Measuring the Voice Onset Time of Navajo and Tlingit Stops

Simone Brown

Department of Linguistics, McGill University

LING 411: Structure of an Indigenous Language

Professor James A. Crippen

Abstract

VOT (Voice Onset Time) is a measurement used to understand a language's phonetic profile, i.e. the unique production of phonemes within a language. This paper will examine two languages in the Na-Dene language family of North America— Navajo and Tlingit— and measure the VOT of stop production in the continuous speech of native speakers. Both languages are unique for their three-way stop contrast between plain voiceless, aspirated voiceless, and ejective stops. Included is a discussion of the various considerations accompanying Indigenous language research, including orthographic variation, transcription bias, and dialectal and idiolectal variation. The VOT measurements show a distinct contrast between stop categories (plain voiceless, aspirated voiceless and ejectives), but evidence suggests there is a range rather than a specific length that speakers target in their production. Both languages demonstrate a preference for plain stops and include several unmeasurable 'unreleased stops.' Tlingit demonstrates a wider range and more accurate VOT targeting than Navajo, while Navajo has significantly longer VOTs, suggesting separate repairing mechanisms in the two languages to manage perception difficulties. The phonological and perceptive processes of ejectives are also discussed, particularly the ejective and non-ejective contrast. Suggestions for further research are provided.

1 Introduction

The Na-Dene language family is well known for its incredibly rich phonetic inventories. Previous research on these sound systems is unfortunately inadequate for building a complete phonetic profile of individual languages or the language family. Many linguists have recognized this deficit in phonemic analysis and attempted to repair it via purposeful research studying the phonetic and phonological processes of Na-Dene languages. The study of consonants, particularly obstruents, reveals intricate phonological processes in Na-Dene languages that affect their production. Particular emphasis has been placed on variable sound length (gemination) in the Athabaskan language family, a subset of the Na-Dene language family encompassing Navajo, Lheidli, Apache, and more. In Lheidli, intervocalic consonant length is substantially longer than consonants in other positions and even the intervening vowels (Bird 2002). Similar proof has been found for intervocalic consonant lengthening in Navajo (Sapir & Hoijer

1967, Young & Morgan 1987, McDonough & Ladefoged 1993). In Apache, evidence suggests a combination of stress placement and morphological position impacts stop length. Still, Apache stop voice onset time is not impacted when placed at a morphological boundary, instead consistently surfacing as longest in phrase-initial positions (Tuttle 2005). Following in the footsteps of these studies, this paper will study sound length in the Na-Dene language family by analyzing the stops in two languages, Tlingit and Navajo, hoping to provide a basis for future research studying its variation and conventions.

Contextualizing this research are two theories surrounding the selection of phoneme contrasts cross-linguistically: the Aperture theory by Keating (1990) and the 'continuum' theory by Cho & Ladefoged (1999). While each addresses how languages define the contrastive nature of voicing, they do so by proposing entirely different mechanisms. Keating's Aperture theory states that all languages universally select their phonemes from three distinct phonetic categories: voiced, voiceless aspirated, and voiceless unaspirated. Each of these distinct phonetic categories manifests as a singular length or 'target' the speaker is trying to produce. Under Keating (1990), Navajo and Tlingit select two of these phonetic categories and ejectives (generally omitted from this theory on voicing universals due to the dramatic differences in their production mechanisms). Each category would have a 'target', in which each singular length represents the type of stop being produced, and any variability in production is due to natural variation in human production, not encoded into the theory itself. Still, this theory fails to address the consistent variability within stop classifications and the language-specific constraints on sound length.

Cho & Ladefoged (1999) propose that the voicing of phonemes is selected from a continuum. This continuum explains the inherent variability in the pronunciation of stops despite their clear categorizations and occurrences of partial voicing, unreleased stops, and similar phenomena. This scale is applied language-specifically or in groupings to explain the relationship between different languages and language families that occupy the same continuum. Under this theory, Navajo and Tlingit would exist on the same continuum, but each language selects different positions on the scale— close enough to be recognized for their voicing category but not uniform.

In contrast to Keating (1990), which neglects the question of ejectives, Cho & Ladefoged (1999) explicitly dismiss the contrast between ejectives and other stops relying on post-release length. Both studies fail to address the adequate similarities between ejectives and other stops for measuring Voice Onset Time (henceforth called VOT) and the possible implications of these measurements in languages where ejective and non-ejective equivalents contrast. Ejectives must therefore be treated as meaningful data to be gathered and analyzed alongside other stops in Navajo and Tlingit, with the caveat of understanding their differences.

Rather than presupposing that Navajo and Tlingit simply distinguish between uniform groups (voiceless, voiceless aspirated and ejective stops), using a spectrum-based theory allows researchers to

determine with more specificity what the manifestations of these stops surface as. By comparing VOT measurements, this paper seeks to determine the phonetic profile of stops in Navajo and Tlingit. The three questions driving this research are: (1) What are the average VOTs for the stops in Navajo and Tlingit? (2) How do VOTs compare across each category (voiceless, aspirated, and ejective)? (3) What might the discerned patterns tell us about the relationship between these two languages and beyond?

2 Language(s) Background

Tlingit (*KLING-IT*) and Navajo (*NAH-VUH-HO*) are two languages of the Na-Dene language family, an Indigenous language family of North America. Tlingit is spoken along the Western coast of Alaska, British Columbia, and the Yukon. The language is considered critically endangered and has only a few hundred living native speakers, spurring intense language documentation and revitalization efforts by Tlingit communities. Historically, Tlingit split from the Na-Dene language family long before Navajo's formation, making them distant relatives descending from a singular Proto-Na-Dene. Navajo is spoken within Southwestern North America, including parts of Arizona, New Mexico and Utah. Comparatively, Navajo is a much younger language and maintains cross-generational transmission, but is still considered highly vulnerable to erasure and requires revitalization interventions.

Tlingit and Navajo share many phonetic characteristics associated with the Na-Dene language family, particularly the use of contrastive tone (high and low), a large consonant inventory and a complete lack of voicing in stops and affricates. Tlingit and Navajo then have a three-way contrast for stops that distinguishes plain voiceless, aspirated voiceless, and ejective voiceless as separate phonemes (Hargus 2010). These languages also contrast regular back obstruents against labialized back obstruents marked with a 'w' diacritic. The distribution of consonants in Na-Dene languages is influenced by the morphological and syntactic constraints of the language; for more information on morphophonology, see McDonough (2013).

Despite these similarities, these two languages are discernable from one another in several key ways. Tlingit has several ejective fricatives (/t', k', k^w' q', q^w'/) that Navajo lacks, while Navajo has a bilabial plosive (/p/) that Tlingit lacks. Tlingit also has very few sonorants— one nasal (/n/) and two approximants (/w, j/)— compared to Navajo's two nasals (/m, n/) and three approximants (/w, j, ɰ/). Consonants in Tlingit prefer dorsal places of articulation, while Navajo is primarily composed of coronal sounds (Hargus 2010).

The voicing contrasts of the stops and affricates are indicated by a unique type of orthography shared by both languages. Traditionally, Na-Dene languages use the IPA symbols for their voiced

counterparts to mark ‘plain’ voiceless stops (/d/ = [t]). The aspirated voiceless stops use the IPA symbol for the equivalent voiceless segment (/t/ = [tʰ]). The ejective voiceless stops use the IPA symbol for the equivalent voiceless segment followed by an apostrophe (/t’/ = [t’]). Despite minor differences in symbol usage between linguistic adaptations of the orthographies of Navajo and Tlingit, this division is currently the most popular writing system. This writing system will be maintained throughout the paper except when the distinction between sound segments necessitates clarification.

	Bilabial	Laterals	Alveolar	Post-alveolar	Palatal	Velar (+Labialized)	Glottal
Unaspirated Stop	(p)		t			k kʷ	ʔ
Aspirated Stop	(pʰ)		tʰ			kʰ kʷʰ	
Ejective Stop			t’			k’	
Unaspirated Affricate		t͡ɬ	t͡s	t͡ʃ			
Aspirated Affricate		t͡ɬʰ	t͡sʰ	t͡ʃʰ			
Ejective Affricate		t͡ɬ’	t͡s’	t͡ʃ’			
Voiceless Fricative			s	ʃ		x	h
Voiced Fricative			z	ʒ		ɣ	
Laterals			ɬ l				
Nasals	(m)		n				
Approximant	w				j		

Figure 1: Navajo consonant inventory (IPA) adapted from Young & Morgan (1980) and McDonough (2013).

	Labial	Alveolar	Post-alveolar	Lateral	Palatal	Velar (+Labialized)		Uvular (+Labialized)		Glottal
Unaspirated Stop	(p)	t				k	k^w	q	q^w	ʔ
Aspirated Stop	(p^h)	t^h				k^h	k^{wh}	q^h	q^{wh}	
Ejective Stop		t’				k’	k^w’	q’	q^w’	
Unaspirated Affricate		t͡s	t͡ʃ							
Aspirated Affricate		t͡s^h	t͡ʃ^h							
Ejective Affricate		t͡s’	t͡ʃ’							
Plain Fricative		s	ʃ			x	x^w	χ	$χ^w$	h
Ejective Fricative		s’				x’	x^w’	χ’	$χ^w$’	
Nasal Stop	(m)	n				(ŋ)				
Approximant				(l)	j	w	ɯ			

Figure 2: Tlingit central consonant inventory (IPA) adapted from Maddieson, Smith & Bessel (2001).

3 Methods

The anatomy of a stop can be divided into three main sections: (1) the closure, where the airflow is obstructed even as pressure mounts in the vocal tract, (2) the release, where the obstruction is removed, creating a sharp relief of the pressure in a ‘popping’ sound, and, depending on the stop, (3) a period of exhalation before the next sound segment begins (Lisker & Abramson 1964). Typically, stops are produced with pulmonic eggressive airflow, with the obstruction originating somewhere in the mouth. Plain voiceless stops lack a lengthy period between the release and the beginning of the following sound. In the brief period following the release, the stops lack energy or ‘noise’ on the spectrogram and the oscillogram, as no exhalation accompanies the release. In contrast to plain stops, aspirated stops have a longer pause between the release and the beginning of the following sounds. This interval contains significant energy, or ‘noise,’ as the speaker exhales following the release of the built-up pressure, called aspiration (Lisker & Abramson 1964). Ejective stops differ from plain and aspirated stops as they utilize a glottalic egressive airstream mechanism, with the sound originating in the contraction of the glottis rather than air travelling from the lungs (Wright, Hargus & Davis 2002).

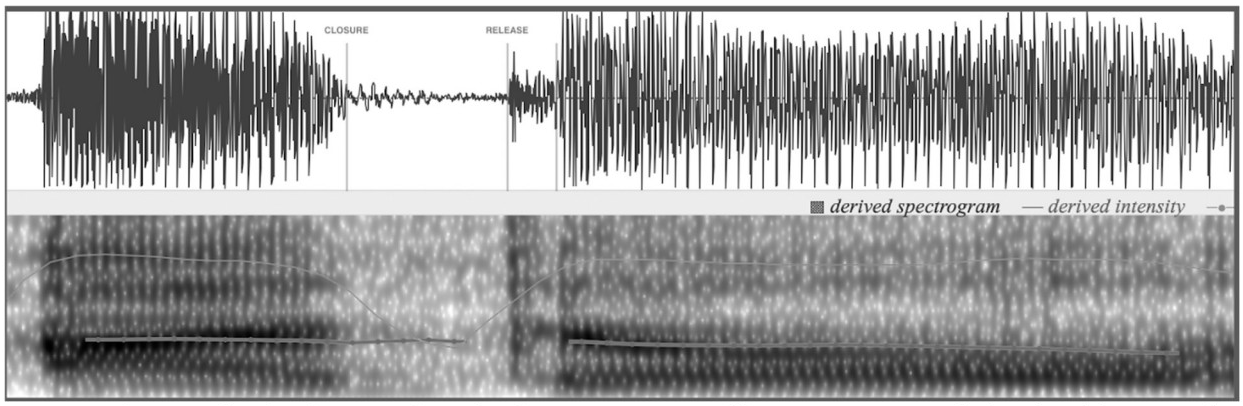

Figure 3: Example of a plain voiceless alveolar stop in Navajo, /t/. Its anatomy is marked with red lines, i.e. the stop's three components: the closure, the release, and the end of the sound.

Since its conception in 1968, Voice Onset Time (VOT) has become a major phonetic tool used to identify contrasts between individual stop tokens and discern larger patterns in the lengths of different stop classifications within a language (Lisker & Abramson 1964). Phoneticians commonly use voicing to reliably distinguish stops from one another within voice recordings or in cases where the larynx activity is not easily measurable. Beyond phonetics, VOT is also an important tool for phonological work as it helps to measure the variable pronunciation of stops depending on their context (Abraham & Whalen 2017). This paper will focus solely on VOT averages for the three contrastive structures of stops within Navajo and Tlingit. Still, it will include a discussion of phonological variability and the next steps for research within the discussion portion. Cho & Ladefoged (1999) describe a continuum for universal voicing that problematizes using VOT to measure the contrast between stops as the variable production can lead to frequent miscategorization of stops. This indicates that VOT is insufficient for addressing the breadth of contrastive stops cross-linguistically, since languages like Hindi or Korean exhibit unusual voicing and aspiration contrasts (Abraham & Whalen 2017).

VOT is a measurement of the time that elapses between the 'release' of the stop and the 'onset' of glottal pulses typically associated with the subsequent sound segment. The release is marked as the 'zero,' or starting point, and the onset is placed somewhere along a negative to positive scale centered around that starting point. Although the negative to positive scale (i.e. whether they occur before or after the starting point) is consistent, classifying stops based on their VOTs depends on traditional patterns for stops within a language. Following standards in phonetic analysis, this paper will assume that: (1) aspiration will lengthen the VOT, (2) the 'release' is crucial to the measurement as a starting point, and (3) in plain stops, the voiced segment will traditionally be longer than its voiceless counterpart (Abraham & Whalen 2017).

In their original paper, Lisker & Abramson (1964) identified the optimal context for measuring VOT as the onset of a CV syllable, and subsequent research identified intervocalic, clustered, and coda-position stops as more likely to experience "bleed"— accidental inclusion of the phonetic properties of surrounding sounds— and provide less accurate data. Previous research in Navajo has also indicated that speakers are likely to target the optimal CV syllable through affixation and epenthesis when necessary to avoid 'marked' syllable structures (Wright 1984). To accurately identify the VOTs of different stops within this study, only stops within CV syllables will be measured. Focusing on stops within CV syllables will ensure that research is not influenced by secondary phonological constraints that limit the inclusion and pronunciation of stops in casual speech.

Also, any stops that do not contain a prominent 'release' are discounted, as there is no accurate way to differentiate the starting point of the sound segment from the closure that precedes it. These sound segments are traditionally called 'unreleased stops' and are characterized by minimal activity in an area where the stop would be typically produced. Despite these unreleased stops lacking the necessary qualities for measuring VOT, they still contain a closure and may hold phonological significance. Across the 83 syllable-initial CV clusters analyzed for each language, each had four unreleased stops. See the 'Discussion' portion for further information on 'unreleased stops' and glottal stops.

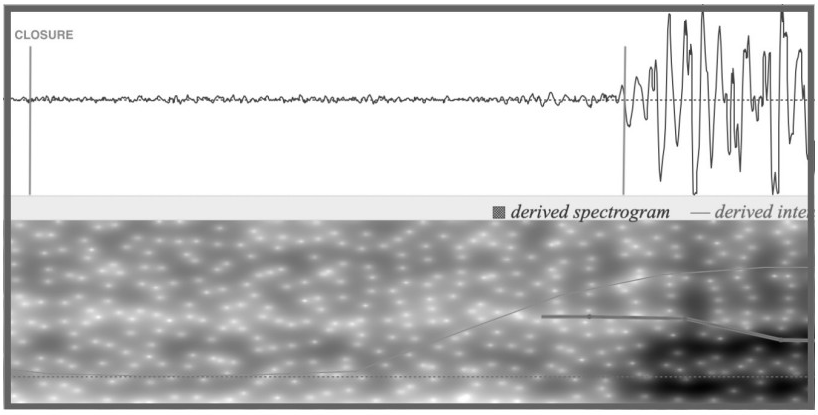

Figure 4: Example of an unreleased stop that is targeting /k^{h}/. Its anatomy is marked with red lines, i.e. the closure and the end of the sound, with a noticeable lack of a release point.

Researchers identified 166 sound segments (83 sound segments in each language) and, excluding the eight unreleased stops, analyzed 158 individual stops encompassing the complete phonetic inventory of each language. These stops were categorized based on language, then further divided by stop type. The unequal distribution of stops is due to the measurement technique and the asymmetrical distribution of the stop types. The first 83 tokens in a 3-minute recording of casual speech were selected to minimize the introduction of research bias through the hand-selection of tokens.

The source for the Tlingit portion is *Four Tlingit Stories,* a collection of oral stories narrated for the October 2014 Tlingit Literacy session at the Yukon Native Language Centre. The storyteller is Sam Johnston, a fluent speaker and leader of the Ishkìtàn clan, and Dr. Jeffry Leer transcribed and translated the audio. The source for the Navajo portion is the *2009 Valentine's Day message* by the 2008-2009 Miss Navajo, Yolanda J. Charley, broadcast on KTNN radio. The audio from the broadcast was later transcribed and translated by Daybreak Warrior on Youtube as a part of his ongoing efforts to increase the accessibility of Navajo.

The sources were selected based on speaker criteria, legibility, accuracy and clarity. Speakers were required to be fluent native speakers of their respective language and communicate with relative confidence in their production. Researchers ensured that the audio had minimal background noise, the speaker was enunciating, the transcription was explicit, and the translation was accurate. The *2009 Valentine's Day message* and *Four Tlingit Stories* were selected based on their ability to meet all necessary criteria. The decision to use one source for each language is partly meant to simplify the data-gathering process and partly due to insufficient sources that met the above criteria. For further discussion about the impacts of single sources on research, see the subsections under 'Consideration' about idiolectal variation and orthography.

The process for the identification and measure of VOT was performed on Praat, a software used for phonemic analysis. All 'stop' tokens in the recordings were identified, and the three parts of the stop were marked and labelled. The beginning of the 'burst' was defined as the first meaningful spike on the oscillogram following a quiet period, i.e. the closure. The end of the sound segment was determined to be the transition in the oscillogram/spectrogram that coincided with the beginning of a vowel. The distance between the spike, i.e. the 'burst' and the transition, i.e. the end of the sound segment, was measured in milliseconds. All measurements were collected at the nearest zero crossing (the boundary between negative and positive measurements) on the oscillogram, in line with phonetic measurement standards. Once the 83 tokens had been measured in their order of appearance, they were recorded in a table and averaged to determine the mean VOT of each stop type for Navajo and Tlingit. Other data calculated from the measurements included the range and distribution of the tokens based on various parameters.

4 Considerations

i) Transcription/Orthographic Bias

In any transcription or translation, the choices made by the transcriber can seriously affect the accuracy of the new written communication with respect to its original intended message by the speaker. This is unfortunately one of the unavoidable side effects of transcription and translation work.

Transcribers are tasked with first interpreting the message through the lens of their own experiences and then relaying that interpretation of the message in a different format. The transcriber's choices can lead to gross misinterpretations of source material, which problematizes the source's usage when seeking to study its qualities. Often in Indigenous storytelling, the author must represent grammatical elements in roundabout or insufficient ways because English has no equivalent. Dauenhauer & Dauenhauer (2011) list several challenges that translators must address in Tlingit, where different representations can entirely change the meaning and composition of the utterance. These include the representation of imperfectives that lack English equivalents and the whole of the aspectual system in grammar for which English has no comparable system. Since Tlingit and Navajo share many grammatical elements that lack an English equivalent, it can be assumed that both translators have repaired these 'gaps' based on their own experiences as bilingual speakers of English and Navajo or Tlingit (Dauenhauer & Dauenhauer 2011). The transcriptive choices are significantly more likely to affect this research, but vigilance of translator bias remains crucial.

Similar to problems created by transcription bias, the choices made in how to represent the sounds of the language best can introduce orthographic bias. An 'orthography' refers to a language's spelling convention associating sounds/segments with certain symbols. Generally, the Indigenous languages of Canada contain much of the same orthography as English, especially for those sounds that English itself has. However, orthographic representation becomes more tricky when representing sounds in Indigenous languages not traditionally considered phonemic in English, such as ejectives or aspirations. The orthographies of Navajo and Tlingit are highly contested among different groups, each following different linguistic standards on how best to represent these unique sound segments.

Some of these concerns can be addressed through a clear methodology that minimizes the opportunity for bias in data collection and analysis. The Navajo speaker was providing a speech on the radio, and the Tlingit speaker was telling a story to a room of people, creating a setting that encouraged projective speech. Similarly, both speakers understood that their voices would be recorded and transmitted through speakers, so their mediums would encourage careful and emphatic speech productions. Such circumstances likely created the conditions for the speakers to project their voices to increase their prominence and enunciate their speech sounds more to increase the recording's intelligibility to listeners. The clarity of speech recordings minimized the transcription bias due to misinterpretation or difficulty discerning the sound segments originating in the recorded conversation's poor quality or casual nature.

The familiarity and experience of the transcribers would also aid in minimizing the bias found within their transcriptions. Daybreak Warrior is a bilingual native speaker of Navajo and English and had several years of experience by this time in transcribing Navajo speech and translating it to English in the process (Daybreak Warrior 2006). Dr. Jeffry Leer is a professor emeritus at the University of Alaska Fairbanks who specializes in studying Na-Dene languages, making him familiar with transcription conventions and research for the language family. He is also a bilingual speaker of both English and Tlingit, giving him ample experience in Tlingit (Alaska Native Language Center 2023).

Despite these interventions, the inevitability of transcription and orthographic bias necessitates the constant consideration of these factors in tandem with any conclusions drawn from the data gathered. The limited amount of sources to choose from when studying Indigenous languages and the limited control by researchers on how these sources are recorded makes strategies such as multimodal transcription difficult to implement in practice (Mondada 2018). Several common orthographic and transcription biases are listed below to prepare readers on what biases are relevant to the results of this study.

Non-verbal information:

Voice recordings, and therefore the researchers who study them, are unable to account for the considerable amount of information that is conveyed using non-verbal communication. Non-verbal communication refers to the information communicated by the speaker separate from the utterance and its properties. Non-verbal communication includes hand gestures, facial expressions, pointing, eye contact, posture, and other movements or body language (Wiener, Devoe, Rubinow & Geller 1972). Without any way to track anything beyond the spoken word, all nonverbal communication is lost with audio recordings. Speakers often use non-verbal communication in tandem with verbal communication, such as pointing to a location being discussed. It can also provide additional context that changes or expands the meaning of what is being said. Since this information is missing from audio recordings, transcribers cannot record or incorporate it into their transcriptions or translations. Two common solutions are providing video recordings of the speaker for the transcriber to watch or having the transcribing individual present for the actual speaking event. With this additional information, the transcriber has more opportunity to consider nonverbal cues and how they may affect the utterance. These solutions still require further revision, considering non-verbal communication is only sometimes easily translatable into words and often up for interpretation, leaving it vulnerable to bias-introduction (Mondada 2018).

Metacommunication:

Much like entirely nonverbal communication, information can also be communicated by secondary qualities of speech that are separate from the actual conventions of utterance building. These qualities include paralinguistic and prosodic information, such as pitch, duration, speaking style (e.g. passive vs active), and stress. Most importantly, they are not a contrastive quality of language. Pitch is a prosodic quality of language that can change the intended meaning of a phrase, while tone is a phonemic quality of language that can change the direct meaning of a sound segment/word. For example, pitch-raising at the end of an utterance may indicate it is a question, while a high tone would change the phoneme being used by the speaker. Transcribers often neglect meta-linguistics since it does not involve observing what is being said but rather the utterance's implications.

Orthographic representation:

The history of orthography invention for Indigenous languages is often intertwined with the increasing politicization and colonization of Indigenous identity, and Navajo and Tlingit are no exception. Much of the early orthographic work was created by colonial settlers and academics using either Cyrillic or Latin conventions, rarely receiving input from native speakers (Sebba 2007). This makes much of the foundation of modern orthographies of Indigenous languages fragile at best and unrepresentative at worst. In recent decades, linguists and native speakers have worked to establish orthographic standards that would better represent the sounds of the language and avoid the colonial bias introduced in earlier manifestations of the writing system. Still, these orthographic standards are not without problems. Often orthographic standards are developed based on a single dialect of a language, then extrapolated onto all other dialects, whether or not it accurately represents their sound system. Orthographic standards are frequently only standardized within certain spheres of usage or among certain individuals, such as academic settings, limiting its accessibility to the language's speakers (Littell et al. 2018). Authors and transcribers are rarely explicit about their orthographic choices, choosing whichever style they have the most experience in without acknowledging its insufficiencies in representing the speaker's pronunciations. The writing conventions of transcribers are also commonly focused on representing what the speaker means to say rather than what they actually say, leading to a misrepresentation of the data. For example, neither Daybreak Warrior or Dr. Leer marked the unreleased stops identified by researchers, representing them instead based on their target sounds. Since researchers are basing their categorizations on the transcribers', these classifications might lack the nuance within the data and therefore be unfairly

represented within this research. While there is little evidence that orthographic bias has significantly impacted this paper's conclusions, it is still an important consideration in examining the data.

ii) Dialectal and Idiolectal Variation

Much of Tlingit's dialectal variation focuses on non-phonological elements or the production of tones or vocoids. There is also ample evidence for dialectal variation in Navajo. Yet dialectal variation is unlikely to affect VOT measurements as it typically concerns lexical change or non-stop phonemes like vowels and fricatives (Saville-Troike 1974). Both languages demonstrate a parallel form of dialectal variation that involves a change in the place of articulation for obstruents. In Tlingit, this arises with some speakers' delabialization of stops word-finally. In Navajo, this materializes as the regular substitution of /t/ → /k/ among certain dialects, including Kiowa Apache and Jiricallia. While these differences affect the pronunciation of stops, there is no evidence that VOT in Navajo is impacted by place of articulation (Saville-Troike 1974). Thus, despite the difference in where each speaker produces stops, there is little possibility that these phenomena would impact VOT stop measurements.

Similarly, the largest idiolectal variation found within Tlingit is regressive palatalization and progressive labialization, and while they target the place of articulation, they affect vowels rather than obstruents (Wolfe 1977). No research into the idiolectal variation of Navajo speakers has been conducted thus far, limiting the opportunity to examine the phenomenon's impact on research. All the same, research shows that on an individual level or within a dialect of these languages, changes in the production of stops concern the place of articulation rather than the stop's manner of articulation or any secondary phonemic qualities. There is no evidence in either language that dialectal or idiolectal variation influences the categorization of stops or their VOT measurement. It is unlikely that these subtle differences between stops will impact data accuracy. New research on the dialectal and idiolectal variation in Na-Dene languages is necessary, but it can be assumed that these changes are not substantial enough to warrant data intervention for the purposes of this study.

5 Results

As mentioned above, 162 sounds were selected, and 156 stops were analyzed following exclusions. These 156 stops were divided equally among Tlingit and Navajo, with each language having 78 analyzed stops. In Navajo, 54 were plain stops, 13 were aspirated, and 11 were ejectives. In Tlingit, 43 were plain stops, 26 were aspirated, and 9 were ejectives. The ratio of stop types (plain, aspirated and

ejective) were proportionate between the two languages, with the number of aspirated stops reduced in Navajo in favour of plain stops.

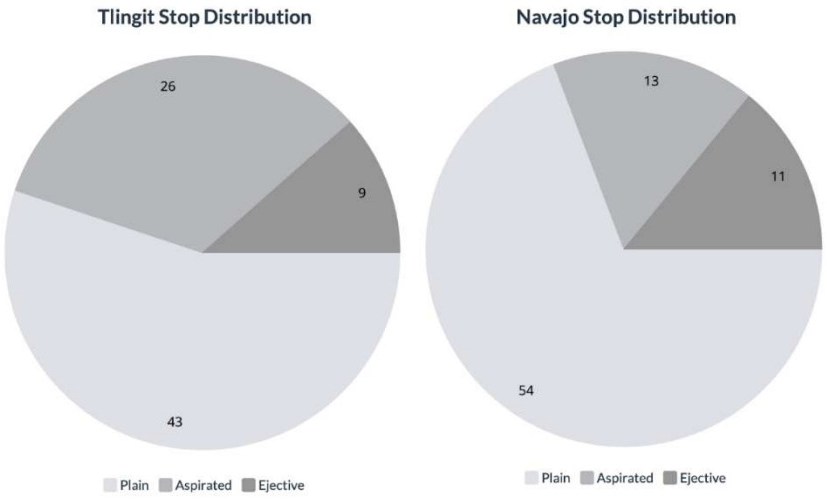

Figure 5: Graphs demonstrating the distribution of stop types (plain, aspirated, ejective) in Navajo and Tlingit.

The average length of VOT for plain stops was 21.1 ms in Navajo and 11.5 ms in Tlingit. The average length of VOT for aspirated stops is 90.5 ms in Navajo and 36.9 ms in Tlingit. The average length of VOT for ejective stops is 76.5 ms in Navajo and 31.8 ms in Tlingit. Ignoring their categorization, the average VOT for all analyzed stops was 48.4 ms in Navajo and 22.3 ms in Tlingit.

	Navajo			Tlingit		
	Plain	Aspirated	Ejective	Plain	Aspirated	Ejective
Avg. VOT (ms)	**21.1**	**90.5**	**76.5**	**11.5**	**36.9**	**31.8**
Avg. VOT (ms)	**48.4**			**22.3**		

Figure 6: Summary of the results divided by stop types and languages.

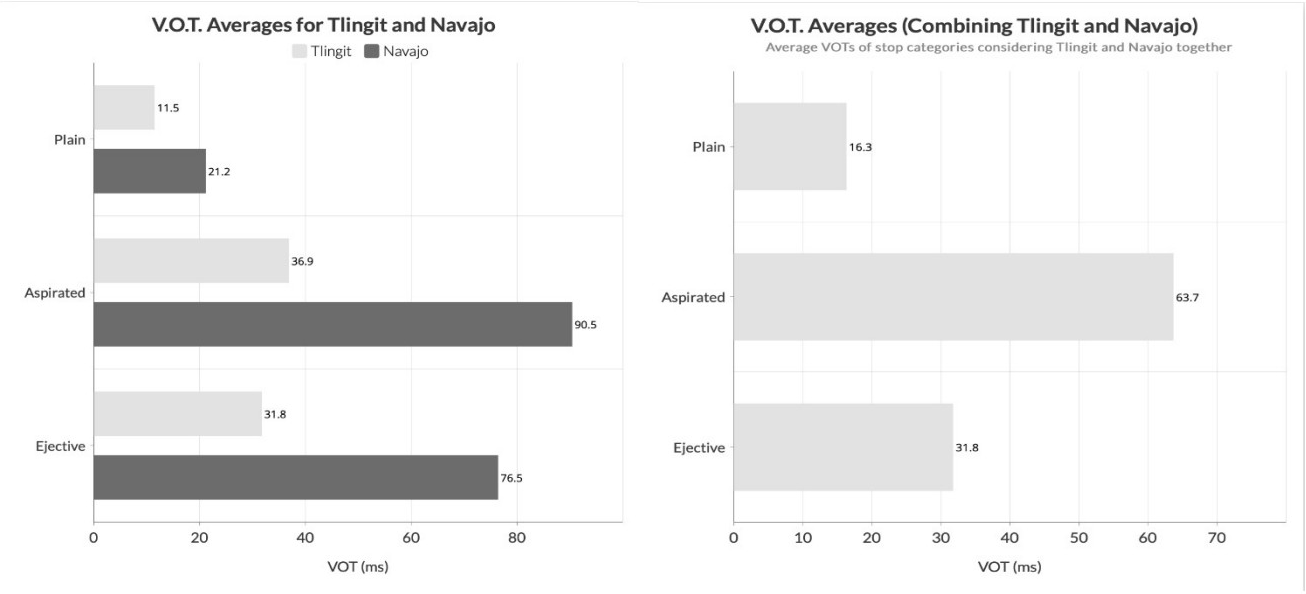

Figure 7: On the left, a graph demonstrates average VOTs compared between the stop categories of Navajo and Tlingit. Figure 8: On the right, a graph demonstrates the average VOT of each stop category, combining Navajo and Tlingit data.

6 Discussion

i) Stop Type Comparison

There is an evident correlation between stop types and VOT measurements: plain stops are associated with shorter VOTs, aspirated stops are associated with longer VOTs, and ejective stops are considered to exist somewhere between these two boundaries. Ejective stops are less uniform in their productions than aspirated and plain stops, leading to higher variability in their VOTs. Ejective tokens vary by 63 ms in Tlingit and 89 ms in Navajo, while plain stops vary by 44 ms and 25 ms, respectively. Aspirated and ejective voiceless stops are closer in their VOTs than plain voiceless stops are to either type of stops.

In both languages, the most common stop is plain, followed by aspirated, the least common being ejectives. The majority of the tokens analyzed were plain stops (62% and 55%), demonstrating a clear preference for plain stops over aspirated (18% and 33%) and ejective stops (14% and 11%). This preference may indicate that plain stops are the least 'marked' of the phonological categories in Na-Dene languages, likely due to their simplicity in production compared to ejective and aspirated stops.

ii) Navajo and Tlingit Comparison

While Navajo and Tlingit may select different points along the VOT 'continuum' as their production target, they still pattern similarly. Speakers consider the distinction between these stop structures crucial to their production, which explains the insignificance of minute differences in the data. Navajo has significantly longer VOTs than Tlingit, averaging 36 ms longer in stop VOTs than Tlingit. The most significant difference in VOT length occurs in aspirated stops; Navajo aspirated stops have the longest VOT at 90.5 ms compared to Tlingit aspirated stops averaging 36.9 ms. This data suggests that, on average, Navajo speakers produce longer stops than Tlingit speakers produce their equivalents.

In general, the range of VOTs is much more narrow in Tlingit compared to Navajo. The data clearly shows a difference in how Tlingit and Navajo arrange their phonemic inventories. Navajo is less accurate in targeting phonemes but repairs this inaccuracy through wider divides between individual sound categories. Tlingit is more accurate at targeting phonemes, so wide distinctions between sound categories are unnecessary. These differences highlight how languages can repair indiscernibility as it arises, either by promoting the differences between phonemes or increasing phoneme production's accuracy.

An unexpected value found in both languages was unreleased stops, which could be identified but not measured, given the lack of the necessary 'release' for VOT measurements. Three unreleased stops were present in the first 81 sounds selected for analysis in both languages. They did not appear in similar contexts among the 81 sounds or contain any significant similarities. Further research is necessary to determine if the equivalent number of unreleased stops indicates a pattern or is coincidental.

iii) Language Family Comparison

Unfortunately, research is currently confined to Tlingit and Navajo, neglecting the many other languages in the Na-Dene family that may contribute to our understanding of VOT. This lack of research poses limitations primarily around the accuracy and form of conclusions about the Na-Dene language family based solely on the data analyzed in this study. Speculations will be made to address this on the assumption that Tlingit and Navajo are at least somewhat representative of the whole language family, with the caveat that this may turn out false, pending new research on the subject.

Since Navajo and Tlingit have similar patterns, we can assume these standards are upheld throughout the Na-Dene language family. The high contrast between plain and aspirated stops is likely to

parallel across all languages simply due to the nature of each stop type, with one requiring significant soundwaves following the release and the other requiring an absence of that activity. The distribution of stops for the two languages is not necessarily demonstrative of a wider pattern for the distribution of stops in the Na-Dene language family since it may be incidental that plain stops are the preferred stop type. The distribution, especially that of Tlingit, is ambiguous enough that it could be debunked through a series of examples of other languages that favour aspirated or ejective stops over plain ones.

This research can also be used to explain the phonological processes surrounding ejectives in the Na-Dene language family. Ejective voiceless stops have the highest degree of variability in their VOTs, with the tokens ranging from 11-74 ms in Tlingit and 20-109 ms in Navajo. Aspirated and ejective voiceless stops are closer in their VOTs than plain voiceless stops are to either type of stops. The ejective/non-ejective contrast for stops is consistently difficult for listeners to discern in other Na-Dene languages (Wright, Hargus & Davis 2002). The lack of distinction may be due to the variability with which ejectives are produced and the sharp difference in their production compared to aspirated stops. If ejectives can be anywhere along the VOT range of the language, then VOT length is no longer a reliable feature in categorizing stops. Similarly, if ejective stops consistently mirror the VOT of aspirated stops, listeners may struggle to accurately identify sounds from one another. While more research is necessary to study the extent to which other languages of the Na-Dene family are affected by the minimal ejective/non-ejective contrast, these VOT measurements may suggest this indiscernibility is due to similar VOT realizations.

iv) Further Research

Following the measurements and conclusions provided in this study, several avenues for future research present themselves: (1) A study into the appearance and mechanisms of unreleased stops, whether at a language or individual-specific level. (2) A study into discernibility between the ejective/non-ejective contrast in Navajo or Tlingit. (3) A study examining the change in VOT length or the distribution of stop types across morphological domains. (4) A study measuring the VOT of stop categories in other languages beyond Navajo and Tlingit in the Na-Dene language family.

7 Conclusion

As discussed in the introduction, three questions guided this research: (1) What are the average VOTs for the stops in Navajo and Tlingit? (2) How do they compare across each category (voiceless,

aspirated, and ejective)? (3) What do the discerned patterns tell us about the relationship between these two languages and beyond? To answer these questions, the stops identified in the continual speech of two native speakers, Sam Johnston, via 'Four Tlingit Stories' and the 2008-2009 Miss Navajo Yolanda J. Charlye, via her Valentine's Day message, were measured.

Stop categorization (voiceless plain, aspirated, and ejective) dictated the length of VOT for sound segments, where longer sound segments were aspirated, and shorter sound segments were plain. The length of ejectives was much more variable in both languages than the production of plain and aspirated stops, likely contributing to the perception difficulties previously identified along the ejective/non-ejective division. The distribution of stops in Navajo and Tlingit shows a preference in both languages for plain stops, with ejectives being the least frequently produced. This preference may be due to the 'unmarkedness' of plain stops compared to the other phonemic stop categories. Navajo and Tlingit show different mechanisms for increasing the discernibility of stop contrasts; Navajo increases the distinction between stop classifications and Tlingit increases accuracy in targeting productions. The similarities between Navajo and Tlingit were predicted to represent wider patterns of the Na-Dene language. Such theories necessitate more data collection and analysis of other languages in the language family.

Returning to the theories in the Introduction, this data supports a perspective on voicing that is more similar to Cho & Ladefoged (1999) than Keating (1990). Despite emulating the unconscious distinction between stop types, the speaker's production of stops is highly variable and resembles a target 'range' rather than a singular point. This suggests that phonemes are selected from a continuum of voicing, with each language selecting ranges that, although similar, show minute differences in their realization of each stop classification. The differences in how Navajo and Tlingit repair indiscernibility further support the soundness of a range-based rather than a point-based theory. Tlingit appears to align more with Keating (1990)'s theory due to its narrower realization of stops, while Navajo aligns more with Cho & Ladefoged (1999) due to its higher production variability. However, Cho & Ladefoged (1999)'s theory is more easily adapted to encompass evidence from Tlingit than Keating (1990)'s is to readily accept Navajo. This research may suggest that the selected 'range' on a continuum is not done uniformly and varies depending on language-specific factors, much like the target ranges each language selects. For example, applying Tlingit to Cho & Ladefoged (1999)'s theory may entail acknowledging that the selections are of much shorter ranges, leading to what appear to be more accurate 'target' productions. The question of ejectives' role in these theories remains unaddressed. Still, increasing the available data may lead to a revision of voicing theories to encompass ejectives as well.

Furthermore, this paper discussed several considerations that may impact phonetic research, including transcriber influence, orthographic bias, and dialectal and idiolectal variation. The weight of these concerns was debated, and several solutions were provided to ensure the accuracy of any conclusions drawn from the data. A brief discussion was provided about the exclusion of glottal stops and the commonality of unreleased stops in the two languages. Finally, further research suggestions were provided to confirm the hypotheses proposed in this paper or expand the degree of analyses available to linguists about the Na-Dene language family.

References

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

Bird, Sonya. 2002. *The phonetics and phonology of Lheidli intervocalic consonants.* University of Arizona: Ph.D. dissertation.

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: 109-115.

Cho, Taehong & Peter Ladefoged. 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27.2: 207-229.

Dauenhauer, Richard & Nora Marks Dauenhauer. 2011. Tense, Aspect, and the Organization and Translation of Tlingit Oral Narrative. In *Working Papers in Athabaskan Languages*. Fairbanks, Alaska: Alaska Native Language Project.

Daybreak Warrior. 2010. *Miss Navajo's Valentine's Day Message.* https://www.youtube.com/watch?v=waELx-wZm0c (Accessed 30 January, 2023).

Edwards, Keri. 2009. *Dictionary of Tlingit.* Juneau, Alaska: Sealaska Heritage Institute.

Hargus, Sharon. 2010. Athabaskan phonetics and phonology. *Language & Linguistics Compass* 4.10: 1019–1040.

Johnston, Sam. 2016. Four Tlingit Stories. *Tlingit Literacy Session.* Jeffry Leer (trans.). Whitehorse: Yukon Native Language Centre. https://ynlc.ca/tlingit/#four-tlingit-stories-by-sam-johnston.

Langford, Peter & Ian Maddieson. 1993. *The Sounds of the World's Languages.* Hoboken, New Jersey: Wiley-Blackwell Press.

Lisker, Leigh & Arthur S. Abramson. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD* 20.3: 384-422.

Littell, Patrick et al. 2018. Indigenous language technologies in Canada: Assessment, challenges, and successes. In *Proceedings of the 27th International Conference on Computational Linguistics:* 2620-2632. Santa Fe, New Mexico: Association for Computational Linguistics.

Maddieson, Ian, Caroline L. Smith & Nicola Bessell. 2001. Aspects of the phonetics of Tlingit. *Anthropological Linguistics* 43.2: 135–176.

McDonough, Joyce. 2013. The Dene verb: how phonetics supports morphology. *18th Workshop on Structure and Constituency in the Languages of the Americas*. University of California, Berkeley.

Mondada, Lorenza. 2018. Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality. *Research on Language and Social Interaction* 51.1: 85–106.

Saville-Troike, Muriel. 1974. Diversity in Southwestern Athabaskan: A historical perspective. *Navajo Language Review* 1.2: 18-35.

Sebba, Mark. 2007. *Spelling and Society: The Culture and Politics of Orthography around the World.* United Kingdom: Cambridge University Press.

Tuttle, Siri G. 2005. Duration, intonation and prominence in Apache. In Sharon Hargus and Keren Rice (ed.), *Athabaskan prosody*, 319–44. Amsterdam: John Benjamins.

University of Alaska Fairbanks. *J. Leer: Comparative Athabaskan Lexicon (CAL), Alutiiq lexicon.* National Science Foundation. https://www.uaf.edu/danl/project-updates/jeff-leer/ (Accessed 20 April, 2023)

Wiener, Morton, Shannon Devoe, Stuart Rubinow & Jesse Geller. 1972. Nonverbal behaviour and non-verbal communication. *Psychological Review* 79.3: 185-214.

Wolfe, Gaelan de. 1977. *Tlingit Phonology in a Generative Framework: An examination of phonological processes and abstract representation*. University of British Columbia: MA thesis.

Wright, Martha. 1984. The CV Skeleton and Mapping in Navajo Verb Phonology. *North East Linguistics Society* 14.28: 461-477.

Wright, Richard, Sharon Hargus, & Katherine Davis. 2002. On the categorization of ejectives: Data from Witsuwit'en. *Journal of the International Phonetic Association* 32.1: 44–77.

Young, Robert & William Morgan. 1980. *The Navajo language: A grammar and colloquial dictionary.* Albuquerque: University of New Mexico Press.