Voice Onset Time in a language without voicing contrast: An acoustic analysis of Blackfoot oral stops

Inge Genee [ORCID]
University of Lethbridge
Inge.genee@uleth.ca

Fangfang Li [ORCID]
University of Lethbridge
Fangfang.li@uleth.ca

This paper presents an acoustic analysis of Voice Onset Time (VOT) in oral stop consonants in Blackfoot, an Algonquian language without contrastive voicing. We focus on VOT as one of the key temporal acoustic correlates of voicing and investigate VOT variation in relation to (i) place of articulation (labial vs. alveolar vs. velar); (ii) length (long vs. short), quality (/a/ vs. /o/), and accent pattern (accented vs. unaccented) of the following vowel; (iii) word position (initial vs. medial); (iv) gender; and (v) age. We analyzed 2096 stop consonant tokens produced by 13 participants, who completed two different tasks: an English-to-Blackfoot translation task and a picture naming task. The key findings are as follows: (i) Blackfoot stop consonants fall into the short-lag range that overlaps with the English voiced category, with mean VOT values ranging from 11.6–32.7 ms; (ii) VOT values become progressively longer as the place of articulation moves to more posterior positions; (iii) VOT values are longer before high vowels than before low vowels; (iv) VOT values are longer before long vowels than before short vowels; (v) no statistically significant effect was found for the linguistic factors word position and accent pattern; (vi) no statistically significant effect was found for the socio-indexical factors age and gender; and (vii) no statistically significant effect was found for the experimental factor task type. The implications of our findings for the Blackfoot writing system and ongoing work on language documentation and revitalization are briefly discussed.

1 Introduction

The purpose of this paper is to examine Voice Onset Time (VOT), one of the major acoustic correlates of voicing, in oral stops in Blackfoot (ISO 639-3: bla), an Algonquian language spoken in Southern Alberta, Canada and Northern Montana, USA (Frantz 2017). Voicing is not distinctive in Blackfoot: all obstruents (oral and glottal stops, fricatives and affricates) are phonemically voiceless, and all nasal stops, approximants and vowels are phonemically voiced (Frantz 2017, Weber 2020). (The full phonemic inventory is presented in Section 2.) The lack of voicing contrast goes back to Proto-Algonquian (Bloomfield 1946;

Goddard 1974, 1979; Thomason 2006) and is well-documented for the daughter languages (e.g. Arapaho and Atsina: Goddard 1974, 1979; Kakadelis 2018; Cree: Mackenzie 1980; Wolfart 1996; Okimāsis & Wolvengrey 2008; Wolvengrey 2011: 9–10; Hodgson 2021; Michif: Rosen et al. 2019). To our knowledge there are very few if any instrumental investigations of the acoustic properties of Algonquian oral stops. Descriptions tend to be impressionistic and usually define them simply as voiceless unaspirated. What little acoustic work there is appears to support this. Recent work on Cree by Hodgson (2021) shows VOT in the speech of one speaker of Plains Cree in the short-lag (voiceless unaspirated) range. Kakadelis (2018) reports similar results for coronal and dorsal stops in Arapaho, with its labial stop best described as phonetically voiced, occurring commonly with negative VOT. Rosen et al. (2019) draw comparable conclusions for Michif; this last language is particularly interesting since Michif is a mixed language with source vocabulary from French and Cree, but the data suggest a phonological system with a single set of oral stops undifferentiated by voicing, as in Cree, rather than a double set, as in French. The only work available on Blackfoot is a preliminary study reported by Hall (2010) based on one speaker, suggesting again that average VOT values are in the short-lag range, fairly close to those for word-initial voiced stops in English. The current investigation is the first systematic acoustic study of the Blackfoot oral stop consonants.

Hodgson (2021) introduces the term 'single series languages' for languages without a stop voicing contrast. Investigations of single-series languages are generally rare compared to those with voicing contrast, presumably because the lack of a phonological contrast somehow makes them less interesting from a descriptive and theoretical point of view. For instance, a recent special issue of *Journal of Phonetics* (Cho, Docherty & Whalen 2018a) entitled *Marking 50 Years of Research on Voice Onset Time* does not include a single language without voicing contrast among the 19 languages discussed in 11 contributions (Cho, Whalen & Docherty 2018b); and an important textbook on acoustic analysis of speech discusses the acoustics of VOT exclusively in the context of its contribution to the voicing distinction (Kent & Read 2002: 150–154). A recent example of a detailed study of three languages without a voicing contrast is Kakadelis (2018), who mentions that another reason for the relative lack of studies of languages without voicing contrast may be the fact that such languages are often 'endangered, under-documented, or not spoken widely' (1).

Nevertheless, there are some obvious theoretical, historical, descriptive and applied reasons for examining VOT in single-series languages. Theoretical reasons include questions about the range of VOT values occurring in such languages, and the effect of the phonetic environment on their production: a central question is whether languages without a voicing contrast have the same or similar VOT values and allophonic variation, or whether such values are language-specific in the same way that languages with a voicing contrast may choose a different range along the VOT spectrum to express their phonemic distinctions (Cho et al. 2018b, Garvin 2018, Kakadelis 2018). Historical reasons include questions about what happens in language contact situations when one of the languages in contact has a voicing contrast and the other one does not (Rosen et al. 2019; Hodgson 2021: 11). Descriptive reasons include the fact that, depending on one's definition of 'voicing contrast', up to 20% of the world's languages have been estimated not to have a voicing distinction (Maddieson 2013), implying that this is not a rare phenomenon and does deserve attention. Applied reasons include questions related to (first and second) language learning and accent acquisition (Riney & Takagi 2008; Kong, Beckman & Edwards 2012; Turner et al. 2014; Netelenbos, Li & Rosen 2016; Kakadelis 2018: 37–38), and orthography design for minority languages in bilingual contexts where the majority language (which tends to be the one whose orthography is adapted to be used for the minority language) has a voicing distinction and the minority one does not (Hinton 2014, Garvin 2018). In fact, the work reported in this paper was originally prompted by a desire to better understand the continuing disagreement in the Blackfoot-speaking community about which letters to use to write Blackfoot oral stops: the <p t k> series, as in the standard orthography developed by linguist Donald G. Frantz

(Frantz 1978, 1993, 2017), or the <b d g> series, as preferred by many community members, at least in some environments (Genee 2020). The frequent preference to use <b d g> in representing Blackfoot stops is partly due to the influence of the English orthography, which denotes its series of phonetically voiceless unaspirated stops with these three letters.

After Lisker & Abramson (1964) introduced VOT as a way of classifying voicing distinctions in oral stops, a large body of work has explored VOT as an acoustic measure that helps distinguish homorganic stops into commonly two or three, or occasionally four phonemically distinct categories in individual languages (Abramson & Whalen 2017, Cho et al. 2018b).[1] From an articulatory perspective, VOT is 'defined as the temporal relation between the moment of the release of the stop and the onset of glottal pulsing' (Abramson & Whalen 2017: 76). Taking the burst associated with the stop release as the zero point, VOT is then measured in milliseconds, with negative values when the voicing begins before the burst (voicing lead), and positive values when it begins after the burst (voicing lag).

Subsequent work has refined the nature and measurement of VOT in word-initial vs. word-medial and word-final position, expanded its application to fricatives and affricates (Abramson & Whalen 2017: 79–81), and identified other factors affecting VOT, such as place of articulation, length and quality of the following sonorant, and prominence. Word-initial VOT values tend to be longer than word-medial ones, especially when the word-medial stop is intervocalic or intersonorant (Flege & Brown 1982, Kakadelis 2018); intervocalic stops are often fully voiced during the closure, with a negative VOT (Davidson 2016). More anterior places of articulation are associated with shorter VOT values (e.g. /t/ shows shorter VOT than /k/, Lisker & Abrahamson 1964, Klatt 1975, Cho & Ladefoged 1999, Abdelli-Berruh 2009, Bijankhan & Nourbaksh 2009, Kakadelis 2018, Chodroff, Golden & Wilson 2019). VOT values are longer before long vowels compared to short vowels (Taff et al. 2001, Hodgson 2021). Following high vowels are sometimes associated with longer VOT (Klatt 1975, Bijankhan & Nourbaksh 2009, Garvin 2018, Hodgson 2021). Presence of prominence on the syllable following the stop is likewise correlated with longer VOT (Davidson 2016).

Language-specific social correlations have also been reported. Social correlates of VOT include gender (e.g. Auzou et al. 2000, Bijankhan & Nourbakhsh 2009, Li 2013, Kharlamov 2018, Kleber 2018) and age (e.g. Auzou et al. 2000, Kleber 2018). Many of these correlations, however, are confounded by speech rate, which varies with age and gender (Swartz 1992, Morris, McCrea & Herring 2008). As Cho et al. (2018b: 59–60) mention, sociophonetic work in the area of VOT is scarce and remains very much a desideratum in the field.

Languages with a stop voicing contrast usually have contrastive voicing associated with two or three VOT ranges, depending on the nature of the contrast. In so-called 'aspirating' two-way languages such as English, German and Persian, both categories are mainly associated with positive VOT: the 'voiced' category is associated with a short-lag positive VOT, while the 'voiceless' category is associated with a long-lag positive VOT (e.g. Lisker & Abramson 1964, Auzou et al. 2000, Bijankhan & Nourbakhsh 2009, Beckman, Jessen & Ringen 2013, Cho et al. 2018b, Kakadelis 2018). In so-called 'true voicing' two-way languages such as French, Spanish, Russian, Hungarian, and many Arabic dialects, the 'voiced' category is associated with negative VOT (pre-voicing) while the 'voiceless' category is associated with short-lag VOT (e.g. Flege & Eefting 1987, Auzou et al. 2000, Beckman et al. 2013, Netelenbos et al. 2016, Al-Tamimi & Khattab 2018, Cho et al. 2018b). Some languages

---

[1] VOT is of course not the only temporal acoustic correlate of the voicing contrast in stops. Other correlates include stop duration, closure duration, release duration, duration of preceding and/or following vowel, fundamental frequency (f0) of the following vowel, and spectral tilt (Abdelli-Beruh 2009, Davidson 2016, Kakadelis 2018, Jesus & Costa 2020, Lousada, Jesus & Hall 2010, Kong et al. 2012, Pape & Jesus 2015). Nevertheless, VOT is highly correlated to the voicing contrast and one of its most salient indicators (Hodgson 2021: 12–13), and as such remains a useful way to begin the investigation of the laryngeal properties of stops, especially in the case of a language in which this has not been investigated before.

such as Swedish use non-adjacent ranges, skipping the short-lag range to contrast negative VOT with long-lag VOT (Helgason & Ringen 2008). Languages with a three-way contrast use all three ranges, distinguishing categories usually called 'voiced', 'voiceless' and '(voiceless) aspirated' by means of voicing lead, short lag and long lag VOT values respectively. Finally, languages with more than three contrasts employ additional laryngeal settings to create the contrasts (Cho et al. 2018b: 54).

The stops in single-series languages theoretically have the entire spectrum from long lead to long lag available, but they are often described as voiceless unaspirated, i.e. as having a short lag VOT; some approaches consider this a kind of default or unmarked setting (Maddieson 1984; Kakadelis 2018: 51–58). However, even these languages show a range of language-specific variation in the actual production of stops. Kakadelis (2018) reports, on the basis of a detailed investigation of three languages without a voicing contrast (Bardi, Nahuatl, and Arapaho, an Algonquian language), that general correlational trends, such as the tendency for more posterior places of articulation to be associated with longer VOT values and a tendency for less prevoicing in phrase-initial compared to phrase-medial or -final contexts, are confirmed in such languages; however, 'the ratio and distribution of tokens with positive versus negative VOT differed significantly' (266), with stops in Bardi being mainly voiced (with voice lead), while in Arapaho and Nahuatl the labial stops have negative VOTs with the coronal and velar stops having positive VOTs, with quite significant differences in the size of the leads or lags. Hodgson (2021: 27) summarizes mean VOT distributions reported in the literature on eight single-series languages, confirming the general correlation with place of articulation alongside a broad language-specific variability in mean VOT ranges.

Our primary goal in this paper is to contribute to the study of VOT in languages without a voicing contrast by describing VOT patterns in Blackfoot. The secondary goal is to provide an in-depth investigation of how the VOT of these stop consonants correlates with different linguistic environments and speaker demographics. Our final goal is to validate our data-collection methods, which needed to be somewhat unorthodox due to the context and current state of the language. From a practical point of view, we also wished to shed light on the common preference of many Blackfoot speakers to use the letters <b d g> to represent Blackfoot stops in writing.

2 The Blackfoot language

Blackfoot (ISO 639-3: bla) is an endangered Algonquian language spoken in four First Nation communities in Alberta, Canada and Montana, USA, which together make up the Blackfoot Confederacy. Its Indigenous name is Niitsi'powahsin ('Real/original language') or Siksikai'powahsin ('Blackfoot language'). There is dialectal variation between the language as spoken in each community (see Figure 1), but the varieties are fully mutually intelligible. A systematic description of dialect variation is lacking.

Estimates regarding speaker numbers range from about 3,000 to about 5,000 speakers, approximately 15% of the total population, with fluent speakers tending to belong to older generations. Among the younger generations there are many so-called 'silent speakers' or 'fluent listeners', who have significant passive understanding but limited conversational fluency. There is little intergenerational transmission and there are very few if any monolingual speakers (Genee & Junker 2018). The standard orthography developed by Frantz (1978, 1993, 2017) was officially adopted by the Canadian Nation schoolboards in 1975 (Frantz 2017: 185) but has not found general acceptance or universal implementation. Fluent elderly speakers tend not to be proficient in it, and several alternative orthographies and spelling systems, including a syllabary, are also in use in the community (Genee 2020).

Blackfoot is a polysynthetic language with extensive inflectional and derivational morphology (Frantz 2017, Genee 2018). This section introduces the Blackfoot phonemic inventory and the main phonological and morphophonological factors that we considered during our data collection and analysis.

Figure 1 Blackfoot Nations in Alberta and Montana: A. = Siksika Nation, B. = Piikani (Peigan) Nation, C. = Kainai (Blood) Nation, D. = Aamsskaapipikani (Piegan Blackfeet, Montana) Nation. Map courtesy Kevin McManigal.

Blackfoot's phonemic inventory is usually given a in Tables 1 and 2 (Frantz 2017, Elfner 2006a, b; but see Weber 2020 for a different inventory; see Derrick & Weber 2023 for a discussion of the lack of consensus regarding the Blackfoot phonemic inventory).

Table 1 Blackfoot consonant system.

	Labial	Alveolar	Palatal	Velar	Glottal
Nasal	m mː	n nː			
Plosive	p pː	t tː		k kː	ʔ
Fricative		s sː		x	h
Affricate		t͡s t͡ːs		k͡s k͡ːs	
Approximant	w		j		

Table 2 Blackfoot vowel system.

i iː	
	o oː
	a aː

Consonants can be singleton or geminate, except for glottal stop, the glottal and velar fricatives, and approximants.[2] Geminate stops do not occur word-initially. /x/ is palatal [ç] after a high front vowel. Vowels can also be singleton or geminate, and have the following common allophones: high front: [ɪ]; high back: [u ʊ]; low: [ʌ].

Two common allophonic processes affected our data collection:

(i) Vowels are voiceless in many contexts, especially short vowels word-finally and between voiceless consonants (Bliss & Glougie 2009, Gick et al. 2012, Prins 2019). These environments were excluded from our analysis because there was no voice onset.

(ii) Assibilation turns all underlying /ti(ː)/ sequences into [t͡si(ː)]. This resulted in a significant systematic gap in our data for /t/, since only sequences of [ta(ː)] and [to(ː)] are possible.

Two other important factors were:

(iii) Orthographic <Vh> sequences do not contain a vowel, but a syllabic dorsal fricative with secondary palatalization (after /i/) or rounding (after /o/) (Miyashita 2017).[3] Both orthographic <ih> and <iih> are pronounced as [s] or [sː] by many speakers.

(iv) Long vowels do not occur before a dorsal fricative. Orthographic <VVh> sequences contain a short vowel followed by a dorsal fricative (Frantz 2017, Miyashita 2017).

Blackfoot is generally described as a pitch accent language (Taylor 1969, Van Der Mark 2002, Weber 2016, Fish & Miyashita 2017, Frantz 2017). Both long and short vowels can be accented or unaccented, and some long vowels are characterized by a falling pitch pattern. Pitch accent is optionally represented in the orthography by means of acute accents.

3 Methodology

3.1 Data

We collected words with stop consonants occurring in different positions, with varying pitch accent patterns on the following vowel, and followed by long vs. short vowels of different quality. Data were collected for non-geminate oral stops in word-initial and word-medial position in all available environments. Geminate stops were excluded due to their limited distribution, in particular their non-occurrence in word-initial position. (Hodgson 2021 excludes geminates from her investigation of Cree for similar reasons). We did not use carrier phrases, so word-initial tokens are in effect utterance-initial tokens. Only the six monophthongs /i iː a aː o oː/ were included, as it was not possible to get a full set of diphthongs in our stimuli collection. Where possible we included accented and unaccented vowels. This meant that for /p/ and /k/ there were twelve distinct following vowels, but for /t/ the environments with following /i iː/ were missing because /t/+/i(ː)/ always assibilates (see Section 2 above).

We aimed to include three different tokens for each environment, but this was not always possible. Table 3 lists the environments with an example of each involving the labial stop /p/.

---

[2] Note that the existence of a length contrast means that Kakadelis (2018) would not consider Blackfoot to be a language without voicing contrast. In this paper we exclude geminate stops because their occurrence is restricted to word-medial contexts.

[3] Derrick (2007) analyzes these as voiceless vowels.

Table 3 List of phonological environments that were controlled for and sample stimulus words elicited for each environment with target stop consonant /p/, in orthographic and broad phonetic transcription. Bold marks the target stop consonant and its pertinent phonological environment.

Environment	Example word (orthographic)	Example word (broad phonetic)
#_V	**pa**nokáínattsi 'paper'	[pʌnʊkɛnʌtːsi]
#_VV	**paa**pó'sin 'lightning, electricity'	[paːpʊʔsɪn]
#_V́	**pá**kkii'p 'chokecherry'	[pʌkːiːʔp]
#_V́V	**p**íítaa 'eagle'	[piːtaː]
(V)V_V	**áápo**tskina 'cow'	[aːpʊtskina]
(V)V_VV	**áápii**kayi 'skunk'	[aːpiːkaji]
(V)V_V́	a**pá**ni 'butterfly'	[apʌni]
(V)V_V́V	ssp**opíí** 'turtle'	[sːpʊpiː]
(C)C_V	s**spo**píí 'turtle'	[sːpʊpiː]
C_VV	apáí**hpii**soká'sim 'skirt'	[apɛçpiːsʊkaʔsɪm]
(C)C_V́	s**spá**tsiko 'sand'	[sːpʌtsikʊ]
(C)C_V́V	aam**skáá**poohtsi 'south'[a]	[aːmskaːpuxtsi]

[a]An example for this environment with /p/ was not available.

3.2 Speakers and data collection procedure

Data were originally collected from 19 adult native speakers of the Kainai (Blood) dialect of Blackfoot.[4] Data from six speakers had to be excluded: two due to insufficient audio quality; three because the recordings were incomplete; and one because English only was reported as the first language on the demographics form. The remaining 13 speakers were between 34 and 77 years old with a median age of 59 years. There were eight female speakers aged 34–69 years, with a median age of 57.5 years, and five male speakers aged 49–77 years, with a median age of 61 years. All speakers self-identified as fluent in Blackfoot and English, and all indicated that their L1 was either Blackfoot alone or Blackfoot and English. Details are presented in Table 4.

Recording sessions took place in the Blackfoot Language Resources lab at the University of Lethbridge or in the speaker's home. The latter was necessary as some speakers were unable to travel or indicated that they preferred to be interviewed in their own environment. Attempts were made to ensure the different locations did not result in significantly differ-ent levels of formality that might affect the speech produced. In both cases, the atmosphere was informal and relaxed. Participants were able to be accompanied by a family member or friend if desired, and were offered refreshments before the recording began; they were seated comfortably and given breaks when needed. The principal fieldworker was a familiar person.

Recordings were made with a Roland Edirol R-09HR handheld recorder and a NexxTech lapel microphone at a sampling rate of 44.1 kHz and 24-bit resolution.

Speakers either sat in front of a laptop computer or were given a paper printout of the stimuli. Many speakers of Blackfoot are not able to read Blackfoot, so we were unable to present the stimuli in written form. We chose not to have the participants simply repeat the stimuli given by the fieldworker so as not to prime the speaker with an initial pronunciation.

---

[4] We attempted to recruit participants with a reasonable spread of age ranges and gender, but beyond that our sample was a convenience sample limited to speakers of the Kainai dialect, as a result of our main field worker being from that community. Further research is required to confirm the results for the other dialects.

Table 4 Speakers.

Speaker code	Gender	Age	Tokens analyzed	Self-reported L1: B(lackfoot) and/or E(nglish)
01	F	56	168	B+E
02	M	77	165	B
03	F	61	190	B
04	M	64	185	B
07	M	57	137	B
08	F	34	157	B+E
11	F	34	139	B+E
12	M	69	168	B
13	M	55	162	B
14	F	60	158	B
15	F	59	162	B
19	F	69	144	B
20	F	47	180	B

Two kinds of stimuli were presented: pictures and written English words/short phrases. Participants were asked to say the Blackfoot word for the pictures and to translate the written English words and short phrases into spoken Blackfoot. They were asked to repeat each target three times, with the second one being used for analysis. There were a total of 79 picture stimuli and 49 translation stimuli. The type of stimulus did not significantly affect VOT (see Table 7 below). All picture stimuli elicited one-word responses. With a few exceptions all translation stimuli also elicited one-word responses in Blackfoot, so that most tokens were single words. Exceptions were four questions containing a question word (*tsimáá* 'where'; *tsá* 'what', *takáá* 'who') plus verb and one phrase consisting of two words (*aapinákosi nitáákssko'too* 'I will come back tomorrow').

Instructions were given in English, but the main fieldworker was an intermediate speaker of Blackfoot and would sometimes whisper a prompt if the speaker could not remember a word. Several rounds of pilot elicitations eliminated stimuli that were likely to be deemed culturally inappropriate or to elicit too many different responses. Even then, some responses had to be discarded because they did not elicit the intended word.

3.3 Acoustic analysis

Praat (version 6.1.41, Boersma & Weenink 2019) was used to segment the recordings into individual words and to mark sound boundaries and acoustic events. For each target stop consonant, two acoustic events were marked: the stop burst and the voice onset (see Figure 2). The burst was defined by the spike or a cluster of spikes of energy in the waveform occurring at the beginning of the closure release. Voice onset is where voicing starts immediately preceding or following the consonant as evident from the presence of the voicing bar in the spectrogram. Voice Onset Time (VOT) was calculated by subtracting the stop burst time from that of the voice onset. Five undergraduate research assistants were trained to mark the acoustic events and discussed their markings in a separate meeting to ensure compatibility and to reach consensus on difficult cases. Their markings had an interrater reliability of above 90% as measured by Pearson's correlation coefficient, and the VOT measurement differences between raters were within 2 milliseconds of each other. Our study thus confirmed the tendency for the reliability of manual VOT measurements to be very high (Auzou et al. 2000: 134).
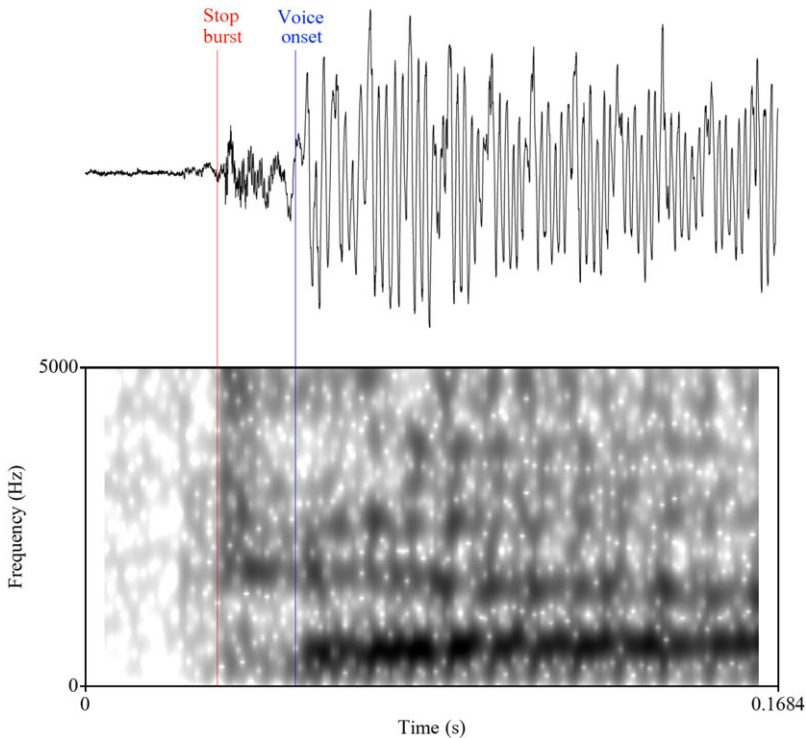
Figure 2 (Colour online) Demonstration of the two acoustic events marked in Praat: Stop burst of /t/ and voice onset of the following vowel /aː/ in the word /piːtaː/ 'eagle' produced by a male Blackfoot speaker.

3.4 Statistical analysis

A combination of linguistic, socio-indexical, and experimental factors was examined in the statistical analysis. As discussed in Section 2, linguistic factors that may affect VOT include: place of articulation (Lisker & Abrahamson 1964, Klatt 1975, Cho & Ladefoged 1999, Abdelli-Berruh 2009, Bijankhan & Nourbaksh 2009, Chodroff et al. 2019); position in the word (Flege & Brown 1982, Davidson 2016); length and quality of the following vowel (Klatt 1975, Cho & Ladefoged 1999, Taff et al. 2001, Bijankhan & Nourbaksh 2009, Garvin 2018); and presence or absence of prominence following the stop (Flege & Brown 1982, Davidson 2016). Socio-indexical factors that may affect VOT include gender and age (Bijankhan & Nourbakhsh 2009, Li 2013, Kharlamov 2018, Kleber 2018, Swartz 1992, Morris et al. 2008). In addition, we added task type as an experimental factor; the goal here was to investigate whether the methods used to collect our data affected the nature of the data.

Using R (R Core Team, 2019), a linear mixed effects model was constructed to identify the effect of linguistic, socio-indexical, and experimental factors on VOT values. The linguistic factors include: the place of articulation of the target stop consonant (three levels: /p/, /t/, and /k/); the following vowel (two levels: /a/ vs. /o/); and the prosodic characteristics of the following vowel such as length (two levels: long vs. short) and pitch accent pattern (two levels: accented vs. unaccented). The socio-indexical factors were speakers' age and gender (two levels: male vs. female). The experimental factor refers to the two types of experiments (picture-naming vs. translation) in which participants were engaged. Thus, the fixed effects components of the model were: target consonant; quality of the following vowel; length of the following vowel; pitch accent pattern of the following vowel; speakers' age (coded as continuous); speaker's gender; and task type. Since the fixed effect factor 'target consonant' has three categories: /p/ (reference level), /t/, and /k/, post-hoc pair-wise

comparison was conducted using the R package *emmeans* (Lenth 2021). With respect to the following vowel, due to the systematic gap in vowel /i/ preceded by /t/, only the vowels /a/ and /o/ were included in the model. Individual speakers and words were included as random intercepts, and a by-participant random slope that varies with target consonant was included to further account for repeated measures. The statistical analysis was conducted using the R package *lme4* (Bates et al. 2015a, b; Table 7 below includes the R code used for the model). Overall model fit was assessed through the pseudo r^2 (Nakagawa & Shielzeth 2013) calculated using the r.squaredGLMM function in the R package MuMIn (Bartoń 2020). In linear modeling, r^2 measures the proportion of variance explained by the model and thus serves as a metric for assessing model fitness. The calculation and interpretation of r^2, however, is less straightforward due to the hierarchical structure of the mixed effects model. Pseudo r^2 is then obtained which computes two r^2 values: marginal r^2 and conditional r^2. Marginal r^2 is the variance explained by the fixed effects in the model while conditional r^2 is the amount of variance accounted for by both the fixed effects and the random effects. r^2 values above 0.06, 0.16, and 0.36 were considered as small, medium and large respectively (Plonsky & Oswald 2014).

4   Results

4.1   Summary statistics and comparison to English

Table 5 lists the summary statistics of VOT values for the three consonant targets followed by the three vowel qualities /a/, /i/ (only for /p/ and /k/), and /o/, for male speakers and female speakers respectively. As shown in the table, VOT values increase as the place of articulation progresses farther back into the oral cavity: the velar /k/ has the longest VOT, followed by the alveolar /t/, with the bilabial /p/ having shortest VOT. Further comparison between Blackfoot and English (Table 6) suggests that the range of VOT values for the Blackfoot stop consonants corresponds most closely to the English voiced stop consonant series.

Table 5   Summary statistics (mean and sd) of VOT values for the three target stop consonants in the three vowel contexts, for each gender group.

		/p/ (n = 667)	/t/ (n = 614)	/k/ (n = 815)
Female	/a/	11.6 (9.3)	14.1 (8.6)	23.9 (13.1)
	/i/	12.6 (8.3)	NA	32.7 (17.8)
	/o/	17.4 (12.0)	16.6 (9.7)	29.0 (15.9)
Male	/a/	14.0 (9.7)	19.1 (13.0)	25.7 (16.9)
	/i/	13.3 (8.3)	NA	31.2 (16.1)
	/o/	17.1 (11.5)	18.6 (11.9)	26.3 (17.8)

Table 6   Mean VOT ranges for English voiceless and voiced stop consonants based on Auzou et al.'s (2000: 139) review of 12 studies of VOT in English as summarized in Kent & Read (2002: 151) compared with Blackfoot VOT ranges found in this study.

English voiced stops	Range of mean VOT (ms)	English voiceless stops	Range of mean VOT (ms)	Range of mean VOT in Blackfoot
/b/	1–20	/p/	46–85	11.6–17.4
/d/	0–21	/t/	65–95	14.1–19.1
/g/	14–35	/k/	70–110	23.9–32.7

4.2 Results of linear mixed effects model: Linguistic factors

Table 7 outlines the statistical results of the linear mixed effects model that systematically tests all factors of interest. First, there is a significant effect of place of articulation. VOTs for velar stops are 10.986 ms longer than for bilabial stops and 9.729 ms longer than for alveolar stops. Post-hoc pairwise comparison did not find significant difference in VOT values between alveolar and bilabial stops. The greater difference between /k/ and /p, t/ (as compared to that between /p/ and /t/) is evident also from Table 5 and Figure 3. The lack of significant difference between /p/ and /t/ and the presence of the significant difference between /k/ and /t, p/ suggests that the increment is not linear. Second, a significant vowel effect was identified. Stop consonants followed by the high back vowel /o/ have an average of 3.293 ms longer VOT values than those followed by the low vowel /a/. Finally, when followed by a long vowel, stop consonants have 1.632 ms longer VOT values than when followed by a short vowel, and this effect is significant, as illustrated in Figure 3. It is to be noted that the effect sizes for vowel quality and length are small and need to be interpreted with caution; this awaits verification with a larger number of participants and a broader range of data in future research. The effects for word position and accent pattern were both substantively small and statistically insignificant. That is, VOT values do not significantly vary with the

Table 7 Results of the linear mixed effects model on VOT (ms). The fixed effects are target stop consonant, the following vowel, task type, word position, the length and pitch accent pattern of the following vowel, as well as speakers' age and sex. Bold in the last column marks statistical significance.

R model:

lmer(VOT∼age+targetConsonant+followingVowel+sex+task+wordPosition+followingVowelLength+
followingVowelStress+(1|word)+(targetConsonant|Participant), data)

Overall model fit: pseudo $r2$ = .14 (marginal) and .36 (conditional)

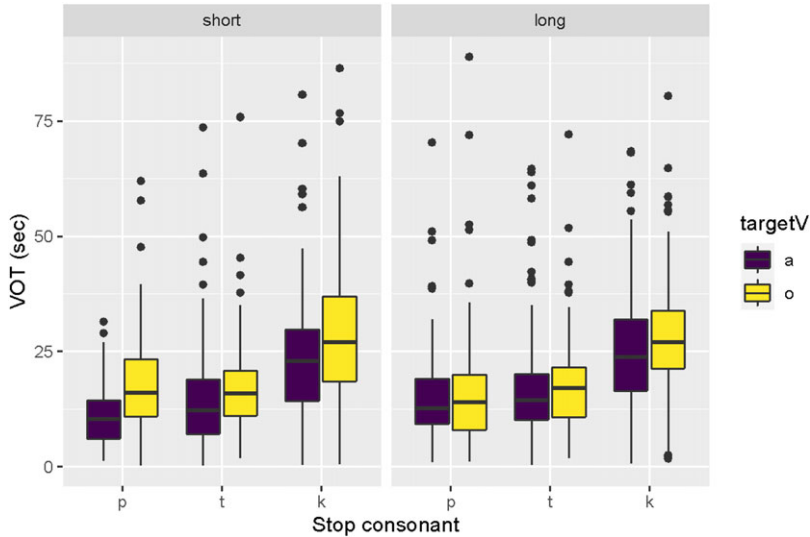|	Estimate	Std. Error	t-value	p-value
Intercept	16.766	4.770	3.514	**.002**
Target consonant /p/ vs. /**k**/ (reference)	−10.986	1.876	−5.855	**< .001**
Target consonant /t/ vs. /**k**/ (reference)	−9.729	2.000	−4.684	**< .001**
Following vowel /o/ vs. /**a**/ (reference)	3.293	0.658	5.007	**< .001**
Following vowel length long vs. short (reference)	1.632	0.636	2.565	**.010**
Following vowel accent pattern unaccented vs. accented (reference)	0.529	0.641	0.825	.410
Word position word-medial vs. word-initial (reference)	0.545	0.848	0.630	.520
Task picture vs. translation (reference)	−0.037	0.717	−0.052	.958
Speaker age (numeric)	0.139	0.074	1.878	.089
Speaker sex ale vs. female (reference)	0.898	1.845	0487	.636

Figure 3 (Colour online) VOT distributions for the three stop consonants followed by long vs. short vowel /o/ and vowel /a/.

position in which the stop consonants occur within a word or with the presence or absence of a following accented vowel.

4.3   Results of linear mixed effects model: Socio-indexical and experimental factors

The results of the linear mixed effects model in Table 7 suggest no effect for the socio-indexical and experimental factors. As illustrated in Figure 4, which displays boxplots for each individual speaker for each gender, there is a wide range of variability in VOT values in both gender groups. No clear differences can be identified between male and female speakers. Age was found to show a substantively small trend that nearly reaches the bar of significance. The specific task in which participants were engaged did not make a significant difference with regard to VOT values; this was a very robust finding, with a substantively very small effect size and very low significance.

5   Discussion

To summarize, our data confirm the traditional impressionistic description of Algonquian stops as phonetically voiceless unaspirated. Table 5 shows that mean VOT values in our study ranged from 11.6 ms for /p/ followed by /a(:)/ to 32.7 for /k/ followed by /i(:)/ (both for female speakers). This is in line with findings from the sparse acoustic work on Algonquian stops. Hall (2010) reports means between 13–24 ms for one Blackfoot speaker from Montana; Kakadelis (2018: 272) reports a generalized mean of 28 ms for Arapaho; Hodgson's (2021: i, 53–54) study of Cree reports means between 11.6–25.6 ms; and Rosen et al. (2019) found means between 18–48 ms for Michif.

These comparisons should be interpreted carefully, however, especially in light of the differences between the data used in the studies. Hall (2010) analyzed an unknown number of tokens selected from target words elicited from one speaker of Piegan Blackfeet (Montana). Kakadelis (2018: 157) analyzed 522 oral stops from a total of about an hour of spontaneous narrative produced by five speakers of Arapaho. Hodgson (2021: 38–43) analyzed 660 tokens selected from about two hours of a publicly available online corpus of narratives produced
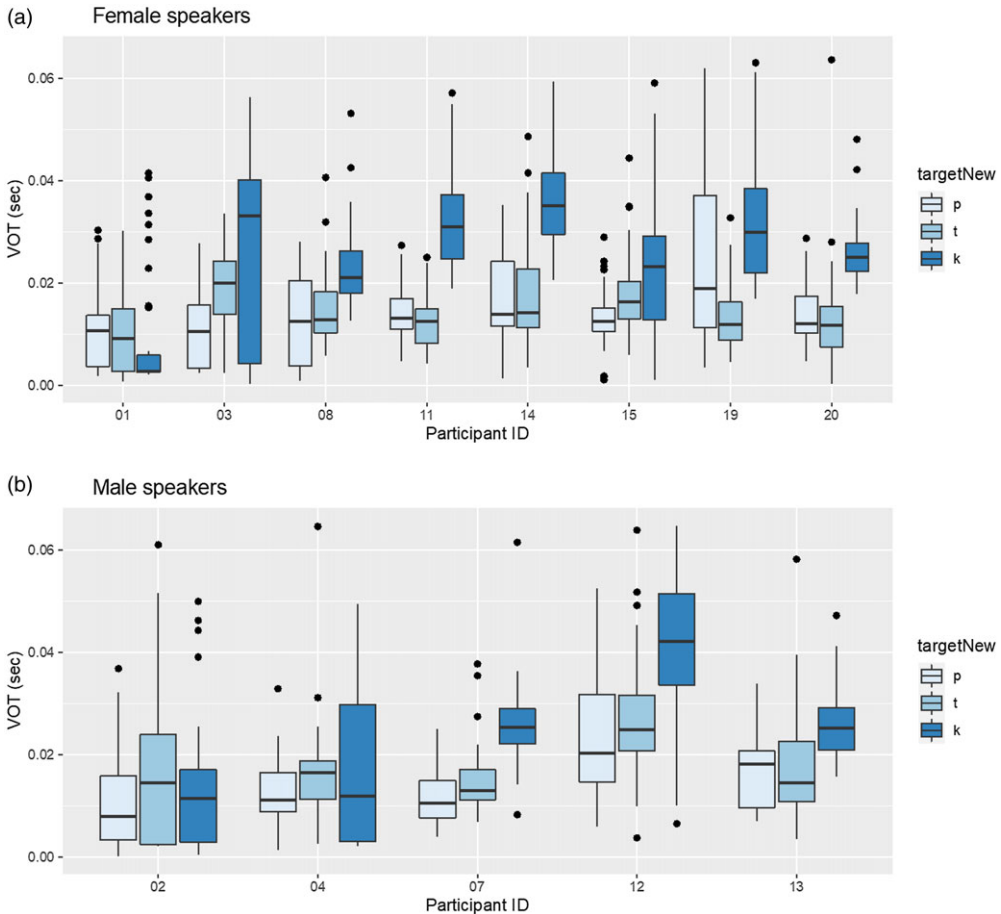
Figure 4 (Colour online) Boxplots of VOT ranges of individual speakers, separated by gender.

by one speaker of Plains Cree. Rosen et al. (2019) analyzed 446 tokens from semi-directed 'Pear Story' narratives produced by 10 speakers of Michif. We analyzed 2096 tokens from 13 different speakers produced in a controlled experimental setting. Interestingly, Rosen et al. (2019) report that the polysynthetic nature of Michif resulted in a dearth of nouns in their data, while we were able to assemble a data set with just over 60% nouns in a more controlled environment. This suggests that, in addition to the expected differences related to levels of formality and genre, different data types and collection methods may also result in differences between the word categories that end up in one's data set. In the case of the Michif data reported in Rosen et al. (2019) this resulted in a shortage of French-origin words which limited their ability to draw conclusions regarding variation related to the Cree vs. French contributions to the phonological system of a mixed language. Other possible effects of unbalanced word categories await further investigations.

The VOT ranges found in our data also agree with the ranges found in other languages without a voicing contrast as reported in some of the literature. Kakadelis (2018), in a detailed study of stops in three languages without voicing contrast, found mean VOT ranges between 23 ms and 28 ms. It has been suggested that there is something basic about the short-lag VOT, perhaps due to the relative articulatory ease of producing stop consonants that do not involve voicing or aspiration (Kong et al. 2012); this may also be related to the fact that the short-lag VOT range seems to be the anchor range from which other contrastive categories emerge in

infants' babbling and children's early speech production in languages with a stop voicing contrast (Macken & Barton 1980; Hitchcock & Koenig 2013). However, such generalizations need to be balanced against the fact that individual single-series languages may employ language-specific ranges that do not overlap with those of other single-series languages, as shown by Hodgson's (2021: 73) comparison of VOT in nine language varieties. An illustrative example is modern Māori: with mean VOT values between 50ms and 66ms (Hodgson 2021: 26), Modern Māori's range falls entirely outside of that covered by Blackfoot and the languages investigated by Kakadelis (2018).

Our study confirms that Blackfoot VOT values are in the range corresponding to English voiced (short lag) stops, reported to have mean ranges between 1–20 ms for /b/, 0–21 ms for /d/, and 14–35 ms for /g/ (Auzou et al. 2000: 139; Kent & Read 2002: 151). This is important as it goes some way to explaining difficulties surrounding the acceptance and implementation of the standard spelling system developed by linguist Donald G. Frantz in the 1970s (Frantz 1978, 1993, 2017). After analyzing the Blackfoot stops as phonetically voiceless unaspirated, Frantz understandably chose the letters <p t k> for the orthography. The orthography he developed was officially adopted by the Canadian Blackfoot Nation schoolboards in 1975, but it continues to lack universal acceptance. Such a situation is by no means unique to Blackfoot; similar disagreements in other North-American Indigenous communities are described by Hinton (2014), who coined the term 'orthography wars'. Writers of Blackfoot who have not been trained in the standard orthography often use the letters <b d g>, digraphs consisting of these letters followed by <h> (<bh dh gh>), or digraphs combining the English voiceless and voiced letters (<bp td kg>), especially in contexts where English voiceless stops would have the long VOT associated with aspiration (Genee 2020). The finding that Blackfoot stops have VOTs that are very similar to those of English voiced (short lag) stops also supports some practices commonly seen in L2 teaching contexts when teachers explain how to pronounce the Blackfoot letters <p t k>, or correct L2 pronunciations of Blackfoot stops pronounced with too much aspiration: the learner is often told to 'make it sound more like a b/d/g'. Our work here shows that producing Blackfoot stops with VOTs closer to those of English voiced (short lag) stops should in fact result in more native-like pronunciations (see Garvin's (2018) comments on Northern Shoshoni for similar points). We hope to investigate this in future work comparing L1 with L2 speakers. We also hope to use these results to develop targeted pronunciation exercises for L2 learners allowing students to compare their own VOT with that of a fluent speaker through self-recording into Praat or a dedicated CALL application based on visualization of Voice Onset Time (compare work by Fish & Miyashita 2017 on the visualization of pitch accent patterns in Blackfoot).

Significant effects were found for three of the five linguistics factors investigated (i.e. place of articulation of the target stop consonant and quality and length of the following vowel), while no significant effects were found for the other two linguistics factors (i.e. pitch accent pattern of the following vowel and word position of the target consonant) nor for the socio-indexical and experimental factors.

The Blackfoot data confirm three cross-linguistic tendencies familiar from the literature on VOT (see Section 1). First, longer VOT values were found to correlate with more posterior places of articulation. As shown in Table 7, this correlation was strong and highly significant, confirming a cross-linguistic trend that is very well established in the literature (Lisker & Abrahamson 1964, Klatt 1975, Cho & Ladefoged 1999, Abdelli-Berruh 2009, Bijankhan & Nourbaksh 2009, Chodroff et al. 2019).

Second, stops followed by high vowels were found to have longer VOTs than those followed by low vowels. This correlation was also highly significant and confirms a well-established cross-linguistic trend (Klatt 1975, Taff et al. 2001, Bijankhan & Nourbaksh 2009). Recall, however, that our findings here are limited by the systematic gap in the data caused by the non-occurrence of [t+i(:)].

Third, longer VOT values were found to correlate with longer following vowels (Klatt 1975, Taff et al. 2001, Bijankhan & Nourbaksh 2009).

These three findings confirm results reported for the Algonquian language Cree, although it must be noted that the correlation of longer VOT with a longer following vowel was only confirmed for /p/ in Hodgson's (2021) study. For Arapaho, the first tendency is confirmed, while the second and third were not addressed in the study by Kakadelis (2018).

Interestingly, our data did not confirm the cross-linguistic tendency for word-initial stops to have longer VOTs than word-medial ones, especially intersonorants (Flege & Brown 1982, Davidson 2016, Kakadelis 2018). In contrast, we found no significant differences between word-initial and word-medial stops, and therefore no correlation with position in the word. This result was a bit surprising, given that it appears to be a robust finding in the literature (Flege & Brown 1982, Davidson 2016; the expected pattern was also confirmed for Cree in Hodgson 2021). Derrick & Weber's (2023) Blackfoot data, obtained from the speech of a single speaker of Kainai Blackfoot, also appear to mirror the cross-linguistic tendency, contrasting with our results. They even found some negative VOTs, mainly in word-medial contexts after a voiced segment, but also a few cases of true word-initial pre-voicing; it should be noted however that their data included all obstruents, including geminates and affricates, and are therefore not fully comparable to ours. Further investigation with more speakers and a wider range of segment types and phonological contexts is warranted. Kakadelis' (2018: 177) study of Arapaho reports that '[t]he phonological context does not seem to significantly affect the direction of VOT across place of articulation', which seems to be more in line with our Blackfoot findings.

The effects of socio-indexical factors such as age and gender on VOT are only rarely studied, and where they are, results appear to be mixed (Swartz 1992, Auzou et al. 2000, Morris et al. 2008, Bijankhan & Nourbaksh 2009, Kharlamov 2018, Kleber 2018). We did not find significant effects for gender or age in our data. Age came closest to reaching significance ($p = .089$). It might be possible to find a stronger effect when including a larger range of different age groups. Auzou et al. (2000: 138–140) suggest that VOT is more variable in very young children and in older adults. Since our study did not include children, we were unable to investigate this aspect. Given the very low rates of home transmission, the window may have closed on being able to study this age group, although it may be still possible to study second language acquisition of VOT in school-aged children, along the lines of comparable studies for French in Canadian immersion contexts (e.g. Netelenbos et al. 2016). We hope to investigate other socio-indexical factors in future work with more participants and speakers from the other dialect areas.

The lack of an effect related to the experimental condition was a particularly robust finding with important methodological implications for research on endangered minority languages. Collecting the data needed for our study was challenging for a number of reasons. First, basic nouns are easier to elicit than verbs, but we were not able to include all required phonological conditions that way and had to include some verbal expressions. However, due to the polysynthetic nature of the language, verb words tend to be very long: the exact form produced in elicitation conditions can be unpredictable and sometimes varied even between successive pronunciations of the same word by the same speaker. Second, pilot sessions led to the elimination of several words that were either not deemed culturally appropriate or whose prompts elicited too many variant responses between speakers. This resulted in the loss of some otherwise very suitable items from our initial target list, which then had to be replaced. Third, limited Blackfoot literacy amongst the participants meant that we could not ask them to read a word list. Fourth, limited Blackfoot fluency amongst the field workers meant we could not use word repetition as an elicitation method. In order to mitigate these challenges we designed two different tasks: a picture elicitation task and an English-to-Blackfoot translation task. The choice between picture and translation task was made on practical grounds: any word that could be represented by a picture in an unambiguous manner was elicited that way, and translation prompts were used for words that could not be elicited with a picture. The fact that no difference was found between these two conditions confirms that the elicitation method is not a confounding factor, and that the pragmatic approach we used is suitable

for collecting data in the often less-than-straightforward conditions surrounding endangered minority languages. Speakers quickly understood what was asked of them, enjoyed both task types, responded positively, and did not get tired or frustrated too quickly. This is important for future research on Blackfoot and other endangered languages with mainly elderly speakers who are not accustomed to these kinds of formal academic activities.

6 Conclusion

This study supports a description of Blackfoot oral stops as voiceless unaspirated, with VOTs falling in the short-lag range. Several cross-linguistic tendencies were confirmed, in particular the correlation of longer VOT values with more posterior places of articulation and with longer and higher following vowels. A tendency for shorter VOT in word-medial position was not confirmed, nor did we find a correlation between VOT and following accent. We also found no correlation with socio-indexical factors or experimental condition. Given the scarcity of socio-phonetic work, potential differences in VOT especially between genders, ages and dialects warrant further investigation, as does the potential for different data types to produce differences in VOT. Further applied work might include the development of targeted pronunciation instruction for L2 learners.

Competing interests

The authors declare none.

References

Abdelli-Beruh, Nassima B. 2009. Influence of place of articulation on some acoustic correlates of the stop voicing contrast in Parisian French. *Journal of Phonetics* 37(1), 66–78.

Abramson, Arthur S. & Doug H. Whalen. 2017. Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics* 63, 75–86.

Al-Tamimi, Jalal & Ghada Khattab. 2018. Acoustic correlates for the voicing contrast in Lebanese Arabic singleton and geminate stops. *Journal of Phonetics* 71, 306–325.

Auzou, Pascal, Canan Özsancak, Richard J. Morris, Mary Jan, Francis Eustache & Didier Hannequin. 2000. Voice onset time in aphasia, apraxia of speech and dysarthria: A review. *Clinical Linguistics & Phonetics* 14(2), 131–150.

Bartoń, Kamil. 2020. MuMIn: Multi-Model Inference. R package version 1.43.17. https://CRAN.R-project.org/package=MuMIn.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015a. Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software* 67(1). doi:10.18637/jss.v067.i01.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015b. *Lme4*: Linear mixed-effects models using Eigen and S4. R package version 1.10-10.

Beckman, Jill, Michael Jessen & Catherine Ringen. 2013. Empirical evidence for laryngeal features: Aspirating vs. true voicing languages. *Journal of Linguistics* 49, 259–284.

Bijankhan, Mahmood & Mandana Nourbaksh. 2009. Voice onset time in Persian initial and intervocalic stop production. *Journal of the International Phonetic Association* 39(3), 335–364.

Bliss, Heather & Jennifer Glougie. 2009. Speaker variation and the phonetic realization of Blackfoot obviation morphology. In Heather Bliss & Raphaël Girard (ed.), *Proceedings of the 13th and 14th Workshop on the Structure and Constituency of Languages of the Americas* (University of British Columbia Working Papers in Linguistics 26), 70–83. Vancouver: University of British Columbia.

Bloomfield, Leonard. 1946. Algonquian. In Cornelius Osgood & Harry Hoijer (eds.), *Linguistic structures of native America*, 85–129. New York: Viking Fund Publications in Anthropology.

Boersma, Paul & David Weenink. 2019. Praat: Doing phonetics by computer [Computer program]. Version 6.1.41, retrieved 25 March 2019 from http://praat.org/.

Cho, Taehong, Gerard Docherty & Doug H. Whalen (eds.). 2018a. *Marking 50 years of research on Voice Onset Time*: Special issue of *Journal of Phonetics* 72.

Cho, Taehong & Peter Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27(2), 207–229.

Cho, Taehong, Doug H. Whalen & Gerard Docherty. 2018b. Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. In Cho et al. (eds.), 52–65.

Chodroff, Eleanor, Alessandra Golden & Colin Wilson. 2019. Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization. *The Journal of the Acoustical Society of America* 145(1), EL109.

Davidson, Lisa. 2016. Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics* 54, 35–50.

Derrick, Donald. 2007. Syllabification and Blackfoot /s/. In Karl S. Hele & J. Randolph Valentine (eds.), *Proceedings of the Northwest Linguistics Conference 22* (NWLC 22), 62–76. Vancouver: Simon Fraser University Working Papers in Linguistics.

Elfner, Emily. 2006a. The mora in Blackfoot. MA thesis, University of Calgary.

Elfner, Emily. 2006b. Contrastive syllabification in Blackfoot. In Donald Baumer, David Montero & Michael Scanlon (eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics* (WCCFL 25), 141–149. Somerville, MA: Cascadilla Proceedings Project.

Derrick, Donald & Natalie Weber. 2023. Káínai Blackfoot. Ms., University of Canterbury, NZ & Yale University.

Fish, Naatosi & Mizuki Miyashita. 2017. Guiding pronunciation of Blackfoot melody. In Jon Reyhner, Joseph Martin, Louise Lockard & Willard Sakiestewa Gilbert (eds.), *Honoring our teachers*, 203–210. Flagstaff: Northern Arizona University.

Flege, James Emil & W. S. Brown Jr. 1982. The voicing contrast between English /p/ and /b/ as a function of stress and position-in-utterance. *Journal of Phonetics* 10, 335–345.

Flege, James Emil & Wieke Eefting. 1987. Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication* 6(3), 185–202.

Frantz, Donald G. 1978. Abstractness of phonology and Blackfoot orthography design. In William C. MacCormack & Stephen Wurm (eds.), *Approaches to language: Anthropological issues*, 307–326. The Hague: Mouton.

Frantz, Donald G. 1993. Blackfoot writing systems. Ms., University of Lethbridge.

Frantz, Donald G. 2017. *Blackfoot grammar*, 3rd edn. Toronto: University of Toronto Press.

Garvin, Karee. 2018. An acoustic outlook on initial stops in Northern Shoshoni. *UC Berkeley Phonetics and Phonology Lab Annual Report (2018)*, 137–167.

Genee, Inge. 2018. Measuring polysynthesis. A Functional Discourse Grammar approach. In Evelien Keizer & Hella Olbertz (eds.), *Recent developments in Functional Discourse Grammar*, 233–272. Amsterdam & Philadelphia, PA: John Benjamins.

Genee, Inge. 2020. "It's written *niisto* but it sounds like KNEE STEW": Handling multiple orthographies in Blackfoot language web resource. *Written Language & Literacy* 23(1), 1–27.

Genee, Inge & Marie-Odile Junker. 2018. The Blackfoot Language Resources and Digital Dictionary project: Creating integrated web resources for language documentation and revitalization. *Language Documentation & Conservation* 12, 274–314.

Gick, Bryan, Heather Bliss, Karin Michelson & Bosko Radanov. 2012. Articulation without acoustics: "Soundless" vowels in Oneida and Blackfoot. *Journal of Phonetics* 40(1), 46–53.

Goddard, Ives. 1974. An outline of the historical phonology of Arapaho and Atsina. *International Journal of American Linguistics* 40(2), 102–116.

Goddard, Ives. 1979. Comparative Algonquian. In Lyle Campbell & Marianne Mithun (eds.), *The languages of Native America*, 70–132. Austin, TX: University of Texas Press.

Hall, Robert. 2010. Blackfoot voiceless stop consonants. Poster presented at the National Undergraduate Research Conference, Missoula, MT.

Helgason, Pétur & Catherine Ringen. 2008. Voicing and aspiration in Swedish stops. *Journal of Phonetics* 36(4), 607–628.

Hinton, Leanne. 2014. Orthography wars. In Michael Cahill & Keren Rice (eds.), *Developing orthographies for unwritten languages*, 139–168. Dallas, TX: SIL International.

Hitchcock, Elaine R. & Laura L. Koenig. 2013. The effects of data reduction in determining the schedule of voicing acquisition in young children. *Journal of Speech, Language, and Hearing Research* 56, 441–457.

Hodgson, Merion Rose. 2021. Voice Onset Time in Plains Cree. Master's thesis, Linguistics, University of Calgary.

Jesus, Luis M. T. & Maria Conceição Costa. 2020. The aerodynamics of voiced stop closures. *EURASIP Journal on Audio, Speech, and Music Processing* 2020(2).

Kakadelis, Stephanie. 2018. *Phonetic properties of oral stops in three languages with no voicing distinction*. Ph.D. dissertation, City University of New York.

Kent, Ray D. & Charles Read. 2002. *Acoustic analysis of speech*. Albany, NY: Thomson.

Kharlamov, Viktor. 2018. Prevoicing and prenasalization in Russian initial plosives. *Journal of Phonetics* 71, 215–228.

Klatt, Dennis H. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* 18, 686–706.

Kleber, Felicitas. 2018. VOT or quantity: What matters more for the voicing contrast in German regional varieties? Results from apparent-time analyses. *Journal of Phonetics* 71, 468–486.

Kong, Eun Jong, Mary E. Beckman & Jan Edwards. 2012. Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics* 40, 725–744.

Lenth, Russell V. 2021. Estimated Marginal Means, aka Least-Squares Means. R package version 1.6.1.

Li, Fangfang. 2013. The effect of speakers' sex on voice onset time in Mandarin stops. *The Journal of the Acoustical Society of America* 133(2), 142–147.

Lisker, Leigh & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *WORD* 20(3), 384–422.

Lousada, Marisa, Luis M. T. Jesus & Andreia Hall. 2010. Temporal acoustic correlates of the voicing contrast in European Portuguese stops. *Journal of the International Phonetic Association* 40(3), 261–275.

Macken, Marlys A. & David Barton. 1980. The acquisition of voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language* 7(1), 41–74.

Mackenzie, Marguerite. 1980. *Towards a dialectology of Cree-Montagnais-Naskapi*. Ph.D. dissertation, University of Toronto.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, Ian. 2013. Voicing in plosives and fricatives. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Miyashita, Mizuki. 2017. Vowel coalescence in Blackfoot. In Margaret Noodin & Monica Macauley (eds.), *Proceedings of the 47th Algonquian Conference*, 117–135. Milwaukee, WI: University of Wisconsin Press.

Morris, Richard J., Christopher R. McCrea & Kaileen D. Herring. 2008. Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics* 36(2), 308–317.

Nakagawa, Shinichi & Holger Schielzeth. 2013. A general and simple method for obtaining R2 from generalized mixed-effects models. *Methods in Ecology and Evolution* 4(2), 133–142.

Netelenbos, Nicole, Fangfang Li & Nicole Rosen. 2016. Stop consonant production of French immersion students in Western Canada: A study in voice onset time. *International Journal of Bilingualism* 20(3), 346–357.

Okimāsis, Jean & Arok E. Wolvengrey. 2008. *How to spell it in Cree*. Regina: miywāsin ink.

Pape, Daniel & Luis M. T. Jesus. 2015. Stop and fricative devoicing in European Portugese, Italian and German. *Language and Speech* 58(2), 224–246.

Plonsky, Luke & Frederick L. Oswald. 2014. How big is "big"? Interpreting effect sizes in L2 research. *Language Learning* 64(4), 878–612.

Prins, Samantha Leigh. 2019. Final vowel devoicing in Blackfoot. Master's thesis, University of Montana.

R Core Team. 2019. R: A language and environment for statistical computing. https://www.R-project.org.

Riney, Timothy J. & Naoyuki Takagi. 2008. Global foreign accent and Voice Onset Time among Japanese EFL speakers. *Language Learning* 49(2), 275–302.

Rosen, Nicole, Jesse Stewart, Michele Pesch-Johnson & Olivia N. Sammons. 2019. Michif VOT. Presented at the 19th International Congress of Phonetic Sciences (ICPhS XIX), Melbourne, Australia.

Swartz, Bradford L. 1992. Gender difference in Voice Onset Time. *Perceptual and Motor Skills* 75(3), 983–992.

Taff, Alica, Lorna Rozelle, Taehong Cho, Peter Ladefoged, Moses Dirks & Jacob Wegelin. 2001. Phonetic structures of Aleut. *Journal of Phonetics* 29, 231–271.

Taylor, Allan R. 1969. *A grammar of Blackfoot*. Ph.D. dissertation, University of California Berkeley.

Thomason, Lucy. 2006. Proto-Algonkian phonology and morpho-syntax. In Ronald E. Asher & Keith Brown (eds.), *Encyclopedia of language and linguistics*, 190–198. Oxford: Elsevier.

Turner, Justin, Nicole Netelenbos, Nicole Rosen & Fangfang Li. 2014. Stop consonant production by French–English bilingual children in Southern Alberta. *Canadian Acoustics* 42(3), https://jcaa.caa-aca.ca/index.php/jcaa/article/view/2709/2432.

Van Der Mark, Sheena. 2002. The acoustic correlates of Blackfoot prominence. *Calgary Working Papers in Linguistics* 24, 169–216.

Weber, Natalie. 2016. Accent and prosody in Blackfoot verbs. In Monica Macaulay, Margaret Noodin & Randolph V. Valentine (eds.), *Papers of the 44th Algonquian Conference*, 348–369. Albany, NY: SUNY Press.

Weber, Natalie. 2020. *Syntax, prosody, and metrical structure in Blackfoot*. Ph.D. dissertation, University of British Columbia.

Wolfart, H. C. 1996. A sketch of Cree, an Algonquian Language. In Ives Goddard (ed.), *Handbook of North American Indians*, vol 17: *Languages*, 391–438. Washington, DC: Smithsonian.

Wolvengrey, Arok E. 2011. *Semantic and pragmatic functions in Plains Cree syntax*. Utrecht: LOT.