UC Berkeley
UC Berkeley PhonLab Annual Report

Title

An Acoustic Outlook on Initial Stops in Northern Shoshoni

Permalink

https://escholarship.org/uc/item/3w05g32t

Journal

UC Berkeley PhonLab Annual Report, 14(1)

Author

Garvin, Karee

Publication Date

2018

DOI

10.5070/P7141042478

Peer reviewed

An acoustic outlook on initial stops in Northern Shoshoni

Karee Garvin

1. Introduction

Shoshoni is a member of the Uto-Aztecan language family and consists of three dialects: Western Shoshoni, Northern Shoshoni, and Eastern Shoshoni. While there has been descriptive research done on the phonetic and phonological properties of all three dialects of the language, little to no acoustic analysis has been done thus far. This paper seeks to begin the discussion of the acoustic properties of Northern Shoshoni. Specifically, the discussed data are from a speaker of Northern Shoshoni from the Shoshone[1]-Bannock Tribes of the Fort Hall Reservation; in this paper I examine the voice onset time of initial stops in Shoshoni.

2. Background on the Shoshoni Language

Shoshoni is a member of the Central Numic branch of the Uto-Aztecan family, along with Comanche and Timbisha. Dialects of Shoshoni are spoken in Idaho, Wyoming, and Nevada, as well as parts of Oregon, Utah, and California (Lewis, Simons, and Fennig 2014). The Northern dialect of Shoshoni is spoken predominantly in Idaho. This particular study focuses on the dialect of Northern Shoshoni spoken on the Fort Hall Reservation in South Eastern Idaho.

It is difficult to estimate the number of Shoshoni speakers as there are no reliable figures for this information. Ethnologue reports 1000 speakers, and the 2010 census reports 2800 speakers (Lewis, Simons, and Fennig 2014; Moseley 2010). However, the number is likely somewhere closer to 4000 or 5000 speakers with a native population of 12,000 (Loether 2006). A vast majority of these speakers are elders. In Fort Hall, there are approximately 5000 members of the Shoshone-Bannock Tribes living both on and off the reservation (personal communication, Drusilla Gould, May 12, 2014). However, only a fraction of these members are fluent speakers of the language. Furthermore, it is estimated that less than 5% of the speakers are under the age of 18 (Loether 2006). Because of the uneven distribution of speakers, UNESCO lists Shoshoni as severely endangered, and Ethnologue lists Shoshoni as threatened (Moseley 2010; Lewis, Simons, and Fennig 2014).

The Shoshoni language has three different orthographies in use. These include the Wick Miller orthography, ISU orthography, and the Tidzump Orthography. The Wick Miller orthography is phonemically based and originated at the University of Utah. This orthography was named after the late Wick Miller and it is the most widely used orthography in publications. The ISU orthography is phonetically based and was created by Drusilla Gould and Christopher Loether at Idaho State University. The Tidzump orthography originated in Wyoming and is used predominantly in that area (personal communication, Christopher Loether, April 3, 2014). The variation in orthography, particularly for the Wick-Miller and ISU Orthography, is important because it can cause a blurring of lines of pronunciation, especially where Shoshoni differs from

---

[1] Shoshoni is spelled with an 'i' when referring to the language and an 'e' when referring to the people and the culture.

English phonetically but is represented similarly to English in terms of orthography. For instance, in the Wick-Miller orthography, the word for daughter is spelled *paite"*, whereas it is spelled *baide'* in the ISU orthography. In these examples, it is unclear whether the initial stop is meant to be voiced, voiceless unaspirated, or voiceless aspirated. The intention is that these two spellings represent the same word phonetically; however, the pronunciation based on these spellings is not necessarily clear in relation to English spelling and pronunciation.

The Shoshone-Bannock Tribes have a high level of interaction with western culture and have had for quite some time. According to Drusilla Gould (personal communication, May 12, 2014), this is primarily for two reasons: 1) the reservation closely borders the cities of American Falls, Blackfoot, and Pocatello, and the native and non-native communities are highly integrated. Tribal members often hold jobs off the reservation, relying on the social infrastructure off the reservation. 2) Many children within the tribes attend western schools off the reservation. This means that at best, children from the Shoshone-Bannock community are bilingual, and at worst, Shoshone children only speak English.

Historically, much of the integration between the native and non-native, both culturally and linguistically, stems from the American Indian boarding schools, which began in the 1800s. Boarding schools have played a major role in the integration of Native American cultures into Anglo-American society throughout the United States, and Shoshoni is no different. Boarding schools produced a variety of attitudes both toward Shoshone, as well as American culture. Gould (personal communication, May 12, 2014) explains that some Native Americans believe that boarding schools had a positive impact because they taught the native people good organization, accountability, and hygiene, as well as giving native people new perspectives and prepared them for the future. Furthermore, boarding schools increased appreciation for native life and their language (personal communication, Drusilla Gould, May 12, 2014). However, boarding schools also produced negative attitudes among tribal members towards both the language and the culture. Boarding schools also damaged the trust between the Native American and western communities, which has diminished the willingness of community members to work with linguists or other outsiders. Additionally, boarding schools made many tribal members feel ashamed or embarrassed of their culture and language, making them reluctant to pass the language on to their children. Moreover, while in boarding schools, many native speakers forgot their language. The impacts of boarding schools have led to a great deal of loss of language and culture. Although there are both positive and negative attitudes associated with boarding schools, the effect of integration between these two cultures is undeniable.

Negative attitudes towards the Shoshoni language and culture also stem from economic considerations. There are a greater number and variety of jobs for which it is necessary to speak English. Children are also more likely to be successful in school if they embrace English and western school, as many Anglo-American expectations clash with Shoshone culture. For example, it is taboo in the Shoshone culture to talk about oneself. Yet, in western culture, it is expected that children will share their accomplishments. Furthermore, a successful job interview is usually dependent on these same skills. Because of this economic dynamic, many tribal

members feel that there is a greater chance for economic success if they embrace western culture (personal communication, Drusilla Gould, May 12, 2014).

The historical and social context of Shoshoni is relevant because it means that all speakers of Shoshoni are bilingual. Some of these speakers may be L1 Shoshoni and L2 English speakers, and others may be L1 English and L2 Shoshoni speakers. Furthermore, among these speakers there is a wide variety of fluency ranging from highly fluent speakers to those who do not speak the language at all. These factors must be taken into consideration when doing linguistic research on the language as the influence of English will likely have affected the language at all levels. Furthermore, the design of any study should be sensitive to the needs and wishes of the community in which the study takes place.

3. Background
3.1. Lisker and Abramson (1964)

Lisker and Abramson (1964) discuss the properties of voicing and aspiration as ways of classifying the world's stops. From an acoustic perspective, voiced stops have periodicity that corresponds with the stop closure. Aspiration is acoustically characterized by a relatively long burst of aperiodic activity following the release of the closure that extends to the onset of vowel voicing. This duration of aperiodic energy is either absent or substantially briefer in an unaspirated stop. In other words, aspirated stops have a longer duration between the release of the closure and the onset of subsequent vowel voicing than unaspirated stops do. This variation in duration before the onset of vowel voicing is call Voice Onset Time (VOT). Aspirated stops are said to have long-lag VOT, while unaspirated stops are said to have short-lag VOT. These durations are quantified by marking the release of the closure as the VOT-on point, or the zero-point, and the onset of voicing as the VOT-off point. However, in prevoiced stops, where voicing begins prior to the release of stop closure, the release of the closure still acts as the zero value, but the onset of prevoicing is marked for the VOT-off point, meaning that voiced stops have a negative VOT. There are two different labels, fortis and lenis, used to categorize these three stops depending on the stop contrast in a language. For instance, in a language that contrasts voiceless aspirated and unaspirated stops, the aspirated stops are referred to as fortis and the unaspirated stops are referred to as lenis. On the other hand, in a language that contrasts voiced and voiceless stops, the voiceless stops are referred to as fortis and the voiced stops are referred to as lenis. Together, these voicing and aspiration features classify the following types of stops: voiceless unaspirated, voiceless aspirated, voiced unaspirated, and voiced aspirated—more commonly referred to as breathy voiced or murmured. Using these categorizations, Lisker and Abramson analyzed eleven languages that fell into three sets: two-category languages (two-way contrast), three-category languages (three-way contrast), and four-category languages (four-way contrast).

Languages with a two-way stop contrast generally have either a voicing contrast, or an aspiration contrast. Puerto Rican Spanish, Hungarian, and Tamil all contrast voiced and voiceless stops. Cantonese and English contrast voiceless aspirated and voiceless unaspirated stops. English is often a subject of debate in terms of how its contrast is classified, as some speakers of English do have some prevoicing in the production of /b, d, g/. However, Lisker and Abramson

found that speakers of English tend toward voiceless unaspirated stops for their production of /b, d, g/, and therefore, they argue that English patterns with Cantonese and has a contrast between short-lag and long-lag stops.

In addition to languages with a two-way contrast, a variety of languages also have three- or four-way stop contrasts. Languages with a three-way stop contrast generally have the categories voiced, voiceless unaspirated, and voiceless aspirated. This pattern can be seen in Thai and Eastern Armenian. Languages with a four-way stop contrast contain all four stop types, utilizing both the voicing and the aspiration contrasts. Marathi and Hindi are both languages with a four-way stop contrast.

Together, these languages paint a reasonably clear picture for the average VOT values for various types of stops. Voiced stops generally had an average of about -100ms with a range of approximately -125ms to -75ms. Voiceless unaspirated stops had an approximate VOT of 10ms—ranging from 1ms to 25ms. Voiceless aspirated stops had an average VOT of about 75ms with a range of approximately 60ms to 100ms (Lisker and Abramson 1964, 407). This was generally true of all languages regardless of how many stop contrasts they had. Despite the overall consistency of the generalizations, there was a fair amount of variation in VOT, much of which seemed to be sensitive to place of articulation.

Overall, Lisker and Abramson found that VOTs for velar stops differed from the averages of the other places of articulation. In the case of the negative VOT, or prevoiced stops, the velars had a lower overall average—meaning that there was less prevoicing in velars than in other places of articulation. For both of the positive VOT categories, the velar stops had significantly longer VOTs than the other place averages and the overall average. In the case of the short-lag VOT stops, the velar value was more than twice as high as the bilabial and alveolar stop VOTs. Based on the average VOTs reported by Lisker and Abramson, Table 1 provides an overview of the VOT values for various languages. These values show that VOT is sensitive to place of articulation, particularly velar stops. This was true of all of the languages regardless of type of stop or type of stop contrast within the language. Therefore, this feature appears to be cross-linguistic.

:

Table 1: VOT averages (in ms) based on data reported in Lisker and Abramson (1964, 392–396)

		Bilabial	Alveolar	Velar
Dutch	Negative	-85	-80	
	Short	10	15	25
Spanish	Negative	-138	-110	-108
	Short	4	9	29
Hungarian	Negative	-90	-87	-58
	Short	2	16	29
Tamil	Negative	-74	-78	-62
	Short	12	8	24
Cantonese	Short	9	14	34
	Long	77	75	87
English	Short/Neg	1/-101	5/-102	21/-88
	Long	58	70	80
Eastern Armenian	Negative	-96	-102	-115
	Short	3	15	30
	Long	78	59	98
Thai	Negative	-97	-78	
	Short	6	9	25
	Long	64	65	100

3.2. Cho and Ladefoged (1999)

Cho and Ladefoged (1999) explain that the variations in VOT based on place of articulation are evident throughout languages and can generally be explained by universally applicable phonetic rules based on physiological and aerodynamic characteristics. Cross-linguistically, velars are known to have significantly longer VOTs than other places of articulation, such as bilabials and alveolars. There are also differences between VOTs for bilabials and alveolars. Cho and Ladefoged provide six principles which contribute to the explanation for these differences.

The first and second principles relate to general laws of aerodynamics where the size of the cavity in front of and behind the closure affects the duration of VOT. The first principle is in regards to the volume of air behind the closure. Because there is a smaller closure behind velar stops, there is an increase in the amount of air pressure. When the closure is released, it takes longer for the pressure behind the closure to fall before the transglottal pressure necessary for voicing may be established—resulting in a longer VOT. The second principle relates to the air in front of the closure. In a velar stop, there is a larger cavity in front of the closure than there

would be in an alveolar or a bilabial stop. Therefore, there is a greater mass of air in front of the velar closure. This mass of air will take longer to fall than in a bilabial or alveolar stop before voicing can begin, once again resulting in a longer VOT (Cho and Ladefoged 1999, 209–211,213).

The air pressure in front of and behind the closure also affects the change in the glottal opening area for aspirated stops. In voiceless stop production, the glottis opens before the release of the closure. Because the pressure in velar stops decreases more slowly, it also takes longer for the glottis to produce the necessary constriction for vocal-fold vibration. This slower time is the result of the outward displacement of the glottal folds. This is also a contributing factor for longer VOTs in velar stops (Cho and Ladefoged 1999, 211–213).

The fourth principle relates to the necessary movements of the articulators. The articulatory movement of the tongue tip is the fastest and the movement of the tongue body is the slowest, with the movement of the lower lip falling somewhere in between. Furthermore, the movement of these articulators is affected by the movement of the jaw. The movement of the jaw is able to accelerate the movement of the lower lip, whereas jaw movement has little effect on the tongue dorsum. Given that faster articulatory velocities result in a more rapid drop in pressure behind the closure, there is a shorter time before voicing can begin for alveolar and bilabial stops (Cho and Ladefoged 1999, 210–211,213).

The fifth principle also relates to the articulators, but in regards to the extent of contact of the articulators. The longer the contact between articulators, the longer the VOT. This is because the articulators are sucked together with a greater force due to the Bernoulli Effect, making the articulators move apart more slowly. Velars have a relatively greater extent of contact area meaning that velars take a greater amount of time to articulate, resulting in a greater amount of time before the onset of voicing (Cho and Ladefoged 1999, 211, 213).

The sixth principle deals with closure duration. Bilabials are generally thought to have a longer closure than other stops. Furthermore, there is a set duration of vocal-fold opening for aspiration. This results in an inverse relationship between closure time and length of aspiration—meaning that the longer the closure, the shorter the VOT (Cho and Ladefoged 1999, 212–213).

Cho and Ladefoged claim that together, these six principles contribute to the differences in VOT cross-linguistically. In addition to their analysis of why there are cross-linguistic differences in VOT based on place of articulation, the authors also survey 18 languages to demonstrate language-specific differences in VOT as a function of place of articulation. Overall, Cho and Ladefoged (1999, 219) report that the average difference in VOT between coronal and velar unaspirated stops is 18.9ms and the value for the corresponding places of articulation for aspirated stops is 16.7ms The specific acoustic and articulatory mechanisms that trigger this difference may vary; however, the authors attribute this difference in velar VOT values, and overall place of articulation sensitivity, to the universal articulatory principles discussed here.

In addition to place of articulation differences, Cho and Ladefoged (1999, 223) discuss cross-linguistic differences in aspiration. When looking simply at the velar stops, the authors

propose four categories of stops based on the VOT values of the languages. The first is around approximately 30ms for unaspirated stops. The second is around 50ms for slightly aspirated stops, the third is around 90ms for aspirated stops, and a fourth category for highly aspirated stops, which tend to fall between 130ms and 160ms. As predicted by the above principles, the bilabial and alveolar stops have VOT durations somewhere below these values, depending on their presence and amount of aspiration. The reasoning behind dividing up the aspiration types this way is that Cho and Ladefoged (1999) propose a feature VOT with values voiced, voiceless unaspirated and aspirated, rather than the features voiced and aspirated features assumed by many, and suggest that different languages adopt different phonetic realizations of those modal values. Given the data provided by Cho and Ladefoged (1999), as well as the data provided by Lisker and Abramson (1964), we combine this data with hypotheses for speech learning models to predict what values we might expect for stops in Shoshoni.

3.3. Speech Learning Model (MacKay et al. 2001; Flege 1995)

Flege (1995) and MacKay et al. (2001) discuss second language acquisition and the Speech Learning Model (SLM). Flege (1995, 239) states that the SLM has four main postulates: First, L2 learning can take place by the same mechanisms and processes used in learning the L1. Second, phonetic categories are created through the storage of language-specific aspects of speech stored in the long-term memory. Third, phonetic categories established during childhood may evolve over the life span. Fourth, bilinguals attempt to maintain contrasts between L1 and L2 phonetic categories. Given these postulates, Flege hypothesizes that new phonetic categories will be formed for the L2 if the phonetic categories in the L1 and L2 are distinct enough that the speaker is able to discern the difference between the two phones. Speakers are more likely to discern a difference between these two phonetic categories if there are greater differences between the two categories. If speakers are unable to discern the difference between the two categories, the speaker will form a single category for the sound in the L1 and the sound in the L2. These two sounds may influence one another and over time the sounds will come to resemble one another (Flege 1995). Additionally, the age at which bilinguals acquire their second language affects the ability of speakers to form two distinct categories, or a single category shared by the L1 and L2. Speakers who learn their L2 at an early age are more likely to produce sounds similar to native speakers of the L2, meaning that distinct L1 and L2 categories have been created, whereas late learners are more likely to form a single category to be shared between the L1 and the L2 (MacKay et al. 2001; Flege 1995).

For stops, the categories at hand are those established by voicing and aspiration features. For instance, in languages with a two-way voicing contrast, /p, t, k/ are realized as voiceless unaspirated. When L1 speakers of a two-way voicing contrast language (such as Italian or French) learn a language with a two-way aspiration contrast (such as English), in which /p, t, k/ are realized as voiceless aspirated, they tend to produce /p, t, k/ with aspiration, but the VOT is too short for native-English-like production. This means that a separate category formation for voiceless aspirated stops in English is likely blocked due to the similarities between the stops in the L1 and the L2. Furthermore, for these speakers, the production of /p, t, k/ in the L1 begins to

resemble the production of /p, t, k/ in the L2, so that, for example, VOTs for voiceless unaspirated stops in L1 become somewhat longer than those of monolingual speakers of the L1.

Mackay et al. (2001) looked at the production of stops in L1 Italian/L2 English speakers for both early and late learners. As predicted by the SLM, speakers who learned English at an earlier age produced stops more similarly to native English speakers. Late learners, on the other hand, were more likely to produce /p, t, k/ in English with a shorter VOT than native English speakers. Furthermore, late learners were more likely to confuse English /b, d, g/ with /p, t, k/ in perception tasks. Furthermore, monolingual Italian speakers produced prevoicing without a pause between voicing and the release burst. English speakers who had prevoicing for /b, d. g/ have a pause between the voicing and the release burst. Italian-English bilinguals still tended to pre-voice /b, d, g/; however, unlike monolingual Italian speakers, the bilinguals were more likely to have a pause between the voicing and the release burst, similar to that of native English speakers. This evidence first demonstrates that the L1 may influence both the production and the perception of the L2, especially where speakers acquire the L2 later on. It further demonstrates that the L1 and L2 influence one another as speakers have more contact with the L2—meaning that the L1 is not static. This model of speech learning, along with the data provided by both Cho and Ladefoged (1999) and Lisker and Abramson (1964), we can use what is already known about Shoshoni phonetics to make predictions for the acoustics of Shoshoni.

3.4. Phonetic Background on Shoshoni

The Shoshoni language is said to have a single series of voiceless unaspirated stops consisting of [p], [t], [k] (Gould and Loether 2002). These are represented in the ISU (Idaho State University) orthography as *b, d, g,* respectively. These stops undergo several distinct phonological processes where they may become [p:], [t:], or [k:], [β], [ɾ], or [ɣ], and [φ], [θ], or [x], depending on the environment in which they appear (personal communication, Christopher Loether, April 8, 2015).

Early transcripts and dictionaries of Shoshoni were often inaccurate because linguists were biased by their L1—generally English—backgrounds (personal communication, Chris Loether, April 3, 2014). Gould explains that most of these records of Shoshoni are unusable because they conflate some segments or completely omit others, making the words unrecognizable to native speakers (personal communication, Drusilla Gould, May 12, 2014). However, more recent work done on several dialects of Shoshoni, including the work started by Wick Miller on the Western Dialect of Shoshoni and the work currently being done by Drusilla Gould and Christopher Loether at Idaho State University, has greatly improved the quality and quantity of materials available to the Shoshone community and the linguistic information regarding the language.

However, there are many elements of Shoshoni phonetic research that may be improved through updated or new information. For instance, Shoshoni textbooks aimed at English speakers teach learners that there is a contrast between initial stops in Shoshoni and initial stops

in English. For example, the pronunciation of the Shoshoni *b/p²* is likened to English in the pronunciation of the *p* in *spy* (Gould and Loether 2002, 13). However, based on the data provided by Lisker and Abramson (1964), the pronunciation of the *p* in *spy* would be the same for most speakers of English as the *b* in *buy,* and therefore it would be most straightforward to tell learners that the Shoshoni *b/p* and the English *b* are pronounced the same. While this Shoshoni Grammar by Loether and Gould makes a distinction between Shoshoni stops and English lenis stops, Loether also explains that the decision to use *b, d, g* in the ISU orthography, rather than *p, t, k* like the Wick Miller orthography, was because he and Gould found that L2 Shoshoni learners were more likely produce the correct sound for Shoshoni stops when the orthography used *b, d, g*. This intuition seems to be based on the same conclusions drawn in Lisker and Abramson, that English lenis stops are in fact voiceless unaspirated stops. However, based on the mismatch between the orthography and the pronunciation explanation in the Shoshoni grammar, this intuition does not seem to be fully realized. Aligning Shoshoni grammars with the intuition that Shoshoni *p, t, k* are most like English *b, d, g* could improve Shoshoni language learning materials.

Given these examples of ways Shoshoni teaching materials can be improved, this study aims to provide information about the acoustics of Shoshoni that may prove useful both in teaching the language and for gaining phonological insight into the language. Additionally, because Shoshoni is a language with a single series of stops and there is little data about single-stop-series languages, this study also seeks to provide insight into the acoustic properties of languages with a single series of stops.

4. Assumptions and Predictions

While Lisker and Abramson (1964) go into detail about what is seen cross-linguistically for two-way, three-way and four-way stop contrasts, they do not give any examples of what we might expect to see in a single-stop-series language. We do see some data on single-stop-series languages in Cho and Ladefoged (1999); unfortunately, they do not necessarily discuss what average values we see across single-stop-series languages, but instead focus on the cross-linguistic differences in VOT by place of articulation. That being said, we get the following values for Wari' and Eastern and Western Aleut:

Table 2: VOT means in ms of Wari' and Eastern and Western Aleut as listed in
Cho and Ladefoged (1999: 219)

	Bilabial	Dental	Alveolar	Velar	Uvular
Wari'	19	26		50-58	
Aleut (Western)			76	95	92
Aleut (Eastern)			59	75	78

These data do provide some information on what we might expect to see in a single-stop-series language; however, they do not paint a very clear picture, predominantly because of the

---

² The first letter in the sequence refers to the ISU Orthography and the second letter in the sequence refers to the Wick Miller Orthography.

significant differences between means for Wari' and the Aleut languages. Single-stop-series languages are generally thought to have voiceless unaspirated stops. This is due to the relative markedness of aspirated and voiced stops. Generally, it is assumed that in order for the marked structure to be present, the unmarked structure must also be present (Jakobson 1968). Therefore, in single-stop-series languages, it is assumed that the stops are voiceless unaspirated; however, this does not seem to be the case in Wari' and the dialects of Aleut. Cho and Ladefoged (1999) go on to categorize these languages as slightly aspirated and aspirated stops, respectively, though these categorizations are not used to contrast between a different category of stops within the languages. Overall, these values don't really serve a predictive function for what we might expect in Shoshoni. However, they do provide additional information about possible characteristics of stops in single-stop-series languages. Depending on the how the values of VOT in Shoshoni stops cluster, we can better conclude whether the stops in Shoshoni are truly unaspirated, or whether they are fall into one of the aspiration categories discussed by Cho and Ladefoged (1999).

In languages with a stop contrast, average VOT values serve as ways to categorize stops. Some values may fall below or above that average, but all values of a certain category must fall within a certain range in order for the correct sound to be perceived. Examples of boundaries in various languages are demonstrated in the following figure:

Figure 1: Category boundaries from various languages (Figure from Brookes and Kempe (2014))
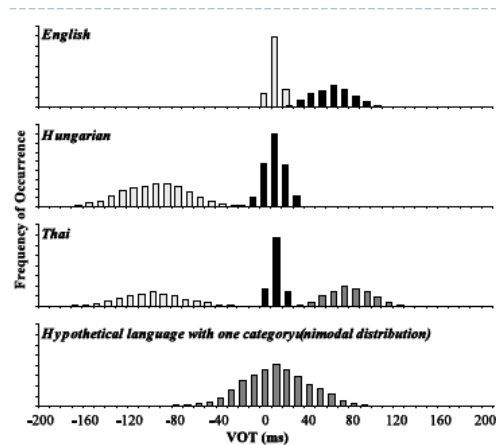
Figure 1: Distribution of VOTs in three languages based on measurements by Leigh Lisko and Arthur Abramson. Bars are colored to indicate different clusters in the input. The bottom panel shows a hypothetical language with only one voicing category.

As shown in the first three distributions in Figure 1 (English, Hungarian, and Thai), there is a divide, or boundary, between the types of stops. These allow the listener to distinguish between types of sounds. However, there is less research on languages with a single stop series. Therefore, the 4th tier demonstrates a hypothesis of how a language with a single stop series might look. Because there is no need to contrast between types of stops, Brookes and Kemp's (2014) hypothesis explains that there might be a broader range of VOTs, as there are no boundaries for the VOT values to push up against. Thus, we may expect to see a wide VOT distribution in Shoshoni.

While this explanation seems plausible, languages with a single stop series are largely under-documented. Therefore, the hypotheses about these languages are often nebulous or speculative and don't seem to be based on concrete data. Another hypothesis would be to assume that if the stops in these languages are truly unaspirated, then we might expect to see a similar value to other voiceless unaspirated stops cross-linguistically—especially since this value seems to hold across two-way, three-way, and four-way contrast languages. This assumption seems to be supported by the fact that Cho and Ladefoged (1999) do not categorize Wari' and the dialects of Aleut as unaspirated, despite the wide-spread assumption that single-stop-series languages have voiceless unaspirated stops—meaning the values in Wari' and the dialects of Aleut are higher than those reported by Lisker and Abramson for voiceless unaspirated stops because they are not languages with a series of voiceless unaspirated stops, but languages with various types of aspiration. Lisker and Abramson (1964) claim that the average VOT for voiceless unaspirated stops is approximately 10ms. Therefore, if Shoshoni truly has a series of voiceless unaspirated stops, we would expect to see values around 10ms. On the other hand, if Shoshoni falls into one of the aspiration categories described by Cho and Ladefoged, we might expect values similar to the values seen in Wari' or one of the Aleut dialects.

Another theory we could draw on to hypothesize potential VOT values for single-stop-series languages is the SLM, especially given the contact between Shoshoni and English and because there are no monolingual speakers of Shoshoni. Although the SLM is discussing speakers that have a two-way contrast in their L1 learning a language with another two-way stop contrast in their L2, there are several predictions that can be made about Shoshoni based on the SLM. These predictions depend both whether the speaker would be an early learner according to the SLM, or a late learner according the SLM. Furthermore, it depends on what we hypothesize happens, according to the SLM, when a speaker with a single-stop contrast learns a language with a two-way stop contrast.

First, if the speaker is considered a later learner of English, the SLM would predict that the speaker is likely to acquire a single category that encompasses both the single series of Shoshoni stops and one series of the English stops. Given that Shoshoni stops are hypothesized to be voiceless unaspirated, it is likely that the merged category would subsume the Shoshoni stops and the lenis stops of English. In this case, we would expect to see similarities between voiceless unaspirated stops in both languages. We would also expect the lenis category in English to be distinct from the English fortis stops. In Shoshoni, we would also expect to see a clustering of values for the Shoshoni stops that is distinct from the values seen for the English voiceless aspirated stops, despite the fact that there is no need for this boundary in Shoshoni. The boundary is not necessary as there is no stop contrast in Shoshoni; however, because the Shoshoni stops have mapped together with the English lenis stops, we would expect the necessary boundary between fortis and lenis stops in English to affect the distribution of stops in Shoshoni.

Another hypothesis is that if the speaker is considered an early learner according to the SLM, then we would expect the speaker to form separate phonetic categories for the two languages. Given this scenario, there are two separate hypotheses we could make about what we

would expect the English and Shoshoni acoustic data to look like. First, in English we would expect phonetic categories similar to a monolingual native English speaker for voiceless aspirated and unaspirated stops in English—meaning around 75ms for fortis stops and either 10ms or -100ms for lenis stops, depending on whether there is prevoicing or not (Lisker and Abramson 1964). In Shoshoni, the predictions would have a little more variability as we have limited information about Shoshoni and other single-stop-series languages; however, we wouldn't expect the English acoustic categories to have much influence on the Shoshoni categories, in which case we might predict something seen in either Lisker and Abramson (1964). where the VOT values cluster around 10ms, or the Brookes and Kemp (2014) model, where there is a wide distribution of VOT values.

A third hypothesis would be that between English and Shoshoni there are three distinct stop categories. For English lenis stops, we might expect prevoicing, or at least a set of positive short-lag values distinct from the set of values we see in Shoshoni. For English fortis stops, we would expect long-lag values that are distinct from the values we see in the Shoshoni stops, and in the middle we would see the category for Shoshoni stops. For this distribution, we might expect to see a model somewhat like the Brookes and Kemp (2014) Thai model, but where the figure represents the superimposition of the Shoshoni and English stop distributions, rather than a single language. One advantage to this hypothesis is that it could provide an explanation for why the Shoshoni grammars often distinguish between Shoshoni *b, d, g* and English *b, d, g*— because there is in fact a distinction between Shoshoni stops and English lenis stops.

Together, these hypotheses present a broad range of possibilities for what we might expect to see in Shoshoni; however, given the general lack of information about Shoshoni, specifically, and single-stop-series languages in general, this is as we might expect. Each hypothesis has its own advantages and disadvantages. It is also possible that some combination of these hypotheses will be borne out by the data. Therefore, this study aims to serve as a preliminary investigation into what single-stop-series languages look like, acoustically.

5. Methods
5.1. Participant

The participant in this study is a member of the Shoshone-Bannock tribe on the Fort Hall Reservation and is a bilingual speaker of English and Shoshoni. The participant's first language is Shoshoni; however, she is somewhat more fluent in English, based on the fact that opportunities and environments to speak Shoshoni are limited. This being said, the speaker is one of the elders in her community and has retained a high level of proficiency in the language by actively participating in the community, which includes teaching Shoshoni to her family members and speaking with them and staying actively involved in the documentation and preservation of Shoshoni. As a young child, the participant spoke Shoshoni at home with her family members; however, she attended an English-speaking elementary school where she learned English at an early age.

5.2. Procedure

The participant completed two production tasks that consisted of one word list in English and another in Shoshoni. These tasks were administered on the same day by the same administrator. The administrator instructed the participant in English to read each item on the list in order, repeating each item twice. The participant was also instructed to read the words at a natural pace. The recordings were done in a sound booth. The participant read the English list first and then the Shoshoni.

5.3. Materials

The English and Shoshoni word lists each consisted of 50 tokens, 30 of which were target items and 20 of which were distractors. In both languages the 30 target words consisted of stops followed by vowels or semi-vowels. For the English words, both the target items and the distracters were monosyllabic. However, there are fewer monosyllabic words in Shoshoni; therefore, there was variation in syllable number. The distractor words were immediately discarded and the rest were separated into individual sound files and analyzed in Praat (Boersma and Weenink 2015). VOT-on was marked at the beginning of the burst and VOT-off was marked at the appearance of clearly defined F2 structure in the following vowel in tokens without prevoicing and at the onset of periodicity corresponding with the initial closure for tokens with prevoicing (Jessen 1998).

A Praat script was then used to extract the VOT durations. Each word was marked with its place of articulation—bilabial, alveolar, or velar—so the data could be analyzed given these distinctions

6. Results
6.1. English

In order to further refine our hypotheses, especially in regards to hypotheses for the SLM, we will first turn our attention to the results for English. The average VOT for lenis stops was 4ms (rounding to the nearest ms), with a range of -120ms to 72ms. This is a fairly substantial range that was affected predominantly by differences in place of articulation. In other words, the average VOT for a lenis bilabial stop was 3ms, with a range of -83ms to 14ms. The average for a lenis alveolar stop was -34ms with a range of -120ms to 13ms, and the average for a lenis velar stop was 43 ms with a range of 31ms to 72ms. There were also differences between values based on place of articulation in the fortis stops, which had an overall average of 93 ms and a range of 27ms to 169ms. The differences among the places of articulation are demonstrated in the following figure:

Figure 2: English VOT means in ms
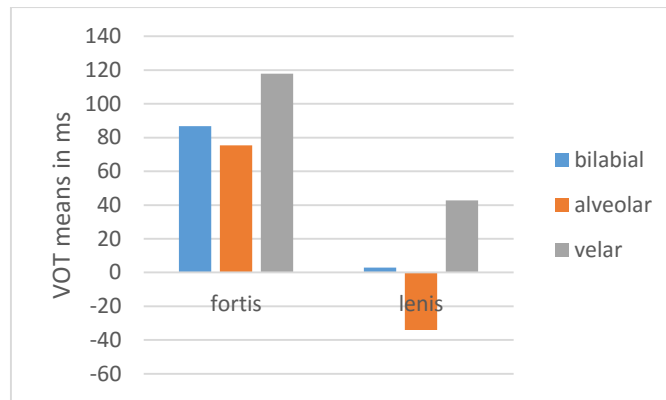
This figure demonstrates the significant differences in VOT averages based on place of articulation, particularly for the lenis stops. These differences distort the overall VOT average. Furthermore, even the averages based on place of articulation are somewhat misleading as there is a fair amount of variation in VOT values within a single place of articulation, especially for the lenis bilabial and lenis alveolar stops where there are both negative and short-lag stops. Distributions by place of articulation for lenis stops are shown in the following figure:

Figure 3: Distribution of English VOTs for lenis stops (in ms)
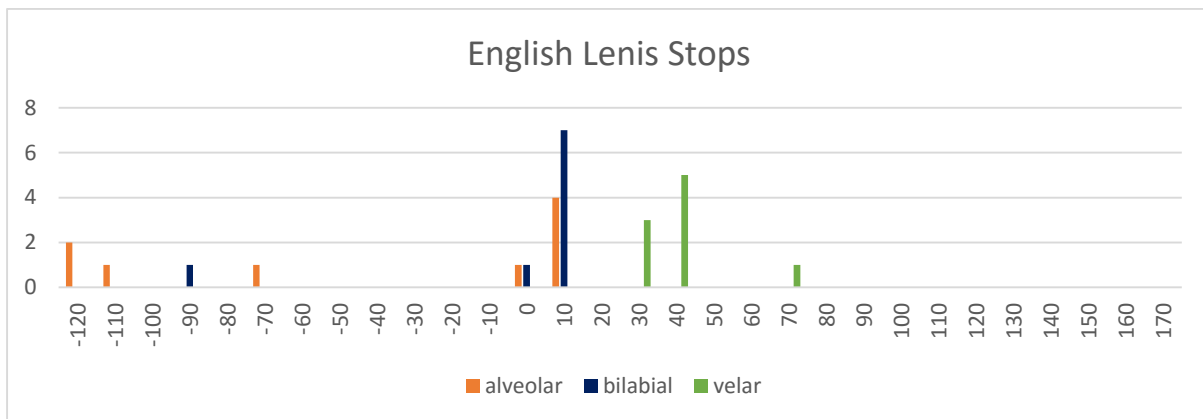
This figure demonstrates both the extreme range of these stops as a collective whole as well as some of the extremes by place of articulation. While the fortis stops have somewhat less variation than the lenis stops, there is still a substantial amount of variation, particularly between the velar stops and the bilabial and alveolar categories. This distribution is demonstrated by the following table:

Figure 4: Distribution of English VOTs for fortis stops (in ms)
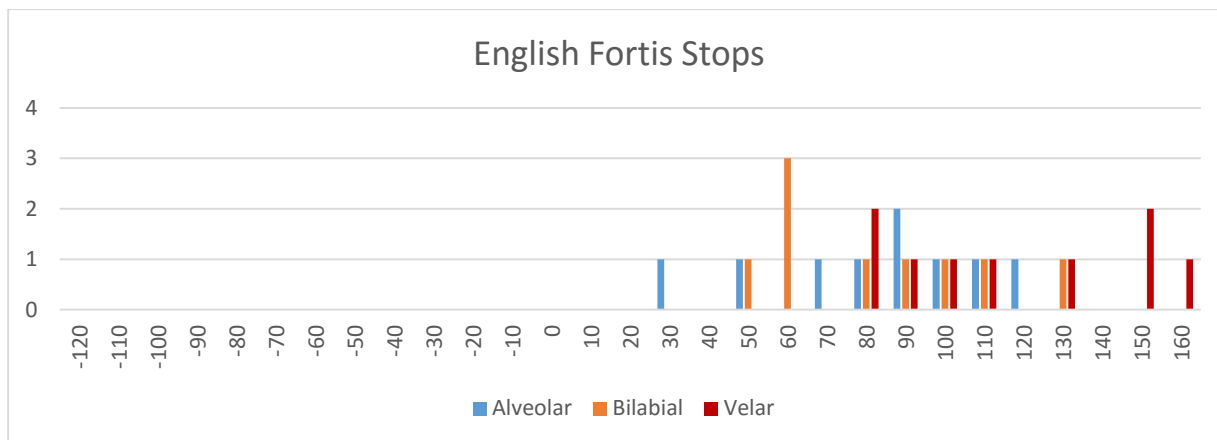
This table, in particular, shows that the distribution of velar stops tends to cluster above the distribution of the velar and alveolar categories. However, there is some overlap between all three places of articulation.

The differences in values by place of articulation are partially explained by the articulatory constraints discussed by Cho and Ladefoged (1999). However, there are also a few complications in these data. The first is in regards to velar stops. In these data, velar stops had especially high VOTs. This is likely due, in part, to the articulatory constraints that tend to affect the duration of velar VOTs universally. However, there are aspects of the velars produced by this participant that are unique and will be discussed further later on in the paper. As previously mentioned, the average VOT for a lenis velar stop was 43 ms—well above what we saw for other categories. Because this average was substantially higher than other VOT averages, these data contributed to a distortion of the overall mean VOT by shifting the value upward.

On the opposite end of the spectrum, the overall average was affected by the negative VOT in prevoiced lenis stops There was not consistent voicing across lenis stops in English; however, there were some instances of prevoicing. Furthermore, the prevoicing that did occur in the data was robust—much more robust than what we often see in the production of prevoicing by monolingual native English speakers. An example of the participant's production of prevoicing is shown in Figure 5:

Figure 5: Full spectrogram and waveform of *boot* in English (-83ms)
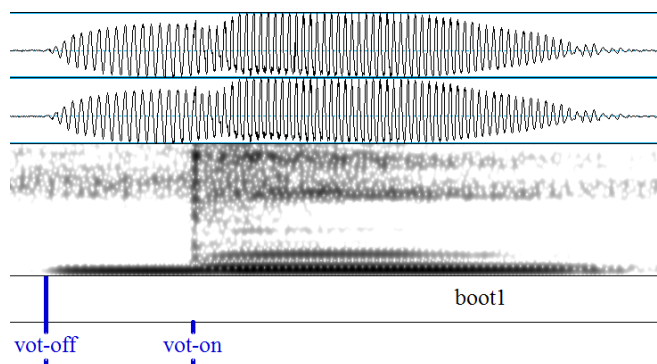
Other examples that are also prevoiced show a similar robustness. However, not all lenis stops in English were produced with prevoicing. In fact, the majority of stops were not prevoiced— only 8.3%. This means that these prevoicing values are pulling down our overall average VOT. This is particularly true of the alveolar category, as most of the prevoiced stops were alveolar. It may be significant to note that often when native English speakers do have prevoicing for lenis stops, the voicing is often intermittent or less robust, and thus this prevoicing does seem potentially relevant. However, given the low ratio of voiceless stops, it seems unlikely that voicing is being used to contrast English stops with the stops in Shoshoni, meaning that a model like the Thai distribution in Figure 1 is unlikely. However, more instances of prevoicing may appear given a larger data set. The overall distribution of VOTs for this speaker's production of English are shown in the following figure:

Figure 6: Histogram of VOT values in English stops
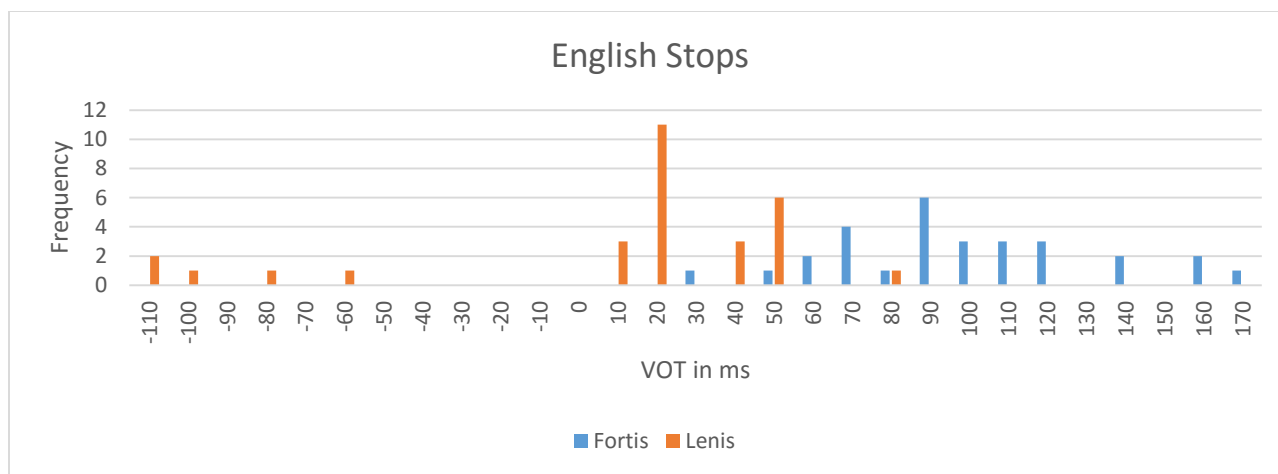
This distribution shows a rather similar distribution to the histogram given in Lisker and Abramson (1964). The mode is a little higher, at approximately 20ms; however, this mode is still well within the range described by Lisker and Abramson (1964), where unaspirated stops ranged from 1 to 25 ms. Additionally, one way to address the wide range presented by the voiced outliers is to separate the averages by voiceless unaspirated and prevoiced stops. In this case, the voiceless-unaspirated-stop average for alveolars was 11 ms and the bilabials was 12 ms. The prevoiced average was -96 ms for alveolar stops and the one prevoiced bilabial was -83 ms. These averages are then strikingly similar to the averages presented by Lisker and Abramson, where the averages were -100 ms, 10 ms. In terms of the fortis stops, the mode for the fortis stops also seems to be shifted somewhat up the scale, with the mode falling at approximately 90ms. Furthermore, we can see from the following figure based on the Lisker and Abramson (1964) data that the overall distribution is quite similar:

Figure 7: Distribution of English VOT based on Lisker and Abramson data—figure from Brookes and Kemp (2014)
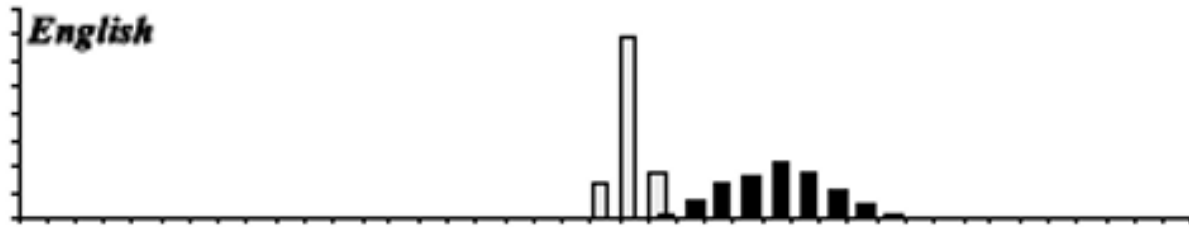
This shows that the distribution of English stops reported by Lisker and Abramson (1964) is similar to the distribution of English stops produced by this speaker—there is more variability in the fortis stop distribution than the lenis stop distribution. Given the similarities both in the mean VOT values and the overall distribution, it would seem that the participant's production of English is largely similar to that of native English speakers. To make further claims about how similar these data are, we would need both more data from this participant, as well as data from monolingual native English speakers in the region to compare values—particularly because VOT values seem to have a lot of variability based on place of articulation and because English VOT tends to vary based on geographical region (Hunnicutt and Morris, n.d.).

6.2. Shoshoni Results

Given these results, we can now turn our attention to the Shoshoni data. When looking at the mode and average for this speaker's production of the Shoshoni stop series, the data appear, in many ways, to be consistent with the English data. The average VOT across all places of articulation, rounded to the nearest ms, was 30ms. Furthermore, when breaking down the data by place of articulation, the values are once again strikingly similar to this speaker's production of English. The average VOT for the bilabial stops was 18ms and the average VOT for the alveolar stops was 17 ms. Though these figures are somewhat higher than the values for voiceless unaspirated stops seen in the Lisker and Abramson (1964) data, they are still lower than what we would expect to see in an aspirated stop series (which were reported at approximately 75ms). Therefore, we might just assume that these fall into the category of voiceless unaspirated stops. However, these values are also fairly similar to the values for Wari' provided by Cho and Ladefoged (1999), which lists an average of 19 ms for bilabial stops and 26 ms for alveolar stops. This may mean that Shoshoni could also be classified as a slightly aspirating language, though this fine-grained distinction may not be necessary for our purposes. The following examples demonstrate typical initial bilabial and alveolar stops in Shoshoni:

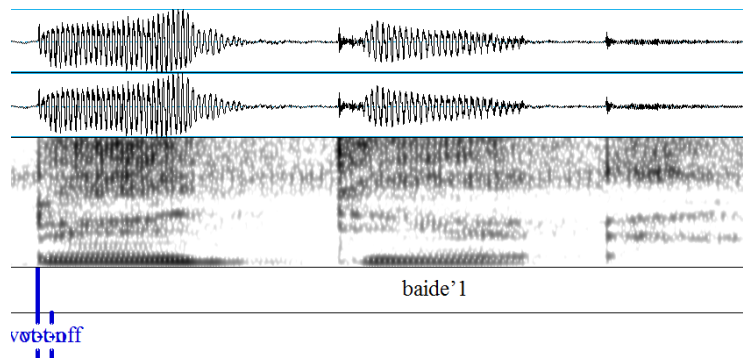Figure 8: Full spectrogram and waveform for the first utterance of *baide'*.

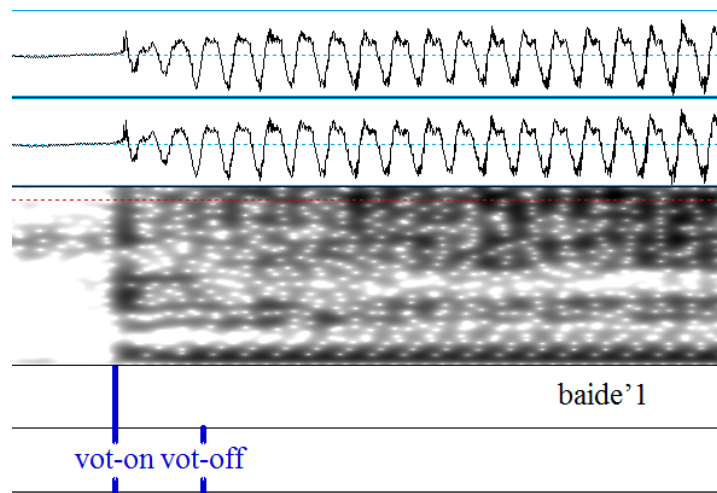Figure 9: Close view, VOT measurement for the initial [p] of *baide'* (11.79 ms)

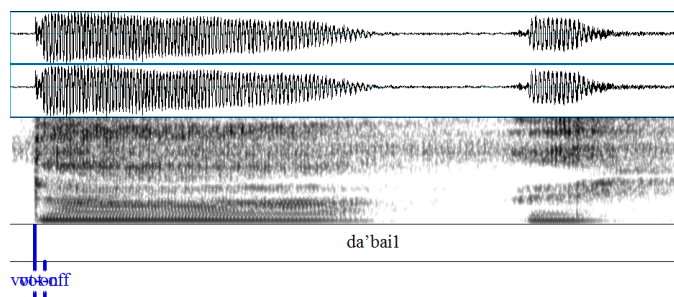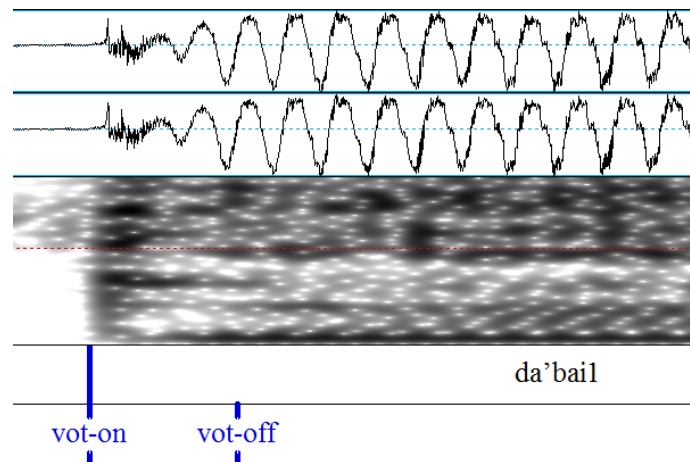Figure 10: Spectrogram and waveform of *da'bai.*

Figure 11: Close view, VOT measurement for the initial [t] of *da'bai* (13.94ms)

These figures represent what we might expect of a typical unaspirated stop, where there is a short release burst followed by the immediate onset of voicing—all occurring within somewhere between 11ms and 14ms in these cases. However, further analysis of the data reveals a somewhat more complicated picture.

For alveolar and bilabial stops, the figures for *da'bai* and *baide'* tend to be representative of the stops for these places of articulation throughout the language. However, there were a few outliers with somewhat longer VOT values. For instance, the following shows the utterance *dei'ape'* which has a VOT of approximately 31 ms:

Figure 12: Full spectrogram and waveform of *dei'ape'*
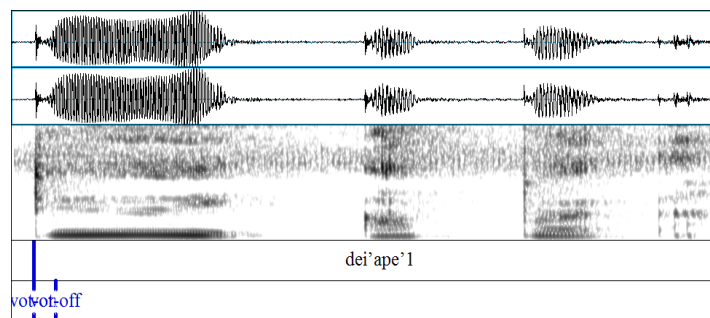
The 31 ms VOT is an outlier for the alveolar category; furthermore, it is slightly higher than the range given by Lisker and Abramson (1964), who reported the range of unaspirated VOTs to fall between 1 and 25 ms. On the other hand, 31 ms is still seems closer to the values provided for the lenis stops than those reported for the fortis stops. Ultimately, this stop is an outlier and should not be used to be make judgements for the category as a whole.

Another outlier example, this time a bilabial stop, is the utterance *buih* which had an initial VOT of 50 ms:

Figure 13: Full spectrogram and waveform of *buih*
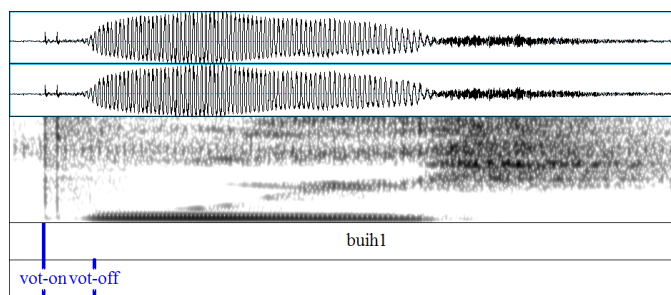
While this example does seem closer to what we would expect for an aspirated stop, it was an extreme outlier overall. The next closest VOT value for a bilabial or alveolar stop was 34ms. One interesting aspect of this particular stop is the instances of what seems to be periodicity following the release of the closure. This will be a relevant and more prevalent feature to discuss for the velar stops, but an isolated occurrence for alveolar and bilabial stops. This outlier is only of importance because of the limited data set. With more data, we may see the gap between the rest of the bilabials and the VOT of this utterance filled in, which would indicate that the Brookes and Kemp (2014) hypothesis, where a wide range of VOT values is predicted for a single-stops-series language, provides a correct prediction for Shoshoni. On the other hand, the stops may continue to cluster around the average we have already established, which would suggest that this utterance is indeed an outlier and potentially some sort of performance error. This means that for now, this outlier serves as a place holder for what could be the range of possible VOT values in Shoshoni. Overall, this segment provides us with something to keep an eye on for future research, but as it is predominantly an isolated occurrence in this data, it should simply be classified as an outlier.

Overall. the alveolar stops are the most prototypical for voiceless unaspirated stops, with values ranging from 13ms—*daiboo'*1 to 31ms—*dei'ape'*1. The range of VOT durations for the bilabial stops spans from 12 ms—*baide'*1 to 50 ms—*buih*1, where *buih*, as previously mentioned, is an outlier. The distribution of VOT values by place of articulation is demonstrated by the following figure:

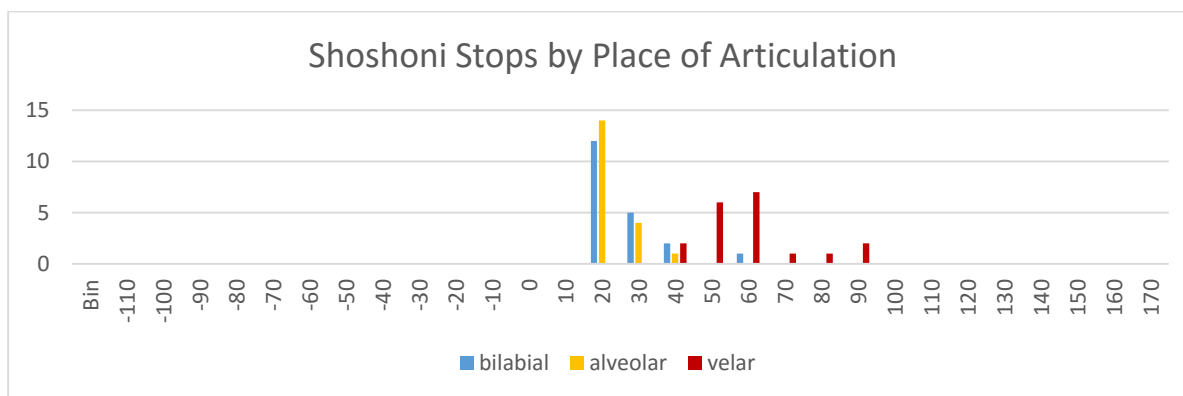Figure 14: Distribution of VOT in Shoshoni stops by place of articulation (in ms)

Figure 14 further demonstrates that the examples given above—*buih* and *dei'ape'*—are both outside the range of what is typical for their respective stop categories. It also shows that the majority of stops cluster around a particular value with very few outliers. Therefore, the data set as a whole does not seem to indicate that a broad range of values is possible for Shoshoni VOT.

The velars, on the other hand, provide a relatively different data set. As shown in both of the above figures, the velars tend to have a higher VOTs that either the alveolar or bilabial stops. The range for the velar stops spans from 37ms— *gaihaiwate* 2 to 86ms—*giiwetawa'*1, with the average falling at 60ms. This average is substantially higher than the other two places of articulation and pulls the overall average up. Without the velars, the average VOT is approximately 19ms—which is not only extremely similar to the mode VOT in the speaker's production of English, but just over 10ms less than the average including the velars. An overview of these averages can be seen in the table below:

Table 3: Average VOT durations (in ms)

bilabial	alveolar	velar	Grand Total
20.47	17.55	51.93	29.98

To some extent, based on the averages reported in both Lisker and Abramson (1964) and Cho and Ladefoged (1999), we expect there to be differences between the alveolar and bilabial stops and the velar stops. As previously mentioned, Cho and Ladefoged (1999) report an average difference of 18.9ms for unaspirated stops. However, the difference between the bilabial average and the velar average in this data is 32 ms, which is a somewhat more substantial difference than we might expect. Furthermore, the values we see in this category would make us suspect that the velar stops were, in fact, aspirated. On the other hand, the velar average in these data seems fairly similar to the velar average provided in Cho and Ladefoged for Wari', which was listed as 50-58ms. If we consider Shoshoni and Wari' to have a similar series of stops, the velar average is somewhat more expected. Overall, the contrast between the bilabial and alveolar categories and the velar category is further demonstrated by the following chart:

Figure 15: Average VOT durations in ms
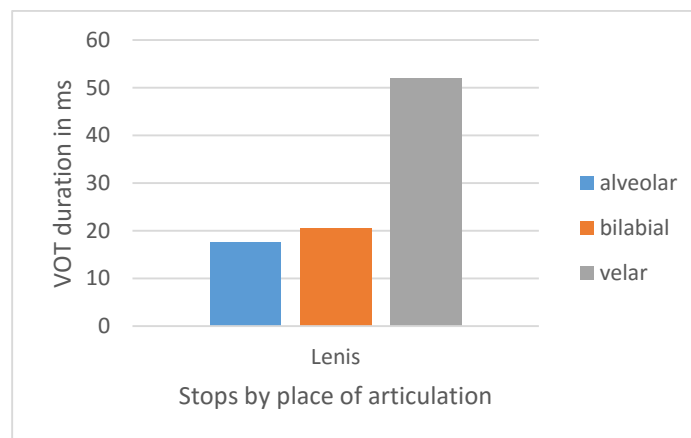
Because of the contrast between the velar category and the other two categories, and the lack of clarity on whether Shoshoni stops should be classified as unaspirated or slightly aspirated, it is necessary to examine the spectrograms and waveforms of these initial stops, particularly those in the velar category, to help determine how Shoshoni stops should be classified.
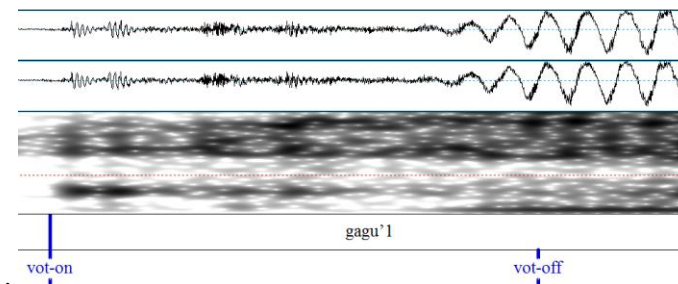
The following figure shows a waveform and spectrogram which is representative of what is seen for velar stops throughout the language:

Figure 16: Full spectrogram and waveform of *gagu'*
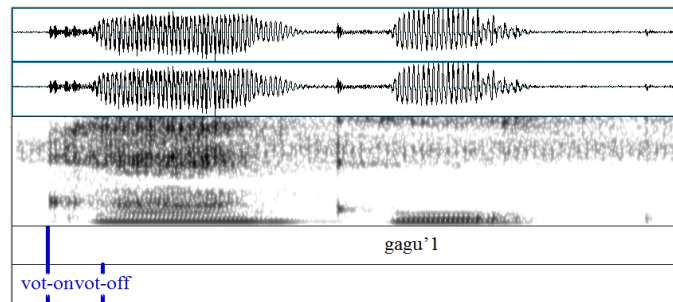
While it's clear that there is a long span of energy before the onset of voicing, it is less clear how this sound should be classified. In some ways, this release of this stop seems to have aperiodic energy, or aspiration. However, we can see from this spectrogram that there seem to be vertical striations that are characteristic of voicing or some other vocal fold movement. This apparent vocal fold movement is further shown from the periodicity shown in the close-view figure of *gagu'*:

Figure 17: Close spectrogram and waveform of *gagu'* (53 ms)

These same instances of apparent voicing or vocal fold movement occur throughout the velar stops in the data. In the following figures, we can see examples of waveforms and spectrograms from the upper values of the Shoshoni velar VOT duration range:

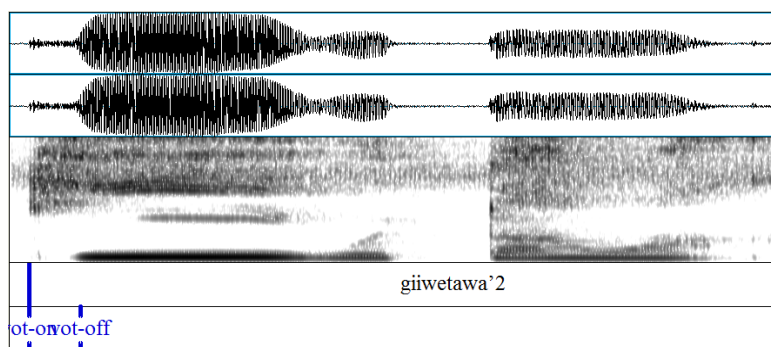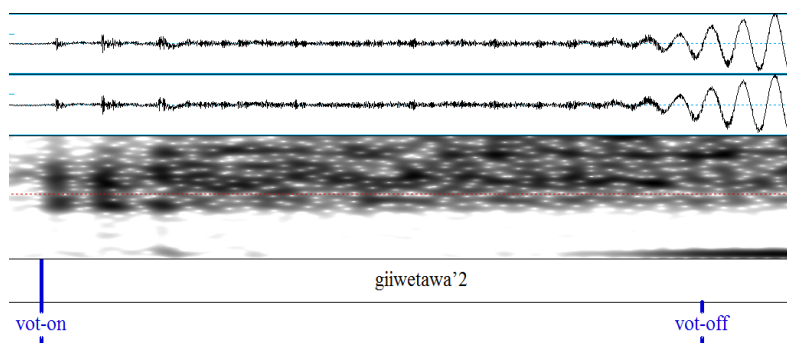Figure 18: Full spectrogram and waveform of *giiwetawa'*

Figure 19: Close view spectrogram and waveform of *giiwetawa'* (86.33ms)

These figures show similar patterns to the wave form and spectrogram of *gagu'*. There are both dark vertical striations in the spectrogram and segments of periodicity in the waveform that seem to set this type of release burst apart from the type of burst we see in voiceless unaspirated stops. The velar stops across Shoshoni exhibit similar patterns of periodicity. Furthermore, the velars have consistently longer VOT durations than the other two places of articulation. Therefore, it seems that the release noise and length of duration seem to be correlated with one another. This correlation is based both on the long VOT of velars as well as the outlier *buih*, which also had a long VOT duration as well as similar instances of what appears to be periodicity. Overall, the distribution of velar VOT values is demonstrated in the following table:

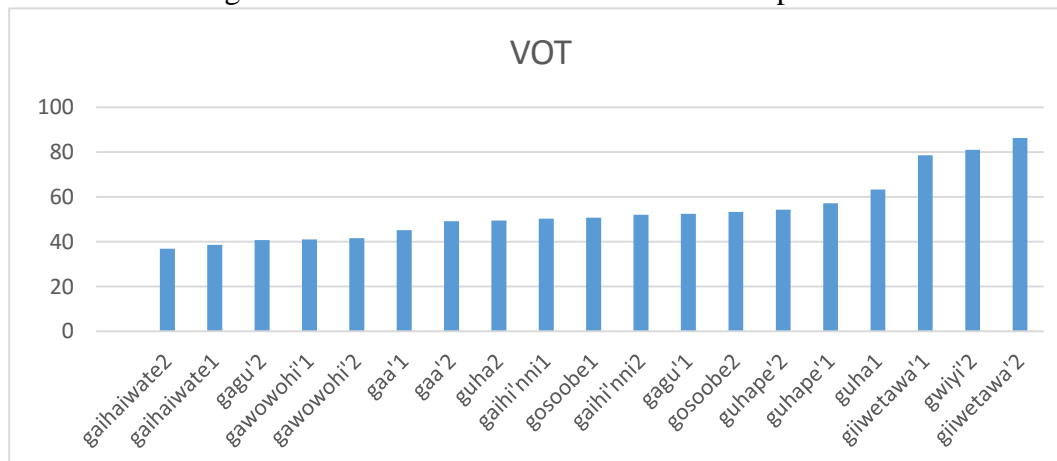Figure 20: VOT durations for initial velar stops in ms

Figure 20 brings up an important trend in the VOT of Shoshoni velar stops. There seems to be a vowel height effect on the VOT durations where high vowels are associated with longer VOT values on preceding stops. According to Esposito (2002), VOT durations increase when followed by a high vowel. This is most likely due to the stops having a slower release before a high vowel. Therefore, some of the especially long durations of velar VOTs in Shoshoni can be attributed to vowel height. However, the patterns of periodicity and vertical striations seen in the spectrogram seem to be a separate characteristic of velars in Shoshoni.

The discussion of velar characteristics in Shoshoni brings us back to the characteristics of velars in English. The patterns of vocal fold movement in Shoshoni velar stops also occur in the English velar stop data. This is shown in the following figure:

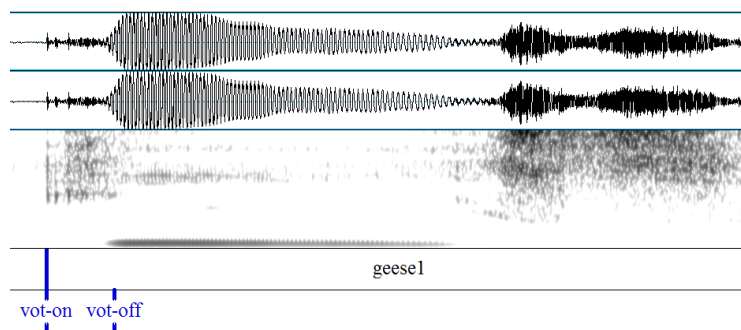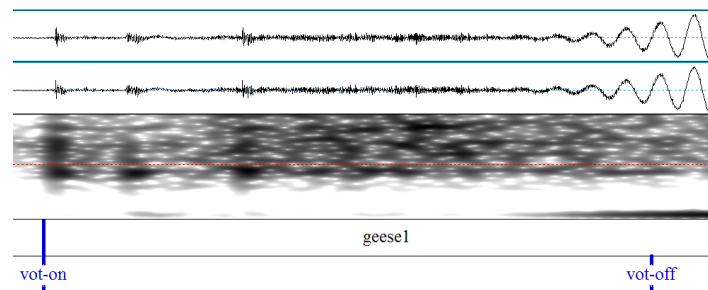Figure 21: Full spectrogram and waveform of *geese*

Figure 22: Close view spectrogram and waveform of *geese* (72 ms)

The VOT for this utterance of geese was 72 ms. Presumably, the high VOT for this utterance is also connected to a high vowel subsequent to the initial stop; however, this cannot be the entirety of the explanation for this high VOT for several reasons. First, we once again see the pattern of vertical striations in the spectrogram and periodicity in the wave form, much like we do for the velar stops in Shoshoni. Second, the duration of this VOT is not just considerably longer than the other English VOT averages, it is in the range of what could be considered a voiceless aspirated stop, when it should be a voiceless unaspirated stop. Lisker and Abramson (1964) reported the average VOT for voiceless aspirated stops to be 75 ms. Furthermore, Cho and Ladefoged (1999) report the average velar VOT for a slightly aspirated stop to be 50 ms and an aspirated stop to be 90 ms. The VOT here, 72 ms, is much closer to these aspirated averages, which account both for place of articulation and VOT variation amongst languages, than it is to an unaspirated stop average.

Furthermore, this same patterning of apparent glottal activity occurs not only in lenis stops, but also in English fortis stops, for example:

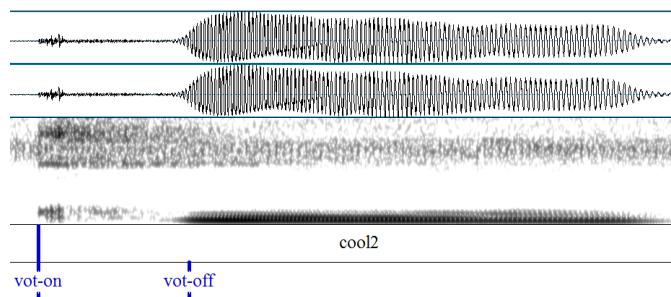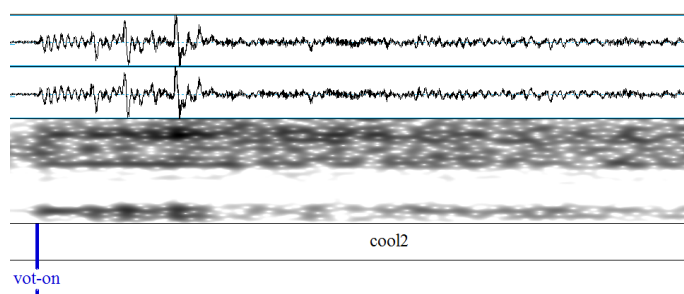Figure 23: Full spectrogram and waveform of *cool*

Figure 24: Close view spectrogram and waveform of *cool* (169 ms)

The VOT of this velar stop is particularly long, 169 ms, well above the 75 ms that is prototypical of voiced aspirated stops. In this utterance we once again see a high vowel; however, the other significant observation afforded by this additional data point, is that the VOT of velar fortis stops, seems well above the VOT of velar lenis stops. This is likely to offer an acoustic contrast between the two types of stops, given that the lenis stop falls within the range of VOT values we would usually expect for a fortis stop.

These two examples were chosen because they exemplify the patterning of release noise in these velar stops. While not all of the data yielded the same, especially long, VOT values, all of the data did display the same patterning of vocal fold movement for velar stops. The overall distribution of VOT stop values in English is demonstrated in the following table:

Figure 25: VOT durations for English velar stops in milliseconds
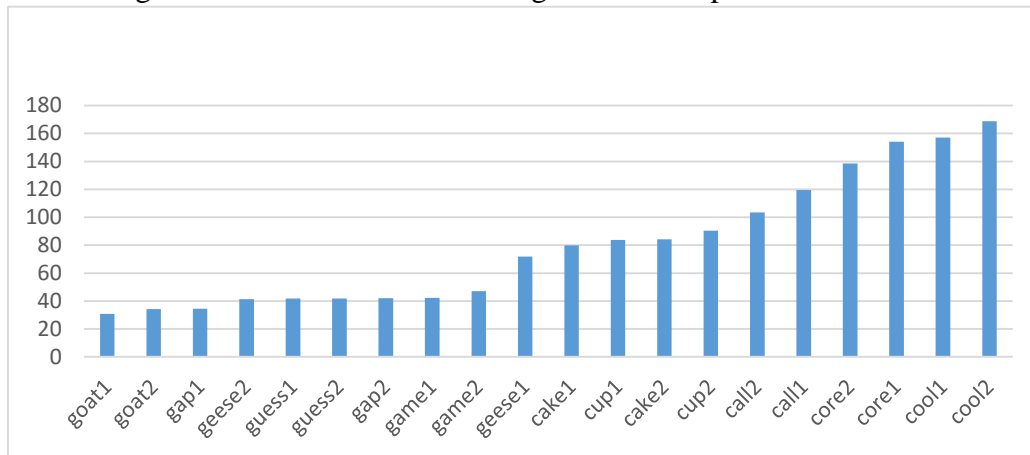
This table elucidates a few of the most important claims discussed thus far. First, there is a clear division of VOT durations between the lenis and the fortis stops in English. As previously mentioned, this is presumably a necessary acoustic division in terms of auditory processing. Second, in this chart, it generally seems to be the case that velars that precede high vowels tend to exhibit longer VOTs than velars that precede low vowels, though not always. Furthermore, even the lower VOT values for velars tend to fall above the VOT for alveolars and bilabials. Ultimately, it seems that the apparent glottal pulses seen after release in both the Shoshoni and English data are a feature of velar stops for at least this speaker, and potentially for Shoshoni.

7. Discussion:

Before turning to the theoretical implications of these data for the language overall, let us first conclude what we know about Shoshoni acoustics. First, based on the average VOTs for the bilabial and alveolar places of articulation, which were approximately 20 ms and 17 ms, respectively, it seems that Shoshoni could be said to have a series of voiceless unaspirated stops, at least for bilabial and alveolar stops. While these values fall slightly above the 10 ms average reported by Lisker and Abramson, the range reported by Lisker and Abramson (1964) was 1-25 ms for voiceless unaspirated stops, which does include the averages reported here. On the other hand, according to the Cho and Ladefoged (1999), we might also classify Shoshoni as having slightly aspirated stops. The conclusions are less clear for the velar stops. While both Lisker and Abramson (1964) and Cho and Ladefoged (1999) do report longer VOT values for velar stops, the differences seen here may extend beyond the previous reports, given the duration of the velar VOTs and the occurrence of apparent vocal fold movement in Shoshoni velar stops. Ultimately, the longer VOTs for velar stops did not serve a contrastive function, meaning that the velar stops characteristics provide an area for further investigation, but overall, the data support the previous assumption that Shoshoni is a single-stop-series language.

We can now turn our attention to broader theoretical implications based on these data by returning to Figure 1:

Figure 1: Category boundaries from various languages (Figure from Brookes and Kempe (2014))
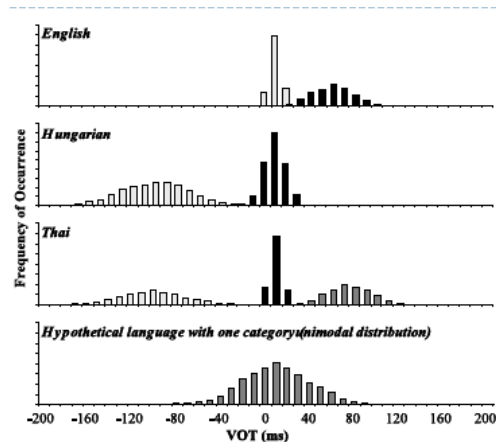

Figure 1: Distribution of VOTs in three languages based on measurements by Leigh Lisk and Arthur Abramson. Bars are colored to indicate different clusters in the input. The bottom panel shows a hypothetical language with only one voicing category.

One original hypothesis that we can rule out based on the data discussed thus far is a distribution similar to Thai, where there is a three-way contrast between the stops in English and the stops in Shoshoni. For this to be true, we would expect there to be three distinct categories of stops. These categories would more than likely either be negative and long-lag VOT stops in English, and short-lag VOT stops in Shoshoni, or unaspirated stops in English and slightly aspirated stops in Shoshoni. In other words, in both cases English lenis stops should remain unaspirated with the Shoshoni stops clustering a distinct category between the two series of English stops. However, as shown in Figure 26, this does not seem to be the case:

Figure 26: VOT distribution for lenis and fortis stops in English and Shoshoni
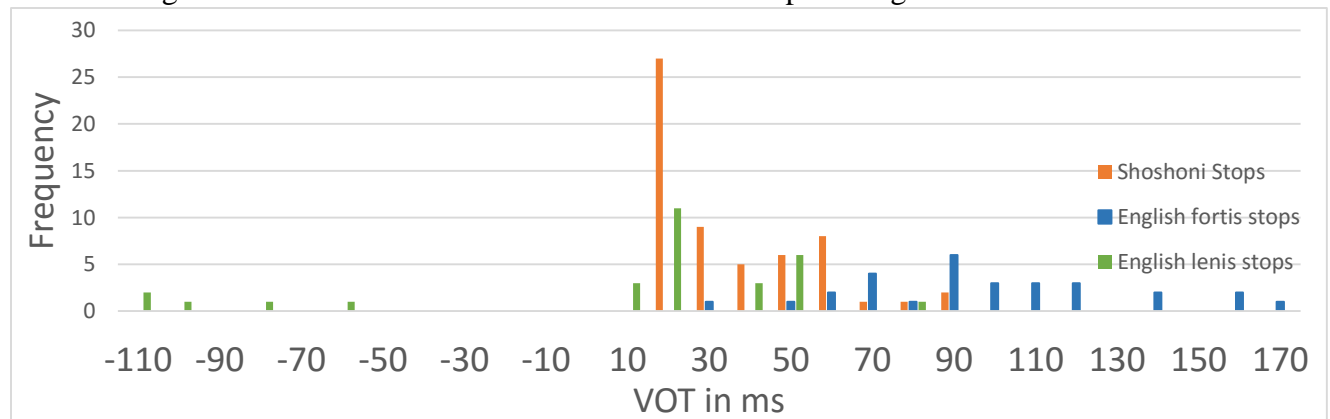
As we can see from this histogram, and as we saw in earlier data, it is not obligatory, at least for this speaker, to voice stops in English. In fact, there is a fair amount of overlap between the lenis stops of English and the single stop series of Shoshoni. This overlap is further shown in Figure 27, which shows the distribution of stops in Shoshoni and lenis stops in English:

Figure 27: VOT distribution for the Shoshoni stops and lenis stops in English
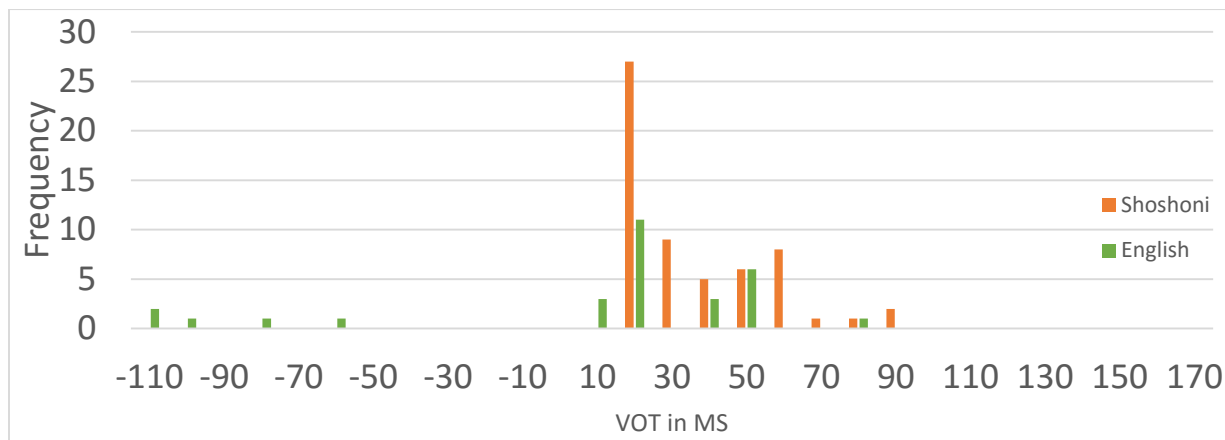
Overall, there does not seem to be a clear distinction between stops in Shoshoni and lenis stops in English. From this information we can conclude there is not a three-way distinction among the stops of the two languages.

The wide distribution prediction provided by Brookes and Kemp (2014) can likely be rejected as well. The distribution displayed here makes it seem unlikely that a wide distribution of VOT values for a single-stop-series language is true of Shoshoni. While there are exceptionally high VOT values for some stops in Shoshoni, these stops are velar. The extreme variation in velar stops may serve as some evidence for a broader distribution of stops than what we might see in another language; however, the English data seem to contradict this, as the lenis distribution was largely the same and the overlap between lenis and fortis categories was resolved by having even higher VOT values for velar fortis stops. Furthermore, the Brookes and Kemp (2014) model predicts the possibility of negative VOT values, which we did not find. An additional argument against the broad distribution of stops is the overall shape of the distributions. The hypothetical distribution proposed by Brookes and Kemp (2014) is unimodal and symmetrical, where the actual distribution is asymmetrical and negatively skewed, as shown by Figures 28 and 29:

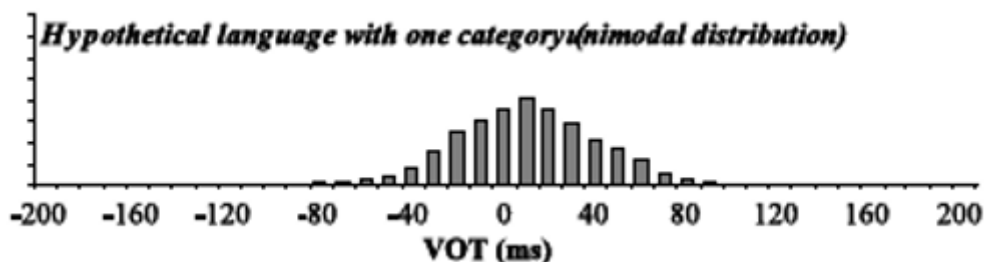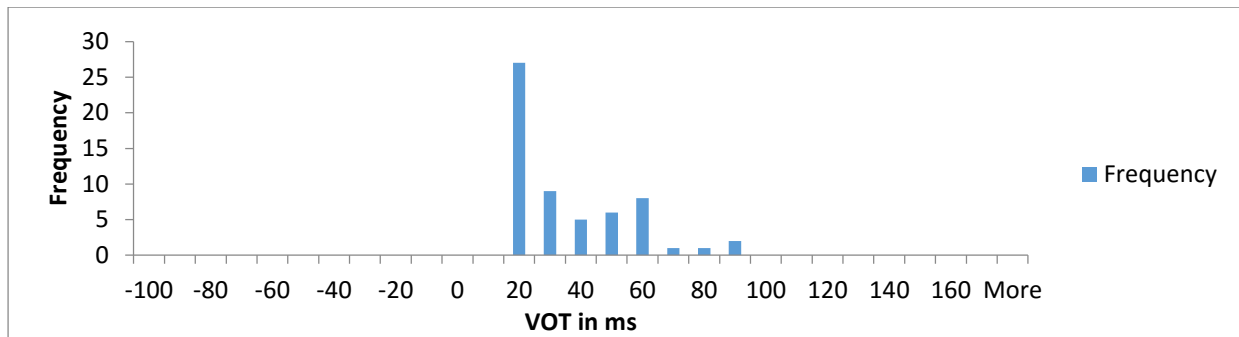Figure 28: Hypothetical VOT distribution (Brookes and Kempe 2014)

Figure 29: Shoshoni VOT distribution in ms

Therefore, this hypothetical distribution does not seem to be accurate for the speaker of Shoshoni. However, this model may still hold true for another single-stops-series language or for other Shoshoni speakers. Potentially, the contact with English has influenced Shoshoni stops, and therefore, altered the distribution of stops to appear more similar to English.

The remaining points we can conclude based on this data are somewhat less clear. While it does seem true that there is a single phonetic category for the lenis stops in English and the stops in Shoshoni, it is not clear what motivates these similarities. One option is that English has influenced Shoshoni and the Shoshoni stops have become more like the English stops. Alternatively, the influence of Shoshoni might have affected this speaker's production of lenis stops in English, causing the English lenis stops to resemble Shoshoni. The transfer of the vocal-fold activity in velar stops from Shoshoni to English and the slightly higher-than-average VOT for voiceless unaspirated stops may be support for this claim. However, it is also possible that there was not enough difference between English and Shoshoni stops for the stops to be mapped separately in the grammars of the two languages. In other words, the Shoshoni distribution of stops here is similar to English because, for this speaker, they were always similar stop categories and always shared properties. More data may be able to help tease apart these ideas; however, it may be harder to ever answer these questions based on the fact that there are no longer any monolingual speakers of Shoshoni.

8. Further Research

There are a variety of directions that this research can go from here. First, for these data to really provide any sort of conclusive insight into the language, more tokens, both from this speaker and from other speakers, are necessary. Finding more Shoshoni speakers may prove to be somewhat difficult as there are a limited number of Shoshoni speakers, many of whom are reluctant to work with linguists—especially on research where, to many, there seems to be little productive output for the community; however, more speakers are needed to show whether the features identified here are idiomatic or features of the language at large. Additionally, it may be important to at least categorize the data between L1 Shoshoni, L2 English speakers and L1 English and L2 Shoshoni speakers, as these two groups of speakers may pattern differently. However, in order for this research to provide any over-arching claims about Shoshoni, more data are necessary. Additionally, in order to make claims about other single-stop-series

languages, it would be useful to compare data from multiple single- stop-series languages, for instance Wari', to see what differences or similarities surface in the data.

Another useful avenue to study may be speaking rate effects. In the study by Kessinger and Blumstein (1997), aspirated stops showed significant VOT-duration differences depending on speaking rate. However, unaspirated stops did not show any significant differences in VOT duration based on speaking rate. Furthermore, according to Beckmen et al. (2011) a language with a single series of stops should not show any rate affect differences if it is a single-stop-series language with a plain voiceless stop series because only cues with specified features undergo change as a function of speaking rate. However, if Shoshoni in fact had a series of slightly aspirated stops or stops with the feature spread glottis, we would expect there to be rate differences in VOT duration. Therefore, rate affect data may provide more information about whether Shoshoni patterns as an aspirated or unaspirated language.

Perception tests might also yield interesting results that would help better define what is true of Shoshoni, as well as the broader theoretical implications. For instance, Shoshoni speakers could be asked to rate the goodness of stops with various synthesized VOT values to test whether or not there is a narrow category of values possible for Shoshoni stops, or if, as hypothesized by Brookes and Kemp (2014), there are a broad range of possible values. The velar stops could also be synthesized with and without the apparent glottal activity following release to see if Shoshoni speakers perceived these stops any differently.

Ultimately, this is a preliminary study on which to base further research. There are a wide range of possibilities for expanding this study that could help provide more insight into Shoshoni and into single-stop-series languages in general.

References

Beckman, Jill, Pétur Helgason, Bob McMurray, and Catherine Ringen. 2011. "Rate Effects on Swedish VOT: Evidence for Phonological Overspecification." *Journal of Phonetics* 39: 39–49.

Brookes, Patricia J, and Vera Kempe. 2014. *Encyclopedia of Language Development*. Thousand Oaks: SAGE Publications.

Cho, Taehong, and Peter Ladefoged. 1999. "Variation and Universals in VOT: Evidence from 18 Languages." *Journal of Phonetics* 27: 207–29.

Esposito, Anna. 2002. "On Vowel Height and Consonantal Voicing Effects: Data from Italian." *Phonetica* 59 (4): 197–231.

Flege, James Emil. 1995. "Second Language Speech Learning: Theory, Findings, and Problems." *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 233–77.

Gould, D, and C Loether. 2002. *An Introduction to the Shoshoni Language: Dammen Daigwape*. University of Utah Press.

Hunnicutt, Ashley, and Paul Morris. n.d. "Prevoicing and Aspiration in Southern American English." *Penn Working Papers in Linguistics* 22 (1).

Jakobson, Roman. 1968. *Child Language, Aphasia and Phonological Universals*. The Hague: Mouton.

Jessen, Michael. 1998. *Phonetics and Phonology of Tense and Lax Obstruents in German*. Amserdam: John Benjamins.

Kessinger, R.H., and S.E. Blumstein. 1997. "Effects of Speaking Rate on Voice-Onset Time in Thai, French, and English." *Journal of Phonetics* 25 (2): 143–68.

Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig, eds. 2014. *Ethnologue: Languages of the World,*. Seventeent. Dallas: SIL International.

Lisker, L, and A S Abramson. 1964. "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements." *Word Journal Of The International Linguistic Association* 20 (3): 384–422.

Loether, Christopher. 2006. *The Sven Liljeblad Collection Shoshoni Language Tapes at the University of Nevada-Reno Translation Project Grant Proposal*. Pocatello.

MacKay, Ian R. A., James Emil Flege, Thorsten Piske, and Carlo Schirru. 2001. "Category Restructuring during Second-Language Speech Acquisition." *The Journal of the Acoustical Society of America* 110 (1): 516–28.

Moseley, Christopher, ed. 2010. *Atlas of the World's Languages in Danger*. 3rd edn. Paris: UNESCO Publishing.