

Phonetic variation in consonants in infant-directed and adult-directed speech: the case of regressive place assimilation in word-final alveolar stops*

LAURA C. DILLEY

Department of Communicative Sciences and Disorders, Department of Psychology, Department of Linguistics and Germanic, Slavic, Asian and African Languages, Michigan State University

AMANDA L. MILLETT

Department of Communication Sciences and Disorders, University of Missouri

J. DEVIN MCAULEY

Department of Psychology, Michigan State University

AND

TONYA R. BERGESON

Department of Otolaryngology – Head & Neck Surgery, Indiana University School of Medicine

(Received 1 November 2011 – Revised 10 August 2012 – Accepted 29 October 2012 – First published online 7 February 2013)

ABSTRACT

Pronunciation variation is under-studied in infant-directed speech, particularly for consonants. Regressive place assimilation involves a word-final alveolar stop taking the place of articulation of a following word-initial consonant. We investigated pronunciation variation in word-final alveolar stop consonants in storybooks read by forty-eight mothers in adult-directed or infant-directed style to infants aged approximately 0;3, 0;9, 1;1, or 1;8. We focused on phonological

[*] This research is supported by NIH-NIDCD R01DC008581 to T. Bergeson. We thank Claire Carpenter, Erin Dixon, Dana Flowerday, Shaina Selbig, Kellie Voss, and Zach Zells for their assistance with data analysis and coding. We also express our gratitude to Jessica Gamache and Evamarie Cropsey for assistance with figures and references, respectively. Moreover, we gratefully acknowledge Chris Heffner, Erin Dixon, Elizabeth Wieland, and Tuuli Morrill for their valuable feedback on an earlier draft of this manuscript. Address for correspondence: Dr Laura C. Dilley, Michigan State University, Department of Communicative Sciences and Disorders, 1026 Red Cedar Road, Rm. 116, Oyer Speech & Hearing Building, East Lansing, MI 48824-1220. tel: 517-884-2255; fax: 517-353-3176; e-mail: ldilley@msu.edu

environments where regressive place assimilation could occur, i.e., when the stop preceded a word-initial labial or velar consonant. Spectrogram, waveform, and perceptual evidence was used to classify tokens into four pronunciation categories: canonical, assimilated, glottalized, or deleted. Results showed a reliable tendency for canonical variants to occur in infant-directed speech more often than in adult-directed speech. However, the otherwise very similar distributions of variants across addressee and age group suggested that infants largely experience statistical distributions of non-canonical consonantal pronunciation variants that mirror those experienced by adults.

INTRODUCTION

In recent decades, a sizeable body of research has identified compelling ways in which adults' speech input to very young children, here termed infant-directed (ID) speech, is clearly distinguished from adult-directed (AD) speech. In particular, when compared to AD speech, ID speech is thought to involve a combination of modifications to prosodic and segmental aspects of the signal. While studies of ID speech prosody are fairly common, studies of phonetic properties of ID speech segments, particularly consonants, are relatively rare. Yet understanding the phonetic properties of consonants in speech to children is important, given evidence that children use distributional phonetic properties of consonants to build their phonological and cognitive representations of speech (Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992; Maye, Werker & Gerken, 2002).

Prior phonetic studies have suggested that both prosodic and segmental attributes can be exaggerated or clearer in ID speech than in AD speech. With respect to prosody, studies of ID speech have shown increased pitch range, slower rate, and longer pauses compared to AD speech (Snow, 1977; Fernald & Simon, 1984; Papoušek, Papoušek & Bornstein, 1985; Fernald, 1992; Bergeson & Trehub, 2002; Bergeson, Miller & McCune, 2006). With respect to segments, a number of studies have suggested that phonemes in ID speech are more carefully articulated relative to those in AD speech (Bernstein Ratner, 1984a, 1984b; Kuhl *et al.*, 1997; Burnham, Kitamura & Vollmer-Conna, 2002). For example, the point vowels /i, a, u/ in ID speech tend to show more extreme formant frequencies relative to those of AD speech (Kuhl *et al.*, 1997; Burnham *et al.*, 2002; Liu, Kuhl & Tsao, 2003); however, see Englund and Behne (2006) for different results. Experiencing clearer or hyperarticulated vowel segments has been proposed to facilitate phonetic categorization ability and the development of phonological categories (Kuhl *et al.*, 1997; de Boer & Kuhl, 2003).

Fewer studies have specifically examined phonetic properties of consonants in ID speech and the findings from these studies are inconsistent.

On the one hand, some studies have reported that ID speech contains exaggerated or clearer cues to consonantal contrasts compared with AD speech. For example, Malsheen (1980) reported a greater contrast in the voice-onset time (VOT) of voiced vs. voiceless stop consonants for ID speech relative to AD speech for speech to children aged 1;3–1;4. Likewise, Englund (2005) reported longer VOT in six Norwegian mothers' ID speech compared with their AD speech for the majority of stop consonants examined. Bernstein Ratner, and Luberoff (1984) examined phonetic cues to the contrast between voiced and voiceless final consonants, and found that vowels tended to be about twice as long before voiced final consonants in ID speech than in AD speech. Phonological reduction of consonants (e.g., *want it*→*wannit*) has also been reported to occur less frequently in ID speech than in AD speech (Bernstein Ratner, 1984b).

On the other hand, some studies have reported no difference in the degree of clarity of consonants in ID speech relative to AD speech, or else have unexpectedly revealed clearer consonantal cues in AD speech compared with ID speech. For example, Shockey and Bond (1980) examined four types of phonological modification – palatalization (did you→[dIɟju]), dental deletion (*want it*→*wannit*), /ð/ deletion (*throw them*→*throw 'em*), and /ts/→/s/ conversion (*that's nice*→*thass nice*) – and found that all four phonological rules were MORE likely to be observed when the eight mothers spoke to children than when they spoke to adults. In addition, though Bernstein Ratner and Luberoff (1984) reported a greater tendency to lengthen vowels before voiced final consonants in speech directed to children compared to that directed to adults, suggesting greater clarity in child-directed speech, they also found that participants deleted or glottalized final consonants more often when speaking to children than to adults; see Shockey and Bond (1980) for similar results.

Studies of the intelligibility of speech to children and speech to adults have likewise not supported the idea that speech to children is generally clearer than speech to adults (Bard & Anderson, 1983, 1994). For example, Bard and Anderson (1983) found that 'intelligibility of parental speech to children was lower, even for matched words, than was intelligibility of speech to an adult listener'. In addition, other studies have found that the VOTs of voiced and voiceless consonants are not more distinctive in ID speech than in AD speech. For example, Sundberg and Lacerda (1999) reported a smaller VOT difference in Swedish speech between voiced and voiceless stop consonants in ID speech compared with AD speech. Furthermore, Baran, Laufer, and Daniloff (1977) found no significant differences in VOT of voiced and voiceless consonants in adult-directed and child-directed conversation produced by their three participants. Indeed, the data of Baran *et al.* (1977) revealed shorter VOT values for the voiceless stops in the child-directed condition than for the adult-directed condition.

One possible reason for the discrepancies across studies of consonant clarity in ID speech is the very different ages of children in prior studies. It has been suggested that mothers may produce different phonetic modifications to consonants as a function of the linguistic development of a child (Bernstein Ratner, 1984b; Sundberg & Lacerda, 1999). One specific proposal is that consonant modification may follow a non-monotonic trajectory from hypoarticulation to hyperarticulation and back to hypoarticulation (Sundberg & Lacerda, 1999). The period of hyperarticulation was proposed by Sundberg and Lacerda (1999) to occur around the time of the first signs of robust infant comprehension, which occur around twelve months of age (Oviatt, 1980). Hyperarticulation of consonant contrasts might be expected to occur in ID speech to older infants, but not to younger infants, since older infants would be better able to make use of this information. Such a non-monotonic trajectory would potentially help explain inconsistent findings about consonant clarity in speech to children at different linguistic stages of development. In support of this possibility, Sundberg and Lacerda (1999) found a smaller contrast in VOT for voiced and voiceless stops for ID speech to children aged 0;3 and a larger contrast in ID speech to children aged 0;11–1;2. Cristia (2009) likewise found a similar result in her study of ID speech to children aged 0;4–0;6 and 1;0–1;2. She examined the contrast between /s/ and /ʃ/ as measured by differences in the first spectral peak of the fricative spectrum and found a smaller contrast between the two fricatives in ID speech to the younger group and a larger contrast in ID speech to the older group.

Another possible reason for conflicting findings about phonetic properties of consonants across studies is lack of representativeness due to inadequate sampling associated with small numbers of participants. For example, the study by Malsheen (1980) included two mother–child dyads for children in each of three age ranges (0;6–0;8, 1;3–1;4, and 2;5–5;2). Likewise, studies by Baran *et al.* (1977), Sundberg and Lacerda (1999), and Englund (2005) examined consonant VOT in speech of three, six, and six participants, respectively at a restricted range of ages. The studies by Shockey and Bond (1980) and Bernstein Ratner (1984b) investigated the speech of eight and seven participants, respectively.

Given that studies of phonetic properties of consonants in speech directed to infants and young children are not only rare but have yielded inconsistent findings, there is good reason to look more carefully at consonants in speech directed to children. Toward this end, the present study investigated phonetic cues associated with a subset of word-final consonants in ID speech, using a much larger sample of mothers ($n=48$) speaking to infants at ages 0;3, 0;9, 1;1, or 1;8. Of particular interest was the phenomenon of regressive place assimilation in word-final alveolar consonants. Regressive place assimilation is the phenomenon whereby the place of articulation

(POA) of a word-final alveolar consonant is altered to match that of the following consonant. For instance, the /n/ at the end of *green* may take the labial place of the following /b/ in the phrase *green boats*, so that *green* appears to be pronounced as *greem*. Assimilation of such consonants has been widely studied in the production and perception of adult-directed speech (Holst & Nolan, 1995; Ellis & Hardcastle, 2002; Dilley & Pitt, 2007) and has figured prominently in theories of adult speech perception (Lahiri & Marslen-Wilson, 1991; Gaskell & Marslen-Wilson, 1998; Gow, 2003).

In the ID speech literature, regressive place assimilation has received little attention, leaving a number of unanswered questions. At the phonological level, little is known about the frequency of assimilation relative to other types of pronunciation variation that might also occur in contexts in which assimilation is possible, that is, in assimilable environments. Some prior work suggests that deleted or glottalized variants can at least occasionally occur for word-final consonants in ID speech (Shockey & Bond, 1980; Bernstein Ratner & Luberoff, 1984).

In this study, we considered several possibilities regarding how assimilation and/or other variant pronunciations may pattern in ID speech compared with AD speech. One possibility is that ID speech contains predominantly canonical (i.e., clear) word-final pronunciations of alveolar stop consonants compared with AD speech, as suggested by findings that pronunciation in ID speech is more careful than in AD speech (Bernstein Ratner, 1984a, 1984b; Kuhl *et al.*, 1997; Burnham *et al.*, 2002).

A second possibility is that, overall, ID speech contains about the same proportion of canonical and non-canonical variant pronunciations as AD speech. In this case, it is possible that the distribution of NON-CANONICAL variant types is either the same or different in the two styles of speech. Prior work has shown that AD speech contains a mixture of canonical, assimilated, deleted, and glottalized variant pronunciations in phonological environments where regressive place assimilation to the word-final alveolar stop could occur (Dilley & Pitt, 2007). In this regard, it is possible that ID speech contains a different distribution of non-canonical variant types than AD speech – for example, perhaps ID speech contains a simpler mixture of canonical and assimilated variant pronunciations of word-final alveolar stop consonants, and few, if any, deleted and glottalized variants. A distribution of pronunciations that was comprised of almost exclusively canonical or assimilated variant pronunciations could readily lend itself to phonological rule-learning and enhanced speech perception and lexical decoding by a child, since assimilated variants (e.g., *green* → [gri:m]) could be easily recovered to their canonical pronunciation forms via application of a simple phonological rule (e.g., ‘restore alveolar place of articulation to a word-final consonant exhibiting labial place of articulation when it is followed by a word-initial labial consonant’).

A final possibility is that ID speech contains FEWER canonical pronunciation variants than AD speech. This is not only a logical possibility but one which found some support from the study of Shockey and Bond (1980). They showed that phonological rules that changed canonical pronunciations to variant forms were more likely to have been applied when mothers spoke to children than when they spoke to adults.

By testing four ages of infants, we additionally were able to address whether the distribution of pronunciations of word-final alveolar stop consonants in ID speech might change over time. If consonant modification follows a non-monotonic trajectory from hypoarticulation to hyperarticulation and back to hypoarticulation, with the period of consonant hyperarticulation predicted to occur around the time that infants exhibit signs of robust comprehension (i.e., around twelve months of age; Oviatt, 1980), then we would expect to see the proportion of canonical variants in ID speech have an inverted U-shape as a function of age.

METHOD

Participants

Forty-eight mother–infant dyads were recruited from the local community. All mothers were from a restricted geographic and dialectal region (i.e., American English speakers from Indiana); they were given \$10 per visit for their time. Infants (19 girls, 29 boys) in the dyads were normally developing and in one of four age groups, with twelve infants per group. The first group of infants was aged approximately 0;3 (mean age 0;3·2; range: 0;2·15–0;4·3; 4 F, 8 M). The second group of infants was aged approximately 0;9 (mean age 0;9·1; range: 0;8·9–0;9·27; 5 F, 7 M). The third group of infants was aged approximately 1;1 (mean age 1;0·24; range: 1;0·3–1;1·24; 4 F, 8 M). The fourth group of infants was aged approximately 1;8 (mean age 1;8·11; range: 1;6·21–1;9·24; 6 F, 6 M). This research and the recruitment of human subjects were approved by the Indiana University Institutional Review Board. Some mother–infant dyads participated in more than one recording session over time; however, each mother–infant dyad selected for analysis was included in only a single infant age group.

Design

Pronunciation variation in word-final alveolar stop consonants (/t/, /d/, /n/) was examined as a function of the infant's age for two speech styles: adult-directed (AD) speech and infant-directed (ID) speech to normally developing infants, with the order of ID and AD speech recordings counterbalanced across mothers. We focused on pronunciation variation for

consonants in phonological contexts where regressive place assimilation could potentially occur (i.e., where a word-final alveolar stop was followed by a word-initial labial or velar consonant).

Apparatus

Speech samples were digitally recorded in a double-walled, copper-shielded sound booth (Industrial Acoustics Company) using an SLX Wireless Microphone System (Shure). This system included an SLX1 Bodypack transmitter with a built-in microphone and a wireless receiver SLX4 which was connected to a Canon 3CCD Digital Video Camcorder GL2, NTSC. The speech samples were recorded directly onto a Mac computer (Apple, Inc. OSX Version 10.4.10) via Hack TV (Version 1.11) software.

Procedure

Each mother's speech was recorded during a single visit to the DeVault Otologic Lab at Indiana University School of Medicine. In the ID speech condition, each mother was asked to read a storybook to her child and to sit with her child in the sound booth and interact with her child as she normally would at home. The text of the storybook is shown in the 'Appendix'; the text was constructed to consist of a range of consonants and vowels of interest in controlled phonetic contexts. The storybook was illustrated with age-appropriate color pictures. In the AD speech condition, each mother was asked to read the same storybook aloud while she was alone in the sound booth as she would to an adult. Each ID and AD session lasted approximately 2–5 minutes. Mothers occasionally deviated from the script; all speech they produced during the recording session was subject to analysis. The order of ID and AD speech recordings was counterbalanced across mothers: approximately half of the mothers completed the AD speech condition first ($n = 22$), while the remaining mothers completed the ID speech condition first ($n = 18$).

Data analysis

Instances of target assimilable environments were identified in recorded speech. Tokens analyzed were limited to word-final alveolar consonants (/t/, /d/, /n/) followed by a word-initial labial (/b/, /p/, or /m/) or velar (/g/ or /k/) consonant where place assimilation could potentially occur. The storybook included twenty-two scripted assimilable environments; the actual number varied slightly by mothers due to spontaneous repetition, omission, etc. Tokens with breaks or pauses between the word-final alveolar consonant and following word-initial segment were eliminated from

analysis based on perceptual evidence of a pause and an observed gap of greater than 100 ms in the spectrogram. This was done because it was reasoned that for such cases the temporal distance between a word-final alveolar consonant and following word-initial non-alveolar consonant would likely be too great to permit regressive place assimilation to occur. Tokens were also eliminated from analysis due to infant vocalizations, microphone static or popping, or other distractions due to recording quality or audibility that impacted the clarity of the token. The mean number of usable tokens in assimilable environments did not differ between the AD speech condition ($M=21.0$, $SD=1.4$) and the ID speech condition ($M=20.3$, $SD=3.9$), $t(47)=0.2$, $p=.70$.

Each word-final alveolar token was given a phonological classification based on labeling conventions established in Dilley and Pitt (2007) and Kiesling, Dilley, and Raymond (2006). Seven undergraduate research assistants were trained in phonetic variant classification. The training involved describing to trainees the four phonological variant categories, including giving specific instructions on how to use key spectrographic evidence along with perceptual evidence to arrive at a reliable classification label for each category. Individual and group practice assignments were given using speech samples from the Buckeye corpus of conversational speech (Kiesling *et al.*, 2006). Labelers were given the opportunity to generate questions and received feedback following practice. Each of the seven labeler's classifications on practice assignments were compared to a consensus classification from the corpus and to the classifications arrived at by the other trained labelers. All met criteria for accuracy before they were accepted as labelers for the present project.

Tokens were assigned to one of four phonological categories on the basis of spectrographic and auditory perceptual evidence; see Figure 1. A word-final /t,d,n/ was classified as ASSIMILATED when there was spectral evidence of a local change in the trajectory of the second formant, F₂, in a sonorant segment just preceding the /t,d,n/ closure such that F₂ fell or rose in the case of a following labial or velar place of articulation, respectively (Figure 1a). Finally, perceptual evidence had to be consistent with a word-final /t,d,n/ adopting a labial or velar POA such that the word-final segment sounded like its POA had changed to that of the following sound. A word-final /t,d,n/ was labeled as GLOTTALIZED if the word-final consonant exhibited glottalization, defined as perceptually creaky voicing accompanied by irregularity in period-to-period durations of successive pitch pulses in the waveform (Figure 1b). Next, a word-final segment was labeled as CANONICAL when it was perceived to be present, unassimilated, and without voicing irregularity. This category was also the default classification when the variant type was uncertain (Figure 1c). Finally, a word-final /t,d,n/ was classified as if there was no auditory or visual

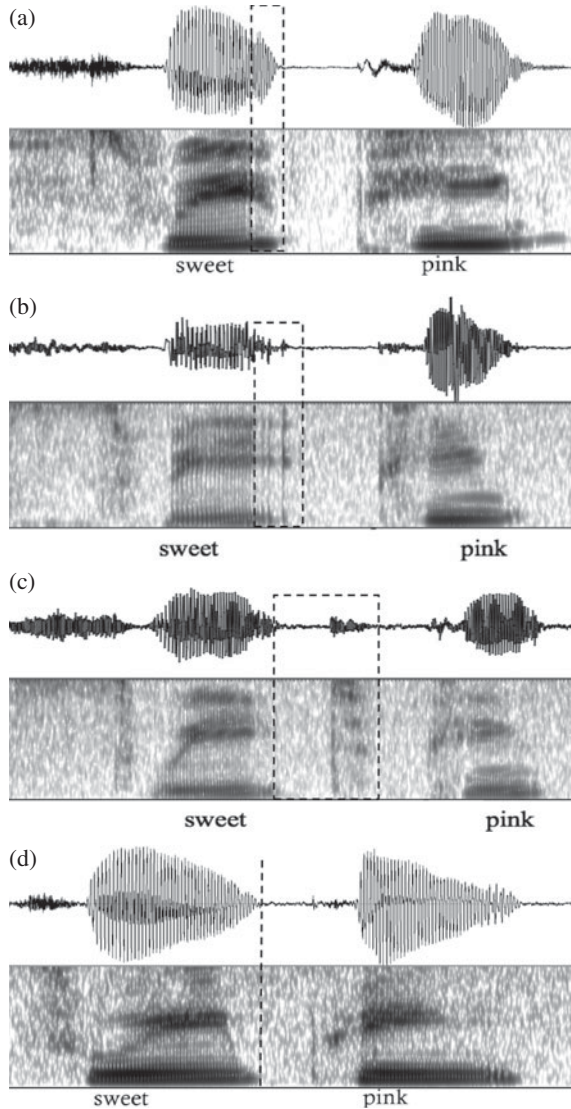


Fig. 1. Speech waveforms and spectrograms showing examples of tokens with underlying word-final /t/ in the phrase *sweet pink* classified in different ways in this study: (a) assimilated, (b) glottalized, (c) canonical, (d) deleted. Dashed rectangles enclose the portions of the acoustic signal corresponding to the /t/ in (a)–(c), while the dashed line in (d) indicates the approximate temporal location of the /t/ segment if it were present for the case when the segment was classified as deleted.

TABLE 1. *Distribution of phonetic variation in mothers' ID and AD speech conditions (A = assimilated, C = canonical, D = deleted, G = glottalized)*

Infant Age (years;months)	ID speech				AD speech			
	A (%)	C (%)	D (%)	G (%)	A (%)	C (%)	D (%)	G (%)
0;3	24	53	5	18	24	51	4	21
0;9	34	44	4	18	38	39	6	17
1;1	25	53	4	18	28	48	4	20
1;8	27	52	5	16	27	49	5	19

evidence in the spectrogram that the segment had been spoken. The word-final /t,d,n/ was to be classified as deleted when it could not be heard when a short context was played and when there was no clear visual evidence in the spectrogram that the segment was present, i.e., the entire closure period for the C#C sequence was relatively short.

Each of the seven trained labelers classified a subset of the tokens. Each token was classified by a total of three labelers out of the set of seven trained individuals; the same three labelers classified tokens in both AD and ID conditions for each mother. For each token, a consensus classification was determined by identifying the category assessed by the majority of labelers. In instances where there was no consensus among the labelers based on independent classification, the token was discussed by all three individuals, who reached a judgment about the best classification for that token. In addition, when one labeler disagreed with the other two, the labeler who did not agree was informed of the disagreement and given the option of leaving his or her classification as is or else changing his/her label. Agreement among labelers was quantified via use of the chance-adjusted Kappa (κ) metric (Carletta, 1996). Reliability was found to correspond to $\kappa = 0.74$ for all phonetic classifications across the four groups; values of κ above 0.6 are typically considered to indicate substantial agreement (Landis & Koch, 1977; Rietveld & van Hout, 1993; Breen, Dilley, Kraemer & Gibson, 2012).

RESULTS

Table 1 shows percentages of each type of pronunciation variant produced by mothers in ID and AD speech conditions for mothers of infants who were aged 0;3, 0;9, 1;1, or 1;8. Canonical speech was the most frequent form across speech style and infant age. Assimilation was consistently the second most frequent form of variation. Glottalized variants were observed next most often, and deleted variants were least frequently observed.

PHONETIC VARIATION IN WORD-FINAL STOP CONSONANTS

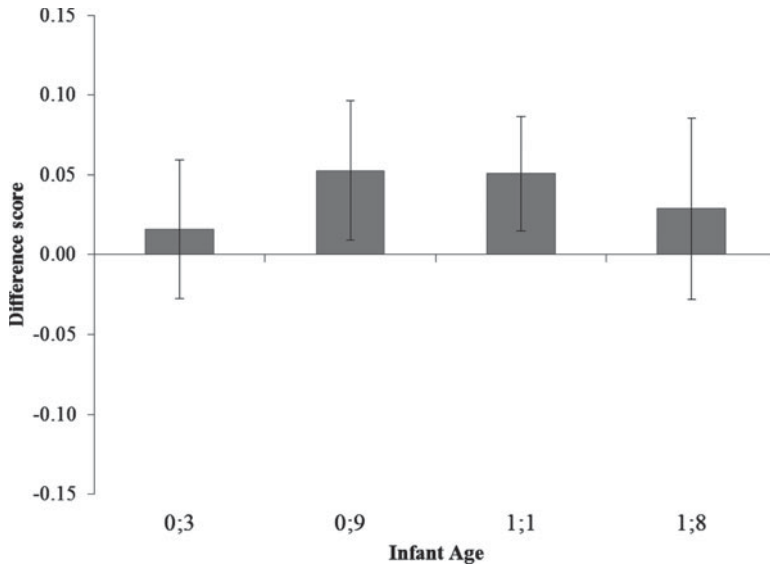


Fig. 2. Difference scores across four infant age groups, calculated as the difference between the proportion of canonical variants produced in the ID condition and the proportion of canonical variants produced in the AD condition. Error bars are ± 1 standard error.

Overall, there were substantial individual differences in the proportion of canonical tokens produced in ID speech ($M=0.51$, $SD=0.22$, range: $0.08-0.95$) and in AD speech ($M=0.47$, $SD=0.19$, range: $0.05-0.95$). Across mothers the proportion of canonical variants in ID speech was highly correlated with the proportion of canonical variants in AD speech ($r=0.73$, $p < .001$, two-tailed), suggesting that the general tendency to produce more or less canonical pronunciations was an attribute of individual mothers that transcended speech style.

A preliminary ANOVA on proportion of canonical variants that included Order (ID speech vs. AD speech first), Addressee (ID, AD), and Infant Age (0;3, 0;9, 1;1, or 1;8) as factors revealed no effect of Order ($p > .25$) and no interactions with Order ($p > .17$); as such, we collapsed over Order for the remainder of the analyses. To take into account the large variation across mothers in production of canonical tokens, we calculated for each mother the difference between the proportion of canonical variants produced in the ID condition and the proportion of canonical variants produced in the AD condition. This difference score was a means of ‘normalizing’ for each mother’s rate of producing canonical variants. Figure 2 shows the mean difference score as a function of Infant Age. Values greater than zero correspond to more canonical variants produced

TABLE 2. *Distribution of phonetic variation in mothers' ID speech separated by the age and gender of the infants (A = assimilated, C = canonical, D = deleted, G = glottalized)*

Infant		ID speech			
Age (years;months)	Sex	A (%)	C (%)	D (%)	G (%)
0;3	M (<i>n</i> =8)	29	43	4	24
	F (<i>n</i> =4)	13	74	6	7
0;9	M (<i>n</i> =7)	37	37	5	21
	F (<i>n</i> =5)	30	54	3	13
1;1	M (<i>n</i> =8)	26	52	4	18
	F (<i>n</i> =4)	25	53	2	20
1;8	M (<i>n</i> =6)	27	54	7	12
	F (<i>n</i> =6)	27	50	4	19

in ID speech than in AD speech, while values less than zero correspond to fewer canonical variants produced in ID speech than in AD speech. Overall, the mean difference score was significantly greater than zero ($M=0.04$, $SD=0.15$, $t(47)=1.67$, $p<0.05$, one-tailed), indicating a modest tendency for ID speech to show an increase in the proportion of canonical variants compared to AD speech. A one-way ANOVA on difference scores revealed no effect of Infant Age ($F(3, 47)=0.15$, $p=0.93$, $\eta_p^2=0.01$).

Next, we conducted a set of non-parametric chi-squared analyses to compare the distributions of the three non-canonical phonetic variant types in ID and AD speech conditions. These analyses revealed no significant difference in the distribution of non-canonical variants in ID speech compared with AD speech ($\chi^2(2, N=1015)=0.08$, $p=0.96$) and no reliable change in the distribution of non-canonical variants across the four infant ages in ID speech ($\chi^2(6, N=491)=4.77$, $p=0.57$). Moreover, there were no significant differences in the distributions of non-canonical variants for each age group considered separately for ID speech compared with AD speech (age 0;3: $\chi^2(2, N=244)=0.53$, $p=0.77$; age 0;9: $\chi^2(2, N=286)=0.69$, $p=0.71$; age 1;1: $\chi^2(2, N=249)=0.03$, $p=0.98$; age 1;8: $\chi^2(2, N=236)=0.28$, $p=0.87$).

Finally, we explored potential differences in ID speech to boys and girls. Table 2 shows the distribution of all four pronunciation variants in the ID speech condition separated by infant gender. Overall, mothers reading to girls tended to produce more canonical variants than mothers reading to boys, especially at 0;3 and 0;9. Chi-squared analyses revealed a significant effect of gender for infants aged 0;3 ($\chi^2(3, N=258)=21.68$, $p<0.001$), a marginal effect of gender for infants aged 0;9 ($\chi^2(3, N=239)=6.53$,

$p = .08$), and no significant effects of gender for either infants aged 1;1 ($\chi^2(3, N = 246) = 0.69, p = .87$) or infants aged 1;8 ($\chi^2(3, N = 232) = 2.74, p = .43$).

DISCUSSION

This study investigated pronunciation variation in word-final alveolar stop consonants in a cross-sectional sample of forty-eight mother–infant dyads in which the infants were aged approximately 0;3, 0;9, 1;1, or 1;8. While the assimilation of word-final alveolar stop consonants to labial or velar place of an upcoming segment has been an important topic in adult speech perception (e.g., Gaskell & Marslen-Wilson, 1998; Gow, 2003), no prior published work has examined this phenomenon in ID speech. Here, spectrographic, waveform, and perceptual evidence was used to classify word-final alveolar stops in assimilable environments as assimilated, canonical, deleted, or glottalized.

Across the four infant age groups, canonical variants were most commonly produced in both ID and AD speech conditions, followed by assimilation, glottalization, and deletion, in that order. Overall, the data provide some support for greater consonantal clarity in ID speech, in that more canonical variants were produced in ID speech than in AD speech after ‘normalizing’ for an individual mother’s rate of canonical variant production. Although there was a trend for this greater clarity ID speech to peak at about 0;9, the effect of infant age was not statistically reliable. Moreover, non-parametric analyses revealed no significant differences in the distribution of produced non-canonical variants in ID and AD speech conditions. Exploratory analyses examining effects of gender provided some evidence that mothers may talk more clearly to girls than to boys at the youngest ages examined here.

Our results are consistent with prior studies that have reported that ID speech contains clearer cues to consonantal contrasts compared with AD speech (Malsheen, 1980; Englund, 2005; Bernstein Ratner & Luberoff, 1984; Bernstein Ratner, 1984b). Our findings thus contrast with those of studies which have shown either no difference in consonantal clarity between ID and AD speech or else less clear consonants in ID speech than AD speech (Baran *et al.*, 1977; Sundberg & Lacerda, 1999). It is noteworthy that in our study, greater clarity in ID speech consonants was only observable after taking into account large individual differences between mothers in canonical variant usage in both speech styles via the use of a difference score; this was true even though we tested substantially more participants and thus had considerably more power to detect differences between AD and ID speech than previous studies. These considerations suggest that any tendency for ID speech to contain clearer consonants is

a rather weak effect, which might explain some other studies' failure to find enhanced consonantal clarity in ID speech relative to AD speech.

Overall, the picture of pronunciation variation in ID speech compared to AD speech appears somewhat more complex for consonants than for vowels. While vowels typically have been reported to show hyperarticulation in ID speech compared with AD speech (e.g., Kuhl *et al.*, 1997; Burnham *et al.*, 2002), consistent with our finding of somewhat more canonical pronunciations for word-final alveolar stop consonants, other studies have shown mixed results in regards to hyperarticulation or consonantal clarity in these two speech styles. It is noteworthy that studies of consonant pronunciation in ID speech have differed on multiple factors, so it is unclear to what extent these studies can legitimately be compared. First, studies of consonant variation in ID speech have investigated different indices of pronunciation (e.g., word-final alveolar stop variants, VOT, phonological rule modification). Moreover, studies of consonant variation in ID speech have differed in the language and/or dialect background of participants. For example, among studies of VOT, Sundberg and Lacerda (1999) examined Swedish, Englund (2005) examined Norwegian, and Baran *et al.* (1977) and Malsheen (1980) examined English; VOT can vary dramatically across different languages and dialects (Lisker & Abramson, 1964; Abramson & Lisker, 1968; Kortmann & Schneider, 2005). Moreover, British English and American English, the dialects which were studied by Shockey and Bond (1980) and Bernstein Ratner (1984b) respectively, are known to differ in their application of phonological rules and distributions of phonetic variants of /t/ and other segments (Celce-Murcia, Brinton & Goodwin, 1996), which could be responsible for their somewhat conflicting findings regarding phonological rule usage in ID speech.

Our study was also able to test the hypothesis that the degree of care in consonant pronunciation follows a non-monotonic trajectory over time from hypoarticulation to hyperarticulation and back to hypoarticulation (Sundberg & Lacerda, 1999). It has been suggested that hyperarticulation of consonant contrasts might be found in ID speech to older infants, but not to younger infants, since older infants would be better able to make use of this pronunciation information. The period of hyperarticulation was predicted to occur around 1;0–1;2, when infants begin to exhibit the first robust signs of comprehension (Oviatt, 1980). This would suggest a kind of U-shaped distribution, which has found some support in prior studies (Malsheen, 1980; Sundberg & Lacerda, 1999). We examined speech to infants of ages selected to span this critical predicted stage in speech development. Although there was a trend for the predicted non-monotonic trajectory in the data, there was no statistical support for the reliability of this trend. As a result, our findings did not provide any evidence to support the hypothesis that the degree of pronunciation care in consonants

follows an inverted U-shaped distribution, as suggested by Sundberg and Lacerda (1999).

A critical aspect of our study design that potentially limits the generalizability of our findings was our choice to use a scripted, storybook read to the child and an (imaginary) adult in an experimental setting in order to elicit phonetic contexts of interest. The nature of the speech input to children used in storybook readings differs in a number of respects from that which infants would typically encounter in more naturalistic circumstances. For one thing, the storybook was scripted, so that our speech materials consisted of a fixed lexicon and syntax; in contrast, in naturalistic interactions, adults tend to tailor the structure of their input to a child's growing comprehension and expressive language skills. In addition, infants do not typically listen to storybook reading most of the day, particularly when they are quite young. Moreover, it may be unnatural for an adult to read a children's story to another adult, or even to an infant in an experimental setting. While all of these factors potentially limit the ecological validity of the present study, the methodology used here still afforded multiple advantages, notably the use of a large sample of mothers producing speech to infants of different ages and to adults in a cross-sectional, counterbalanced design, and the use of materials with controlled phonological properties designed to elicit multiple observations of phonetic phenomena of interest.

Additional investigation will be necessary to determine if our findings of more frequent canonical (i.e., clear) pronunciations in ID speech relative to AD speech generalize to more naturalistic materials and settings. However, it is reasonable to think that ID spontaneous speech consonants might show a different patterning than ID read speech, since previous work on adult-directed speech suggests that under conditions of casual speaking styles, as well as at faster speech rates (e.g., Barry, 1992), place assimilation is more common. Moreover, adult-directed spontaneous speech is known to show different acoustic-phonetic attributes from adult-directed read speech, including greater gestural overlap, a higher degree of segmental deletion, and different strength of consonantal gestures (e.g., Browman & Goldstein, 1990; Shockey, 2003; Johnson, 2004). Indeed, overall variant rates observed in this study of ID read speech were quite different from those in the American-English AD spontaneous speech examined in Dillely and Pitt (2007); this is likely to be a function of the high rate of function words in their corpus, compared with the high rate of content words examined here.

A minor methodological limitation of the present work is that it was not designed to fully take into account the distributions or types of prosodic boundaries in the speech. Large prosodic boundaries have been argued to block application of certain phonological rules (e.g., assimilation) (Shattuck-Hufnagel & Turk, 1996), and as such may explain some of the

distributional facts about consonantal variant usage obtained here. Our methodology attempted to exclude the largest prosodic boundaries (e.g., intonation phrase and utterance boundaries) by eliminating tokens where the word-final alveolar consonant and following labial or velar segment were separated by a silence. Ultimately, the extent to which prosodic boundary distributions may help to explain patterns of consonant variant distribution across AD and ID speech is a topic for further study.

The present findings have implications for understanding children's growth of knowledge of pronunciation variation of words. Recall that one view of development of children's segmental knowledge which motivated the current research was the view that children first develop a prototypical feature category based on clear, canonical consonant exemplars; this knowledge then progresses to dealing with more and more varied consonant forms as a child ages. Such a view of development of segmental knowledge has grown out of previous findings that vowels in ID speech tend to show more careful, canonical pronunciations early on, which become more variable as children get older (Kuhl *et al.*, 1997; Burnham *et al.*, 2002). This view would have been consistent with findings that ID speech contained a restricted or simplified set of phonetic variants relative to AD speech. Two possible ways that ID speech might have been simplified relative to AD speech are (i) if ID speech were to consist wholly (or almost wholly) of canonical, 'veridical' segmental pronunciations in phonological environments where assimilation could occur, or (ii) if ID speech were to consist of a simple alternation between canonical pronunciations and assimilated forms in these environments. In case of (ii), i.e., a simple alternation between variant types (e.g., between assimilated vs. canonical forms), it would be reasonable to posit a cognitive apparatus whereby child listeners could accommodate phonetic variation using a straightforward phonological rule (e.g., Kager, 1999) which enabled them to 'undo' phonological variation (e.g., assimilation) and thereby recover the underlying form (i.e., a canonical variant).

In contrast to these possibilities, our results show that ID speech is not greatly simplified in the pronunciations of consonants relative to AD speech. Rather, ID speech contains a mixture of pronunciation variants in restricted phonological environments (the environment where assimilation is expected to occur). As a result, the present findings do not provide much support for the notion that children's pronunciation knowledge starts by first developing a prototypical feature category based on clear, canonical consonant exemplars. Rather, the present findings are more consistent with the alternative hypothesis that children deal with pronunciation variation by being sensitive to the probabilistic distribution of phonetic cues in their environments. It is not the case that phonological environments where regressive place assimilation is possible are associated with a single type

of pronunciation variation, i.e., assimilation. Rather, deletion and/or glottalization of the word-final stop segment can also occur, indicating greater complexity in the process of uncovering the intended segmental string than is implied by many phonological treatments.

In contrast, the present findings suggest that children's knowledge of pronunciation variation may be probabilistic in nature, such that children may pick up on distributional information of consonants. Prior findings have similarly suggested that children use distributional phonetic properties of consonants to build their phonological and cognitive representations of speech in order to bootstrap their speech perception abilities (Kuhl *et al.*, 1992; Maye *et al.*, 2002). Adult learners have been shown to be sensitive to distributional phonetic information in learning about consonants (Clayards, Tanenhaus, Aslin & Jacobs, 2008). The present pattern of results therefore is consistent with the idea that the distribution of pronunciation variants of word-final stop consonants which children are exposed to would prepare them well for the distribution of word-final stop consonants that they would be exposed to as adults. This varied pattern of pronunciation may be necessary early in life to facilitate children's perception of the variation present in typical adult speech.

The present findings may also have clinical implications for interventions for children diagnosed with speech sound disorders. Clinicians often focus on providing clear, canonical examples of segmental contrasts to children of words in isolation, under the assumption that children's knowledge of segmental categories and contrasts develops around canonical pronunciations (Bernthal, Bankson & Flipsen, 2009; Bowen, 2009). The present findings are consistent with the idea that development of consonantal pronunciation for typically developing children is guided by exposure to a wide variety of variant pronunciations which are comparable in type and distribution to those seen in adult speech. This suggests that exposure to and knowledge of variable pronunciation forms may be helpful or necessary for optimal speech development in children, so that clinical intervention might be more effective if focused on mirroring the consonant pronunciation variation seen in the adult language. More research is needed to investigate this possibility.

Although the pattern of results shown here did not support the hypothesis that children's typical early speech development is guided by exposure to a predominance of canonical pronunciations of word-final alveolar stop consonants in ID speech, the possibility still exists that particular patterns of pronunciation might be more beneficial for children's speech development. For example, one possibility is that those few children who are exposed to a predominance of canonical pronunciations in early exposure receive benefits through speeded learning of the clear, canonical segmental categories, thereby providing an 'anchor' for variant pronunciations.

One reason we are interested in this possibility is our long-term goal of investigating how individual differences in phonetic input correlate with differential achievement in children's speech and language skills, including those with hearing impairment. Previous findings suggest a link between the differential nature of speech input to children and their speech–language achievement later on (Hart & Risley, 1995; Kuhl *et al.*, 1997; Hurtado, Marchman & Fernald, 2008). In particular, the pronunciation of segments (i.e., vowels) in infant-directed speech is one predictor of later speech skill abilities, at least for normal-hearing children (Liu *et al.*, 2003).

Overall, our measure of degree of assimilation was conservative, such that cases marked as assimilation were those where assimilation was generally very clear. This means that the estimate provided by our 'assimilated' category of the degree to which partial assimilation occurred is likely to be somewhat low. Much evidence exists that assimilation can be partial or incomplete and that there are residual cues to the intended place of articulation, even when assimilation has occurred. For example, an alveolar stop which assimilates to a following labial or velar consonant often shows acoustic or articulatory evidence of both the alveolar place of articulation, as well as the non-alveolar place of articulation (e.g., Kohler, 1990; Ohala, 1990; Barry, 1992; Holst & Nolan, 1995; Nolan, Holst & Kuehnert, 1996; Gow, 2001; Ellis & Hardcastle, 2002; Gow, 2002; Gow, 2003). In cases where traces of alveolar place of articulation remain, listeners can use the associated perceptual cues to recover the alveolar place of articulation (Gow, 2001, 2002, 2003).

In conclusion, the present research showed that mothers produce clearer (i.e., canonical) consonants somewhat more often when talking to infants than when talking to adults. However, we also showed that from the earliest developmental stages infants largely experience statistical distributions of non-canonical consonantal pronunciation variants that mirror those experienced by adults. These findings inform our understanding of how speech input – and in particular, variant segmental pronunciations and connected speech processes – shapes the development of a child's linguistic system. In this regard these findings help ensure that theories of speech–language development and spoken word recognition will be able to adequately account for children's acquisition of language.

REFERENCES

- Abramson, A. S. & Lisker, L. (1968). Voice timing: cross-language experiments in identification and discrimination. *Journal of the Acoustical Society of America* **44**(1), 377.
- Baran, J. A., Laufer, M. Z. & Daniloff, R. (1977). Phonological contrastivity in conversation: a comparative study of voice onset time. *Journal of Phonetics* **5**, 339–50.
- Bard, E. G. & Anderson, A. H. (1983). The unintelligibility of speech to children. *Journal of Child Language* **10**, 265–92.

- Bard, E. G. & Anderson, A. H. (1994). The unintelligibility of speech to children: effects of referent availability. *Journal of Child Language* **21**, 623–48.
- Barry, M. (1992). Palatalisation, assimilation, and gestural weakening in connected speech. *Speech Communication* **11**, 393–400.
- Bergeson, T. R., Miller, R. J. & McCune, K. (2006). Mothers' speech to hearing-impaired infants and children with cochlear implants. *Infancy* **10**, 221–40.
- Bergeson, T. R. & Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science* **13**, 72–75.
- Bernstein Ratner, N. (1984a). Patterns of vowel modification in mother-child speech. *Journal of Child Language* **11**, 557–78.
- Bernstein Ratner, N. (1984b). Phonological rule usage in mother-child speech. *Journal of Phonetics* **12**, 245–54.
- Bernstein Ratner, N. & Luberoff, A. (1984). Cues to post-vocalic voicing in mother-child speech. *Journal of Phonetics* **12**, 285–89.
- Bernthal, J. E., Bankson, N. W. & Flipsen, P., Jr. (2009). *Articulation and phonological disorders: speech sound disorders in children*. Boston: Allyn & Bacon.
- Bowen, C. (2009). *Children's speech sound disorders*. Chichester: John Wiley & Sons.
- Breen, M., Dilley, L. C., Kraemer, J. & Gibson, E. (2012). Inter-transcriber agreement for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory* **8**, 277–312.
- Browman, C. P. & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics* **18**, 299–320.
- Burnham, D., Kitamura, C. & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science* **296**(5572), 1435.
- Carletta, J. (1996). Assessing agreement on classification tasks: the Kappa statistic. *Computational Linguistics* **22**, 249–54.
- Celce-Murcia, M., Brinton, D. & Goodwin, J. (1996). *Teaching pronunciation: a reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.
- Clayards, M., Tanenhaus, M. K., Aslin, R. & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* **108**(3): 804–809.
- Cristia, A. (2009). Individual variation in infant speech processing: implications for language acquisition theories. Unpublished doctoral dissertation, Purdue University.
- de Boer, B. & Kuhl, P. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online* **4**, 129–34.
- Dilley, L. C. & Pitt, M. (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**, 2340–53.
- 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**, 373–96.
- Englund, K. (2005). Voice onset time in infant directed speech over the first six months. *First Language* **25**, 220–34.
- Englund, K. & Behne, D. (2006). Changes in infant directed speech in the first six months. *Infant and Children Development* **15**, 139–60.
- 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*, 262–82. Cambridge: Cambridge University Press.
- Fernald, A. & Simon, T. (1984). Expanded intonation contours in mother's speech to newborns. *Developmental Psychology* **20**, 104–113.
- Gaskell, G. M. & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* **24**, 380–96.
- Gow, D. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language* **45**, 133–59.

- Gow, D. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance* **28**, 163–79.
- Gow, D. (2003). Feature parsing: feature cue mapping in spoken word recognition. *Perception and Psychophysics* **65**, 575–90.
- Hart, B. & Risley, R. T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes.
- 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*, 315–33. Cambridge: Cambridge University Press.
- Hurtado, N., Marchman, V. A. & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speech and vocabulary size in Spanish-learning children. *Developmental Science* **11**, F31–F9.
- Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (eds.), *Spontaneous speech: data and analysis. Proceedings of the 1st Session of the 10th International Symposium*, 29–54. Tokyo: National International Institute for Japanese Language.
- Kager, R. (1999). *Optimality Theory*. Cambridge: Cambridge University Press.
- Kiesling, S., Dilley, L. C. & Raymond, W. (2006). The Variation in Conversation (ViC) Project: creation of the Buckeye Corpus of Conversational Speech. Department of Psychology. Ohio State University. Available at: www.buckeyecorpus.osu.edu.
- Kohler, K. J. (1990). Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In *Proceedings of the NATO Advanced Study Institute on Speech Production and Speech Modelling*, 69–92, Bonas, France. Dordrecht: Kluwer Academic Publishers.
- Kortmann, B. & Schneider, E. W. (eds.) (2005). *A handbook of varieties of English*. Berlin: Mouton de Gruyter.
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U. & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* **277**, 684–86.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* **255**, 606–608.
- Lahiri, A. & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* **38**, 245–94.
- Landis, J. & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–74.
- Lisker, L. & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word* **20**, 384–422.
- Liu, H.-M., Kuhl, P. K. & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science* **6**, F1–F10.
- 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*, 173–84. New York: Academic Press.
- Maye, J., Werker, J. & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* **82**, B101–B111.
- Nolan, F., Holst, T. & Kuehnert, B. (1996). Modelling [s] to [ʃ] accommodation in English. *Journal of Phonetics* **24**, 113–37.
- Ohala, J. J. (1990). *The phonetics and phonology of aspects of assimilation*. New York, NY: Cambridge University Press.
- Oviatt, S. (1980). The emerging ability to comprehend language: an experimental approach. *Child Development* **51**, 97–106.
- 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*, 269–97. Norwood, NJ: Ablex.

- Rietveld, T. & van Hout, R. (1993). *Statistical techniques for the study of language and language behavior*. Berlin: Mouton de Gruyter.
- Shattuck-Hufnagel, S. & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* **25**, 193–247.
- Shockey, L. (2003). *Sound patterns of spoken English*. Cambridge: Blackwell.
- Shockey, L. & Bond, Z. S. (1980). Phonological processes in speech addressed to children. *Phonetica* **37**, 267–74.
- Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language* **4**, 1–22.
- Sundberg, U. & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica* **56**, 186–99.

APPENDIX

‘LOOK WHAT I FOUND’ BY BRITTNIE AND HEATHER

The sweet pink kitten went for a walk and saw the cool green turtle. The cool green turtle found a little green key. Who did it belong to? The cool green turtle wanted the sweet pink kitten to help in finding who the key belonged to. As they were walking, the sweet pink kitten saw a small green ball. The sweet pink kitten and the cool green turtle were not sure who it belonged to. They picked up the small green ball and the little green key and kept walking. Along the way, the cool green turtle found a pretty blue crystal. Once again, the sweet pink kitten and the cool green turtle wanted to know who the pretty blue crystal belonged to. They picked up the pretty blue crystal along with the little green key and the small green ball and kept walking. Then they saw the cute brown dog. He looked very sad! The cute brown dog said, ‘Oh no! I have lost my little green key, my small green ball and my pretty blue crystal. I dropped them and cannot find them anywhere!’ The sweet pink kitten and the cool green turtle were very happy that they found who the little green key, small green ball and pretty blue crystal belonged to. The cute brown dog wanted his things returned. The sweet pink kitten and the cool green turtle were glad to return them, and this made the cute brown dog very happy.