DURATIONAL CUES TO PLACE AND VOICING CONTRASTS IN AUSTRALIAN ENGLISH WORD-INITIAL STOPS

Julien Millasseau[1], Laurence Bruggeman[1,2], Ivan Yuen[1], Katherine Demuth[1]
[1]Department of Linguistics & ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, Australia; [2]The MARCS Institute & ARC Centre of Excellence for the Dynamics of Language, Western Sydney University
Julien.Millasseau | Ivan.Yuen | Laurence.Bruggeman | Katherine.Demuth@mq.edu.au

ABSTRACT

This study examined the durational characteristics of word-initial voicing and place of articulation (PoA) contrasts in Australian English-speaking adults, providing the first comprehensive study of this kind. Voice onset time (VOT) and closure duration (CD) for /b, d, g, p, t, k/ were measured in word-initial position of CVC words. The results showed that voiced stops have a short VOT and a long CD, while voiceless stops show the opposite pattern. In addition, the results showed that VOT alone failed to contrast /t/-/k/, and CD failed to contrast /d/-/g/. This might be related to the difficulty of maintaining appropriate aerodynamic conditions for voicing and distinguishing alveolars from velars. The results suggest that there is a trading relationship between VOT and CD, and that both cues are necessary in Australian English to contrast for all oral stop contrasts.

Keywords: VOT, closure duration, voicing, place of articulation, Australian English.

1. INTRODUCTION

The acoustic parameters used to signal phonological contrast may vary from one dialect to another. In English, voice onset time (VOT) can be used to contrast voiced and voiceless oral stops (e.g., /b/ - /p/), with shorter VOT for voiced stops than for voiceless stops [7, 17]. VOT may also be used to differentiate places of articulation (PoA; bilabial – alveolar – velar) [e.g., 17]. Despite consistency in voicing organisation across English dialects, the use of VOT to signal PoA is not always consistent [8]. Typically, the VOT of alveolar and velar stops is longer than that of bilabial stops, and alveolar stops may not differ from velars in VOT [10, 21], though this may depend on voicing [3]. For instance, some studies of American [6, 8, 17], Canadian [25] and British English [9] report a significantly longer VOT for velars compared to alveolars, irrespective of voicing, while other British [10] and Scottish English [21] studies found no difference for voiceless stops (i.e., /t/ = /k/). Thus, different English dialects may vary in the extent to which VOT can be used as a cue to PoA.

To our knowledge, only two studies have previously reported VOT values for Australian English (AusE) stop consonants in word-initial position [1, 13] as produced by adults. Voiceless stops in [13] had longer VOT than voiced stops, /b/ was realized with pre-voicing, and no difference was found between the VOTs of /t/ and /k/. However, these results were obtained from only four speakers (all females) of a variety of AusE spoken in Katherine, Northern Territory, raising questions about their generalizability. [1] examined the production of /Ca/ syllables (/ba-da/; /pa-ta/) by AusE speakers in New South Wales. Like [13], /b/ was realized with pre-voicing. However, the study did not include /g-k/ and the test items were not real words.

In addition to VOT, the preceding closure duration (CD) of a stop consonant has been suggested as a potential cue for both voicing and PoA in English [6, 16, 19, 20]. However, most studies have only looked at CD in word-medial position. In this position, voiced stops appear to have a shorter CD than voiceless stops [16, 19]. Counter to the pattern in word-medial position, two studies suggested that voiced stops have a longer CD than voiceless stops in word-initial position [11, 24]. Unfortunately, no statistical analyses were performed, and only the production of bilabials was examined in both studies.

Studies which have looked at CD with respect to PoA contrasts have reported consistent results when CD is measured in word-medial position [19, 20]. The CD of alveolar stops in American English is reported to be shorter than that for both bilabials and velars, with the CD of bilabials being the longest. However, no study has yet investigated the production of CD in regard to PoA contrasts in word-initial position.

Previous research has suggested that voicing and PoA influence one another and therefore should not be assessed independently. For instance, [3] showed that phonological voicing moderates the use of VOT to contrast for PoA. Thus, the productions of voicing and PoA contrasts have to be examined in relation to one another since these two may interact.

The present study therefore systematically assessed the production of voicing and PoA contrasts in word-initial stops in AusE-speaking adults, with a focus on how AusE speakers use VOT and CD to make phonemic contrasts. We predicted that voicing would be contrasted using VOT and CD, with shorter VOT and longer CD for voiced than voiceless stops. We also predicted that most PoAs would be contrasted using both VOT and CD.

2. METHODS

2.1. Participants

Ten adults (5M, 5F) participated in this study (mean age: 33 years; range: 21-40 years). Participants were recruited from the participant recruitment system of Macquarie University, Sydney. All participants were born in Australia to Australian-born parents and had only been exposed to Australian English during childhood. The study was approved by Macquarie University's Human Ethics Panel. Each participant received course credit for their participation.

2.2. Stimuli

The stimuli consisted of 18 high-frequency CVC near minimal pair words with a mean frequency of 4.5 Zipf in the Subtlex-UK CBeebies pre-schooler corpus [23]. All 6 English stops were included in word-initial position in three different vowel contexts of CVC words (Table 1). Each noun target word (n = 16) was embedded in the carrier sentence "See this X". Verb target words (n = 2) were embedded in the sentence "These mice X". The audio stimuli were produced by a 25-year-old female native speaker of AusE.

Table 1: List of target words.

	/b/	/d/	/g/	/p/	/t/	/k/
/ɪ/	bib	dig	give	pig	tip	kid
/ɐ/	bug	duck	gut	puff	tub	cup
/ɔ/	bomb	dog	god	pot	top	cob

2.3. Procedure

Testing was conducted in a sound-attenuated room at Macquarie University. Participants were seated at a table facing an iPad. A microphone (AKG C535 EB) was placed 30cm from the participant's mouth and was connected by an XLR cable to a computer in a control room via a pre-amplifier (Sound Devices, USBPre2). The stimuli were presented to participants on an iPad Air using Keynote software. Participants were instructed to touch the picture on the screen to start the presentation and to advance from one trial to another. When participants touched the screen, the sentence associated to the picture was played, and participants were instructed to repeat after each sentence. Three practice trials were used to familiarize the participants with the procedure. Items were pseudo-randomized to create five different blocks. Each participant produced a total of 15 tokens per stop (3 vowel contexts * 5 blocks) for a total of 90 tokens (6 stops * 3 vowels * 5 blocks) per participant. The task took about 15 minutes to complete. Recordings were captured and encoded as mono WAV files using Audacity with a 44.1 kHz sampling rate and 16-bit quantization.

2.4. Acoustic coding and Analysis

Acoustic analysis was conducted in Praat [4] by the first author. VOT was measured from the first peak of the release burst to the beginning of vowel periodicity, as marked by a strong F2. In cases where participants produced multiple bursts, VOT was measured from the first peak of the first release burst. CD was measured from the end of the preceding segment (i.e., /s/) until the first peak of the release burst. This measurement allowed us to detect potential pre-voicing since it is reported that AusE voiced stops are sometimes realised with lead voicing [1, 13]. A total of 856 tokens were analysed for VOT and 866 tokens for CD, after excluding 44 outliers from the former (5.4%) and 34 outliers from the latter (4%). Outliers were defined as tokens +/- 2 standard deviations from the means of VOT and CD.

3. RESULTS

3.1. VOT

Mean VOTs for each stop are displayed in Table 2. Figure 1 shows the distribution of VOTs for each PoA across Voicing categories.

Table 2: Mean VOT in ms (SD) for each stop.

/b/	/d/	/g/	/p/	/t/	/k/
9 (4)	15 (4)	24 (6)	63 (15)	78 (16)	77 (13)

VOT measurements were analysed with a linear mixed-effects regression model using the packages lme4 [2] and lmerTest [14] in R [22]. Fixed factors were all main effects of and the interaction between Voicing and PoA. Both fixed factors were coded using Helmert contrasts with Voiced = 1 and Voiceless = -1 for Voicing, and PoA-1 (= mean of bilabials vs. mean of non-bilabials) and PoA-2 (= mean of alveolars vs. mean of velars) for PoA. By-subject and by-item random intercepts were added, as well as random slopes by subject for Voicing and PoA.

Results (as shown in panel a of Table 4) showed a main effect of Voicing, PoA-1 (i.e., mean of bilabials vs. mean of non-bilabials) and PoA-2 (i.e., mean of alveolars vs. mean of velars).

Figure 1: VOT distribution for each PoA across voicing (yellow dot represents the mean).

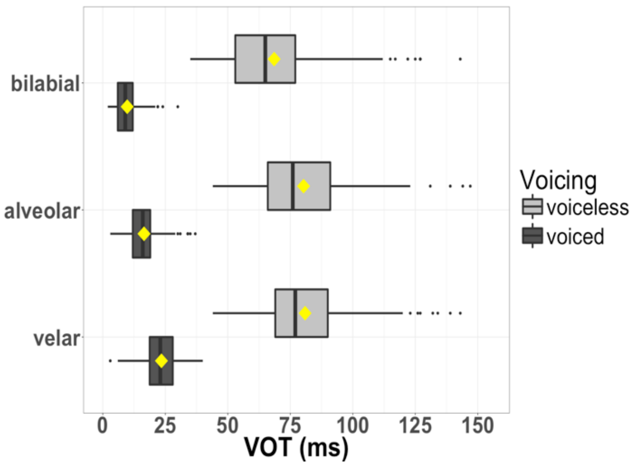
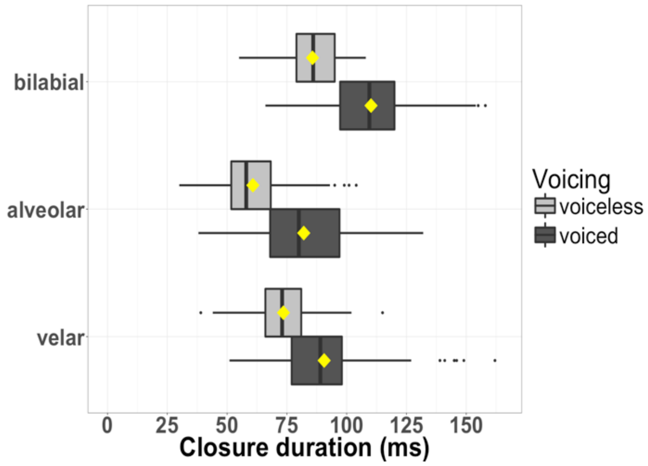


Figure 2: CD distribution for each PoA across voicing (yellow dot represents the mean).



In addition, the model revealed a 2-way interaction between Voicing and PoA-2. Tukey (adjusted α-levels) post-hoc comparisons [14] were conducted and showed that the VOT of voiced stops was significantly shorter than for voiceless stops across each PoA, in line with our predictions.

As regards PoA, the VOT of bilabials was significantly shorter than that of non-bilabials in both voicing conditions. However, the Voicing * PoA-2 interaction indicates that the mean VOT of alveolars is shorter than that of the velars in the voiced condition, but not in the voiceless condition, in line with our predictions.

3.2. Closure duration

Mean CDs are displayed in Table 3 for each stop. Figure 2 shows the distribution of CD for each PoA across the two voicing categories. A linear mixed-effects model similar to the one fitted for VOT in section 3.1. was fitted on the CD data. Results are displayed in Table 4 (b). Effects of Voicing, PoA-1 (i.e., mean of bilabials vs. mean of non-bilabials) and PoA-2 (i.e., mean of alveolars vs. mean of velars) were observed. Since PoA is a three-level factor, we ran a post-hoc analysis (Tukey adjusted α-levels) to determine the significance levels of all pairwise comparisons. The post-hoc revealed that all pairwise comparisons were significant except for /d/-/g/.

Table 3: Mean CD in ms (SD) for each stop.

/b/	/d/	/g/	/p/	/t/	/k/
110 (14)	85 (17)	88 (15)	87 (9)	58 (10)	73 (9)

4. DISCUSSION

The present study investigated adults' production of voicing and PoA contrasts in AusE by looking at VOT and CD. Our results showed that both VOT and CD are used to contrast voicing, with shorter VOT and longer CD for voiced stops. Our VOT results are consistent with previous studies of voicing in other dialects of English [e.g., 6, 8, 10, 21]. The present results on CD showed that in word-initial position, voiced stops have a longer CD than voiceless stops. This is in line with previous predictions on CD measured in word-initial position and our findings therefore provide empirical evidence for such descriptions as postulated in [11, 24]. They also corroborate similar findings in a study on AusE-speaking children acquiring voicing [5]. Although CD may have been conflated with a pause (if present) between the carrier phrase and the target word, the values we observed (especially for voiceless stops) are similar to those found in a corpus study by [12]. It is therefore not likely that our findings were affected by pauses.

Table 4: Results of the linear mixed-effects models of VOT (panel a) and of CD (panel b).

Effects	**a) VOT**			**b) CD**		
	Est.	**SE**	***t*-value**	**Est.**	**SE**	***t*-value**
(Intercept)	44.60	1.34	33.25*	83.26	2.44	34.12*
Voicing	28.38	0.51	55.42*	-10.72	0.69	-15.50*
PoA-1 (bilabial vs. non-bilabial)	-12.76	1.08	-11.81*	22.60	1.47	15.34*
PoA-2 (alveolar vs. velar)	-4.06	1.25	-3.26*	-10.95	1.69	-6.48*
Voicing * PoA-1 (bilabial vs. non-bilabial)	-1.86	1.08	-1.72	-1.99	1.47	-1.35
Voicing * PoA-2 (alveolar vs. velar)	4.61	1.25	3.70*	-3.89	1.69	-1.70

Voicing contrasts in AusE exhibited short VOT and long CD for voiced stops, but long VOT and short CD for voiceless stops. This suggests a trading relationship between VOT and CD for the voicing distinction in word-initial stops. Thus, the discrepancy between the present study and previous reports might be related to the difference in sentence position, as most previous reports on CD were based on stops in intervocalic post-stress position [e.g., 16, 19].

As regards PoA contrasts in AusE, it is interesting to note that VOT did not contrast voiceless /t/ from /k/, stops that both have fairly long periods of aspiration. Similarly, CD failed to contrast voiced /d/ from /g/, stops with a PoA that involves a short closure period. This might be related to the difficulty speakers experience in maintaining appropriate aerodynamic conditions for voicing and distinguishing alveolars from velars at the same time. However, the present study observed a trading relationship between CD and VOT, suggesting that both cues may be needed to distinguish PoA in onset position. These results provide a much-needed baseline for further investigation of how children (including those with hearing loss) acquire voicing and PoA contrasts, which are known to be a challenge, and how both children and adults learning English as a second language master these contrasts as well.

5. ACKNOWLEDGMENTS

This work was funded by the ARC Laureate Fellowship (FL130100014) awarded to the last author and by the ARC Centre of Excellence in Cognition and its Disorders (CE110001021). We thank our participants, and the members of the Child Language Lab for helpful feedback and comments.

6. REFERENCES

[1] Antoniou, M., Best, C. T., Tyler, M. D., Kross, C. 2010. Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code- switching, *J. Phon*. 39, 558-570.

[2] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4, *J. Stat. Softw*. 67, 1-48.

[3] Benkí, J.R. 2001. Place of articulation and first formant transition pattern both affect perception of voicing in English. *J. Phon.* 29, 1-22.

[4] Boersma, P., Weenink, D., 2018. Praat: doing phonetics by computer [Computer software]. Version 6.0. 37.

[5] Bruggeman, L., Millasseau, J., Yuen, I., Demuth, K. 2018. The production of voicing and place of articulation contrasts by Australian English-speaking children, *Proc. 17th SST,* Sydney.

[6] Byrd, D. 1993. 54,000 American stops. *UCLA Work. Papers Phonet.* 83, 97-116.

[7] Cho, T., Ladefoged, P. 1999. Variation and universals in VOT: evidence from 18 languages. *J. Phon.* 27, 207–229.

[8] Chodroff, E., Wilson, C. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *J. Phon*. 61, 30-47.

[9] Deterding, D., Nolan, F. 2007. Aspiration and voicing of Chinese and English plosives, *Proc. 16th ICPhS*, Saarbrücken, 385-388.

[10] Docherty, G. J. 1992. *The timing of voicing in British English obstruents* (Vol. 9), Berlin: Walter de Gruyter.

[11] Flege, J.E., Brown, W.S. 1982. The voicing contrast between English/p/and/b/as a function of stress and position-in-utterance. *J. Phon.* 10, 335-345.

[12] Ghosh, P.K. and Narayanan, S.S., 2009. Closure duration analysis of incomplete stop consonants due to stop-stop interaction., *J. Acoust. Soc. Am.* 126, pp.1-7.

[13] Jones, C., Meakins, F. 2013. Variation in Voice Onset Time in stops in Gurindji Kriol: picture naming and conversational speech, *Aust. J. Linguist* 33, 196–220.

[14] Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. 2017. lmerTest package: Tests in linear mixed effects models, *J. Stat. Softw.* 82, 1-26.

[15] Lenth, R.V. 2016. Least-Squares Means: The R Package lsmeans, *J. Stat. Softw.* 69, 1-33.

[16] Lisker, L. 1957. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, 33, 42–49.

[17] Lisker, L., Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384–422.

[18] Lisker, L., Abramson, A. S. 1970. The voicing dimension: some experiments in comparative phonetics. Proceedings of the Sixth International Congress of Phonetic Resources. Prague, 563-567.

[19] Port, R. F. 1979. The influence of tempo on stop closure duration as a cue for voicing and place. *J. Phon.* 7, 45–56.

[20] Repp, B. H. 1984. Closure duration and release burst amplitude cues to stop consonant manner and place of articulation, *Lang. Speech* 27, 245-254.

[21] Stuart-Smith, J., Sonderegger, M., Rathcke, T. 2015. The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian, *Lab. Phonol*. 6, 505-549.

[22] R Core Team. 2018. R: A Language and Environment for Statistical Computing.

[23] Van Heuven, W., Mandera, P., Keuleers, E., Brysbaert, M. 2014. Subtlex-UK: A new and improved word frequency database for British English, *Q. J. Exp. Psychol.* 67, 1176-1190.

[24] Westbury, J.R., Keating, P.A. 1986. On the naturalness of stop consonant voicing. *J. Linguist.* 22, 145-166.

[25] Yu, V. Y., De Nil, L. F., Pang, E. W. 2015. Effects of age, sex and syllable number on Voice Onset Time: evidence from children's voiceless aspirated stops, *Lang. Speech* 58, 152–167.