Voicing contrast in Nepali infant-directed speech

Sujal Pokharel

MASLP, Mangalore University, 2012

Department of Linguistics

Faculty of Human Sciences

Macquarie University

A thesis submitted in the fulfilment of the requirements for the degree

Master of Research (Linguistics)

10th October, 2016

Resubmitted: 22nd November, 2016

Abstract

Studies investigating voicing in oral stops using voice onset time (VOT) in IDS and ADS, present mixed results regarding the presence of hyper-articulation of VOT in IDS. However, there have been no studies of VOT in IDS in a language with a four-way contrasts of voicing in stops such as Nepali.

Sixteen Nepali speaking mother-infant dyads were recruited. Four target pictures of minimal pair objects, each contrasting in onset consonant voicing (/ɡa.ɖa/ 'bullock cart', /ɡʱa.ʈi/ 'neck', /ka.ʈa/ 'hairpin', and /kʰa.na/ 'food') were selected as targets. Mothers were asked to play with their infant using the target pictures, thereby eliciting IDS. To elicit ADS, mothers were asked to interact with the adult experimenter. Acoustic analysis was then carried out for all word initial stops in the target words that occurred in sentence initial position or in isolation, in both IDS and ADS. Voicing cues were measured as lead time and lag time. In addition, the occurrence of devoicing (complete absence of lead time) in voiced consonants was recorded as a binary variable. Vowel duration was measured to control for speaking rate. The aim of the current study was to test two hypotheses. The first hypothesis was that there would be hyper-articulation of devoicing, lead time, and lag time in IDS compared to ADS. The second hypothesis was that the hyper-articulation might be a side-effect of speaking rate differences.

The results showed the absence of hyper-articulation of devoicing, lead time, and lag time in IDS. Rather, the higher rate of devoicing and shorter lag time led to poorer voicing contrasts between the stop categories in IDS compared to ADS. Further, this difference between the categories could not be explained by speaking rate differences between the registers. The longer vowels in IDS suggest that mothers focus more on vowels while talking to infants, and this appears to have the effect of shortening the onset consonants.

To our knowledge, this is the first time that the voicing contrast in IDS is being looked at this way. This then has wide ranging theoretical implications for better understanding the nature of IDS and its effects on language learning

Declaration

I hereby declare that this thesis has not been submitted for a higher degree to any other university or institution. I have made every effort to clearly indicate the sources of information used and acknowledge the extent to which the work of others has been used in the text. The research presented in this thesis has been approved by the Macquarie University Faculty of Human Sciences Research Ethics Sub-Committee (ref: 5201600078).

Acknowledgement

This thesis would not have been possible without the help from number of people. First of all, I would like to express my profound gratitude to my supervisors, Professor Katherine Demuth and Dr Titia Benders, for their support, encouragement and positive feedback throughout this research project. This thesis would not have been possible without their precious inputs and kind help. I would also like to thank Dr. Elaine Schmidt, for her assistance to set up this project.

I would like to express my gratitude to Dr. Ivan Yuen, for all his assistance throughout this project, especially with acoustic analysis. I would also like to thank all Child Language Lab and Phonetics Lab members for very helpful comments and suggestions. I am highly grateful to members of writing group for their feedback and suggestions on my writing. I would also like to thank Amy Hanna, Louise Ratko, and Robina Anne Day for their help on writing. Especial thanks to Laurence Brooks for her comments and encouragements.

This project would not be complete without the participants. I would like to thank all my participants. Especial thanks to Mrs. Mira Neupane for her assistance during recruitment.

I would like to thank all my friends and colleagues here at Macquarie. Special thanks to Bec (Rebecca Holt) for her very helpful comments, encouragement and patience. I am thankful to Ping (Ping Tang) for all the moral support and his assistance with scripts and data analysis.

Finally, I would like to thank my family, especial thanks to Mom, Dad, Uncle (Ani) and Aunty (Minu) without whose support and encouragement none of this would have been possible at first place.

Table of Contents

1. Introduction

Studies of infant-directed speech (IDS) often report hyper-articulation of phoneme contrasts compared to adult-directed speech (ADS) (Kuhl et al., 1997). This hyper-articulation is suggested to play an important role in infant language acquisition (Kuhl et al., 1997; Soderstrom, 2007). However, many of these studies have focused on vowels.

Studies investigating oral stops in IDS and ADS have analysed voice onset time (VOT), a significant acoustic correlate for voicing contrasts in word-initial stops (Lisker and Abramsons, 1964). These studies present mixed results regarding the presence of hyper-articulation of VOT in IDS. Some report a hyper-articulation in VOT in IDS compared to ADS, with IDS VOT being longer in duration. This is thought to assist infants in learning voicing contrasts (Malsheen, 1980; Sundberg, 2001; Englund, 2006; Burnham et al., 2013). However, other studies report no hyper-articulation of VOT in IDS compared to ADS (Synnestvedt et al.; Narayan & Yoon, 2011; McMurray et al., 2013). Furthermore, the study from McMurray et al. (2013) suggested the hyper-articulation observed in IDS may be due to a secondary effect of lower speaking rate in IDS leading to longer VOT.

However, when previous studies found prevoiced tokens, these were treated as outliers, since these languages do not have obligatory prevoicing in stops (Englund, 2006; Synnestvedt et al., 2010; Burnham et al., 2013). This may have biased the findings, particularly if voicing hyper-articulation in IDS is realized through prevoicing.

To date, however, there have been no studies investigating VOT in IDS in a language which has a phonemic distinction between prevoicing (negative VOT) and voicing lag (positive VOT). Additionally, there have been no studies of VOT in IDS in a language which has a four-way contrast in voicing, such as that found in Nepali. It therefore remains unclear how this type of VOT contrast may be realized in IDS. For this reason, the current study investigats IDS in Nepali, a language with a four-way contrast of voicing in oral stops: prevoicing (voiced unaspirated and aspirated) and voicing lag (voiceless unaspirated and aspirated).

In the current study, instead of using VOT, voicing cues will be measured as lead time and lag time, similar to what Davis (1994) used for Hindi stops. The realization of lead time and lag time of all word initial stops in target words in sentence initial position or in isolation will be analysed to investigate two hypotheses. The first is the *hyper-articulation hypothesis*, which predicts the lead time and lag time differences in stop categories in IDS will be larger than in ADS. The second is the *speaking rate hypothesis*, which predicts longer lead time and lag time in the register with a slower speaking rate (IDS).

The current thesis is organized as follows: the review of literature is presented first (section 2), leading to the hypotheses (section 2.4). Secondly, section 3 describes the methodology employed in the study, followed by the results (section 4), discussion (section 5) and conclusion (section 6).

2. **Literature review**

2.1 Language experience in infants

The language experience of infants in most cultures is not the same as adults, as adults in most cultures are reported to modify their speech while talking to infants (Ferguson, 1977). This modified register is said to serve three possible functions: communicating affect, eliciting attention, and language learning (Fernald and Simon, 1984; for detail see Soderstrum, 2007). This special register is known by various terms such as motherese, baby talk, child-directed speech, or infant-directed speech. Hereafter, we will refer to this special register as infant-directed speech (IDS).

IDS is reported to have simple syntactic structure and to be lexically simpler, with the use of more affect words (for detail see Ferguson et al., 1977). In terms of acoustics, it is also reported to show differences at both prosodic (Fernald et al., 1989) and segmental (vowels and consonants) levels of structure (Kuhl et al., 1997; Sundberg & Lacerda, 1999) compared to adult-directed speech (ADS). The following section will first review the literature on the prosodic characteristics of IDS, and then the vowels and then the consonants in IDS.

2.1.1 Prosodic characteristics of IDS

IDS typically has higher pitch, a wider pitch range, slower rate of speech, and longer pauses compared to ADS (Fernald et al., 1989). The prosodic aspects of IDS are observed cross-linguistically (Fernald et al., 1989) and are considered as one of the most salient differences between the two registers (Fernald & Simon, 1984).

It has been suggested that the exaggerated prosody in IDS serves various functions. It is thought that it helps in communicating affect, as high pitch and wider pitch range conveys positive affect (Fernald & Simon, 1984). The prosodic changes in IDS are also shown to assist in the elicitation and maintenance of attention (Fernald & Simon, 1984). It is used as a feature to regulate the arousal level of the infants (Fernald & Simon, 1984). Furthermore, it suggested to play an important role in language learning as the exaggerated prosodic boundaries may assist in segmentation of continuous speech stream (Fernald & Simon, 1984).

2.1.2 Vowels in IDS

The studies investigating vowels have usually used vowel space as a measure to show the difference between the two registers (Kuhl et al., 1997; Benders, 2013). For the vowel space measurements the first formant (F1) and second formant (F2) frequencies of the three cardinal vowels (/a/, /i/, and /u/) are typically measured. Then the F1 and F2 of vowels are plotted in a graph on the X and Y axes respectively to produce an F1× F2 vowel space.

Cross-linguistically, vowel space in IDS is reported to be larger in comparison to ADS (Kuhl et al., 1997). The increased vowel space is suggested to increase the contrast between the vowel categories. This cross-linguistic phenomenon, referred as hyper-articulation, was suggested to be universal feature of IDS (Kuhl et al., 1997). Kuhl et al. (1997) reported from their cross linguistic study of point vowels (/a/, /i/, and /u/) in American English, Russian, and Swedish a hyper-articulation of vowels in IDS compared to ADS. They found that, across all languages, vowels in IDS showed exaggerated formant values, leading to larger contrast between vowels in IDS compared to ADS.

The hyper-articulation in IDS is suggested to play an important role in infants' learning of language (Kuhl et al., 1997), where the increased separation of vowel categories in IDS allows the speaker to produce more variability in vowels with less risk of overlap between the categories. Thus, hyper-articulation makes it is easier for the learner to learn these vowel categories, as the smaller the degree of overlap, the greater the contrast between the vowel categories.

This idea of hyper-articulation and language learning facilitation is further supported by studies such as Liu et al. (2003) who associated hyper-articulation in IDS with better speech discrimination in infants. Song, Demuth, and Morgan (2010) also have suggested that vowel hyper-articulation may facilitate word learning.

However, the idea that hyper-articulation is a universal feature of IDS is contested. Studies of Dutch vowels (Benders, 2013), Norwegian vowels (Englund & Behne, 2005), and Japanese vowels (Martin et al., 2015) have reported no hyper-articulation in IDS compared to ADS. Further, Dutch vowels show greater separation in ADS than IDS. In addition, Cristià & Seidl (2014) suggest that even within a language, hyper-articulation may be a phoneme specific phenomenon.

The absence of hyper-articulation and reports of hypo-articulation of contrasts in IDS in some studies raises questions about the importance of hyper-articulation for language learning. Furthermore, there are suggestions that hyper-articulation is an unintended side-effect of smiling when talking to infants (Englund & Behne, 2005). Englund and Behne (2005) suggested that smiling raises the formant frequencies, giving the impression of hyper-articulation in vowels.

In summary, there is evidence suggesting hyper-articulation of vowels in IDS (Kuhl et al, 1997). However, there is also contradicting evidence suggesting the absence of vowel hyper-articulation in IDS (Englund & Behne, 2005; Benders, 2013). Additionally, Benders (2013) reported vowels in ADS show enhanced contrasts in comparison to IDS. Hence, it is unclear whether or not acoustic contrasts between vowels in IDS are universally enhanced.

2.1.3 Consonants in IDS

Unlike vowels, studies comparing the acoustic phonetic features of consonants between IDS and ADS have received less attention. Sibilant and plosive consonant contrasts have been studied. Readers interested in sibilant contrasts are referred to Cristià (2010). In this paper we focus on studies investigating stops (specifically oral stops) which have compared voicing contrasts between the two registers using voice onset time (VOT).

Voice onset time is the durational difference between the initiation of laryngeal voicing and the release of articulatory closure (Likser & Abramson, 1964). VOT is a well-established acoustic correlate which differentiates word initial stops of various languages such as English, Spanish, and Cantonese with a two-way contrast, and Thai and Armenian with a three-way contrast, in both production (Lisker & Abramson, 1964) and perception (Abramson & Lisker, 1970).

An example of VOT of word initial velar stops in English is shown in Figure 1. As shown in Figure 1, if the laryngeal vibration starts before the articulatory release (left side of 0ms point), the VOT value is negative: this is also referred as *lead time* or *pre-voicing*. If the laryngeal vibration starts after the articulatory release (right side of 0ms point), the VOT value is positive: this is also referred to as *lag time*. In the case of English, both the phonologically voiced stops /b, d, g/ and the phonologically voiceless stops /p, t, k/ have positive VOT, meaning the laryngeal vibration starts after the articulatory release. In addition, the difference in VOT of voiced vs. voiceless stops is approximately 30ms.

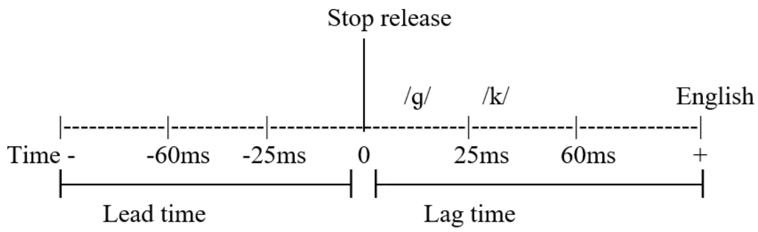


Figure 1: VOT of word initial velar stops in English (Lisker & Abramsons, 1964). The horizontal axis represents the time in milliseconds (ms) which ranges from negative to positive. The 0ms point refers to the point of onset of articulatory release of a stop consonant. Negative VOT, onset before 0ms point, is referred as lead time. Likewise, positive VOT, onset after 0ms point, is referred as lag time.

2.1.3.1 VOT studies in IDS

Studies investigating VOT in IDS compared to ADS present mixed results. These mixed results can be classified into three groups, with one reporting hyper-articulation of VOT in IDS (Malsheen, 1980; Sundberg, 2001; Englund; 2005), the second reporting no hyper-articulation of VOT in IDS (McMurray et al., 2013), and the third reporting reduced VOT contrasts in IDS to compared to ADS (Narayan & Yoon, 2008; Synnestvedt et al., 2010). However, except for one study with a three-way contrast (Korean- Narayan & Yoon, 2011), all the previous studies were based in two-way contrast languages (American English- Malsheen, 1980; Burnham et al., 2013; McMurray et al., 2013; Synnestvedt et al., 2010; Swedish- Sundberg and Lacerda, 1997; Norwegian- Englund, 2005). It is not known how voicing contrasts manifest in a language with a four-way contrasts.

Studies from Norwegian (Englund, 2005), Swedish (Sundberg and Lacerda, 1997; Sundberg, 2001) and English (Malsheen, 1980; Burnham et al., 2013) have supported the idea that VOT contrasts in IDS are hyper-articulated compared to ADS. This means that the two-way voicing of stops in these languages shows enhanced contrast of VOT between the stop categories in IDS compared to ADS. However, studies investigating VOT in IDS compared to ADS in American English (McMurray et al., 2013) suggest the absence of hyper-articulation in IDS. Furthermore, the Narayan & Yoon (2011) study on Korean stops and the Synnestvedt et al. (2010) study on American English stops have shown hypo-articulation of VOT in IDS compared to ADS.

Each study has their own explanation regarding the presence or absence of hyper-articulation in their findings. The studies explaining hyper-articulation base their assumptions on the effect of infant's age on IDS (Malsheen, 1980; Sundberg & Lacerda, 1999). In contrast, the studies suggesting no hyper-articulation in IDS suggest only a prosodic difference between the two registers (McMurray et al., 2013). These explanations are discussed further in the following section.

2.1.3.1.1 Hyper articulation of VOT

The first explanation for hyper-articulation of voicing contrasts was proposed by Malsheen (1980). Malsheen (1980) suggested that the hyper-articulation of VOT in IDS is due to the mother's intention of teaching language to their infants. This explanation derives from her cross-sectional study comparing VOT in word initial stops in mothers' speech addressed to infants with mothers' speech addressed to an adult. She reported that among the three groups of mother-infant dyads (6-8 months; 15-16 months; 2.5-5.2 years), only the mothers talking to their infant aged 15-16 months showed hyper-articulation of VOT.

Malsheen (1980) proposed that mothers tailor their speech to facilitate the learning of voicing contrasts. Thus, hyper-articulation is observed in IDS only to infants who are expected to be learning such contrasts (15-16 months) and not in the IDS to pre-linguistic infants (6-8 months), who are not expected to learn voicing contrasts, and not to older children (2.5-5.2 years) who have already acquired voicing contrasts.

However, a recent study in American English investigating the difference in VOT in IDS and ADS has reported an absence of correlation between VOT in IDS and mothers' linguistic expectations (Synnestvedt et al., 2010). This longitudinal study compared the VOT of 15 American English-speaking mother-infant dyads with infants aged 7.5-11 months in IDS and ADS. They reported longer VOT only for voiced stops in IDS throughout this period. Furthermore, they reported no correlation between the VOT in IDS and mother's linguistic expectations of her infant, measured using MacArthur-Bates Communicative Development Inventories (MCDI). In conclusion, the study reported reduced VOT contrasts in IDS due to the longer VOT lag time of voiced stops, resulting in an increased overlap between the voicing categories.

In contrast, Sundberg and Lacerda (1999) and Sundberg (2001) proposed that the acoustic phonetic features of IDS, such as vocalic features (e.g. formants of vowels), consonantal features (e.g. VOT of oral stops), and prosodic features (e.g. pitch) are related to the communicative intention of the mothers, which changes with the age of the infant. Thus, Sundberg and Lacerda (1999) reported the absence of hyper-articulation in six Swedish speaking mothers' speech to their 3-month-old infants compared to speech to an adult. In contrast, Sundberg (2001) reported hyper-articulation of VOT for both voiced and voiceless stops in the speech of six Swedish mother-infant dyads of infants aged 11-14 months compared to ADS. The authors argued that this difference in VOT in IDS across the two age groups was due to mothers' adaptation of speech depending on her communicative intent. While talking to pre-verbal infants (3-month olds), mothers intend to communicate affect and direct the infant's attention and arousal. Thus, the phonetic features of IDS addressed to young infants shows exaggerated prosodic characteristics such as higher pitch and wider pitch range associated with affect. Mothers communicating with pre-verbal infants have no intention of teaching language. This may have resulted in lower differentiation of VOT. However, as the age of the infant increases in age, the mothers' communicative intent gradually focuses on language teaching. This change towards language teaching is reflected iin the hyper-articulation of voicing contrasts in the IDS of older infants (11-14 months) to facilitate the learning of voicing contrasts.

However, Englund (2005) conducted a longitudinal study investigating the word initial stops of six Norwegian mothers' interacting with their infants and adults throughout the first six months of the infants' lives. She reported that hyper-articulation of VOT in IDS was observed throughout the study. Furthermore, she reported that the observed difference in VOT between the two registers was independent of speaking rate differences between IDS and ADS.

2.1.3.1.2 No difference in VOT contrasts

However, McMurray et al. (2013) suggested that the hyper-articulation observed in IDS compared to ADS may be due to the speaking rate difference between the two registers. They examined the VOT of 18 mother-infant dyads (9-13 months) of American English-speaking mothers reading to their infants compared to mothers reading to an adult. They reported hyper-articulation, which was due to longer VOT for the voiceless stops in IDS than in ADS. However, they also reported that the observed hyper-articulation was due to the slower speaking rate in IDS. This slower speaking rate led to an increased VOT of stops, giving an impression of hyper-articulation. Thus, when the two registers were normalized for speaking rate, there was no hyper-articulation effect. Based on this evidence, McMurray et al. (2013) argued that the reported hyper-articulation may be due to the speaking rate differences between the two registers. However, this explanation fails to account for the longer VOT observed in Norwegian mothers' speech to their infants, where there was no difference of speaking rate when mothers were speaking to their infants compared to an adult (Englund, 2005).

In conclusion, there is disagreement regarding the presence of hyper-articulation of voicing contrasts in IDS. Studies from Sundberg and Lacerda (1999) and Malsheen (1980) suggest that voicing contrasts in IDS are hyper-articulated. However, the study from McMurray (2013) suggests that the difference in speaking rate between the two registers can account for the difference in voicing contrasts. This highlights the need for further research to determine the factors that give rise to these conflicting results.

Except for Korean (Narayan & Yoon, 2011), which has a three-way stop voicing contrast, all the previous studies of voicing contrasts in IDS compared to ADS have been carried out in languages with only a two-way contrast (American English- Burnham et al. 2013; McMurray et al. 2013; Synnestvedt et al., 2010; Malsheen, 1980; Norwegian- Englund, 2005; Swedish- Sundberg & Lacerda, 1999; Sundberg, 2001). None of the studies have reported the voicing contrasts in IDS in a language with a four-way voicing contrasts in stops, such as Nepali. Furthermore, when previous studies found prevoiced tokens, they were treated as outliers, since these languages do not have obligatory prevoicing in stops (Malsheen, 1980; Englund, 2006; Synnestvedt et al., 2010; Burnham et al., 2013). If it was the case that voicing exaggeration in IDS is realized through prevoicing, then the exclusion of prevoiced tokens may lead to a misrepresentation of the effect, biasing the results. To date, however, there have been no studies of VOT in IDS in a language which has a phonemic contrast between prevoicing (negative VOT) and voicing lag (positive VOT). It therefore remains unclear how this type of voicing contrast may be realized in IDS. For this reason, the current study investigates IDS in Nepali, a language with a four-way contrast of voicing in oral stops: prevoicing (voiced unaspirated and aspirated) and voicing lag (voiceless unaspirated and aspirated).

In languages with a four- way voicing contrasts such as Hindi (Lisker & Abramson, 1964), Bengali (Lisker & Abramson, 1964), and Nepali (Poon & Mateer, 1985) (see Figure 3), VOT does not contrast the four categories. The reports from these studies have shown that VOT shows a clear contrast with distinct VOT values for three categories (voiced unaspirated, voiceless unaspirated, and voiceless aspirated). However, the VOT of voiced aspirated stops largely overlaps with the VOT of voiced unaspirated stops in Hindi and Bengali (Lisker & Abramson, 1964), and overlaps with both voiced unaspirated and voiceless unaspirated stops in Nepali (Poon and Mateer, 1985). The following section presents a brief overview of the Nepali language, with implication in research design.

2.2 Nepali

Nepali is the lingua franca of Nepal. There are 42 million speakers of Nepali word-wide ("Nepal-Ethnologue"-2016). Of the total population of Nepal, 44.6% speak Nepali as a mother tongue, and the rest of the population speaks it as their second language (National census (2011), c.f. statistical year book of Nepali (2013)). The following section will present a brief descriptions of Nepali phonology and morphology (for details see Pokharel (1989); Acharya (1991); Manders (2007)).

2.2.1 Phonology

2.2.1.1 Consonants

Nepali, similar to other Indo-Aryan languages such as Hindi and Bengali, has a series of stops and affricates (Table 1) which show a contrast in both aspiration (extent of glottal opening) and voicing (laryngeal vibration) (/dʱam/ 'pilgrimage' vs. /dam/ 'money' vs. /tʰam/ 'stop' vs. /tak/ 'button').

Table 1: Consonants in Nepali (Khatiwada, 2009).

	Bilabial	Dental	Alveolar	Retroflex	Palatal	Velar	Glottal
Stops	p b pʰ bʱ	t d tʰ dʱ		ʈ ɖ ʈʰ ɖʱ		k g kʰ gʱ	
Affricate			ts dz tsʰ dzʱ				
Nasals	m	n				ŋ	
Tap or Flaps			r				
Fricative			s				ɦ
Lateral			l				
Approximant	(w)				(j)		

2.2.1.1.1 Voicing contrasts in Nepali

Voicing contrasts, measured for all word initial stops using VOT, show contrast only for three stop categories (Poon and Mateer, 1985). Poon and Mateer (1985) investigated VOT in word initial stops of ten adult male speaker of Nepali from a repeated elicitation of citation words. The study reported that voiced aspirated stops had longest negative VOT (-67.55ms), whereas the longest positive VOT was reported for voiceless aspirated stops (83.56ms). The mean VOT for both voiced unaspirated (24.47ms) and voiceless unaspirated stops (26.79ms) were positive (Figure 2). Note the resulting partial VOT overlap between voiced aspirated stops and the two voiceless stop categories. In addition, the study reported devoicing, i.e. the absence of (optional) pre-voicing, in voiced aspirated stops. Thus, both prevoicing (negative VOT or lead time) and aspiration (positive VOT or lag time) need to be measured to provide a complete picture of the acoustic characteristics of voiced aspirated stops.

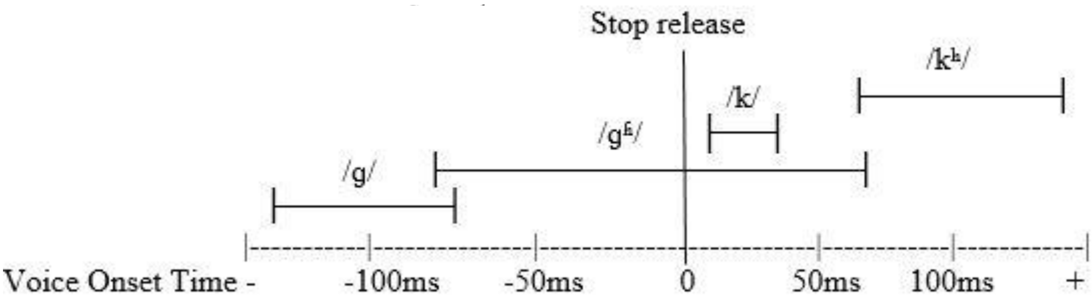


Figure 2: Mean VOT values (ms) for Nepali word initial velar stops (Poon and Mateer, 1985): ‘kh’- voiceless aspirated stop, ‘k’- voiceless unaspirated stop, ‘gh’- voiced aspirated stop, ‘g’- voiced unaspirated stop.

As VOT is a one-dimensional measure it either measures the prevoicing or the aspiration of a stop, and can never characterize both simultaneously. Thus, in cases of voiced aspirated stops, which show both prevoicing and aspiration, VOT fails to capture both aspects. Davis (1994) therefore measured voicing cues in Hindi stops using the two independent duration cues of *lead time* and *lag time*, which measure prevoicing and aspiration, respectively. The present study will also utilise lead and lag time instead of VOT to measure voicing in Nepali stops.

2.2.1.1.2 Other phonological features of Nepali

There are 11 contrastive vowels in Nepali (/i/, /u/, /e/, /^/, /o/, /a/, /ĩ/, /ẽ/, /ʌ̃/, /ũ/, /ã/). Syllable structure of Nepali is (C) (G) V (G) (C) (Acharya, 1991), where “C” stands for consonants, “V” for vowels (monothongs and dipthongs) and “G” for glides. Stress in Nepali is non-distinctive. There is fixed stress on the word initial syllable (Acharya, 1991).

2.2.2 Morpho-syntax

Nepali is a morphologically rich language. Various grammatical morphemes are used in Nepali to indicate properties such as case, gender and number on nouns and verbs (for details see Chen, 2014, Acharya (1991)). Nepali can be also considered a ‘free word’ order language. Due to case marking, all word orders such as SOV, SVO, VOS, VSO, OSV, and OVS (where “S” refers to subject, “O” refers to object and “V” refers to verb) are allowed

(Chen, 2014). Case marking is usually shown by addition of a suffix to the noun. For example:

hari-le dzants^{h}ʌ gʌr-∅
Hari-ERG[1] go-1SG.NPST[2] house-NOM[3]
'Hari eats rice'

2.3 Role of linguistic experience in language learning

Infants are thought to learn language from their language experience (Kuhl et al., 2006; Maye et al. 2002). There are studies suggesting that mere exposure to input may be adequate for learning (Maye et al., 2002). Other studies suggested that not only the exposure, but the social interaction between the infant and adult plays a role in infants' language learning (Kuhl, Tsao, and Liu, 2003). Kuhl, Tsao, and Liu (2003) exemplified the role of social interaction in infants' language development. They investigated phonetic perception in infants learning English and Mandarin between six and twelve-months. The findings suggested that English-learning infants exposed to Mandarin in an interactive session could perceive the Mandarin contrasts (an alveolo-palatal affricate /tɕ^h / vs an alveolo-palatal fricative /ɕ/). However, the English-learning infants exposed in a non-interactive session for the similar duration performed no better than the infants who were not exposed to Mandarin. Therefore the study suggests that IDS, due to its interactive nature, plays a vital role in language learning.

2.3.1 Language experience and acquisition of voicing contrasts

Studies suggest that 10-11-month-old infants may use VOT cues to discriminate between the voicing contrasts in stops in their native language (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Furthermore, laboratory learning experiments suggests that infants are sensitive to the distribution of VOT in the language input they receive (Maye et al., 2002). This suggests that it is vital to investigate how these voicing contrasts are represented in the input to the infants, and the implications this may have for learning voicing contrasts. A study of Hindi voicing acquisition reported that, even by six years of age, children could not produce all the four-way voicing contrasts (Davis, 1995). This may indicate that children learning Nepali may face similar difficulty in learning Nepali voicing contrasts. Given that infants are sensitive to voicing cues in the input (Maye et al., 2002), this suggests that infants learn voicing contrasts

[1] ERG-ergative case
[2] SG- singular and NPST- not a past tense
[3] NOM- nominative case

in their language based on the distribution of voicing cues in input (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). It is therefore of significant interest to look at infants' input in Nepali, as it may inform us about the acquisition of Nepali voicing contrasts.

2.4 Hypothesis of the current study

The aim of the current study was therefore to investigate the realization of four-way voicing contrasts of Nepali oral stops in IDS compared to ADS. Voicing cues were measured as lead time and lag time. In addition to voicing cues, the occurrence of devoicing in voiced stops and the duration of vowel following the target stop was measured. These measures were then used to test the following two hypotheses.

1. The *hyper-articulation hypothesis* predicts that voicing contrasts in IDS will be enhanced compared to ADS. This hypothesis has the following predictions for each measure of devoicing, lead time, and lag time.

The occurrence of devoicing will be lower in IDS, since the presence of lead time signals the contrast between voiced and voiceless stops. If lead-time is hyper-articulated in IDS, we should observe less devoicing of voiced stops in IDS than in ADS.

Lead time will be longer in IDS. Additionally, the lead time contrast between the voiced aspirated and unaspirated stops should be larger in IDS. Again, as the presence of lead time signals the difference between voiced and voiceless stops, if IDS enhances these contrasts this should lead to longer lead time and a larger contrast between voiced consonants in IDS compared to ADS.

Lag time contrasts in IDS will be larger for voiced and voiceless and aspirated and unaspirated stops. Lag time is the other cue that signals the contrast between voiced and voiceless stops as well as aspirated and unaspirated stops. Hyper-articulated IDS will have larger difference of lag time between the voiced and voiceless as well as between the aspirated and unaspirated stops in IDS compared to ADS.

2. The *speaking rate hypothesis* predicts no enhancement of voicing contrasts in IDS compared with ADS conditions. This hypothesis attributes the longer lead time and lag time, which may give an impression of hyper-articulated voicing contrast, to slower rate of speech. Thus, according to this hypothesis, when the two registers are normalized for speaking rate, there will be no difference in lead time and lag time between the registers.

If this hypothesis is true, we should observe that the register with longer vowel duration (attributed to lower speaking rate) will have longer lead time as well as lag time. The difference between the registers for lead time and lag time will disappear when normalized by vowel duration.

3. Method

In this study, 16 mother-infant and mother-adult dyads were recorded. At first, mothers were familiarized with the selected target pictures. Later, mothers were asked to play with their infant using pre-selected target pictures to elicit IDS. Finally, to elicit ADS, mothers were asked to interact with the adult-experimenter. All word initial stops in target words in sentence initial position or isolation in both registers (IDS and ADS) were analysed for lead time, lag time, and vowel duration.

3.1 Participants

Participants included 16 mother-infant dyads. Mothers who grew up as monolingual eastern dialect Nepali speakers (Acharya, 1991) were selected. The age of the infants ranged from 10 to 18 months (M= 14.1 months; SD= 2.4 months; 8 males and 7 females). Mothers ranged from 26 to 34 years (M= 30 years; SD= 2.8 years) (see Appendix A for details of participants). An additional six mother-infant dyads were excluded from analysis. The reasons for exclusion were as follows: mother with different mother tongue (n=1), mother who grew up as simultaneous bilingual (n=1), mother with premature infant (n=1), and dyad did not yield enough target tokens (i.e. fewer than five repetitions) at utterance initial position for data analysis (n=3).

Participants were recruited via advertisement in Sunday community schools across the Sydney area and on the Facebook page of the Macquarie University Nepali Student Association. Interested participants contacted the experimenter through the email address advertised.

3.2 Design

3.2.1 Stimuli and other materials

Target words were four disyllabic CV.CV words of child-friendly picturable object nouns (see Table 2 for details). Each target word had a velar stop at onset position followed by low back vowel /a/ in the first syllable. The onset of second syllable was never an approximant consonant. In addition, there were two other words /gi.dza/ (gums) and /gu.p^{h}a/ (cave) with different vowels which were not of interest to this study.

Table 2: List of target words.

	Unaspirated	Aspirated
Voiced	/ga.ɖa/ (bullock cart)	/gʱa.ʈi/ (neck)
Voiceless	/ka.ʈa/ (hair pin)	/ kʰa.na/ (meal)

Pictures were used in order to facilitate spontaneous elicitation of target words across subjects while simultaneously limit the possible word choices available. The provision of near minimal pairs of picturable object nouns and the allophonic variation in Nepali stops limited the choice of target words to words with initial velar stops with following vowel /a/. Although Nepali has a four-way voicing contrast in oral stops in bilabial, alveolar, retroflex, and velar place of articulation, only words with initial velar stops can yield a near minimal pair with child-friendly pictures. In addition, in bilabial position there is free variation between /pʰ/ (bilabial voiceless aspirated stops) and /f/ (labiodental voiceless fricative) (Pokharel, 1989). Hence, VOT of /pʰ/ cannot reliably be measured as these are not always realised as stops.

An approximant was avoided in the onset position of the second syllable of target words to facilitate segmentation of the vowels' acoustic cues. This is because in spectrogram analysis, the formants and the amplitude of vowels show minor changes for the initiation of approximants (Ladefoged & Maddieson, 1996). This suggests an increased difficulty in the identification of a vowel boundary preceding an approximant. Hence, in order to facilitate the extraction of acoustic information of vowels in the spectrogram, approximant consonants were avoided in the onset position of the second syllable.

Three different pictures for each target word were selected (see Appendix B for the pictures). In total, eighteen pictures (i.e. six words including two filler words x three pictures each) were used. The use of different pictures for each word aimed to make it more interesting for the participants in order to elicit more repetitions.

Pictures were selected based on the ratings of a series of pictures by ten adult Nepali speakers. Each picture was rated for the representativeness of the target word on a seven-point rating scale where a score of one referred to ‘does not represent the target word’ and a score of seven referred to ‘exactly represents the target word’. In each category of target word, the three pictures with the highest score were selected. In addition, a questionnaire was used to collect information of participants such as demographic data, language proficiency, and language use history (see Appendix C).

3.2.2 Procedure

3.2.2.1 Testing room and Equipment

All mother-infant dyads were recorded in the Macquarie University Child Speech Production Lab. To help reduce participants’ possible anxiety from being in an unfamiliar place, items typically found in Nepali living rooms, such as pictures, rugs, and table cloths were used to alter the recording room ambience. The entire recording was carried out using a head-set microphone (AKG-C520) worn by the mother. The microphone was connected to a recorder (Marantz Professional-PMD661 MKII) which recorded an uncompressed WAV file with a 44.1 kHz sampling rate and 16-bit quantization.

3.2.2.2 Familiarization

Initially, mothers were familiarized with the pictures being used. This was done to facilitate the production of the target word instead of other synonyms, picture descriptions, or onomatopoeic words. Put differently, the familiarisation aimed to ensure that the same words were used across speakers. During familiarisation, only the experimenter and the mother were present in the recording room. For this task, all the pictures were placed in a bag, and were picked one at time for naming. The mother was asked to name each of the eighteen pictures presented one at a time (see Appendix D for details of instructions). The order of presentation of pictures was not controlled. Feedback was then provided regarding the selected name for each picture. This task was recorded and the mother’s production of target tokens were used in analysis of adult-directed speech.

3.2.2.3 Infant-directed speech

The familiarisation was followed by elicitation of infant-directed speech, in which mothers were asked to play with their infants using the target pictures. For this task, the target pictures were divided into three bags, each contained one picture from each word (i.e. one bag had six pictures). Bags were used to assist in randomizing the order of presentation of target words. Mothers were instructed to interact as naturally as they would do at home (see Appendix D for details of instructions). Mothers were allowed to use any bag and any picture for the interaction. They were not informed that their speech will be analysed in order to reduce mother's focus on her speech. During this task, only mother and infant were present in the recording room.

3.2.2.4 Adult directed session

For the third task, the experimenter interacted with mothers to elicit adult directed speech (see Appendix D for details of instructions). The experimenter interacted with the mother using topics and questions which would facilitate the production of target words by the mother (see Appendix E for examples of topics and questions).

3.2.2.5 Language background questionnaire

After the recording sessions, the mother was then asked to complete the language background questionnaire (see Appendix C for the questionnaire).

On average it took around 45 minutes for the entire recording including all three tasks. For their participation mothers were given $40 and a balloon animal was presented to the infant.

3.3Analysis

The entire acoustic analysis of the data was carried out in Praat (Boersma & Weenink, 2016). Initially, the recording was screened for the selection of the target word. This was then followed by the acoustic annotation of word initial stops and following vowel of each selected target word.

3.3.1 Data selection and screening

All tokens selected for analysis were disyllabic target words in isolation or sentence initial position. In Nepali, voiced consonants can have lead voicing. Thus, in situations where a voiced consonant has a continuous voicing from previous segment, it becomes difficult to delineate the onset of voicing. As the investigation of lead voicing is a part of the focus of current study, target tokens which were not in isolation or sentence initial position were excluded from analysis (1019 tokens excluded for this reason; see Appendix H for details). This was done to facilitate the segmentation of lead voicing. The exclusion of words in sentence medial and final position also implied that the sentence position, which could have effects such as phrase final lengthening, did not impact the lead time and lag time measures. The other criteria for exclusion from the analysis were: tokens with overlap of vocalizations from infant or experimenter (40), targets in isolation or sentence initial position tokens with continuous voicing from the previous segment (20), tokens not produced with modal voice (20), tokens produced in sing-song (5), tokens with no observable bursts (10), and tokens that are case-marked by addition of another suffix (155). All repetitions of tokens produced by each participant which passed the screening criteria were selected for further analysis (in total 1423 tokens).

3.3.2 Acoustic annotations

As discussed in Section 2.2.1.1.1, similar to Hindi (Davis, 1994), VOT alone fails to capture the voicing in Nepali voiced stops which shows both prevoicing and aspiration. Hence, in the present study, voicing cues in stops were measured as the lead and lag time similar to Davis' (1994) use for Hindi. In addition, the occurrence of devoicing in voiced consonants (voiced consonants without lead voicing) was noted. The duration of the vowel following the target stop consonants was measured to control for speaking rate (McMurray et al., 2013). Hence, for each selected target token devoicing and three duration measures (lead time, lag time, vowel duration) were measured.

Lead time was measured as the time between the first zero crossing of the pre-voicing period until the beginning of the burst (see Figure 3A). If there was a break in voicing before the burst release, lead time was measured as the duration between the first zero crossing of the pre-voicing period until the last zero crossing of voicing.

Lag time was measured as the duration between the onsets of burst release to the onset of the second formant of the following vowel (see Figure 3B.). In the case of stops with double bursts, measurements were taken from the start of the first burst release. Likewise, onset and offset of the second formant (F2) of the vowel following the target stop was annotated to measure vowel duration (see Figure 3).

Annotations were made by examining the spectrograms and waveforms simultaneously with the auditory percept of each segment to increase the accuracy of the measurement. Furthermore, in the case of difficulty with annotation, such tokens were flagged and later examined by an expert phonetician. Inter-rater reliability and intra-rater reliability was assessed by the remeasuring one third of the tokens by the experimenter in the former and another trained phonetician in the latter. Inter-rater reliability was found to be within 3.5 ms and the intra-rater reliability was found to be within 4.5 ms.

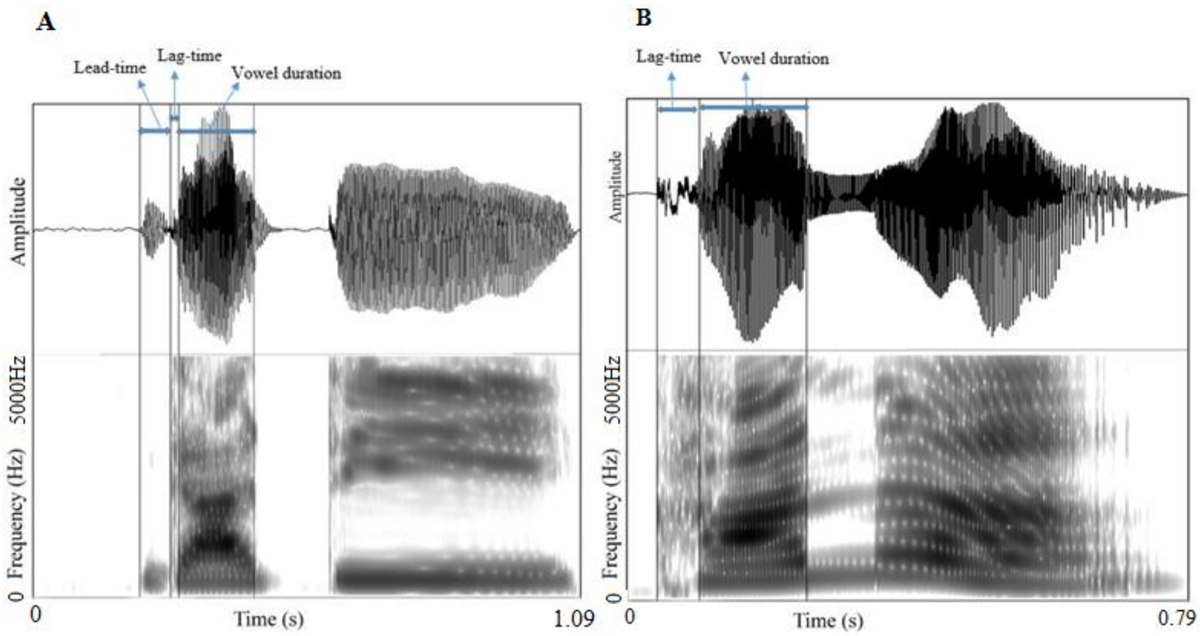


Figure 3: Figure 'A' is an example of the measurement of lead time, lag time, and vowel duration in a token of the word /ga.ɖa/. As /ga.ɖa/ starts with a voiced unaspirated stop, measures for both lead and lag time can be taken. Figure 'B' in an example of the measurement of lag time and vowel duration in a token of the word /kʰa.na/. As /kʰa.na/ starts with a voiceless stop, no measurement for lead time is taken. In both Figures A and B, the x-axis shows the time (seconds). The upper panel shows the acoustic waveform, with amplitude on the y-axis. The lower panel shows the spectrogram, with frequency (Hz) on the y-axis.

3.4 Statistical analysis

Prior to running the statistics, the outliers were removed. Any repetition with a value 1.5 times below the interquartile range or above the interquartile range for a given token within a given stop category within the register was judged to be an outlier and was removed from the analysis. In total less than 5% of the data was removed from the analysis (see Appendix F for the details).

The data were analysed using hierarchal linear regression, using the lmer function in the lme4 package (Bates, Maechler, & Bolker, 2012) in the R (R development Core Team, 2004) statistical programming environment. Mixed-effects regression models are considered best suited for these data as they make use of the multiple tokens per speaker rather than aggregating to a mean. Each dependent measure (i.e. devoicing, lead time, lag time, and vowel length) was analysed separately to test various hypothesis resulting in following four analysis.

3.4.1 Devoicing analysis

First the frequency of occurrence of devoicing (absence of lead time) in voiced tokens (voiced aspirated and voiced unaspirated) across the register was analysed. This was done to test the hypothesis that devoicing in IDS will be less than devoicing in ADS. As occurrence of devoicing is a binominal variable (devoiced: 0 or voiced: 1), it was analysed in a fully crossed logistic regression linear mixed-effects model. The fixed factors were voiced consonant categories (voiced unaspirated: -1 & voiced aspirated: 1) and register (ADS: -1 & IDS: 1). Random effects are addressed in separate section below.

3.4.2 Lead time analysis

Lead time of voiced consonant categories in both registers was analysed to test two predictions of the hyper-articulation hypothesis. First, it was predicted that the lead time in IDS will be longer compared to ADS. Secondly, the difference in lead time between the voiced aspirated and voiced unaspirated was predicted to be larger in IDS compared to ADS. Voiced tokens produced with devoicing were excluded from this analysis. To test these predictions, lead time was modelled in a fully crossed linear mixed-effects model. The fixed factors were voiced consonant categories (voiced aspirated: -1 & voiced unaspirated: 1) and register (ADS: -1 & IDS: 1). Random effects are addressed in separate section below.

3.4.3 Lag time analysis

Lag time of all four stop consonant categories in both registers was analysed to test two hypotheses. First, it was predicted that the lag time difference between voiced and voiceless stops in IDS will be larger compared to ADS. Secondly, the lag time difference between the aspirated and unaspirated stops was predicted to be larger in IDS compared to ADS. To test these predictions, lag time was modelled in a fully crossed linear mixed-effects model. The fixed factors were voicing categories (voiceless: -1 & voiced: 1), aspiration categories (unaspirated: -1 & aspirated: 1), and register (ADS: -1 & IDS: 1). Random effects are addressed in separate section below.

3.4.4 Vowel duration analysis

The duration of the vowel following the target stops in the first syllable of the target words in both registers was analysed to test the hypothesis that the register with the longer lead and lag time would be spoken at a slower rate as reflected by longer vowel duration. To test this hypothesis, the log vowel duration was modelled in a fully crossed linear mixed-effects model. The fixed factors were voicing categories (voiced: 1 & voiceless:-1), aspiration (aspirated: 1 & unaspirated:-1), and register (ADS: -1 & IDS: 1). Random effects are addressed in separate section below.

3.4.5 Testing significance of fixed factors

To test the significance of a fixed factors, the full model was compared to a reduced model which was created by eliminating from the full model the fixed factor of interest. Difference in fit was computed and evaluated against a χ^2 (Chi-square) distribution with one degree of freedom using ANOVA function. The factor of interest was considered significant only if the p-values associated with removing the fixed effect was lower than the alpha level $p= 0.05$. In cases, when the reduced model did not converge, the fixed effects were considered significant only when p_{mcmc} values were lower than $p_{mcmc} = 0.05$.

3.4.6 Selection of random factors

Random factors for a model were selected on the basis of model convergence. First, a full model with random intercept by subject was fitted. If the model converged, then by subject random slopes for all fixed factors (including interactions) were added. If the new full model converged, then it was used for the analysis. However, in the case with three fixed predictors (i.e. voicing, aspiration, and register) the full models with by-subject random intercepts and all by-subject random slopes did not converge. The selection of random factors for the full model with three fixed predictors were then based on the trials with variable combination of random slopes. These trials were carried out in the following order until model convergence was found.

1. Full model with by-subject random intercept and by-subject random slopes of all the effects (main effects, two way interactions, and three way interaction).

2. Full model with by-subject random intercept and by-subject random slopes of main effects (voicing, aspiration, and register) and the two way interaction effects. The three way interaction was excluded.

3. Full model with by-subject random intercept and by-subject random slopes of main effects (voicing and aspiration) and the two way interaction effects. The main effect of register, two way interaction of register with aspiration and voicing, and three way interaction were excluded.

4. Full model with random intercept of subject and random slopes of main effects (aspiration and register) and the two way interaction effects. The main effect of voicing, two way interaction of voicing with aspiration and register, and three way interaction were excluded.

5. Full model with by-subject random intercept and by-subject random slopes of main effects (voicing and register) and the two way interaction effects. The main effect of aspiration, two way interaction of aspiration with voicing and register, and three way interaction were excluded.

In cases where models converged with two random main effects (models 3, 4, and 5) but not with three random main effects (models 1 and 2), the model with random factor with random intercept of subject and random slope of aspiration and register by subject was preferred. The choice was made based on hypothesis testing; as the study investigates the difference between the register, the register was used as random slope. Likewise, the random slope of aspiration was selected instead of voicing, as there was more variability in regards to aspiration than voicing.

4. Results

The aim of the current study was to compare Nepali ADS and IDS on the realization of a four-way stop contrast, which contrasts for both aspiration and voicing. The voicing cues of word initial stops were measured as lead time in voiced consonants and lag time in both voiced and voiceless consonants. In addition, the occurrence of devoicing (complete absence of lead time) in voiced consonants was recorded as a binary variable. Duration of the vowel following word initial stops in the target words was measured to assess speaking rate. These measures were used to test two hypotheses.

1. The *hyper-articulation hypothesis*: which predicts that IDS would be hyper-articulated compared to ADS. If lead time is hyper-articulated in IDS, we should observe less devoicing of voiced stops in IDS than in ADS. Additionally, an overall, longer lead time in voiced categories, a larger lead-time difference between voiced aspirated and unaspirated stops. If lag time is hyper-articulated in IDS, we should observe larger lag time difference between voiced and voiceless stops, as well as between aspirated and unaspirated stops.

2. The *speaking rate hypothesis*: IDS will have a lower speaking rate, which explains that this register has a longer lead time, and a longer lag time.

The first subsection 4.1 presents the descriptive results. The second section presents the results from the inferential statistical models that are aimed at testing the predictions from the two hypotheses. The final section presents the results from an exploratory analysis that compares the consonant-to-syllable ratio between IDS and ADS.

4.1 Descriptive results

In both the registers, devoicing of voiced consonants was observed (Table 3). Devoicing was more frequent in voiced aspirated than voiced unaspirated stops, and more frequent in IDS than ADS.

Table 3: Count and percentage of devoiced voiced unaspirated and devoiced voiced aspirated tokens in ADS and IDS.

		ADS		IDS	
		Count	Percentage	Count	Percentage
Devoiced	Unaspirated	3	2.61	15	8.77
	Aspirated	39	37.14	68	47.55
	Overall	42	19.88	73	28.16

In Figure 4, the lead time distributions of voiced unaspirated and voiced aspirated stops show strong overlap in both ADS and IDS. Although the peaks of the distribution are not far apart in ADS, the lead time of voiced aspirated stops is on average 21ms shorter than the lead time of voiced unaspirated stops (Table 4). In IDS, the distributions of voiced aspirated and voiced unaspirated stops show considerable overlap with a mean difference of only 3ms.

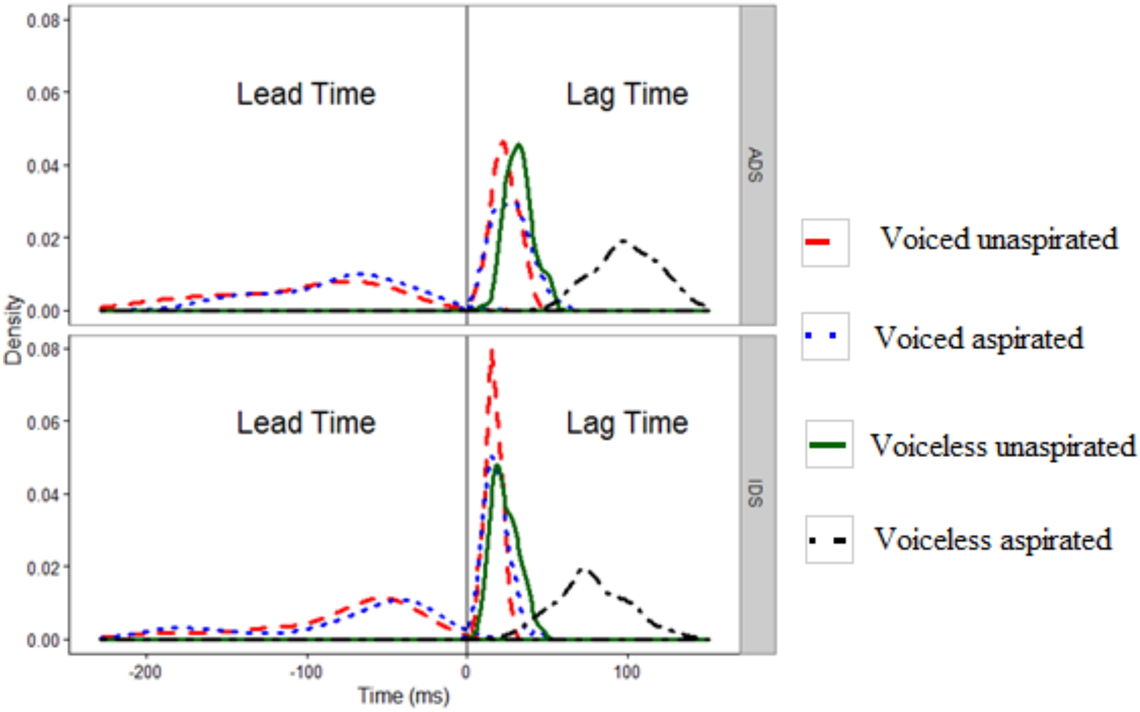


Figure 4: Distribution of lead time and lag time across voicing categories between registers. The x-axis shows two time scales in milliseconds. The centre line at 0 ms in horizontal scale refers to point of burst release. The time prior to 0ms on the negative scale (0 to -240ms) shows the measure of lead time; the time after ms on the positive scale (0 to 130ms) shows the measure of lag time. The y-axis shows the density of the number of observations. The upper panel shows ADS and the lower panel shows IDS. As there are only two categories with lead time, there are only two distribution on the lead time scale, whereas all four stops had lag time and are represented with a distribution on the lag time scale.

The lag time distributions of all four stop consonant categories in both ADS and IDS show two general peaks: for the voiceless aspirated stops and for the other three categories combined. The lag time distribution of the voiceless aspirated stops has a mean value of 100.3ms in ADS and a mean of 77.58ms in IDS (Table 5). The distribution of voiced unaspirated, voiced aspirated, and voiceless unaspirated showed considerable overlap in both ADS and IDS. The mean lag times of these three consonants fell within a close range of 23-32ms in ADS and 16-23ms in IDS.

Overall, the distributions of lead time and lag time across the four stop categories are similar in IDS and ADS, although both lead times and lag times are lower in IDS compared to ADS. In terms of vowel duration, the average vowel duration in IDS was 39ms longer than vowel duration in ADS (Table 6).

Table 4: Mean lead time (ms) and SD of voiced stops in ADS and IDS.

		ADS		IDS	
		Mean	SD	Mean	SD
Voiced	Unaspirated	-104.75	50.97	-79.04	48.52
	Aspirated	-82.43	40.92	-76.41	56.01

Table 5: Mean lag time (ms) and SD of stops in ADS and IDS.

		ADS		IDS	
		Mean	SD	Mean	SD
Voiced	Unaspirated	23.19	8.11	16.27	5.06
	Aspirated	27.04	11.78	18.19	7.98
Voiceless	Unaspirated	32.24	8.54	23.56	8.47
	Aspirated	100.03	20.33	77.58	22.70

Table 6: Mean vowel duration (ms) and SD of vowel following stops in target words in ADS and IDS.

		ADS		IDS	
		Mean	SD	Mean	SD
Voiced	Unaspirated	177.60	38.16	221.09	117.97
	Aspirated	130.96	22.24	169.47	75.02
Voiceless	Unaspirated	117.75	27.33	155.53	109.98
	Aspirated	88.01	28.51	127.56	102.45

4.2 Statistical results

This section presents the results from the statistical models that tested how the dependent variables (occurrence of devoicing, lead time, lag time, and vowel duration) differed between the stop categories (voicing and aspiration) and registers (IDS and ADS). The first section (section 4.2.1) presents the results from the analysis of devoicing, lead time, and lag time across the stop categories and the registers. The second section presents the analysis of vowel duration. The final section presents the results from the exploratory analysis investigating the consonant-to-syllable ratio between IDS and ADS.

4.2.1 Register difference in devoicing, lead time, lag time, and vowel duration

4.2.1.1 Devoicing of voiced stops between registers

The hyper-articulation hypothesis predicts that voiced consonants are realized with fewer instances of devoicing in IDS compared to ADS and would be confirmed by a significant main effect of register.

A logistic regression model was run with the dependent variable devoicing (devoiced: 0 & voiced: 1), and the fixed factors aspiration (voiced unaspirated: -1 & voiced aspirated: 1) and register (ADS: -1 & IDS: 1). The model contained a by-subject random intercept and a by-subject random slope of register.

The results (Table 7) showed a significant main effect of aspiration (β= -2.50; SE= 0.38; p<0.001): the likelihood of devoicing was larger in voiced aspirated than in voiced unaspirated stops. There was also a significant main effect of register (β= -0.98; SE= 0.43; p_{mcmc}=0.02): the likelihood of devoicing in IDS was higher than in ADS. Finally, the interaction between aspiration and register was also significant (β= 0.81; SE= 0.38; p_{mcmc}=0.03). Further post hoc analyses was carried out by testing the effect of register in the 'aspirated' and 'unaspirated' subsets, and then testing the effect of aspiration is the 'ADS' and 'IDS' subsets (see Appendix G for details on the post hoc analysis).

Within both 'aspirated' and 'unaspirated' subsets, neither test showed a significant difference of register between the rate of devoicing. On the other hand, within both 'ADS' and 'IDS' subsets, a significant effect of aspiration was observed (IDS: β= -1.37; SE= 0.31; p<0.001. ADS: β=-2.98; SE= 1.44; p= 0.4). This indicates that in both registers, the likelihood of devoicing was higher in voiced aspirated than voiced unaspirated stops (cf. Figure 5). The post-hoc tests thus fail to identify the source of the Register X aspiration interaction. Numerically, it appears as if the interaction may be due to a particularly high rate of devoicing of voiced aspirated stops in IDS, but we cannot draw final conclusions.

Table 7: Results from logistic regression model examining devoicing between the two voiced stop categories and between registers.

Devoicing~ Aspiration*Register+(1+Register|Subject)							
	β	SE	z	p_{mcmc}	χ^2	df	p
Intercept	2.50	0.65	3.81	<0.001***			
Aspiration	-2.24	0.38	-5.83	<0.001***	134.48	1	<0.001***
Register	-0.98	0.43	-2.29	0.02*	15.04	Model failed to converge	
Register x Aspiration	0.81	0.38	2.17	0.03*		Model failed to converge	

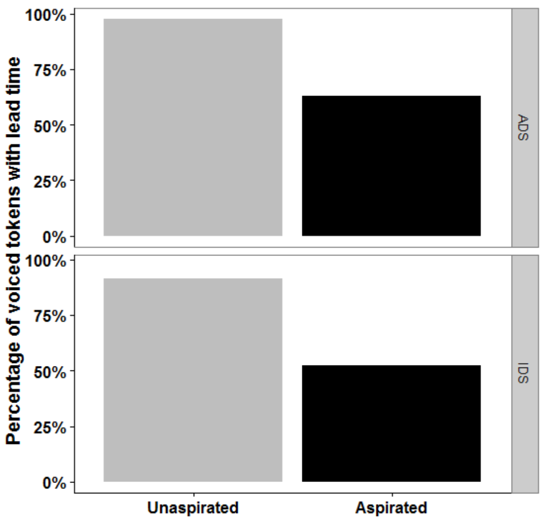


Figure 5: Percentage of voiced aspirated and voiced unaspirated stop tokens with lead time in ADS and IDS.

At the level of devoicing in individual participants, we observed that the 9 out of 16 participants showed at least one instance of devoicing in both the voiced aspirated and the voiced unaspirated stops. Across the entire sample of 16 participants, we observed a positive correlation between the devoicing rate for voiced aspirated and voiced unaspirated stops (r^2=0.72; *p*=0.002) (Figure 6).

Overall, the finding that voiced consonants are realized with more instances of devoicing in IDS compared to ADS contradicts the hyper-articulation hypothesis.

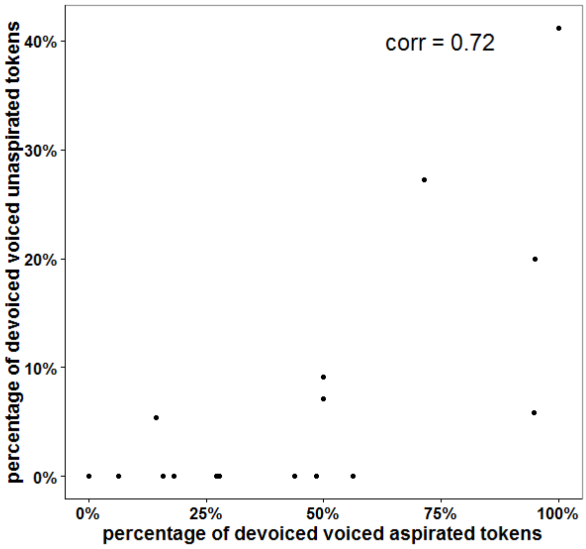


Figure 6: Scatterplot showing relation between the devoicing in voiced aspirated with devoicing in voiced unaspirated across entire sample of 16 participants. The x-axis shows the percentage of devoiced voiced aspirated tokens. The y-axis shows the percentage of devoiced voiced unaspirated tokens. The corr shows the result of Pearson correlation coefficient which is equal to 0.72.

4.2.1.2 Lead time of voiced stops between registers

The hyper-articulation and the speaking rate hypotheses both predict that voiced stops are produced with longer lead time in IDS compared to ADS. The hyper-articulation hypothesis additionally predicts a larger lead time difference between voiced unaspirated and voiced aspirated stops in IDS compared to ADS. The corresponding predictions are a significant main effect of register, and a significant interaction of register with aspiration.

A linear mixed effect model was run with the dependent variable lead time and the fixed factors aspiration (voiced unaspirated: -1 & voiced aspirated: 1) and register (ADS: -1 & IDS: 1). The model contained a by-subject random intercept and by-subject random slopes of aspiration, register, and the interaction between aspiration and register.

The results (Table 8) showed a significant main effect of aspiration ($\beta = 9.50$, SE= 2.61, p=0.004), where lead time in voiced unaspirated stops (Mean= -89.83ms; SD= 51.07ms) was longer than in voiced aspirated stops (Mean= -79.16ms; SD = 46.84ms). There was no significant main effect of register, nor a significant interaction between aspiration and register, which is also observed in Figure 7.

Overall, the hyper-articulation hypothesis and the speaking rate hypothesis are not supported by the present results, as these do not show a lead time difference between IDS and ADS predicted by both hypotheses, nor a larger difference between voiced unaspirated and voiced aspirated stops in IDS compared to ADS predicted by the hyper-articulation hypothesis.

Table 8: Results from the linear mixed effects model examining lead time between two stop categories and between registers.

LeadTime~Aspiration*Register+(1+Aspiration*Register|Subject)							
	β	SE	t	p_{mcmc}	χ^2	df	p
Intercept	-80.99	10.21	-8.00	<0.001***			
Aspiration	9.50	2.61	3.65	0.004**	8.98	1	0.003**
Register	5.23	3.75	1.39	0.19	1.82	1	0.18
Aspiration x Register	-3.10	1.99	-1.56	0.15	2.24	1	0.13

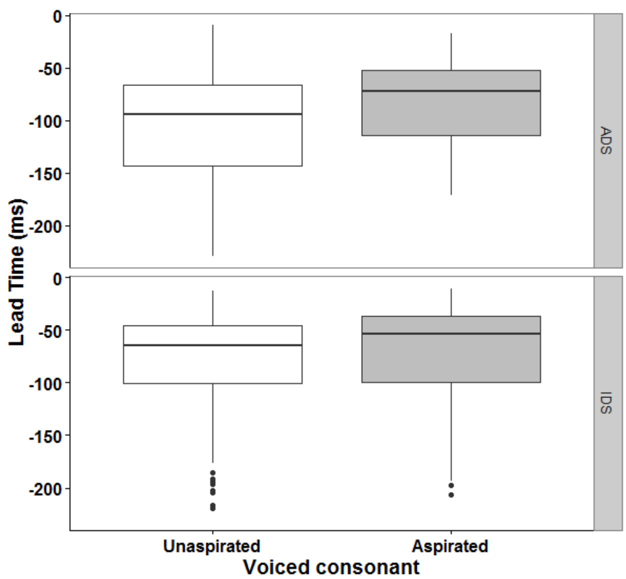


Figure 7: Boxplots showing lead time (ms) in y-axis of each voiced consonants: unaspirated and aspirated in x-axis for ADS (upper panel) and IDS (lower panel).

4.2.1.3 Lag time of all four stop categories

The hyper-articulation hypothesis predicts a larger lag time difference between the voiced and voiceless categories, as well as between the aspirated and unaspirated categories. The corresponding predictions are a significant main effect of register, a significant interaction of voicing and register, and a significant interaction of aspiration and register. The speaking rate hypothesis predicts all four stop categories are produced with a longer lag time in IDS compared to ADS and would be confirmed by a significant main effect of register.

A linear mixed effect model was run with the dependent variable lag time and fixed factors aspiration (unaspirated: -1 & aspirated: 1), voicing (voiceless: -1 & voiced: 1), and register (ADS: -1 & IDS: 1). The model contained a by-subject random intercept and by-subject random slopes of aspiration, register, and the interaction between aspiration and register.

The results (Table 9) showed a significant main effect of aspiration (β = 16.31, SE= 1.02, p<0.001): aspirated stops (Mean=61.21ms; SD= 36.34ms) had longer lag time than unaspirated stops (Mean = 23.26ms; SD= 9.27ms). There was a significant main effect of voicing (β= -18.22, SE= 0.36, p_{mcmc}<0.001): voiceless stops (Mean= 55.86ms; SD=34.45ms) had longer lag time than voiced stops (Mean= 20.39ms; SD= 9.13ms). There was a significant main effect of register (β = -5.28, SE= 0.63, p<0.001): lag time in ADS (Mean= 47.01ms; SD= 34.86ms) was longer than in IDS (Mean= 39.26ms; SD= 30.79ms). All 2-way and 3-way interactions were significant as well.

Table 9: Results from the linear mixed effects model examining lag time between all four stop categories and between registers.

LagTime~ Aspiration*Voicing*Register+(1+ Aspiration*Register|Subject)							
	β	SE	t	p_{mcmc}	χ^2	df	p
Intercept	40.18	1.35	29.82	<0.001***			
Aspiration	16.31	1.02	15.88	<0.001***	46.26	1	<0.001***
Voicing	-18.22	0.36	-50.00	<0.001***	Model did not converge		
Register	-5.28	0.63	-8.42	<0.001***	27.63	1	<0.001***
Aspiration x Voicing	-14.14	0.36	-38.85	<0.001***	1014.7	1	<0.001***
Aspiration x Register	-1.70	0.43	-3.91	<0.001***	11.00	1	<0.001***
Voicing x Register	2.27	0.36	6.23	<0.001***	38.44	1	<0.001***
Aspiration x Voicing x Register	1.80	0.36	4.94	<0.001***	24.18	1	<0.001***

To identify the source of the 3-way interaction, altogether twelve post-hoc analyses were conducted, four each evaluating the effect of aspiration, voicing, and register. First results from post-hoc analyses investigating the effect of aspiration are presented, followed by results from post-hoc analyses investigating the effect of voicing and register (see Appendix G for details of post hoc).

Four post-hoc analyses were conducted that evaluated the effect of voicing in the four subsets of unaspirated stops in ADS, aspirated stops in ADS, unaspirated stops in IDS, and aspirated stops in ADS.

Within all four subsets ‘ADS unaspirated’, ‘ADS aspirated’, ‘IDS unaspirated’, and ‘IDS aspirated’ subsets, a significant effect of voicing was observed (‘ADS unaspirated’: β= -4.82; SE= 0.71; p<0.001. ‘ADS aspirated’: β= -36.041; SE=2.14; p<0.001. ‘IDS unaspirated’: β= -3.91.62; SE= 0.59; p<0.001. ‘IDS aspirated’: β= -29.79; SE= 1.61; p<0.001): voicing led to a shorter lag time in both unaspirated and aspirated stops.

Four post-hoc analyses were conducted that evaluated the effect of register in the four subsets of unaspirated stops in voiced stop categories, aspirated stops in voiced stop categories, unaspirated stops in voiceless stop categories, and aspirated stops voiceless stop categories.

Within all four subsets ‘voiced unaspirated’, ‘voiced aspirated’, ‘voiceless unaspirated’, and ‘voiceless aspirated’ subsets, a significant effect of register was observed (‘voiced unaspirated’: β= -2.96; SE= 0.70; p=0.001. ‘voiced aspirated’: β= -4.50; SE=0.78; p<0.001. ‘voiceless unaspirated’: β= -4.00; SE= 0.71; p<0.001. ‘voiceless aspirated’: β= -11.14; SE= 1.09; p<0.001): IDS led to a shorter lag time in both unaspirated and aspirated stops.

Four post-hoc analyses were conducted that evaluated the effect of aspiration in the four subsets of voiced *stops in ADS, voiceless stops in ADS, voiced stops in IDS, and voiceless stops in ADS.*

Within both the ‘ADS voiced’ and ‘ADS voiceless’ subsets, a significant effect of aspiration was observed (‘ADS voiced’: β= 2.30; SE= 0.95; p= 0.03. ‘ADS voiceless’: β= 33.512; SE=1.831; p<0.001): aspiration led to a longer lag time in both voiced and voiceless stops. Within the ‘IDS voiced’ and ‘IDS voiceless’ subset, a significant effect of aspiration was observed only for the ‘IDS voiceless’, and not for 'IDS voiced' (IDS voiced: β= 1.28; SE= 0.58; p=0.05. IDS voiceless: β= 27.35; SE= 1.59; p<0.001). The 3-way interaction between Aspiration x Voicing x Register can thus be ascribed to the fact that aspiration may not have an effect on lag time in the voiced stops of IDS, whereas it has an effect in voiceless stops of IDS as well as across both voiced and voiceless stops in ADS (cf. Figure 8).

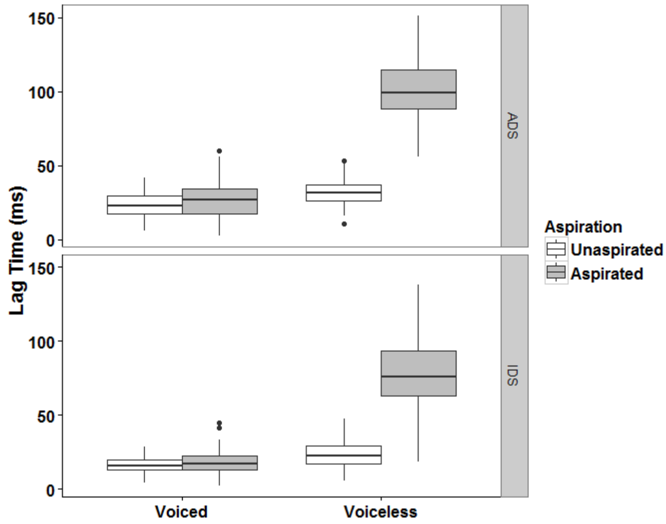


Figure 8: Boxplots showing lag time (ms) along the y-axis of all four stop categories. The x-axis shows the voicing categories (voiced & voiceless). The grey scale shows the aspiration categories (white=unaspirated & grey=aspirated). The upper panel shows the ADS and the lower panel shows IDS.

Overall, the hyper-articulation hypothesis and the speaking rate hypothesis are not supported by the present results. The data do not show a larger lag time difference between the voiced and voiceless categories, between aspirated and unaspirated categories, nor do they show longer lag time in IDS compared to ADS.

4.2.1.4 Vowel duration between registers

The speaking rate hypothesis predicts longer lead and lag times in IDS as a result of the slower rate of speech in this register (McMurray et al., 2013). Even though the previous analyses did not reveal the predicted lead time differences between IDS and ADS and actually revealed unpredicted *shorter* lag time in IDS, it was decided to still analyse the speaking rate measure to fully evaluate the speaking rate hypothesis. The main assumption of the speaking rate hypothesis is that speaking rate is lower in IDS than in ADS. In the current study, vowel duration was measured as proxy for speaking rate. The assumption that speaking rate is lower in IDS than in ADS, would be supported by a significant main effect of register on vowel duration.

A linear mixed effect model was run with the dependent variable log vowel duration and fixed factors voicing (voiceless: -1 & voiced: 1), aspiration (unaspirated: -1 & aspirated: 1), and register (ADS: -1 & IDS: 1). The model contained a by-subject random intercept and by-subject random slopes of aspiration, voicing, and the interaction between aspiration and voicing.

The results (Table 10) showed a significant main effect of aspiration (β = -0.14, SE= 0.03, *p*<0.001): log vowel duration was longer in unaspirated (Mean= 5.01ms; SD= 0.46ms) than in aspirated stops (Mean= 4.72ms; SD=0.51ms). There was a significant main effect of voicing (β = 0.22, SE= 0.01, *p*<0.001): log vowel duration was shorter in voiceless (Mean= 4.71ms; SD= 0.51ms) than voiced stops (Mean= 5.11ms; SD= 0.38ms). There was also a significant main effect of register (β = 0.1, SE= 0.03, *p*<0.01): log vowel duration was shorter in ADS (Mean= 4.78ms; SD= 0.36ms) than in IDS (Mean= 4.91ms; SD= 0.56ms). There were no significant 2-way or 3-way interactions.

These results support the assumption of slower speaking rate in IDS, as vowel duration is longer in IDS than ADS. However, the key prediction of the speaking rate hypothesis, which predicted longer lead and lag time in IDS as a result of the slower speaking rate is not supported by the results, as these are shorter in IDS than in ADS.

Table 10: Results from the linear mixed effects model examining log vowel duration between all four stop categories and between registers.

LogVowelLength~ Aspiration*Voicing*Register+(1+ Aspiration*Register|Subject)							
	β	SE	t	p_{mcmc}	χ^2	df	p
Intercept	4.90	0.04	133.90	<0.001***			
Aspiration	-0.14	0.02	-5.73	<0.001***	18.651	1	<0.001***
Voicing	0.22	0.01	19.68	<0.001***	337.33	1	<0.001***
Register	0.10	0.03	3.07	0.008**	7.86	1	<0.001***
Aspiration x Voicing	0.01	0.01	0.96	0.34	0.92	1	0.34
Aspiration x Register	0.01	0.02	0.52	0.61	0.29	1	0.59
Voicing x Register	0.00	0.01	0.29	0.77	0.01	1	0.77
Aspiration x Voicing x Register	0.00	0.01	0.45	0.64	0.21	1	0.64

4.2.1.5 Summary of confirmatory results

In summary, the two voiced stops showed a higher rate of devoicing in IDS than in ADS, but there was no difference in lead time between IDS and ADS. Lag time was overall shorter for all four stop categories in IDS than in ADS. In addition, the lag time difference between voiced and voiceless stops, as well as between aspirated and unaspirated stops was smaller in IDS compared to ADS. These findings do not support and in some instances contradict the hyper-articulation hypothesis, which predicted that IDS would be hyper-articulated compared to ADS by having a lower rate of devoicing, an overall longer lead time in voiced categories, a larger lead-time difference between voiced aspirated and unaspirated stops, and larger lag time differences between voiced and voiceless stops, as well as between aspirated and unaspirated stops.

According to the speaking rate hypothesis, longer lead times, longer lag times, and a larger lag time difference between aspirated and unaspirated stops, in IDS may be an automatic side effect of the generally lower speaking rate in this register. Vowel duration, which was measured as a proxy for speaking rate, was longer in IDS, suggesting that IDS had the expected lower speaking rate. Despite this low speaking rate in IDS, this register did not have the predicted longer lead time than ADS, and even had an opposite-to-predicted shorter lag time, and smaller lead-time difference between voiced aspirated and unaspirated stops. Both current hypotheses of segmental adaptations in IDS thus fail to explain the current data.

4.2.2 Consonant-to-syllable duration ratio

As discussed above, IDS is characterized by shorter lag times and longer vowel durations than ADS. This raises the possibility of a different trading relationship between consonant and vowel duration in IDS compared to ADS. The present section therefore explores whether IDS and ADS differ in regards to the proportion of time in a syllable that caregivers dedicated to the consonant rather than the vowel. Given the observed shorter lag times and longer vowel durations in IDS, it could be expected that within a syllable, proportion of time allocated to consonants would be smaller in IDS.

To test this possibility, the ratio of consonant duration to the total syllable duration was measured. This ratio indicated the proportion of consonant duration within a syllable. The computation of consonant duration for voiced and voiceless stops was necessarily different, as our analysis was based on tokens only in isolation and sentence initial position. In isolation and sentence initial position, except for voiced tokens with lead time, the closure duration of stops could not be captured for the devoiced voiced stops and the voiceless stops. Therefore, the data were separated into two subsets based on voicing (voiced and voiceless). In addition the analysis of voiced stops excluded any devoiced tokens.

4.2.2.1 Voiced-stops-to-syllable ratio between registers

The analysis of voiced consonants excluded any devoiced consonants because the absence of prevoicing made it impossible to capture the closure duration of the stop. The computation was as follows:

Voiced-stop-to-syllable ratio= voiced consonant duration / syllable duration

Voiced consonant duration = lead time + lag time

Syllable duration = lead time+ lag time+ vowel duration

The consonant duration of all voiced stops with lead time was measured from the onset of voicing until the onset of F2 (lead time + lag time). The syllable duration was measured from the onset of voicing until the end of F2 (lead time+ lag time+ vowel duration). *This is proportional measure, which thus corrects for speaking rate differences between the register.* Inspection of Figure 10 suggests a lower consonant ratio in IDS than ADS.

A linear mixed effects model was run with the dependent variable voiced-stop-to-syllable ratio and fixed factors aspiration (unaspirated: -1 & aspirated: 1) and register (ADS: -1 & IDS: 1). The model contained a by-subject random intercept and by-subject random slopes of aspiration, register, and the interaction between aspiration and register.

The results (Table 11) showed a significant main effect of register (β= -0.06; SE= 0.01; p<0.001): the proportion of time within a syllable allocated to the consonant is smaller in IDS (Mean= 0.28, SD=0.10) compared to ADS (Mean= 0.34, SD=0.8). There was no significant effect of aspiration and no significant interaction between aspiration and register (cf. Figure 9).

These findings reveal that speakers in IDS allocate relatively less time within a syllable to the voiced consonant, and conversely, allocate relatively more time to the vowel. This finding suggests that there is a trading relationship between the duration of the consonant and the duration of the vowel within the syllable, and that these vary as a function of register.

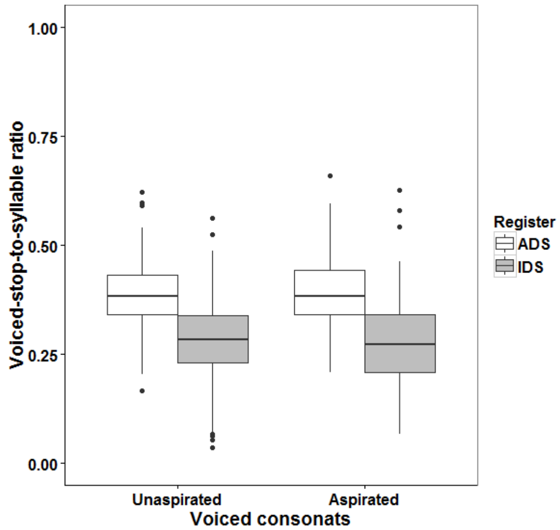


Figure 9: Boxplots showing voiced-stop-to-syllable duration ratio in y-axis. The x-axis shows the stop categories (voiced unaspirated & voiced aspirated). The colour shows the register (white- ADS & grey-IDS).

Table 11: Results from the linear mixed effects model examining voiced-stop-to-syllable ratio between two stop categories and between registers.

Voiced-stop-to-syllable_ratio~ Aspiration*Register + (1+Aspiration*Register|Subject)							
	β	SE	t	p_{mcmc}	χ^2	df	p
Intercept	0.33	0.12	29.94	<0.001***			
Aspiration	<0.001	0.001	0.40	0.70	0.11	1	0.73
Register	-0.06	0.01	-6.96	<0.001***	23.04	1	<0.001***
Aspiration x Register	-0.001	<0.001	-1.23	0.23	1.17	1	0.28

We further wanted to explore whether within the duration of a voiced consonant, speakers allocated time differently to the lead and lag time, and whether this allocation was different between the registers. This question relates to the hyper-articulation hypothesis as lead time is cue to voicing, relatively longer lead time within the consonant duration in IDS would show that parents use the restricted consonant time economically for the more distinctive cue.

To explore this, we first computed the lead-to-consonant ratio:

Lead-time-to-voiced-consonant duration ratio= lead time / voiced-consonant duration

A linear mixed effect model was run with the dependent variable lead-time-to-voiced-consonant ratio and fixed factors aspiration (unaspirated: -1 & aspirated: 1) and register (ADS: -1 & IDS: 1). The model contained a by-subject random intercept and by-subject random slopes of aspiration, register, and the interaction between aspiration and register. The results (Table 12) showed a significant main effect of aspiration (β= 0.02; SE= 0.01; p = 0.01): the lead-to-consonant ratio was larger for unaspirated stops (Mean= -0.74, SD=0.13) than aspirated stops (Mean= 0.71, SD=0.14). There was no significant difference between the registers nor an interaction between aspiration and register.

These results show that lead time was longer than lag time in both voiced aspirated and voiced unaspirated stops, but the trade-off differed between the categories, resulting in a longer lead-to-consonant ratio in voiced unaspirated stops. However, this trade-off is consistent between the registers, which does not support the hyper-articulation hypothesis. If there was a hyper-articulation, we would have seen that the mothers when talking to infant show the contrasts in voiced consonants with higher proportionate duration of lead time within a consonant duration.

Table 12: Results from the linear mixed effects model examining lead time ratio between two stop categories and between registers.

Lead-time-to-voiced-consonant_ratio~ Aspiration*Register + (1+Aspiration*Register|Subject)							
	β	SE	t	p_{mcmc}	χ^2	df	p
Intercept	-0.72	0.02	-41.64	<0.001***			
Aspiration	0.02	0.01	2.81	0.02*	6.28	1	0.01*
Register	-0.01	0.02	-0.81	0.44	0.68	1	0.41
Register x Aspiration	-0.01	0.01	-0.83	0.43	0.72	1	0.40

4.2.2.2 Voiceless-stop-to-syllable ratio between registers

For the analysis of voiceless consonants, consonant to syllable duration ratio was computed as follow:

Voiceless-stop-to-syllable ratio = voiceless consonant duration/ syllable duration

Voiceless consonant duration = lag time

Syllable duration= lag time + vowel duration

The consonant duration of voiceless stops was measured from the onset of burst release until the onset of F2 (lag time), syllable duration was measured from the onset of burst release until the end of F2 (lag time+ vowel duration).

A linear mixed effect model was run with voiceless-stop-to-syllable ratio as the dependent variable. Fixed factors were aspiration (unaspirated: -1 & aspirated: 1) and register (ADS: -1 & IDS: 1). The model contained a by-subject random intercept and also by-subject random slopes of aspiration, register, and the aspiration and register interaction.

The result (Table 13) showed a significant main effect of aspiration (β= 0.15; SE= 0.01; $p<0.001$): voiceless-stop-to-syllable ratio was smaller in voiceless unaspirated stops (Mean= 0.37, SD=0.93) compared to voiceless aspirated stops (Mean= 0.41, SD=0.91). There was a significant main effect of register (β= -0.04; SE= 0.01; p<0.001): voiceless-stop-to-syllable ratio was smaller in IDS (Mean= 0.29, SD=0.17) compared to ADS (Mean= 0.38, SD=0.18). There was a significant interaction between aspiration and register (β= -0.01, SE= -2.28, $p<0.001$).

Table 13: Results from the linear mixed effects model examining voiceless-stop-to-syllable-ratio between all four stop categories and between registers.

Voiceless-stop-to-syllable_ratio~ Aspiration*Register + (1+Aspiration*Register|Subject)							
	β	SE	t	p			
Intercept	0.33	0.01	35.43	<0.001***			
Aspiration	0.15	0.01	24.69	<0.001***	59.44	1	<0.001***
Register	-0.04	0.01	-6.66	<0.001***	22.12	1	<0.001***
Aspiration x Register	-0.01	0.00	-2.28	0.03*	4.56	1	0.03*

Inspection of Figure 10, suggests the effect of register was larger for aspirated stops than unaspirated stops. Further, post hoc analysis was done by separating the data first into an 'aspirated' and an 'unaspirated' subset to analyse the effect of register, and then separate the data into an 'ADS' and 'IDS' subset to analyse the effect of aspiration (for details on post hoc, see Appendix G). Within both the ‘aspirated’ and the ‘unaspirated’ subset, there was a significant effect of register: The voiceless-stop-to-syllable ratio was smaller in IDS compared to ADS for both voiceless aspirated stops (β= -0.06; SE= 0.01; p<0.001) and voiceless unaspirated stops (β= -0.03; SE= 0.00; p<0.001) (cf. Figure 10).

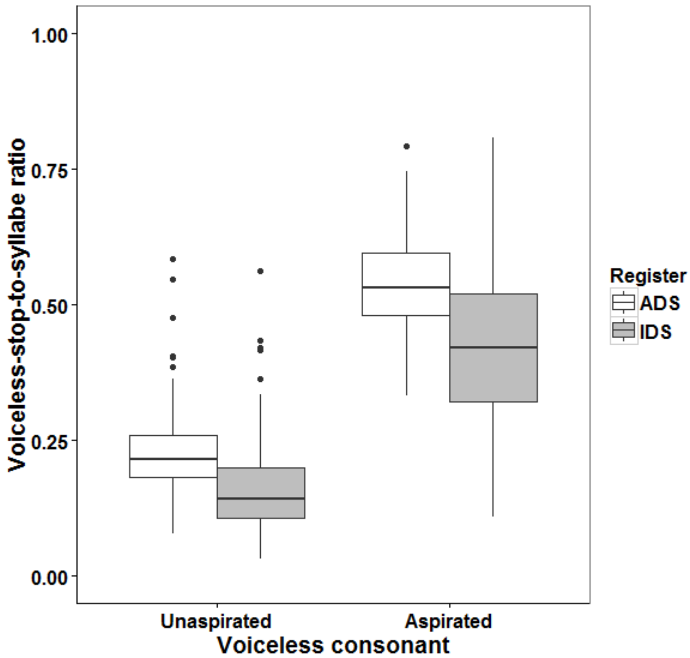


Figure 10: Boxplots showing voiceless-stop-to-syllable ratio in y-axis of voiceless stop categories. The x-axis shows the stop categories (voiceless unaspirated & voiceless aspirated). The colour shows the register (white- ADS & grey-IDS).

4.2.2.3 Summary of exploratory results

In summary, both voiced and voiceless stops had a smaller consonant-to-syllable ratio in IDS compared to ADS. This result indicate that there is a trading relationship between the duration of the consonant and vowel with the syllable. The longer vowel duration in IDS may be responsible for the shorter stop durations in IDS than in ADS. The full implications of this finding will be addressed in the discussion.

5. Discussion

This study investigated the difference between the ADS and IDS in the realization of the four-way voicing contrast in Nepali oral stops. These voicing cues, which contrast in both voicing and aspiration, were measured using lead time and lag time. The aim of the study was to test two hypotheses. The first hypothesis predicted there would be hyper-articulation of devoicing, lead time and lag time in IDS compared to ADS. The second hypothesis predicted that these effects might be a side-effect of speaking rate differences between the two register.

In this discussion, the results from devoicing, lead time, and lag time differences of stop categories between the register will be discussed, addressing the hyper-articulation hypothesis. The second section discusses the findings from vowel duration analysis addressing the speaking rate hypothesis. This is then followed by a discussion of results from the further analysis of consonant/vowel duration ratios. Finally, the last section of the discussion presents the limitations of current study and possible directions for further research.

5.1 Register difference in devoicing, lead time, lag time, and vowel duration

Recall that the hyper-articulation hypothesis predicted that consonant durations might be longer for IDS than for ADS. Except for lead time, all other variables showed a clear effect of register. However, devoicing was higher in IDS compared to ADS, and IDS had shorter lag times for all the stop consonant categories compared to ADS. As might be expected from other studies, the vowel duration in IDS was longer than in ADS. These results are discussed below in detail.

5.1.1 Devoicing between the registers

In the current study, devoicing of voiced consonants was observed to be higher in IDS than in ADS, making them more similar to the voiceless consonants. According to hyper-articulation hypothesis, the voicing contrasts in IDS should be larger than in ADS. This implies that the devoicing in IDS should be lower than the devoicing in ADS. However, results from the current study showed that devoicing of voiced consonants was higher in IDS than ADS. This raises many question for how the learner might distinguish voiced from voiceless categories.

The increased devoicing in IDS, suggests that the variability in voicing of voiced consonants in IDS is higher compared to ADS. In addition, although small, it was also observed that mothers who devoice more in one consonant also devoice more in the other, meaning that these speakers do not contrast voiced consonants via lead voicing consistently. This may present challenges for the infant, who when exposed to other speaker, such as the father or aunty, may confuse the number of consonant categories present in the language. This suggests that this type of IDS, with its increased variability, may not be advantageous for the infant trying to learn voicing contrasts in stop consonant categories, especially in this complex type of four-way voicing system.

5.1.2 Lead time between registers

The results of the lead time analysis between registers, showed no significant difference in the lead time between the registers. This can be taken as negative evidence for the hyper-articulation hypothesis which predicted longer lead time, and a larger difference of lead time between the voiced aspirated and voiced unaspirated.

5.1.3 Lag time between registers

Lag time showed a clear effect of IDS. IDS led to shorter lag time in IDS compared to ADS for all stop consonant categories. Lag time was shorter for both voiced and voiceless as well as aspirated and unaspirated stops in IDS compared to ADS. Furthermore, the effect of register difference was larger for voiceless than voiced stops, and larger in aspirated than in unaspirated stops.

These results again go counter to the hypothesis proposing hyper-articulation of voicing contrasts in IDS. The hyper-articulation hypothesis predicted a larger lag time difference between the voicing categories in IDS compared to ADS. However, the shorter lag time in IDS resulted in lower contrasts of voicing categories in IDS compared to ADS. These lower contrasts in IDS suggest that the separation between the voicing categories is larger in ADS than in IDS. These smaller contrasts in IDS may again have an implication for the learner.

The smaller separation of stop categories in IDS compared to ADS suggests an increased chance of overlap between the categories. High variation in the production of stop categories, may lead to significant overlap. Although the density plot in Figure 4 does suggest overlap between the categories in both registers, it does not measure the variance in the production of register. It will be valuable to have this information, which is suggested for future research.

5.1.4 Vowel duration between registers

In the current study, vowel duration was measured as a control for speaking rate differences between the two registers. The speaking rate hypothesis states that the register with a lower speaking rate will have longer lead time and lag time. So, perhaps the reduction in lag time in IDS is the result of faster speaking rate in IDS. However, our analysis of vowel duration indicated lower speech rate in IDS, as IDS had longer vowel duration than ADS. This is contrary to speaking rate hypothesis. Thus, despite the fact that the mothers exhibited lower speaking rate to their infants, lag time are shorter compared to the mothers faster speech to the adult.

Overall, the results showed that neither the hyper-articulation nor the speaking rate hypothesis can account for the current observations. The lag time was shorter and vowel duration was longer in IDS. On the other hand, lag time was longer and vowel duration was shorter in ADS.

These findings raise an interesting question as to why consonant duration is shorter and vowel duration longer in IDS? To investigate the possibility that proportionate duration of consonants within a syllable is shorter in IDS compared to ADS, we measured the ratio of consonant to syllable duration. Due to the difference in computation of consonant to syllable duration for voiced and voiceless stop categories, separate analyses were carried out on each subset of the data. In the following section, findings from voiced categories are discussed first, then the findings from voiceless categories.

5.2 Voiced-stop-to-syllable ratio between registers

Within a syllable, the proportional duration of the consonant was shorter in IDS compared to ADS. This indicated that, within the syllable, vowel duration was enhanced more than consonant duration in IDS compared to ADS. The results indicate that, within a syllable, mothers allocated less time to consonants than vowels in IDS.

This raised another interesting issue; within consonant duration, do mothers allocate more time in lead voicing than lag voicing? Given that lead voicing is an important cue to stop contrasts, if mothers are hyper-articulating, we expected to observe a longer proportion of lead time in IDS. The results showed no difference in the proportion of duration of lead voicing within the consonant duration of voiced stops. However, between the voiced consonants, realization of voiced unaspirated consonants had larger proportion of lead time than the voiced aspirated. This suggests that, with a voiced consonant, either a focus is on lead voicing or on lag voicing.

There seems to be a trade-off between lead and lag voicing. This trading relationship suggest the variability in the realization within the voiced stops. This then raises a question about the learning of the voicing contrasts between voiced and voiceless. For a learner, the two voiced categories may seem inherently different as they have different realization.

5.3 Voiceless-stop-to-syllable ratio between registers

Similar to voiced categories, within the syllable, the proportional duration of the consonant was shorter in IDS compared to ADS. This indicates that, even for voiceless categories, vowel duration was enhanced at the cost of consonant in IDS. This indicates that, within the syllable, mothers allocated less time to consonants than to vowels.

Overall our results suggest that mothers focus more on the vowel while talking to infants. This appears to have the effect of shortening the onset consonants. Vocalic features of IDS are reported to be exaggerated to convey affect and direct attention (Fernald & Simon, 1984). This suggest that Nepali mothers may do the same, focusing more on the communication of affect or in directing their infant’s attention.

To our knowledge, this is the first time that the voicing contrast in IDS is being looked at this way. This then has major theoretical implications for better understanding the nature of IDS and its implication for the learning of consonant contrasts, and for language learning more generally.

5.4 Alternative interpretations of the data

An alternative possibility is that difference in lead time and lag time observed between the registers may be attributed to the method of elicitation of tokens between the registers. This would mean that mothers during ADS produced the target tokens in focus position, which may have resulted in longer voicing lead and lag times. However, if this were the case, then the vowels should also have been longer in ADS as well, whereas vowels were in fact longer in IDS.

As a follow up, we tested the possibility that the inclusion of tokens produced during the familiarization task might have contributed to the above results. In the familiarization, mothers were asked to name the pictures presented to them one at a time. The possibility that mothers noticed that a key difference between the picture names was the onset of their first syllables, and that this might have led to exaggeration of voicing contrasts in ADS (resulting in longer lead and lag time) is improbable for two reasons. First, the order of presentation of pictures was not controlled, pictures were pulled out from the bag and neither the experimenter nor the mother was aware of the next picture. In addition, the stimulus set contained two additional words with different vowels in the first syllable. Furthermore, the comparison of lead time and lag time between the tokens elicited during the familiarization and the mother-experimenter interaction after IDS revealed no difference. Lastly, it was observed that mothers were naming pictures in focus position to their infants during IDS. The possibility that adults were producing the token words in focus position in ADS and not IDS will thus not account for this phenomenon.

A second possible explanation for the lack of clear voicing contrasts in IDS is the possibility that lead time and lag time are not the main voicing cues used for distinguishing voicing categories in Nepali. It may be that the contrast, although very clear in ADS, may be realized by some other means in IDS. Compare for example Korean, where pitch is now used rather than simply VOT (Silva, 2006). To explore this possibility, future research must look at other cues to voicing contrasts in Nepali. This is therefore a limitation of the current study and an obvious area for future research.

A third possibility may be that the context of recording, i.e. a laboratory recording with the instruction to label objects instead of a naturalistic spontaneous recording of IDS in the home, lead to a reduced stop contrast in IDS.. However, numerous studies comparing IDS and ADS elicited in laboratory settings using labelling of objects have reported hyper-articulation in IDS for both vowels (Kuhl et al., 1997) and consonants (Sundberg, 2001, Burnham et al., 2013). This previous literature suggests that hyper-articulation instead of hypo-articulation can be observed in IDS that is recorded in a laboratory environment. Although the preference may be to record natural mother-child interaction in the home environment, this is also a challenge due to issues such as controlling the acoustic environment and controlling the linguistic context in which the target segments appear. Future studies can investigate the possible difference of voicing cues between laboratory and home elicitations.

5.5 Devoicing, lead time, and lag time in ADS

The devoicing in voiced consonants showed lower likelihood of devoicing in voiced unaspirated that voiced aspirated. Additionally, the lead time analysis of voiced consonants showed longer lead time of voiced unaspirated compared to voiced aspirated stops. Thus, the two voiced consonant categories were different in-terms of devoicing and lead time.

In terms of lag time, voiceless stops had longer lag time than voiced stops, and aspirated stops had longer lag time than unaspirated stops. Furthermore, the effect of aspiration was larger in voiceless stops than voiced stops and in aspirated stops than unaspirated stops. Among the four stop consonant categories, the lag time of voiceless unaspirated stops was longer than the other three categories (i.e. voiced unaspirated, voiced aspirated, and voiceless unaspirated), which showed a smaller lag time difference.

The pattern of voicing contrasts shown by the results of current study is similar to the Nepali pattern shown by Poon and Mateer (1985). This was observed despite the fact that the current study measures voicing cues as lead time and lag time and the Poon and Mateer (1985) study measured voicing cues as VOT. In both the studies, voiced unaspirated consonants had the longest negative voicing cue (lead time and negative VOT), and voiceless aspirated had the longest positive voicing cue (lag time and positive VOT). In addition, the devoicing in voiced aspirated consonants was higher than in voiced unaspirated.

The voiced aspirated stops in Nepali show high inconsistency with lead time, as well as a weaker aspiration effect compared to Hindi. This may suggest that voiced aspirated stops in Nepali may be unstable in terms of pre-voicing. This inconsistency in pre-voicing of voiced aspirated stops was also observed within the same speaker. Previous literature suggests that unstable contrasts in a language tend to be lost in the language over time (Hombert, Ohala, & Ewan, 1979). This is especially interesting given that a four-way voicing contrast of stops in Punjabi has been reduced to a three-way contrast due to a loss of the voiced aspirated consonants (Kanwal & Ritchart, 2015). This raises many interesting questions about how the Nepali voiced aspirated stops are perceived by both adults and infants, and if this contrast is now being lost. It will be valuable to address this in further research.

5.6 Limitations

The current study targeted only velar stop consonants in word-initial position of disyllabic words. In addition, only target tokens in isolation or sentence initial position which were not case marked with additional suffixes were analysed (see Appendix H for details of words in all sentence position). The focus on word-initial stops in utterance-initial position was motivated by the need to facilitate the segmentation of voicing lead, which can occur in Nepali, voiced stops. When a stop with voicing lead has continuous voicing from previous segment, it becomes difficult to delineate the onset of voicing. As voicing lead was a part of the focus of current study, target tokens which were not in isolation or sentence initial position were excluded from analysis. This restriction to word-initial velars in disyllabic word in only one utterance position leaves open the possibility that the realization of voicing lead time and lag time may be different in stops at other places of articulation, stops in word with different number of syllable, or other sentence positions.

With respect to the measurement of voicing cues, we have limited ourselves to lead time and lag time. There may be other cues to voicing such as burst duration, change in fundamental frequency or voice quality of the following vowel (Lisker, 1986). There is a possibility that the voicing contrasts may be hyper-articulated in the other cues, especially if the lead time and lag time are not the main cues to voicing (Cristià, 2010). However, the current study utilized the durational cues of lead and lag time due to two reason. Firstly, there are no perception studies reporting voicing cues in Nepali, so there is no information about the main cue to voicing. Secondly, the only study reporting the realization of voicing contrasts in Nepali has measured VOT which is reported to be inefficient in showing Nepali four-way contrast. For this reason, lead time and lag time which are suggested to capture the voicing contrasts in all four stop categories (Davis, 1994), were used in the present study.

In addition to only measuring a subset of the possibly relevant voicing cues, only the mean lead time and lag time of stop categories were compared between IDS and ADS. The variation in these cues was not directly assessed. We concluded that IDS is not hyper-articulated because we observed a reduced difference between the means in IDS compared to ADS. However, the variation in the production of stop categories may be lower than ADS. Such lower variation may imply that the stop contrasts are still enhanced in IDS than ADS.

The input children hear is the full distributions which are characterized by means and variance. Hence to fully characterize the input, variances are required. Moreover, contrasts are defined by both means and variances (Cristià & Seidl, 2013). Maybe variances are equal in IDS and ADS- then the conclusions about difference actually mean differences in contrast. Maybe variances are larger in IDS as well, as reported by Kuhl et al. (1997). If so, then the contrasts in IDS are even more reduced than we think on the basis of present data. This further implies the contrasts in IDS may not be hyper-articulated or even hypo-articulated. The final conclusion about hyper-articulation in Nepali IDS (or the reverse thereof) requires further research.

A final limitation of this study relates to the selection of participants. The study was conducted in Sydney, Australia. These speakers may probably all use English on a daily basis. This might have an impact on their Nepali VOT (see for example Flege, 1987). It would therefore be interesting to replicate this study with Nepali-speaking mothers residing in Nepal.

The Nepali community in Sydney has only a relatively small number of native speakers of Nepali. As a result, it was difficult to get a sample of infants that were all the same age, and we reported on a sample in which infant age ranged from 10-17 months. One of the limitations of this study is we did not explore the possibility of age effects. With the reports that age of the infants may affect hyper-articulation (Sundberg & Lacerda, 1999), there is a possibility that infant's age may have an effect on the results.

5.7 Future directions

Given the limitations identified above, there are many future directions which could be explored. The present study is the first to suggest that the trading relationship between consonant and vowel duration in a syllable is different in IDS than in ADS. It would be important to establish to what extent this difference in the trading relationship also exists in other languages than Nepali.

A related area for future research would be to analyse vowel space in Nepali IDS. Especially, given that vowels are longer in IDS, it would be valuable to know, we anticipate that vowel space may be hyper-articulated. If the vowels are hyper-articulated, it would show not only a trade-off in time spent in the duration of various constituents of the syllable, but it would actually show that they choose to hyper-articulate vowels at the expense of the hyper-articulation of consonants. This would mean that hyper-articulation is not an all-or-nothing situation, but may be subject to trade-offs.

To determine whether age of infant has an effect on the lead time and lag time, follow up study with large number of infants from different age groups could be compared. Given findings from Sundberg & Lacerda (1999) that the acoustic phonetic feature of IDS change with the age of the infant, it would be interesting to explore if the lead time and lag time show any effect of infants age. This could be examined with further analysis here and/or with a more restricted age group in Nepal.

It would also be interesting in future research to compare the amount of variance in production of the stop categories between the two registers. Although the results suggests smaller separation of stop categories in IDS compared to ADS, there may be less variability in the production of stops in IDS. If this is the case, then voicing contrast in IDS may still be enhanced compared to ADS.

Follow up research investigating infant's perception of voicing contrast would shed light on how Nepali infants acquire voicing contrasts. It would also be interesting to investigate when infants learn to produce these contrast.

It may be the case that mothers are enhancing the stop contrasts using other acoustic cues. Thus, it would also be interesting to examine other acoustic cues to voicing contrasts, such as fundamental frequency or vowel quality (cf. Clements and Khatiwada, 2007) in future studies.

Eliciting target words in isolation or in sentence initial position was more difficult in ADS than in IDS. In IDS mothers spontaneously repeated numbers of target words in isolation or in sentence initial position while interacting with their infants. However, in ADS target words were usually uttered in sentence medial or sentence final position. Moreover, the target tokens in sentence medial or final position were case marked, leading to additional suffixes and thus, were excluded from the analysis. However, children will typically encounter many words in utterance medial position or case marked, and therefore many stop consonants they hear will not occur at the beginning of utterances. Thus, the acoustics of these stop categories will vary across these different contexts. In further research it would be interesting to explore what the acoustics of such items are in mother's speech in IDS and how they are perceived and produced by children.

6. Conclusion

In summary, this study investigated the difference between the ADS and IDS in realization of a four-way voicing contrast in Nepali stops. Voicing cues in Nepali stops, which contrast in both voicing and aspiration, were measured using lead time and lag time. In addition, the occurrence of devoicing (complete absence of lead time) in voiced consonants was recorded as a binary variable. Vowel duration was measured to control for speaking rate. The aim of the current study was to test two hypotheses. The first hypothesis was that there would be hyper-articulation of devoicing, lead time, and lag time in IDS compared to ADS. The second hypothesis was that the hyper-articulation might be a side-effect of speaking rate differences between the register.

The results showed the absence of hyper-articulation of devoicing, lead time, and lag time in IDS. Rather, the higher rate of devoicing and shorter lag time led to hypo-articulation of voicing contrasts between the stop categories in IDS compared to ADS. Further, this difference between the categories could not be explained by speaking rate differences between the registers. The longer vowels in IDS suggest that mothers focus more on vowels while talking to infants, and this appears to have the effect of shortening the onset consonants.

To our knowledge, this is the first time that the voicing contrast in IDS is being looked at this way. This then has a wide ranging theoretical implications for better understanding the nature of IDS and its effects on language learning.

References

Abramson, A. S., & Lisker, L. (1970b). Discriminabiity along the voicing continuum: cross-language tests. *Proceedings of the Sixth International Congress of Phonetic Sciences. Prague, 1967, 569-573. Prague:Academia*.

Acharya, J. (1991). *A descriptive grammar of Nepali and an analysed corpus*. Washington, D.C.: Georgetown University Press.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. http://CRAN.R-project.org/package=lme4

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. doi: http://dx.doi.org/10.1016/j.infbeh.2013.09.001

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.14, retrieved 20 February 2016 from http://www.praat.org

Burnham, E., Gamache, J. L., Bergeson, T., & Dilley, L. (2013). Voice-onset time in infant-directed speech over the first year and a half. *Proceedings of Meetings on Acoustics, 19*(1), 060094. doi: doi:http://dx.doi.org/10.1121/1.4800072

Chen, Guan-yu (2014). *Case and agreement in Nepali*. Unpublished master's dissertation, National Tsing Hua University, Taiwan.

Cristià, A. (2010). Phonetic enhancement of sibilants in infant-directed speech. *The Journal of the Acoustical Society of America, 128*(1), 424-434. doi: doi:http://dx.doi.org/10.1121/1.3436529

Cristià, A., & Seidl, A. (2013). The hyperarticulation hypothesis of infant-directed speech*. *Journal of Child Language, 41*(4), 913-934. doi: 10.1017/S0305000912000669

Davis, K. (1994). Stop voicing in Hindi. *Journal of Phonetics*, 22, 177-193.

Davis, K. (1995). Phonetic and phonological contrasts in the acquisition of voicing: voice onset time production in Hindi and English. *Journal of Child Language*, 22(02), 275-305.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science, 171*(3968), 303-306.

Englund, K. T., & Behne, D. M. (2005). Infant Directed Speech in Natural Interaction—Norwegian Vowel Quantity and Quality. *Journal of Psycholinguistic Research, 34*(3), 259-280. doi: 10.1007/s10936-005-3640-7

Ferguson, C. A. (1977). Baby talk as a simplified register. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: language input and acquisition* (pp. 341-369). New York: Cambridge University Press.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*(1), 104-113. doi: 10.1037/0012-1649.20.1.104

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*(03), 477-501. doi: doi:10.1017/S0305000900010679

Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic Explanations for the Development of Tones. *Language, 55*(1), 37-58. doi: 10.2307/412518

Kanwal, J., & Ritchart, A. (2015). An experimental investigation of tonogenesis in Punjabi. Proceedings of ICPhS 18.

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*(5326), 684-686. doi: 10.1126/science.277.5326.684

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13-F21.

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-Language Experience in Infancy: Effects of Short-Term Exposure and Social Interaction on Phonetic Learning. *Proceedings of the National Academy of Sciences of the United States of America, 100*(15), 9096-9101.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Likser, L., & Abramson, A. S. (1964). A Cross- Lanugage Study of Voicing in Initial Stops: Acoustical Measurements. *Word, 20*, 384.

Lisker, L. (1986). “Voicing” in English: A catalogue of acoustic features signaling/b/versus/p/in trochees. *Language and speech*, *29*(1), 3-11.

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. doi: 10.1111/1467-7687.00275

Malsheen, B. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.), *Child phonology* (Vol. 2). New York: Academic Press.

Manders, and Christopher, J. (2007). *A foundation in Nepali grammar*. Bloomington: AuthorHouse.

Martin, A., Schatxz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., and Cristià, A. (2015). Mothers Speak Less Clearly to Infants Than to Adults: A Comprehensive Test of the Hyperarticulation Hypothesis. *Psychological Science,* 1-7.doi: 10.1177/0956797614562453

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101-B111. doi: http://dx.doi.org/10.1016/S0010-0277(01)00157-3

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*(2), 362-378. doi: http://dx.doi.org/10.1016/j.cognition.2013.07.015

Narayan, C., & Yoon, T.-J. (2011). VOT and F) in Korean Infant-Directed Speech. *Canadian Acoustics, 39*(3), 152-153.

Nepali- Ethnologue (2016, 1-08-2016). Retrieved 1-8-2016, from http://www.ethnologue.com/17/country/NP/

Pokharel, M. P. (1989). *Experimental analysis of Nepali sound system.* Unpublished doctoral dissertation, University of Pune, India.

Poon, P. G., & Mateer, C. A. (1985). A Study of VOT in Nepali Stop Consonants. *Phonetica, 42*(1), 39-47.

R Development Core Team. (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.http://www.R-project.org

Silva, D.J. (2006). Acoustic evidence for the emergence of a tonal contrast in contemporary Korean. *Phonology* 23: 287-308.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27*(4), 501-532. doi: http://dx.doi.org/10.1016/j.dr.2007.06.002

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognitiona). *The Journal of the Acoustical Society of America, 128*(1), 389-400. doi: doi:http://dx.doi.org/10.1121/1.3419786

Statistical year book of Nepal (13^{th} ed.) (2013). Central bureau of statistics, Ramshahpath, Nepal.

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. . *Working Papers, 29*, 148-151.

Sundberg, U., & Lacerda, F. (1999). Voice Onset Time in Speech to Infants and Adults. *Phonetica, 56*(3-4), 186-199.

Synnestvedt, A., Bernstein Ratner, N., & Newman, R. (2010). Voice onset time in infant-directed speech at 7.5 and 11 months. *The Journal of the Acoustical Society of America, 127*(3), 1853-1853. doi: doi:http://dx.doi.org/10.1121/1.3384386

Appendix A

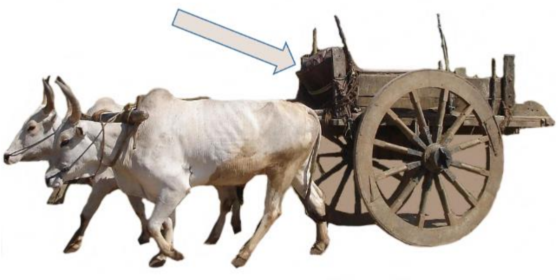
Table: Participant mother age (years) and infants age (months) and gender

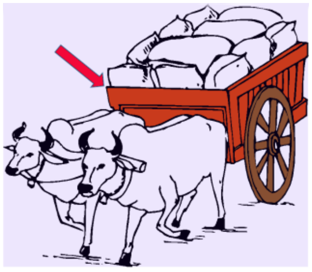
Subject	Mother's age	Infants age	Infant's gender
1	26.5	15.5	Male
2	31.6	10.12	Male
3	26	14.2	Male
4	33	14.3	Female
5	31	12.26	Male
6	29.1	14.3	Male
7	33.6	15.19	Female
8	26.8	15	Male
9	28.5	17.5	Female
10	34.8	14.22	Female
11	29.11	17.4	Female
12	32.5	17.28	Male
13	34.4	14.4	Male
14	28.6	10.21	Female
15	28.7	10.22	Female
16	32.1	12.08	Female

Appendix B

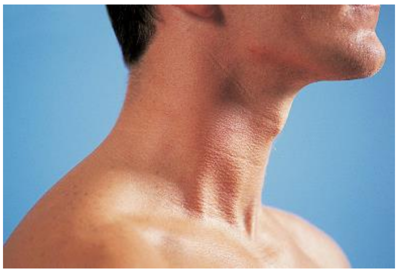
List of picture stimuli used

Bullock cart





Neck

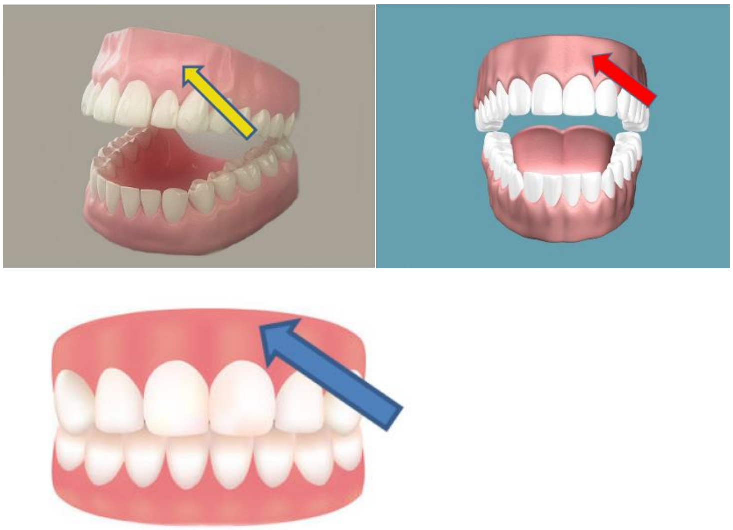


Hair pin

Meal

Cave

Gums



Appendix C

Questionnaire in English version

Language History Questionnaire

Participant ID:	Gender:	Date of Test:

Section I- Information about participant

1. What is your date of birth? ____________________________
2. Where do you live? (suburb and postcode)_________________________
3. Where is your birth place? (city and country)_____________________________
4. Which other countries than Nepal you lived in, and for how long? (in chronological order):

   a.____________________________

   b.____________________________

   c.____________________________
5. What is your native language? _______________
6. Which language(s) did you speak when you were growing up? ___________________
7. Which language(s) do you speak at home? _____________________
8. Please fill the following table regarding your language background.

What languages do you speak?	At what age did you begin to learn?	How often do you use it? (1(rarely) to 9 (very often))	How fluent are you in this language? (1(not fluent at all) to 9(very fluent)

9. Please tick the appropriate boxes for education attainment

	Mother
Primary School	
High School (year 10)	
High School (year 12)	

TAFE Certificate	
University Degree	
Postgraduate Degree	

10. Do you/your child have any hearing difficulties? ___________
11. Have you/your child had any ear infections? _______________
12. Do you/your child have any language difficulties? ___________

Section II- Information about participants' child

13. What is your child's date of birth? __________________
14. What is the gender of your child? _________________
15. What language(s) do you speak to the child?

    Mother: _______________________ Father: ________________________

16. Who is the child's main caregiver during the daytime? (I.e. father, mother, grandmother, etc.) ___________________________
17. If the main caregiver is not the mother/father, what language(s) do they speak to the child?

    ____

18. If your child is exposed to languages other than Nepali, which ones, and how many hours per week?
    Language 1: _________________ Hours/week: _______________

    Language 2: __________________ Hours/week: _______________

19. Please fill the following table regarding child's siblings.

How many siblings does your child have?	What is the age of sibling?	What language does the sibling use with the child?

20. What date (approx.) did your child begin attending childcare? ________________________

21. How many hours per week is your child in childcare? ____________________

Questionnaire in Nepali version

भाषा सम्बन्धि जानकारी फारम

खण्ड १- सहभागी आमा सम्बन्धि जानकारी

1. जन्म मिति :- ______________________________

2. ठेगाना :-___________________________

3. जन्म स्थान (सहर र देश): _______________________________

4. नेपाल बाहेक अरु कुन देशमा कति अवधिको लागि बस्नुभयो (क्रमसंग उल्लेखित गर्नु होस):

   a.______________________________

   b.______________________________

   c.______________________________

5. मात्रीभाषा:- ________________

6. हुर्किदा कुन- कुन भाषा प्रयोग गर्नु हुन्थेयो? ___________________

7. घरमा कुन भाषाको प्रयोग गर्नु हुन्छ? _____________________

8. भाषा प्रयोगसंग सम्बन्धित जानकारी पाउनका लागि तल दिएका प्रस्नहरुको कृपया जवाफ दिनुहोला |

हजुर कुन कुन भाषा प्रयोग गर्नुहुन्छ?	कुन उमेर देखि सिक्नु भएको ?	कतिको प्रयोग गर्नु हुन्छ (१ (एकदमै कम) देखि ९ (एकदमै धेरै))?	यो भाषा कति को आउछ (१ (त्यति आउदैन) देखि ९ (मज्जाले आउछ)) ?

9.आफ्नो शिक्षा अनुसार तलको कोठामा सहि चिन्ह लगाउनुहोस् |

	आमा
१ देखि ९ कक्षासम्म पढेको	
S.L.C. सम्म पढेको	
+2 सम्म पढेको	
TAFE बाट प्रमाण पत्र लिएको	
Bachelors गरेको	
Masters गरेको	

10. के हजुर वा हजुरको बच्चालाई सुनाई सम्बन्धि कुनै समस्या छ? ___________
11. के हजुर वा हजुरको बच्चाको कहिले कान पाक्ने समस्या थियो? ______________
12. के हजुर वा हजुरको बच्चालाई भाषा वा बोलि सम्बन्धि कुनै समस्या छ? _________

खण्ड २- सहभागी बच्चा सम्बन्धि जानकारी

13. जन्म मिति :- __________________
14. लिङ्ग :- _________________
15. आमा-बुवाले बच्चासंग कुन- कुन भाषा बोल्नुहुन्छ?

    आमा: ____________________ बुवा: _____________________

16. प्रायजसो दिउसो बच्चाको हेर विचार कसले गर्छ? (जस्तै : आमा, बुवा, हजुर-आमा,)

    ____

17. यदि बच्चाको प्रमुखहेर विचार गर्ने मान्छे आमा बुवा बाहेक अरु कोइ हो भने, उहाहरुले बच्चासंग कुन -कुन भाषा प्रयोग गर्नु हुन्छ?

    ____

18. के तपाइको बच्चा नेपाली बाहेक अरु भाषा बोल्ने वतावरणमा हुन्छ? हुन्छ भने, कुन भाषा र हप्ताको कति घण्टा हुन्छ उल्लेखित गर्नु होस

    भाषा १: _________________ हप्तामा कति घण्टा: ______________

    भाषा २: __________________ हप्तामा कति घण्टा: ______________

19. बच्चाको दाजु भाई दिदि बहिनी सम्बन्धि जानकारी

बच्चाको कति जना दाजु भाई दिदि बहिनी छन्? (क्रमसंग लेखनुस)	उनीहरुको उमेर कति भयो?	उनीहरुले बच्चासंग कुन भाषा प्रयोग गर्छन?

20. के बच्चा child/day care जाने गरेको छ? छ भने, कुन तारिखबाट जान थालेको हो?

____

21. हप्तामा जम्मा कति घण्टा बच्चा chid/day careमा हुन्छ?

Appendix D

Instructions

1) Familiarization

Instruction: "Now, I am going to show you some pictures. There are several names of these pictures. However, for the purpose of this study I want all the mothers to use the same name of these selected pictures while interacting with their infants. Hence, I have already selected a name for each. Now, I am going to show you each of these pictures one by one, I want you to name each picture when I show it to you. I will let you know if the selected name of the picture is what you have said or is it different than what you have said. Please remember there is no correct or wrong name, for the purpose of this study we want all the mothers to use the same name."

2) IDS

"I have placed the pictures into these three bags. I want you to play with your child using these pictures. You can open the bag and take out one picture at a time or go as you like. Introduce these to your child as you would when you are introducing a new item or a new person to your child at home or at any other situations. You can share your past experiences about these. Try to be as natural as possible. Take your own time and pace but try to use as much pictures as you can while you are interacting. We don't expect the child to produce these words in this session. Our aim is to look how infant in Nepali culture react to learning via pictures. "

3) ADS

"This is the last recording task. I would like to ask you few questions regarding your experience today."

Appendix E

Examples of questions used in ADS

1. Did you like the pictures? Which picture you liked the most?

2. Which pictures do you think you child liked the most?

3. Do you think a child raised in Australia will learn about all the words? If not, which are the ones s/he will have difficulty knowing?

4. Among these pictures, can you name three which you think your child likes the most?

5. Among these pictures, can you name three which you think your child did not like?

6. Among these pictures, can you name those which you think a child growing up in Australia will have difficulty learning compared to child in Nepali?

7. Among these pictures, can you name those which you think as a child growing up you had learnt earlier?

8. Among these pictures, can you name those which you think as a child growing up you had learnt later?

9. What do you think about these teaching language from these pictures? Which pictures are clear and which are not?

10. Do you think you child liked any of these pictures?

Appendix F

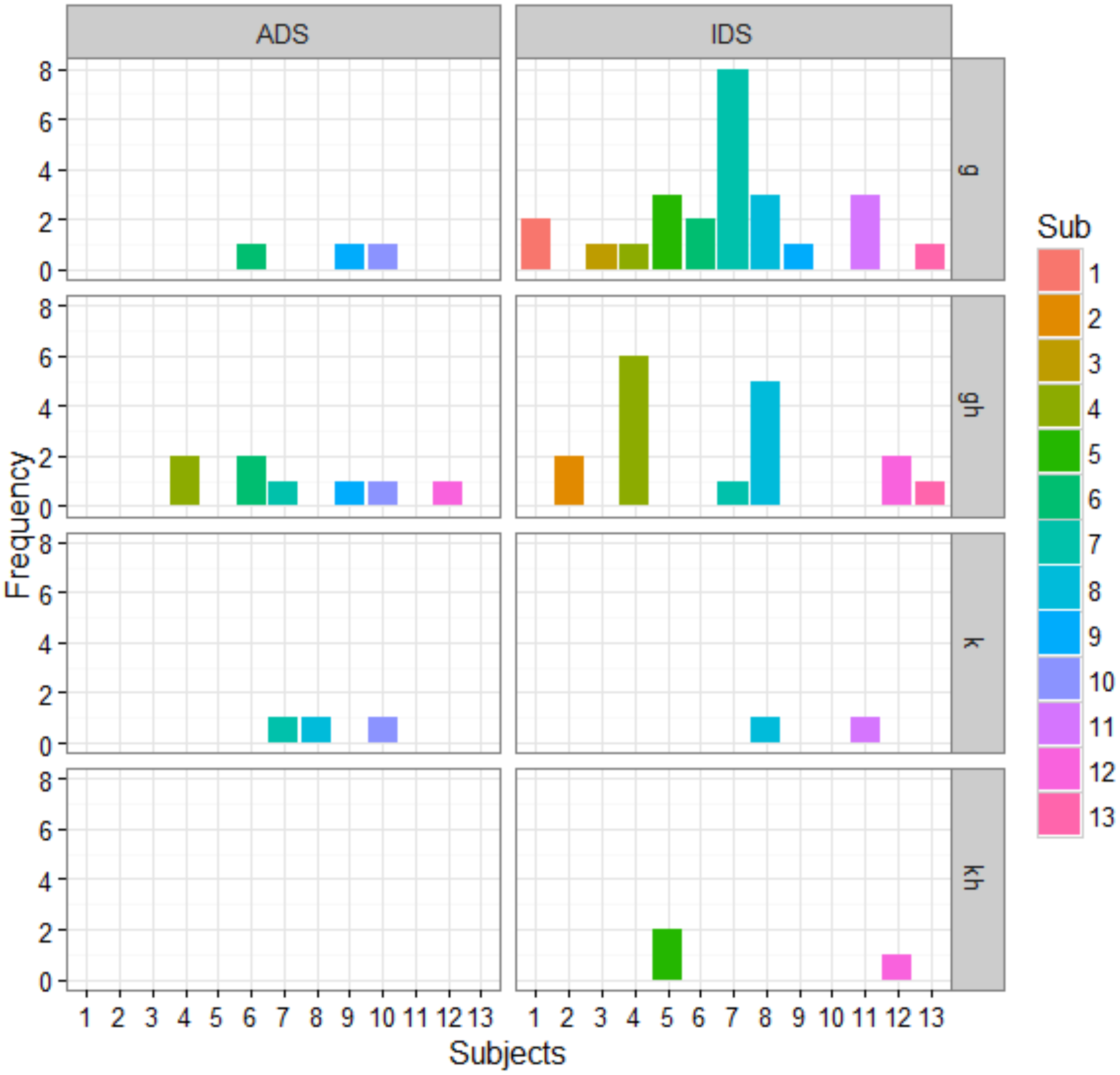


Figure showing number of outliers across 13 subjects between the two registers for each token. The panel on the left shows ADS and the panel on right shows IDS. The grid on top shows voiced unaspirated stops '/g/', followed by voiced aspirated /gʱ/, followed by voiceless unaspirated '/k/', and finally voiceless aspirated'/kʰ/' on the bottom. X-axis in both the panel shows number of subjects. Y-axis shows the number of outliers. The three subjects: '14', '15', and '16' did not have any outliers.

Appendix G

Post hoc analysis of interaction between aspiration and register

Results from 'aspirated' subset

Devoicing~ Register+(1+Register|Subject)							
	β	SE	z	p_{mcmc}	χ^2	df	p
Intercept	0.33	0.56	0.61	0.54			
Register	-0.19	0.31	-0.60	0.55			

Results from unaspirated

Devoicing~ Register+(1+Register|Subject)							
	β	SE	z	p_{mcmc}	χ^2	df	p
Intercept	4.09	1.56	3.15	0.002^{**}			
Register	-1.94	1.38	-1.40	0.16			

Results from 'ADS'

Devoicing~ Aspiration+(1+Aspiration|Subject)							
	β	SE	z	p_{mcmc}	χ^2	df	p
Intercept	3.49	1.64	2.12	0.03^{*}			
Aspiration	-2.98	1.44	-2.074	0.04^{*}			

Results from 'IDS'

Devoicing~ Aspiration+(1+Aspiration|Subject)							
	β	SE	z	p_{mcmc}	χ^2	df	p
Intercept	1.49	0.44	3.41	$<0.001^{***}$			
Aspiration	-1.37	0.31	-4.413	$<0.001^{***}$			

Three way interaction

Post hoc analysis of three-way interaction between aspiration, register, and voicing

'ADS voiced'

LagTime~ Aspiration+(1+Aspiration|Subject)							
	β	SE	t	p_{mcmc}	χ^2	df	p
Intercept	24.91	1.43	17.43	<0.001***			
Aspiration	2.30	0.95	2.42	0.03*			

'ADS voiceless'

LagTime~ Aspiration+(1+Aspiration|Subject)							
	β	SE	t	p_{mcmc}	χ^2	df	p
Intercept	65.83	1.73	37.98	<0.001***			
Aspiration	33.51	1.83	18.30	<0.001***			

'IDS voiced'

LagTime~ Aspiration+(1+Aspiration|Subject)							
	β	SE	t	p_{mcmc}	χ^2	df	p
Intercept	17.70	0.84	21.20	<0.001***			
Aspiration	1.28	0.59	2.18	0.05			

'IDS voiceless'

LagTime~ Aspiration+(1+Aspiration|Subject)							
	β	SE	t	p_{mcmc}	χ^2	df	p
Intercept	51.29	2.15	23.89	<0.001***			
Aspiration	27.35	1.59	17.24	<0.001***			

Voiceless-stop-to-syllable ratio

Post hoc analysis of interaction between aspiration and register

Results from 'aspirated' subset

Voiceless-stop-to-syllable ratio ~ Register+(1+Register|Subject)

	β	SE	t	p_{mcmc}	χ^2	df	p
Intercept	0.48	0.01	30.49	<0.001***			
Register	-0.06	0.01	-9.19	<0.001***			

Results from unaspirated

Voiceless-stop-to-syllable ratio ~ Register+(1+Register|Subject)							
	β	SE	t	p_{mcmc}	χ^2	df	p
Intercept	0.19	0.01	23.74	<0.001***			
Register	-0.03	<0.01	-8.70	<0.001***			

Appendix H

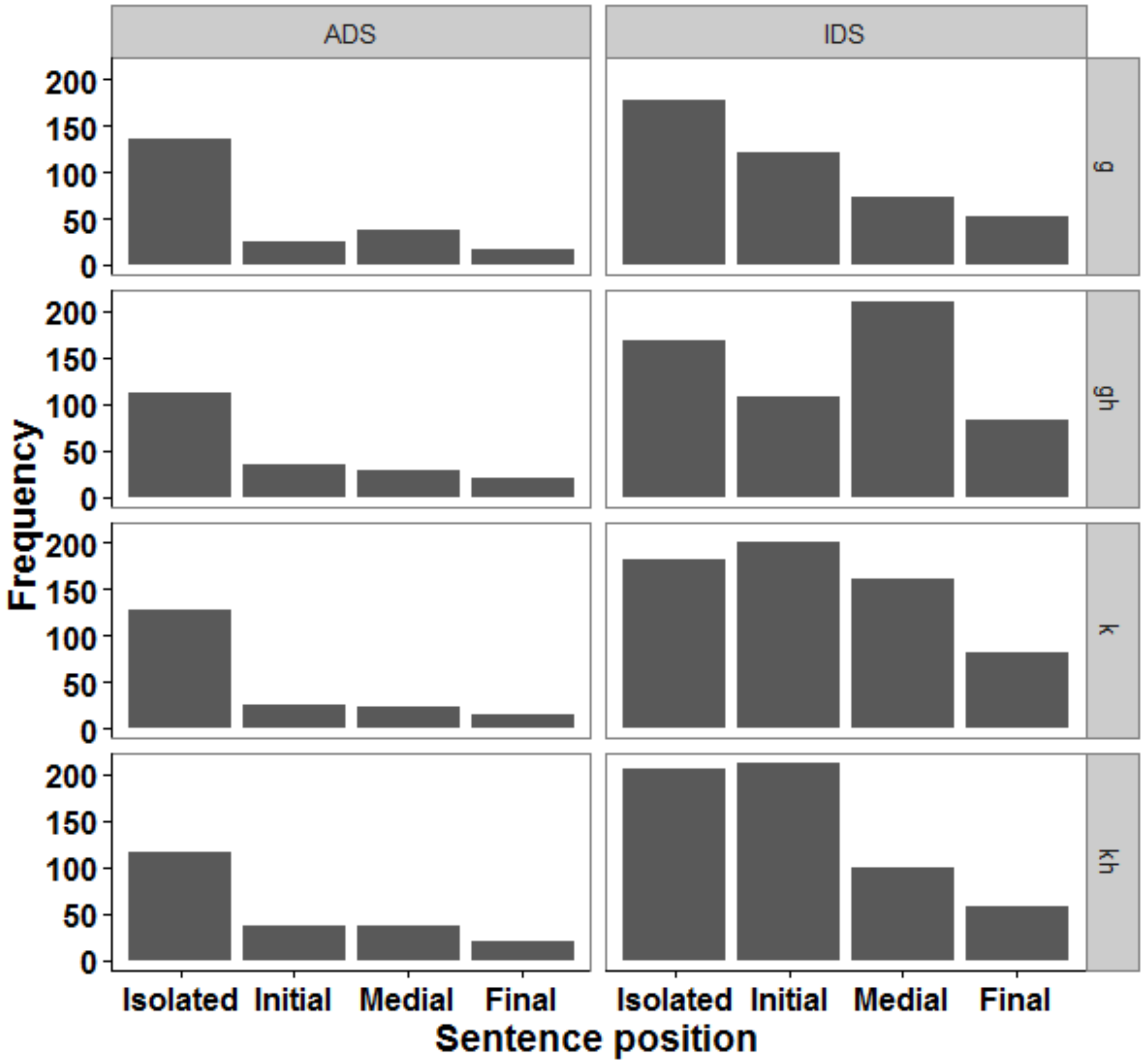


Figure showing number of target tokens across various sentence position (isolation, initial, medial, and final) between the two registers for each token. The panel on the left shows ADS and the panel on right shows IDS. The grid on top shows voiced unaspirated stops '/g/', followed by voiced aspirated /gʱ/, followed by voiceless unaspirated '/k/', and finally voiceless aspirated'/kʰ/' on the bottom. X-axis in both the panel shows the sentence position. Y-axis shows the number of words.

Appendix I

Dear Professor Demuth,

Re: "Acquisition of geminates and VOT contrasts for stops in Nepali-speaking children" (5201600078)

Thank you very much for your response. Your response has addressed the issues raised by the Faculty of Human Sciences Human Research Ethics Sub-Committee and approval has been granted, effective 5th April 2016. This email constitutes ethical approval only.

This research meets the requirements of the National Statement on Ethical Conduct in Human Research (2007). The National Statement is available at the following web site:

http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e72.pdf.

The following personnel are authorised to conduct this research:

Dr Anne Titia Benders
Mr Sujal Pokharel
Ms Elaine Schmidt
Professor Katherine Demuth

Please note the following standard requirements of approval:

1. The approval of this project is conditional upon your continuing compliance with the National Statement on Ethical Conduct in Human Research (2007).

2. Approval will be for a period of five (5) years subject to the provision of annual reports.

Progress Report 1 Due: 5th April 2017
Progress Report 2 Due: 5th April 2018
Progress Report 3 Due: 5th April 2019
Progress Report 4 Due: 5th April 2020
Final Report Due: 5th April 2021

NB. If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. If the project has been discontinued or not commenced for any reason, you are also required to submit a Final Report for the project.

Progress reports and Final Reports are available at the following website:

http://www.research.mq.edu.au/current_research_staff/human_research_ethics/application_resources

3. If the project has run for more than five (5) years you cannot renew approval for the project. You will need to complete and submit a Final Report and submit a new application for the project. (The five year limit on renewal of approvals allows the Sub-Committee to fully re-review research in an environment where legislation, guidelines and requirements are continually changing, for example, new child protection and privacy laws).

4. All amendments to the project must be reviewed and approved by the Sub-Committee before implementation. Please complete and submit a Request for Amendment Form available at the following website:

http://www.research.mq.edu.au/current_research_staff/human_research_ethics/managing_approved_research_projects

5. Please notify the Sub-Committee immediately in the event of any adverse effects on participants or of any unforeseen events that affect the continued ethical acceptability of the project.

6. At all times you are responsible for the ethical conduct of your research in accordance with the guidelines established by the University. This information is available at the following websites:

http://www.mq.edu.au/policy

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/policy

If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.

If you need to provide a hard copy letter of approval to an external organisation as evidence that you have approval, please do not hesitate to contact the Ethics Secretariat at the address below.

Please retain a copy of this email as this is your official notification of ethics approval.

Yours sincerely,

Dr Anthony Miller
Chair
Faculty of Human Sciences
Human Research Ethics Sub-Committee

Faculty of Human Sciences - Ethics
Research Office
Level 3, Research HUB, Building C5C

Dear Professor Demuth,

RE: 'Voicing contrasts in Nepali infant-directed speech ' (Ref: 5201600078)

Thank you for your recent correspondence regarding the amendment request.

The request has been reviewed and the amendment has been approved.

Please accept this email as formal notification of approval and find the attached for your records. Please do not hesitate to contact us in case of any further queries.

All the best with your research.

Kind regards,

FHS Ethics
****
Faculty of Human Sciences - Ethics
Research Office
Level 3, Research HUB, Building C5C
Macquarie University
NSW 2109