THE CAP'S OUT OF THE BAG: PLACE ASSIMILATION IS COMMON IN INFANT-DIRECTED SPEECH

Helen Buckler, Huiwen Goy, Julie Kow, Elizabeth K. Johnson

University of Toronto, Department of Psychology
helen.buckler@utoronto.ca, huiwen.goy@utoronto.ca, kowjulie@gmail.com, elizabeth.johnson@utoronto.ca

ABSTRACT

Adult-directed speech (ADS) contains many words that deviate from their canonical form due to connected speech processes such as coronal place assimilation (e.g. 'cat' realized as 'cap' in the phrase 'cat box'). Here, we ask how often this occurs in Infant-directed speech (IDS), which is often believed to be articulated more clearly than ADS. Mothers of 18-month-olds were recorded producing pairs of phrases that did or did not license assimilation (e.g., *cat box*/*cap box*) while addressing either their infant (IDS) or another adult (ADS). Both scripted and unscripted utterances were collected. Key phrases were extracted and presented to adults for identification in a forced choice task, with the expectation that listeners would more accurately identify tokens with less assimilation. Surprisingly, adults identified phrases more accurately in ADS than IDS, suggesting that assimilation occurs frequently in IDS, as it does in ADS, regardless of whether the speech is scripted or not.

Keywords: infant-directed speech, language acquisition, coronal place assimilation

1. INTRODUCTION

In connected speech, it is common for one phoneme to masquerade as another in its acoustic-phonetic realisation, potentially giving rise to lexical ambiguity. Coronal place assimilation in English is a prime example of such a process. Coronal segments, such as /t/, /d/ and /n/ can assimilate to the following labial or velar segment, that is to say, they take on the place of assimilation of the following segment [8]. For example, the word 'green' in the phrase 'green beans' may be realised as 'greem,' or 'cat' in the phrase 'cat box' may be realised as 'cap.' Remarkably, even when the result is lexically ambiguous, these neutralizations do not disrupt communication. A wealth of studies show that adults are adept at rapidly and automatically undoing assimilations, hearing 'green' in the first example, and 'cat' in the second [e.g. 15,17,22]. In contrast to adult listeners, young infants do not possess sophisticated phonological or linguistic knowledge necessary to realise that different tokens with varying characteristics refer to the same underlying word. Acoustic-phonetic variation that arises through coronal place assimilation may pose a problem for infants' comprehension of speech.

Surprisingly little is known about how children learn to use, or compensate for, such variation. Newton and Wells [24], following the speech production of one British English-learning child, attested that assimilations first appeared in his speech at around two-and-a-half years of age, and by his third birthday they were appearing regularly. In terms of comprehension, 7 and 8 year-olds display adult-like compensation for assimilation patterns native to their language [4,21], and even 2 and 3 year-olds have been found to exhibit some compensation abilities [26,27].

In this study we focus on the input that children receive to try and better understand how the ability to cope with assimilation might develop. That is, how children learn to map the acoustic signal with the lexicon, when the signal deviates from the canonical form. Infants glean much of their knowledge about their native language from their caregivers, who adopt a specific register when talking to infants and young children. Indeed, it is often claimed that one of the key functions of Infant-Directed Speech (IDS) is to facilitate language acquisition [e.g. 5,19,28] and experimental evidence suggests that IDS may aid speech segmentation and word learning [25,29,32].

Following the notion that IDS is a teaching tool, it is often believed to be 'clearer' than Adult-Directed Speech (ADS) [14]. Speakers will expand their vowel space [2,7,19,35], increase the VOT contrast in plosives [13], increase vowel duration contrast before word-final obstruents [1], and enhance the sibilant contrast between /s/ and /ʃ/ [9], all in a hypothesised attempt to emphasise phonemic contrasts. In terms of connected speech processes, if IDS is 'clearer' than ADS, then it follows that it would be slower and contain fewer examples of assimilation or other sources of variation, as adults exert more effort in unambiguously signalling phonemic contrasts and prosodic boundaries. These hypotheses have been supported in the literature; compared to ADS, IDS has a slower speech rate [18] and fewer reductions [3].

However, not all data supports the notion that IDS is primarily a learning aid, rife with clear, exaggerated contrasts that make the task easy for the infant. Reports of segmental exaggeration have been widely disputed, and many of the above findings have been contested in different data sets [e.g. 10,12,31]. Furthermore, in a recent paper, it was found that there are an equal number of reductions in IDS as ADS [20].

We tested the hypothesis that IDS should contain fewer instances of regressive place assimilation than ADS. To date, only one study we are aware of has previously addressed this question [11], and they found more canonical pronunciations in IDS than ADS. However, they only analysed read speech (and not unscripted speech), and conducted acoustic classifications rather than perceptual measures.

In the present study, parents of 18-month-old infants were recorded reading a story containing phrases that either did or did not licence place assimilation (e.g. *cat box/cap box*) to either their infant (IDS), or the experimenter (ADS). In addition, they were recorded retelling the story using cue words to both of the listeners. In this way we elicited both IDS and ADS, and scripted and unscripted speech. Subsequently, in a forced choice task, adult listeners were required to identify these phrases when presented in isolation. We predicted that there would be less assimilation in IDS than ADS, and accordingly listeners would be more accurate in identifying tokens in IDS than ADS. Similarly, as speakers tend to be more cautious when reading (to adults), and exhibit more articulatory effort [23,30,34] than when speaking spontaneously, we expected to find fewer assimilations, and higher accuracy, in the scripted than unscripted condition.

2. METHODS

2.1. Recording speech materials

Twelve mothers of 18-month-old children were recorded producing fluent speech to their infant (IDS condition) or the experimenter (ADS condition). Recordings were made on a Zoom Handy H4n digital audio recorder positioned 30cm from the mother's lips. Mothers sat in a quiet room with their child on their lap. In the scripted (or read) condition they were provided with a storybook containing pictures and sentences with two-word phrases that did or did not licence assimilation (e.g. *eight babies/ape babies*). Mothers were instructed to read the story aloud to their child. In the unscripted condition mothers were given a version of the same storybook that only contained key phrases and were

asked to retell the story to their child. In the ADS condition mothers were instructed to read a different book to the experimenter as if they were reading aloud from a newspaper, and then to retell the story to the experimenter using just the key words. There were two different storybooks to avoid practice effects triggered by reading the same story multiple times. Parents were randomly assigned which of the two books they received in the IDS and ADS conditions. Seven pairs of key phrases were elicited across speakers: *ape babies/eight babies; beam picker/bean picker; cap burglar/cat burglar; comb maker/cone maker; grape pies/great pies; Jem Pickles/Jen Pickles; team bears/teen bears.* Key phrases were extracted from the recordings and equalised on rms level. Phrases were excluded that contained disfluencies, reading errors or noise (e.g. from the child).

2.2 Listeners

52 undergraduate students (M=18.8 years; 39 females) participated in the listening task. All had acquired English by the age of 5 and spoke English as their dominant language.

2.3 Listening task procedure

The total stimuli set was halved into two sets (A & B), due to the large number of tokens collected. Half of the listeners heard set A, and half set B. Listeners heard 224 phrases, which consisted of an equal number of ADS and IDS and an equal number of scripted and unscripted speech samples. There were an equal number of samples from each speaker in each set. Listeners were seated in a sound-attenuating booth and individual phrases (e.g. *ape babies*) were presented over closed headphones. Stimulus presentation was controlled using a Praat experimental script [6]. Listeners saw the orthographic representation of two phrases on screen (e.g. *ape babies* or *eight babies*) and were required to identify which version of the phrase they heard. Accuracy, confidence ratings (1-4 where 1 indicated 'not at all sure' and 4 indicated 'very sure') and response time data were collected. Listeners were permitted to repeat each speech sample up to three times. They received no feedback on the accuracy of their response.

3. RESULTS

Individual trials were excluded from analysis if listeners' response time was greater than 10s or they responded before they had heard the complete target phrase. An accuracy score was calculated for each

listener based on their number of correct responses in each of the four conditions, intended addressee (IDS or ADS) and scriptedness (scripted vs. unscripted speech).

Accuracy was analysed in ANOVA with the within-subject factors Intended Addressee (IDS vs. ADS) and Scriptedness (scripted vs. unscripted) and the between-subject factor Stimuli Set (A vs. B). Overall accuracy was low, at just 65%, however, there was a significant, albeit small, main effect of Intended Addressee ($F(1,50)=12.26$, $p<.05$, $\eta^2=.03$); somewhat surprisingly, listeners were more accurate in identifying phrases spoken in ADS ($M=0.66$) than IDS ($M=0.64$). There were no significant effects of Scriptedness or Stimuli Set.

Figure 1: Mean proportion of correctly identified phrases of Adult- and Infant-directed speech in scripted or unscripted speech by adult listeners.
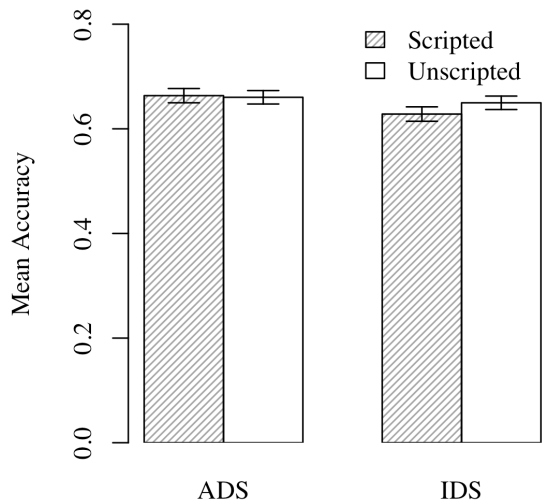

ANOVAs were also conducted using the same within- and between-subject factors, but taking Response Time and Confidence as the dependent variables. With regard to Response Time we find a significant effect of Intended Addressee ($F(1,50)=22.86$, $p<.05$, $\eta^2=.02$); participants were faster to make a decision on ADS ($M=1.83s$) than IDS tokens ($M=1.91s$). With regard to Confidence, we find significant main effects of Intended Addressee ($F(1,50)=14,33$, $p<.05$, $\eta^2=.006$) and Scriptedness ($F(1,50)=5.73$, $p<.05$, $\eta^2=.002$), indicating that participants were more confident in their judgments of ADS than IDS (ADS, $M=3.0$; IDS, $M=2.94$), and unscripted speech than scripted speech (Unscripted, $M=2.99$; Scripted, $M=2.95$). There were no other significant effects or interactions.

We are currently conducting supplementary acoustic analyses to address whether acoustic evidence supports the perceptual data in indicating that assimilation is as common in IDS as ADS. To date, two phonetically trained adults have classified all phrases that license assimilation (e.g. *eight babies*) as Assimilated, Canonical or Other. The definitions of these criteria were based on [11]. We find almost identical distributions of the three categories in each speech register. There were 56 tokens each of IDS and ADS. In ADS 23 tokens were classified as assimilated, compared to 22 in IDS. Furthermore, an equal number of tokens were classified as Canonical in IDS and ADS (31).

Figure 2: Distribution of Assimilated pronunciations in ADS and IDS, based on acoustic classification information. The category "Other" here includes both Canonical and any other pronunciation.
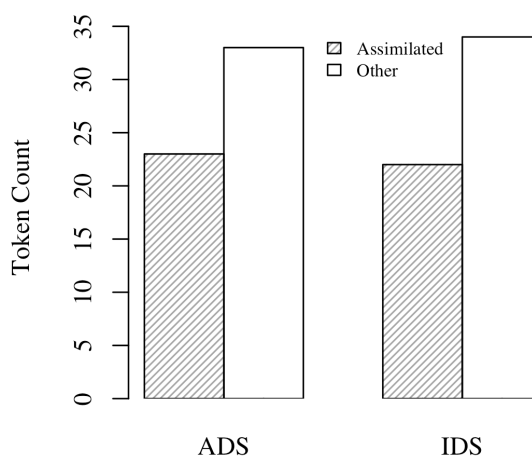
4. DISCUSSION

A prevalent view in the field of language acquisition is that IDS has a pedagogical function; adult speakers increase the clarity of their speech to infants in order to assist them in identifying the crucial elements of their native language. This predicts that IDS should contain fewer connected-speech processes, such as coronal place assimilation, and the underlying target should be easier to identify in decontextualized IDS than ADS. Accordingly, in the present study we predicted that adult listeners would exhibit better performance in identifying phrases in IDS than ADS. Surprisingly, results of all three dependent measures show that adult listeners found it easier to identify the intended utterance in ADS than IDS. They were more accurate, faster and more confident in their judgment of ADS than IDS, however, in all cases the difference was very small. That adults did not perform better with IDS than ADS tokens indicates that assimilations are similarly frequent in both registers, thereby not supporting the hypothesis that adult speakers vastly simplify their speech and use many more canonical forms, in order to support the infant's linguistic development.

In this task adult listeners were used as a proxy for measuring the degree of assimilation in IDS compared to ADS. We are following up this task by investigating whether accuracy and confidence was lower for IDS tokens because adults, specifically university undergraduate students, are not accustomed to listening to this speech register. Increasing their exposure to IDS by extending the length of the target utterances to include whole sentences (and not just the two-word phrase) and blocking trials by register may provide them with sufficient experience to adapt to the speech style and be more confident in their ability to judge IDS.

It could be argued that adult perception data is not as objective as acoustic analyses. However, it is not uncommon in psycholinguistic tasks for adult perception to be taken as the dependent measure of an acoustic contrast in ADS [e.g. 22, 33], or as a measure of relative acoustic salience in IDS or ADS [16]. In some sense listeners can be seen as more reliable than acoustic measures, as it is not always clear which cues listeners base their judgements on, or the weighting they assign to different cues. We supplemented the perceptual task with an acoustic classification analysis of the speech tokens. This analysis corroborated the finding of the adult listening task; neither analysis supports the hypothesis that regressive place assimilation is more frequent in ADS than IDS. Both analyses indicate that assimilations are at least as frequent in both registers. At first glance this seems to speak against the conclusion of Dilley et al. [11], who claim that IDS contains more canonical pronunciations than ADS. However, in their study they included infants across a range of ages. At 1;8, the closest age to our 18-month-olds, they also find equal numbers assimilated pronunciations in IDS and ADS.

In the present task familiarity was added as an inadvertent additional factor. In the ADS condition the addressee was an experimenter, an unfamiliar adult. In the IDS condition it was the adult's own child, a familiar infant. It has previously been found that adults adopt a more formal register when speaking to an unfamiliar adult than to a familiar adult e.g. their spouse or parent [18]. It is possible that the ADS we collected was more careful speech, even when unscripted, containing fewer assimilations than ADS to a familiar adult would. However, this does not detract from our core result that the IDS collected contained many assimilated tokens, as evidenced by listeners' poor ability to accurately identify target phrases in IDS and preliminary acoustic analyses.

In sum, the current study contributes to the debate surrounding the function of IDS in language acquisition, and the lengths that adults will go to when adapting their speech to talk to infants. Our data do not support the hypothesis that IDS is always acoustically simpler than ADS, specifically, we find no evidence that there are fewer examples of coronal place assimilation in IDS than ADS. That is to say, we do not find that adults avoid connected speech processes that would result in realisations that deviate from the canonical form when talking to their infant. We nevertheless believe that IDS has an important role to play in language acquisition, but suggest that (over)simplification or exaggeration of the language in IDS does not always create the optimal situation for the child to learn the complexity of their native language's phonological system.

5. REFERENCES

[1] Bernstein Ratner, N., Luberoff, A. 1984. Cues to post-vocalic voicing in mother-child speech. *Journal of Phonetics.* 12, 285-289.

[2] Bernstein Ratner, N. 1984. Patterns of vowel modification in mother-child speech. *Journal of Child Language.* 11, 557-578.

[3] Bernstein Ratner, N. 1984. Phonological rule usage in mother-child speech. *Journal of Phonetics.* 12, 245-254.

[4] Blomert, L., Mitterer, H., Paffen, C. 2004. In search of the auditory, phonetic and/or phonological problems in dyslexia. *Journal of Speech, Language and Hearing Research.* 47, 1030-1047.

[5] De Boer, B., Kuhl, P. 2003. Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online.* 4, 129-134.

[6] Boersma, P., Weenink, D. 2011. *Praat: Doing Phonetics By Computer.*

[7] Burnham, D., Kitamura, C. What's new, pussycat? On talking to babies and animals. *Science.* 296, 1439.

[8] Chomsky, N., Halle, M. 1968. *The Sound Pattern of English.* New York: Harper and Row.

[9] Cristia, A. 2010. Phonetic enhancement of sibilants n infant-directed speech. *J. Acoust. Soc. Am.* 128, 424-434.

[10] Cristia, A., Seidl, A. 2014. The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language.* 41, 913-935.

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

[12] Englund, K., Behne, D. 2006. Changes in infant-directed speech in the first six months. *Infant and Child Development.* 15, 139-160.

[13] Englund, K. 2005. Voice onset time in infant-directed speech over the first six months. *First Language.* 25, 219-234.

[14] Ferguson, C. 1964. Baby talk in six languages. *American Anthropologist.* 66, 103-114.

[15] Gaskell, M. G., Marslen-Wilson, W. D. 1996. Phonological variation and inference in lexical access. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 144-158.

[16] Golinkoff, R. M., Alioto, A. 1995. Infant-directed speech facilitates lexical learning in adults hearing Chinese: Implications for language acquisition. *Journal of Child Language.* 22, 703-726.

[17] Gow, D. W. 2001. Assimilation and anticipation in continuous spoken word recognition. *J. Mem. Lang.* 45, 133-159.

[18] Johnson, E. K., Lahey, M., Ernestus, M., Cutler, A. 2013. A multimodal corpus of speech to infant and adult listeners. *J. Acoust. Soc. Am.* 134, EL534-540.

[19] Kuhl, P. K., Andruski, J. E., Chistovich, I. A. et al. Cross-language analysis of phonetic units in language addressed to infants. *Science.* 277, 684-686.

[20] Lahey, M., Ernestus, M. 2013. Pronunciation variation in infant-directed speech: phonetic reduction of two highly frequent words. *Language Learning and Development.* 10, 308-327.

[21] Marshall, C., Ramus, F., van der Lely, H. 2011. Do children with dyslexia and/or specific language impairment compensate for place assimilation? Insight into phonological grammar and representations. *Cognitive Neuropsychology.* 27, 563-586.

[22] Marslen-Wilson, W., Nix, A., Gaskell, G. 1995. Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes*, 10, 285-308.

[23] Nakamura, M., Iwano, K., Furui, S. 2008. Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech and Language.* 22, 171-184.

[24] Newton, C., Wells, B. 2002. Between-word junctures in early multi-word speech. *Journal of Child Language.* 29, 275-299.

[25] Singh, L., Nestor, S., Parikh, C., Yull, A. 2009. Influences of infant-directed speech on early word recognition. *Infancy.* 14, 654-666.

[26] Skoruppa, K., Mani, N., Peperkamp, S. 2013. Toddlers' processing of phonological alternations: Early compensation for assimilation in English and French. *Child Development.* 8, 313-330.

[27] Skoruppa, K., Mani, N., Plunkett, K., Cabrol, D., Peperkamp, S. 2013. Early word recognition in sentence context: French and English 24-month-olds' sensitivity to sentence-medial mispronunciations and assimilations. *Infancy.* 18, 1007-1029.

[28] Soderstrom, M. 2007. Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review.* 27, 501-532.

[29] Song, J. Y., Demuth, K., Morgan, J. 2010. Effects of the acoustic properties of infant-directed speech on infant word recognition. *J. Acoust. Soc. Am.* 128, 389-400.

[30] Strik, H., van Doremalen, J., Cucchiarini, C. 2008. Pronunciation reduction: How it relates to speech style, gender and age. *Proc. Interspeech 2008* Brisbane, Australia, 1477-1480.

[31] Sundberg, U., Lacerda, F. 1999. Voice Onset Time in speech to infants and adults. *Phonetica.* 56, 186-199.

[32] Thiessen, E. D., Hill, E., Saffran, J. 2005. Infant-directed speech facilitates word segmentation. *Infancy.* 7, 53-71.

[33] Tuinman, A., Mitterer, H., Cutler, A. 2011. Perception of intrusive /r/ in English by native, cross-language and cross-dialect listeners. *J. Acoust. Soc. Am.* 130, 1643-1652.

[34] Warner, N., Tucker, B. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *J. Acoust. Soc. Am.* 130, 1606-1617.

[35] Xu, N., Burnham, D., Kitamura, C., Vollmer-Conna, U. 2013. Vowel hyperarticulation in parrot-, dog- and infant-directed speech. *Anthrozoos.* 26, 373-380.