD speech was not observed for either /d/ or /n/.

Further, mixed effects multinomial logistic regression modeling revealed that the increased rate of canonical variants for /t/ in ID compared with AD involved a statistical tradeoff with one other specific variant type: glottal variants. While the rate of canonical variants for /t/ increased in ID, compared with AD speech, the rate of glottal variants for /t/ decreased in ID, compared with AD speech. By contrast, deleted and assimilated variants did not differ in relative frequency for /t/ in ID compared with AD speech. Importantly, additional multinomial logistic regression modeling revealed that these different distributions of glottal variants were not due to different distributions of sonorant contexts in ID compared with AD.

These findings show the novel result that early in life, children hear speech directed to them which is characterized by a predominance of canonical /t/ consonant variants, at least in assimilable contexts. These data lend important new findings about the distributional properties of early consonants, in keeping with a broad range of evidence that early exposure to the statistical properties of phonetic information shapes

speech perception (Maye et al., 2002; Werker & Curtin, 2005; Werker et al., 2007). A variety of evidence has shown consistent phonetic modifications in ID speech relative to AD speech, for example, expansion of the vowel space in ID compared with AD speech (Hartman et al., 2017; Kuhl et al., 1997; Wieland et al., 2015). Such vowel variation cannot be uniformly understood as hyperarticulation (Burnham et al., 2015; Cristia & Seidl, 2013; Kondaurava, Bergeson, & Dilley, 2012; Martin et al., 2015; McMurray et al., 2013; Miyazawa et al., 2017). The picture of phonetic variation in consonants has been even more complicated (Baran et al., 1977; Bernstein Ratner, 1984; Fish et al., 2017; Lahey & Ernestus, 2013; Malsheen, 1980; McMurray et al., 2013; Shockey & Bond, 1980; Sundberg & Lacerda, 1999; Sundberg, 2001), with an emerging consensus that characterizing consonant variability in ID speech necessarily requires careful control of phonological variables, along with consideration of factors such as age and gender of the child.

Why might this tradeoff in relative frequencies of canonical and glottal variants for ID vs. AD speech specifically occur for /t/? Of the phonemes examined, /t/ alone is phonologically voiceless. Further, among voiceless obstruent stops in English – /p/, /t/, and /k/ – both /t/ and /p/ can show glottal voicing (Pierrehumbert, 1994; Foulkes et al., 2005; Huffman, 2005; Seyfarth & Garellek, 2015), with /t/ evidencing particularly high glottal variant rates (Dilley & Pitt, 2007; Huffman, 2005; Pitt et al., 2011). Note that the distribution of glottal and other variant types in our AD data was very similar to that found in Dilley and Pitt (2007). Below, we consider evidence that these frequent glottal variant types lead to differential robustness vs. confusability in word perception, relative to other /t/ variants. We then put forward a proposal regarding how this influences mothers' variant productions in ID speech.

Sumner and Samuel (2005) examined how recognition of words ending in final /t/ was affected by glottal vs. canonical /t/ variants. Using semantic priming, they showed that canonical, assimilated, and glottal variants were equally good at activating a related word. However, in a task of recognition of newly vs. previously presented words, there was a strong advantage for canonical /t/ on forming accurate lexical memories, showing that canonical /t/ created considerably stronger activation of abstract representations over glottal /t/ for variants.

Garellek and colleagues (Chong & Garellek, 2018; Garellek, 2015) have investigated how word-final glottalized voicing helps or hinders word recognition. Chong and Garellek (2018) found that glottalization did not hinder recognition for /t/ and /p/, but it did hinder recognition for the voiced stops /d/ and /b/. Their findings indicated that use of glottalization as a cue to phoneme identity is complex and requires consideration of phonological voicing for correct recognition. Further, Garellek (2015) examined whether listeners could distinguish /t/ glottalization in word pairs like *button* vs. *bun* and *dent* vs. *den*, from creaky voicing due to prosodic structure, since non-modal phonation is also common at intonation phrase endings (Huffman, 2005; Redi & Shattuck-Hufnagel, 2001). Listeners could tell apart phrase-level creak from /t/ glottalization but were generally poorer and less confident at identifying words with /t/-glottalization when they occurred with phrase-final creak than when no creak was present. Listeners showed particularly poor accuracy and low confidence for recognizing words which potentially ended in a final /t/, like *dent* vs. *den*, when in the presence of phrase-final creaky voice. These findings indicated high perceptual confusability of acoustic cues that signal a glottal variant /t/ vs. those that signal phrase endings.

Given these findings, we propose that maternal caregivers' greater usage of canonical /t/ (and decreased usage of glottal /t/) in ID speech compared with AD speech involves unconscious structuring of input in a manner that aids infants to generate correct inferences about phonemic categories, i.e., /t/ (Eaves et al., 2016). Consistent with this, most maternal caregivers are expected to be literate and to have explicit knowledge of phoneme categories (Bowers et al., 2016; Dijkstra, Roelofs, & Fiews, 1995; Eaves et al., 2016; Kazanina, Bowers, & Idsardi, 2018; Ohala & Shattuck-Hufnagel, 1986). Further, regardless of their literacy status, adult caregivers have implicit knowledge developed over their lifetimes regarding how pronunciation factors may enhance or detract from lexical activation. We propose that this lifetime of linguistic knowledge – including explicit knowledge about /t/ as a phonemic category, together with implicit knowledge about /t/ pronunciations – drives the different distributions of canonical /t/ produced by mothers in ID speech. Regarding implicit knowledge, for familiar dialects mothers are expected to have internalized representations of statistical distributions of variant types for /t/, /d/, and /n/ – as well as knowledge that these distributions across various phonological contexts in AD are not the same. Mothers are also expected to have implicit knowledge that canonical /t/ enhances lexical recognition (Sumner & Samuel, 2005) and that glottal variants of stops – which mothers would implicitly “know” to be common for /t/ but uncommon for /d/ or /n/ – are particularly confusable with acoustic markers of phrase-finality due to the prevalence of creaky voice in that prosodic position (Garellek, 2015).

Our proposal builds on a recent formal mathematical theoretic approach to teaching and learning in ID speech advanced by Eaves et al. (2016). They proposed, following the social learning literature, that IDS is an *ostensive* cue, i.e., a social cue that engages stricter learning mechanisms in its target (Gergely, Egyed, & Király, 2007). They further define teaching data as data generated with the learner in mind (Shafto, Goodman, & Griffiths, 2014) and their computational

simulation implemented a model of human teaching that has accurately captured human learning in a range of scenarios. Eaves et al.'s account shows plausibly how ID speech can reflect the top-down, knowledge-based linguistic structuring of a maternal “teacher” while displaying certain phonetic properties that have been argued to *not* be consistent with teaching, including non-corner vowels being hypoarticulated (Cristia & Seidl, 2013; Kirchoff & Schimmel, 2005) and within-phoneme variability increasing for some vowels (de Boer & Kuhl, 2003; McMurray et al., 2013). Frequent lexical contexts could also bootstrap learning that different variants correspond to the same phonetic category (cf. Swingley & Alarcon, 2018). Consistent with this approach, the greater proportion of canonical tokens in ID compared with AD for /t/ is proposed to reflect an unconscious tendency of mothers when speaking to their infants to structure infants' input by providing acoustically clearer (i.e., less potentially confusable) tokens of /t/. On this view, /d/ or /n/ do not show this distinct pattern in ID, because unlike /t/, they are not particularly prone to lexical and perceptual confusions.

Different kinds of ID modifications might reflect different core causal factors. Not all modifications need necessarily reflect implicit knowledge e.g., of perceptually or attentionally salient phonetic forms (cf. Adriaans & Swingley, 2017). Rather, some modifications may stem principally from evolutionary adaptations or social conventions for conveying emotional connection and/or appear non-threatening (Benders, 2013; Englund & Behne, 2005; Kalashnikova, Carignan, & Burnham, 2017). In spite of this, it is unclear how certain phonetic factors reported in ID – such as enhanced breathiness recently reported for ID in Japanese – could account for these results (Miyazawa et al., 2017).

The present data might be relatable to a long-standing puzzle about speech perception regarding the exceptionality of canonical /t/ with respect to processing of variant pronunciations. Multiple studies have found in general that the most frequently-occurring variant pronunciation in corpora is also the variant type that shows the most robust lexical representations (Connine, 2004; Dilley & Pitt, 2007; Pitt et al., 2011). The case of canonical /t/ pronunciations has been a puzzling exception to this pattern; while prior corpus studies have found it to occur less frequently than other variants (Connine, 2004; Dilley & Pitt, 2007; Pitt et al., 2011), a number of studies have shown it to have more robust lexical representations (Ernestus & Baayen, 2007; McLennan, Luce & Charles-Luce, 2003, 2005; Pitt et al., 2011; Tucker & Warner, 2007). Learning to read cannot fully account for the exceptional processing of canonical /t/ (Dijkstra et al., 1995; Mitterer & Reinisch, 2015; Seidenberg & Tanenhaus, 1979), suggesting the value of searching for other explanations. The present data support that canonical /t/ may be frequent – and indeed a predominant form – in early language input, at least for some phonological contexts. This predominance – shown here for assimilable environments – is consistent with the idea that canonical /t/ is more frequent in early language input than in adult conversational forms of the language. As such, these data provide some measure of support for the speculation that the apparent exceptionality of canonical /t/ in adult language processing – whereby the less frequent canonical variant form is better processed than other forms – might potentially reflect vestiges of

the distributional statistics of early language input, which our data indicate may show a relative preponderance of clear phonemic exemplars. It should be cautioned that we have only demonstrated a preponderance of canonical /t/ for a single phonological environment – assimilable contexts – so it is still not clear whether other phonological environments where /t/ occurs will show a preponderance of canonical /t/ in early input. Further, any findings further showing proportionately greater incidence of canonical /t/ in early language input would merely constitute correlational evidence supportive of our hypothesis, and thus cannot inform our understanding of causal factors the way that careful, controlled experiments can. It is thus too early to say whether listener speech perception behavior for canonical /t/ later in life may reflect a predominance of canonical /t/ forms early in life. However, the present evidence provides one small measure of evidence in support of this hypothesis, which may be fruitfully investigated in future research.

These findings also inform our understanding of how listeners recover the intended words in contexts where assimilation is possible, a topic of much theorizing and empirical study (Dilley & Pitt, 2007; Ellis & Hardcastle, 2002; Gaskell & Marslen-Wilson, 1998; Lahiri & Marslen-Wilson, 1991). The present data from spontaneous ID speech confirm earlier findings from read ID storybook speech (Dilley et al., 2014) that in contexts where assimilation is possible, more variant types are observed than just a simple alternation of assimilated and canonical (Gaskell & Marslen-Wilson, 1998; Gow, 2003). The present data suggest that early speech listening experience involves exposure patterns in which phonetic characteristics of spoken words bear more faithful correspondence between segments that are present in their lexical representations, at least for words ending in /t/ in these contexts, lending insight into the nature of development of the early lexicon.

The present study of variant consonant pronunciations in assimilable contexts differed in critical ways from two earlier studies of this topic (Buckler et al., 2018; Dilley et al., 2014). As mentioned earlier, our study was distinguished by using spontaneous speech, clearly separable ID and AD conditions, and all typically-developing children. The present study also used a larger corpus than Buckler et al. (2018). However, the most important difference for the present study compared with earlier work was that neither Dilley et al. (2014) nor Buckler et al. (2018) examined the influence of the specific word-final phoneme (/t/, /d/, or /n/). A separate analysis of each segment turned out to be critical for characterizing pronunciation variation in assimilable contexts, revealing a very different picture of variation for /t/ compared with either /d/ or /n/. By extension, the usage of read speech in Dilley et al. (2014) is also likely part of the reason why that study, but not Buckler et al. (2018), showed more canonical variants of /t/ in ID compared with AD.

Moreover, we found effects of the content vs. function word distinction and word frequency on variant usage, consistent with other work (see e.g., Ernestus & Warner, 2011), but these did not interact with speech style (ID vs. AD) or any other factor examined (e.g., the distribution of sonorant vs. obstruent contexts). Finally, there were no effects of gender or age of the child on pronunciation variation in our data, in contrast to at least one other study which showed different distributions of

the glottal variant in speaking to male vs. female infants in the Tyneside dialect of British English (Foulkes et al., 2005). We note that the present study focused on factors affecting pronunciation in the Midwestern dialect of American English for talkers in Indiana, which is considered a standard, mainstream variety. It is expected that the frequency of variant usage will differ as a function of the language and dialect being spoken.

In summary, we have shown for the first time clear evidence of a relative statistical predominance of canonical /t/ in speech directed to infants, in contrast to speech directed to adults, specifically for assimilable contexts. These results show that ID speech can be considered clearer than AD speech – at least for word-final /t/ in contexts where assimilation may occur, which are considered a theoretically significant subset of all /t/ contexts. These findings raise the possibility, still speculative, that the exceptional patterns of canonical /t/ variants with respect to speech processing may partly reflect a relatively more frequent occurrence of canonical /t/ in early language input as compared to later language input. To further test this hypothesis, it will be necessary to examine distributions of phonetic variants, especially of /t/, in a wider variety of phonological contexts in ID compared with AD speech, and ideally conduct controlled experiments. In any case, the present findings shed light on distributional characteristics of early speech to infants, particularly for consonants, showing distinct distributions of /t/ variants in ID compared with AD register. Future research may address whether the particulars of these consonant variant distributions in early exposure may contribute to children's development of robust, adult-like lexical perception abilities.

### Acknowledgments

This research is supported by NIH-NIDCD grant 5R01DC008581-08 to D.M. Houston and L. Dilley and NIH-NIDCD grant R01DC008581 to T. Bergeson. Thanks to Mark Pitt for valuable feedback. We thank Claire Carpenter, Dan Chabala, Evamarie Cropsey, Erin Dixon, Dana Flowerday, Devin McAuley, Amanda Millett, Shaina Selbig, Zach Zells for their assistance with data analysis, organization, and/or coding. We thank Madeleine McAuley, Leif McAuley, Jason Cristini, and Dominic Cristini for supporting of the research effort.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.wocn.2019.05.004>.

### References

- Adriaans, F., & Swingle, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *Journal of the Acoustical Society of America*, 141(5), 3070–3080. <https://doi.org/10.1121/1.4982246>.
- Agresti, A. (2012). *Categorical data analysis*. Hoboken, NJ: John Wiley & Sons Inc.
- Andruski, J. E., Kuhl, P. K., & Hayashi, A. (1999). The acoustics of vowels in Japanese women's speech to infants and adults. *Paper presented at the 14th International Congress of Phonetic Sciences, San Francisco*.
- Baayen, H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- 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.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>.

- Barth, D., & Kapatsinski, V. (2018). Evaluating logistic mixed-effects models of corpus-linguistic data in light of lexical diffusion. In D. Speelman, C. Heylen, & D. Geeraerts (Eds.), *Mixed-effect regression models in linguistics* (pp. 99–116). Berlin: Springer.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior and Development*, 36(4), 847–862.
- Bergeson, T. R., Miller, R. J., & McCune, K. (2006). Mothers' speech to hearing-impaired infants and children with cochlear implants. *Infancy*, 10(3), 221–240. [https://doi.org/10.1207/s15327078in1003\\_2](https://doi.org/10.1207/s15327078in1003_2).
- Bergeson, T. R., & Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science*, 13(1), 72–75. <https://doi.org/10.1111/1467-9280.00413>.
- Bernstein Ratner, N. (1984). Phonological rule usage in mother-child speech. *Journal of Phonetics*, 12, 245–254.
- Bolstad, W., & Curran, J. (2016). *Introduction to Bayesian statistics* (3rd ed.). Hoboken, New Jersey: Wiley.
- Bowers, J. S., Kazanina, N., & Andermane, N. (2016). Spoken word identification involves accessing position invariant phoneme representations. *Journal of Memory and Language*, 87, 71–83.
- Breen, M., Dilley, L. C., Kraemer, J., & Gibson, E. (2012). Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory*, 8(2), 277–312. <https://doi.org/10.1515/cllt-2012-0011>.
- Buckler, H., Goy, H., & Johnson, E. K. (2018). What infant-directed speech tells us about the development of compensation for assimilation. *Journal of Phonetics*, 66, 45–62.
- Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science*, 296(5572), 1435.
- Burnham, E., Wieland, E., Kondaurova, M. V., McAuley, J. D., Bergeson, T. R., & Dilley, L. C. (2015). Phonetic modification of vowel space in storybook speech to infants up to 2 years of age. *Journal of Speech, Language, and Hearing Research*, 58(2), 241–253.
- Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, 89, 68–86.
- Cho, T., & McQueen, J. M. (2008). Not all sounds in assimilation environments are perceived equally: Evidence from Korean. *Journal of Phonetics*, 36(2), 239–246.
- Chong, A. J., & Garellek, M. (2018). Online perception of glottalized coda stops in American English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 9(1), 4.
- Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review*, 11(6), 1084–1089.
- Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics*, 70(3), 403–411.
- Cooper, R. P., & Aslin, R. N. (1990). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development*, 65, 1663–1677.
- Cooper, F. S., Delattre, P., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *The Journal of the Acoustical Society of America*, 24, 597. <https://doi.org/10.1121/1.1906940>.
- Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2017). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*. <https://doi.org/10.1111/cdev.12974>.
- Cristia, A., & Seidl, A. (2013). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, 41(4). <https://doi.org/10.1017/S0305000912000669>.
- Davies, M. (2017). NOW Corpus (News on the Web) Retrieved from <https://corpus.byu.edu/now/>.
- de Boer, B., & Kuhl, P. K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online*, 4(4), 129–134. <https://doi.org/10.1121/1.1613311>.
- Deelman, T., & Connine, C. M. (2001). Missing information in spoken word recognition: Non-released stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 656–663.
- Dijkstra, T., Roelofs, A., & Fieus, S. (1995). Orthographic effects on phoneme monitoring. *Canadian Journal of Experimental Psychology*, 49(2), 264–271.
- Dilley, L. C., Millett, A., McAuley, J. D., & Bergeson, T. R. (2014). Phonetic variation in consonants in infant-directed and adult-directed speech: The case of regressive place assimilation in word-final alveolar stops. *Journal of Child Language*, 41(1), 155–175. <https://doi.org/10.1017/S0305000912000670>.
- Dilley, L. C., & Pitt, M. A. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *Journal of the Acoustical Society of America*, 122(4), 2340–2353. <https://doi.org/10.1121/1.2772226>.
- Eaves, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review*, 123(6), 758–771.
- Ellis, L., & Hardcastle, W. J. (2002). Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics*, 30, 373–396. <https://doi.org/10.1006/jpho.2001.0162>.
- Englund, K. (2005). Voice onset time in infant directed speech over the first six months. *First Language*, 25(2), 220–234. <https://doi.org/10.1177/0142723705050286>.
- Englund, K., & Behne, D. M. (2005). Infant directed speech in natural interaction – Norwegian vowel quantity and quality. *Journal of Psycholinguistic Research*, 34(3), 259–280.
- Englund, K., & Behne, D. M. (2006). Changes in infant directed speech in the first six months. *Infant and Child Development*, 15, 139–160. <https://doi.org/10.1002/icd.445>.
- Ernestus, M., & Baayen, H. (2007). The comprehension of acoustically reduced morphologically complex words: The roles of deletion, duration, and frequency of occurrence. *Paper presented at the Proceedings of the International Congress of Phonetic Sciences, Saarbrücken*.
- Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39(3), 253–260.
- Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papoušek, U. Jürgens, & M. Papoušek (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 262–282). Cambridge: Cambridge University Press.
- Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20(1), 104–113. <https://doi.org/10.1037/0012-1649.20.1.104>.
- Fish, M. S., Garcia-Sierra, A., Ramirez-Esparza, N., & Kuhl, P. K. (2017). Infant-directed speech in English and Spanish: Assessments of monolingual and bilingual caregiver VOT. *Journal of Phonetics*, 63(19–34).
- Fong, Y., Rue, H., & Wakefield, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics*, 11(3), 397–412.
- Foulkes, P., Docherty, G., & Watt, D. (2005). Phonological variation in child-directed speech. *Language: Journal of the Linguistic Society of America*, 81(1), 177–206.
- Garellek, M. (2015). Perception of glottalization and phrase-final creak. *Journal of the Acoustical Society of America*, 137(2), 822–831. <https://doi.org/10.1121/1.4906155>.
- Gaskell, G. M., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 380–396. <https://doi.org/10.1037/0096-1523.24.2.380>.
- Gergely, G., Egyed, K., & Király, I. (2007). On pedagogy. *Developmental Science*, 10(1), 139–146.
- Gomez, R. (2007). Statistical learning in infant language development. In G. M. Gaskell (Ed.), *The Oxford handbook of psycholinguistics*. New York, NY: Oxford University Press.
- Gow, D. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, 45, 133–159. <https://doi.org/10.1006/jmla.2000.2764>.
- Gow, D. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 163–179. <https://doi.org/10.1037/0096-1523.28.1.163>.
- Gow, D. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception and Psychophysics*, 65(4), 575–590. <https://doi.org/10.3758/BF03194584>.
- Hadfield, J. (2010a). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33(2), 1–22.
- Hadfield, J. (2010b). MCMCglmm: Markov chain Monte Carlo methods for Generalised Linear Mixed Models. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.192.1018&rank=1> on 5/11/19.
- Hartman, K. M., Ratner, N. B., & Newman, R. S. (2017). Infant-directed speech (IDS) vowel clarity and child language outcomes. *Journal of Child Language*, 44(5), 1140–1162. <https://doi.org/10.1017/S0305000916000520>.
- Holst, T., & Nolan, F. (1995). The influence of syntactic structure on [s] to [ʃ] assimilation. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence: Papers in laboratory phonology IV* (pp. 315–333). Cambridge: Cambridge University Press.
- Huffman, M. K. (2005). Segmental and prosodic effects on coda glottalization. *Journal of Phonetics*, 33, 335–362.
- Jaeger, F. T. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446. <https://doi.org/10.1016/j.jml.2007.11.007>.
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34, 485–499. <https://doi.org/10.1016/j.jwocn.2005.08.004>.
- Kalashnikova, M., Carignan, C., & Burnham, D. (2017). The origins of babytalk: Smiling, teaching or social convergence? *Royal Society open science*, 4(8) 170306.
- Kazanina, N., Bowers, J. S., & Idsardi, W. (2018). Phonemes: Lexical access and beyond. *Psychonomic Bulletin and Review*, 25(2), 560–585. <https://doi.org/10.3758/s13423-017-1362-0>.
- Kirchhoff, K., & Schimmel, S. (2005). Statistical properties of infant-directed versus adult-directed speech: Insights from speech recognition. *The Journal of Acoustical Society of America*, 117(4 Pt 1), 2238–2246.
- Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, 25(4), 463–492.
- Kleinschmidt, D. F., & Jaeger, F. T. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203. <https://doi.org/10.1037/a0038695>.
- Kondaurova, M. V., Bergeson, T. R., & Dilley, L. C. (2012). Effects of deafness on acoustic characteristics of American English tense/lax vowels in maternal speech to infants. *Journal of the Acoustical Society of America*, 132(2), 1039–1049. <https://doi.org/10.1121/1.4728169>.
- Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93–107.
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., ... Lacerda, F. (1997). Cross-language analysis of phonetic units in

- language addressed to infants. *Science*, 277, 684–686. <https://doi.org/10.1126/science.277.5326.684>.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the World's Languages*. Languages: Wiley-Blackwell.
- Lahey, M., & Ernestus, M. (2013). Pronunciation variation in infant-directed speech: Phonetic reduction of two highly frequent words. *Language Learning and Development*, 10(4), 308–327. <https://doi.org/10.1080/15475441.2013.860813>.
- Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245–294. [https://doi.org/10.1016/0010-0277\(91\)90008-R](https://doi.org/10.1016/0010-0277(91)90008-R).
- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. <https://doi.org/10.2307/2529310>.
- Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, 6(3), F1–F10. <https://doi.org/10.1111/1467-7687.00275>.
- Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: A longitudinal study. *Journal of Child Language*, 36, 909–922.
- Malsheen, B. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 2: Perception, pp. 173–184). New York: Academic Press.
- Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America*, 125(6), 3962–3973.
- Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis. *Psychological Science*, 26(3), 341–347.
- Matuschek, H., Kiegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>.
- Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101–B111. [https://doi.org/10.1016/S0010-0277\(01\)00157-3](https://doi.org/10.1016/S0010-0277(01)00157-3).
- McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29, 539–553.
- McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2005). Representation of lexical form: Evidence from studies of sublexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 1308–1314.
- McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition*, 129, 362–378.
- McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30(6), 1113–1126. [https://doi.org/10.1207/s15516709cog0000\\_79](https://doi.org/10.1207/s15516709cog0000_79).
- Mitterer, H., Chen, Y., & Zhou, X. (2011). Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science*, 35(1), 184–197. <https://doi.org/10.1111/j.1551-6709.2010.01140.x>.
- Mitterer, H., Cho, T., & Kim, S. (2016). How does prosody influence speech categorization? *Journal of Phonetics*, 54, 68–79. <https://doi.org/10.1016/j.jwocn.2015.09.002>.
- Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, 34, 73–103.
- Mitterer, H., Kim, S., & Cho, T. (2013). Compensation for complete assimilation in speech perception: The case of Korean labial-to-velar assimilation. *Journal of Memory and Language*, 69, 59–83.
- Mitterer, H., & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 244–263.
- Mitterer, H., & Reinisch, E. (2015). Letters don't matter: No effect of orthography on the perception of conversational speech. *Journal of Memory and Language*, 85, 116–134. <https://doi.org/10.1016/j.jml.2015.08.005>.
- Mitterer, H., Reinisch, E., & McQueen, J. M. (2018). Allophones, not phonemes in spoken-word recognition. *Journal of Memory and Language*, 98, 77–92. <https://doi.org/10.1016/j.jml.2017.09.005>.
- Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, 129(2), 356–361. <https://doi.org/10.1016/j.cognition.2013.07.011>.
- Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H., & Mazuka, R. (2017). Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition*, 166, 84–93. <https://doi.org/10.1016/j.cognition.2017.05.003>.
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395. <https://doi.org/10.1037/0033-295X.115.2.357>.
- Ohala, J. J., & Shattuck-Hufnagel, S. (1986). The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech. *Phonology*, 3, 117–149.
- Oviatt, S., MacEachern, M., & Levow, G. A. (1998). Predicting hyperarticulate speech during human-computer error resolution. *Speech Communication*, 24(2), 87–110.
- Papoušek, M., Papoušek, H., & Bornstein, M. H. (1985). The naturalistic vocal environment of young infants: On the significance of homogeneity and variability in parental speech. In T. Field & N. Fox (Eds.), *Social perception in infants* (pp. 269–297). Norwood, N.J.: Ablex.
- Patterson, D. J., & Connine, C. M. (2001). A corpus analysis of variant frequency in American English flap production. *Phonetica*, 58(4), 254–275.
- Patterson, D. J., LoCasto, P., & Connine, C. M. (2003). A corpus analysis of schwa vowel deletion frequency in American English. *Phonetica*, 60, 45–68.
- Picheny, M. A., Durlach, N. I., & Braidia, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech Language and Hearing Research*, 29, 434–446.
- Pierrehumbert, J. (1994). Knowledge of variation. In *Papers from the parasession on variation, 30th meeting of the Chicago Linguistics Society* (pp. 232–256). Chicago: Chicago Linguistic Society.
- Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46(2–3), 115–154.
- Pitt, M. A. (2009b). The strength and time course of lexical activation of pronunciation variants. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 896–910.
- Pitt, M. A. (2009a). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language*, 61, 19–36.
- Pitt, M. A., Dilley, L. C., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics*, 39, 304–311. <https://doi.org/10.1037/e520592012-253>.
- Pitt, M. A., Dilley, L. C., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). Buckeye Corpus of Conversational Speech (2007; Final release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor). The Ohio State University. Columbus, OH.
- Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives* Technical Report DAMTP 2009/NA06. Cambridge, UK: Centre for Mathematical Sciences, University of Cambridge.
- Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, 57, 273–298.
- Redi, L. C., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, 29(4), 407–429.
- Reinisch, E., & Mitterer, H. (2016). Exposure modality, input variability and the categories of perceptual recalibration. *Journal of Phonetics*, 55, 96–108. <https://doi.org/10.1016/j.jwocn.2015.12.004>.
- Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics*, 45, 91–105. <https://doi.org/10.1016/j.jwocn.2014.04.002>.
- Rietveld, T., & van Hout, R. (1993). *Statistical techniques for the study of language and language behavior*. Mouton de Gruyter.
- Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, 12, 110–114.
- Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, 41(3), 249–263.
- Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, 52, 183–204.
- Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, 5(6), 546–554.
- Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133, 140–155. <https://doi.org/10.1016/j.cognition.2014.06.013>.
- Shafiq, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, 71, 55–89.
- Shockey, L., & Bond, Z. S. (1980). Phonological processes in speech addressed to children. *Phonetica*, 37, 267–274. <https://doi.org/10.1159/000259996>.
- Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 36(1), 195. <https://doi.org/10.1037/a0016803>.
- Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language*, 4, 1–22. <https://doi.org/10.1017/S0305000900000453>.
- Stam Casasanto, L. (2008). *Experimental investigations of sociolinguistic knowledge* (Ph.D.). Stanford.
- Stent, A. J., Huffman, M. K., & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication*, 50(3), 163–178.
- Stevens, K. N. (2000). Diverse acoustic cues at consonantal landmarks. *Phonetica*, 57, 139–151.
- Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111(4), 1872–1891.
- Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, 52, 322–338.
- Sumner, M., & Samuel, A. G. (2009). The effects of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60(4), 487–501. <https://doi.org/10.1016/j.jml.2009.01.001>.
- Sundberg, U. (2001). Consonant specification in infant-directed speech. Some preliminary results from a study of Voice Onset Time in speech to one-year-olds. *Paper presented at the Lund University Working Papers*.
- Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica*, 56, 186–199. <https://doi.org/10.1159/000028450>.
- Swingle, D., & Alarcon, C. (2018). Lexical learning may contribute to phonetic learning in infants: A corpus analysis of maternal Spanish. *Cognitive Science*. <https://doi.org/10.1111/cogs.12620>.
- Seyfarth, S., & Garellek, M. (2015). In *Coda glottalization in American English*. In *The Scottish Consortium for ICPHS, Proceedings of the 18th International Congress of*

- Phonetic Sciences. 978-0-85261-941-4. Glasgow, UK: the University of Glasgow. Paper number 807 retrieved from <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0807.pdf>.
- Th. Gries, S. (2015). The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, 10(1), 95–125.
- Th. Gries, S. (2016). *Quantitative corpus linguistics with R: A practical introduction* (2nd ed.). London: Routledge.
- Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7(1), 53–71.
- Tucker, B. V., & Warner, N. (2007). Inhibition of processing due to reduction of the American English flap. *Paper presented at the Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken*.
- Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, 49(1), 2–7.
- Wang, Y., Bergeson, T. R., & Houston, D. M. (2017). Infant-directed speech enhances attention to speech in deaf infants with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 60(11), 3321–3333.
- Wang, Y., & Seidl, A. (2015). The learnability of phonotactic patterns in onset and coda positions. *Language Learning and Development*, 11(1), 1–17.
- Wang, Y., Seidl, A., & Cristia, A. (2015). Acoustic-phonetic differences between infant- and adult-directed speech: The role of stress and utterance position. *Journal of Child Language*, 42(4), 821–842.
- Warner, N., Fountain, A., & Tucker, B. V. (2009). Cues to perception of reduced flaps. *The Journal of the Acoustical Society of America*, 125(5), 3317–3327. <https://doi.org/10.1121/1.3097773>.
- Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, 130(3), 1606–1617.
- Wassink, A. B., Wright, R. A., & Franklin, A. D. (2007). Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics*, 35, 363–379.
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143–2152. <https://doi.org/10.1177/0956797613488145>.
- Werker, J., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, 1, 197–234.
- Werker, J., Ferran, P., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, 103, 147–162.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganisation during the first year of life. *Infant Behavior and Development*, 7, 49–63.
- Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science*, 21(4), 221–226.
- Wieland, E. A., Burnham, E. B., Kondaurova, M. V., Bergeson, T. R., & Dilley, L. C. (2015). Vowel space characteristics of speech directed to children with and without hearing loss. *Journal of Speech, Language and Hearing Research*, 58(2), 254–267. [https://doi.org/10.1044/2015\\_JSLHR-S-13-0250](https://doi.org/10.1044/2015_JSLHR-S-13-0250).
- Zellou, G., & Scarborough, R. (2015). Lexically conditioned phonetic variation in motherese: Age-of-acquisition and other word-specific factors in infant- and adult-directed speech. *Laboratory Phonology*, 6(3–4), 305–336. <https://doi.org/10.1515/lp-2015-0010>.