The Changing Sounds of Exceptionally Aspirated Diné Stops

Kayla Palakurthy*

1 Introduction

Many studies have analyzed the phonetic measurement of voice onset time (VOT), and previous research has identified several linguistic and social factors that influence its realization (Flege 1991, Yao 2007, Nagy and Kochetov 2013). Further, VOT is recognized as a frequent site of contact-induced phonetic interference in multilingual communities: the VOT of bilinguals often differs from that of monolinguals based on the phonemic stop categories of the languages in contact, the linguistic experiences of the bilingual speakers, and the social evaluation of variation in VOT or aspiration (Fowler et al. 2008, Michnowicz and Carpenter 2013, Newlin-Łukowicz 2014).

Diné bizaad, or Navajo, a Southern Dené/Athabaskan language spoken in the North American Southwest, plays a particularly important role in typological discussions of VOT due to the exceptionally long values reported for its voiceless aspirated stops, /t^h/ and /k^h/ (McDonough and Ladefoged 1993, Cho and Ladefoged 1999). However, extant VOT measurements were recorded 25 years ago, and an increase in English usage among Diné speakers has since ensued. Additionally, prior to the current state of widespread bilingualism, scholars noted variation in the aspirated releases, some of which was said to be socially meaningful (Reichard 1945, Saville-Troike and McCreedy 1980).

Given this background, this paper describes a study of phonetic variation in the release of voiceless stops, /t^h/ and /k^h/, measured acoustically with VOT, as well as spectral center of gravity (CoG) to compare the frication of the releases. Through quantitative analysis, this study seeks to address whether measurements of aspiration pattern synchronically with social predictors, and whether these exceptionally long releases are changing due to increased contact with English.

2 Background

Diné bizaad is a polysynthetic language spoken by over 150,000 speakers (Census 2010), mostly in and around the 25,000 square mile Navajo Nation, which spans Arizona, New Mexico and Utah. Though Diné bizaad remains the most spoken indigenous language in North America, the current vitality of the language is threatened by rapid intergenerational shift (House 2002).

The Diné phoneme inventory includes labial, alveolar, and velar stops with a three-way laryngeal distinction: voiceless unaspirated, voiceless aspirated, and voiceless aspirated ejective (McDonough 2003). Earlier studies report a VOT value of 154ms for /k^h/ and 130ms for /t^h/ (McDonough and Ladefoged 1993), and Diné stops are also notable for the strength of their aspirated releases (McDonough 2003). Some researchers analyze the stops as affricates /tx/ and /kx/ due to the timing of segmental gestures, with release periods that are phonetically similar to the plain velar fricative /x/ (McDonough and Wood 2008). However, other descriptions cite differences between the segments whereby only /tx/ is considered an affricate (Young and Morgan 1987).[1] Table 1 displays documented mean VOT values for Diné stops /tx/ and /k^h/, alongside the VOT of similar phonemes in English.

In addition to differences between /tx/ and /k^h/, the supralaryngeal constriction of the aspirated stops is said to vary across speakers: "consonants are aspirated, by some speakers very weakly, by others so strongly as to form consonant clusters—tx, kx" (Reichard 1951:19). Diné speakers are reportedly aware of the variation: "Navaho who do not emphasize the breathiness refer to those who

*This research was supported by an NSF Graduate Research Fellowship (2014178334) and an NSF DEL DDRIG grant (1713793). Thank you to the 51 Diné participants, to Barsine Bennally, Louise Ramone, and Melvatha Chee for their help in recruiting, and to the Navajo Nation Historic Preservation Department for their support. I would also like to thank Marianne Mithun, Matthew Gordon, and Lorene Legah for feedback on earlier versions of this analysis and Lal Zimman for sharing Praat scripts.

[1] I follow Young and Morgan 1987 in referring to these segments as /tx/ and /k^h/.

do as x-speakers, and mimic them by articulating the affected sounds almost as if they were coughed" (Reichard 1945:160). In other descriptions, stronger aspiration has been associated with Diné speakers from the Western region in Arizona, who are described as sounding "harsher than in New Mexico" on account of the force of their aspiration (Saville-Troike and McCreedy 1980:33). Strong aspiration has even been the focus of comedic imitation in performances of the Diné comedian, Vincent Craig, known for his portrayal of an exaggerated Navajo accent (Jacobsen 2017).

	Voiced unaspirated	Voiceless unaspirated	Voiceless aspirated	Voiceless aspirated ejective
Diné		/t/ 6	/tx/ 130	/t'/ 108
		/k/ 45	/k^h/ 154	/k'/ 94
English	/d/ 5		/t/ 70	
	/g/ 21		/k/ 80	

Table 1: VOT of Alveolar and Velar stops in Diné bizaad from McDonough and Ladefoged 1993 and English from Lisker and Abramson 1964.

3 Data and Methods

3.1 Data

Data for this analysis come from interviews recorded in 2016 and 2017 with 51 bilingual Diné bizaad/English speakers as part of a broader project investigating sociolinguistic variation and language attitudes. Table 2 presents the distribution of participants by relevant social factors. These factors serve as independent predictors in the statistical models described in Section 3.3.

Factors	Levels	Number of participants
Region	Western (Arizona)	n=29
	Eastern (New Mexico)	n=22
Gender	Men	n=20
	Women	n=31
Age	Younger speakers (18-38)	n=14
	Middle-aged speakers (39-58)	n=22
	Older speakers (59-78)	n=15

Table 2: Background of participants.

Following methods outlined in Cho and Ladefoged 1999 and McDonough 2003, I used an oral translation task to elicit tokens of targeted phonemes /tx/ and /k^h/ in stem-initial position, evenly distributed before different vowels. Participants repeated each word twice. Figures 1 and 2 display spectrograms containing /tx/ and /k^h/ respectively.

3.2 Acoustic Methods

Voice onset time (VOT) and center of gravity (CoG) were targeted as acoustic measures of aspiration to quantify both the length and the mean frication frequency of each release as an acoustic correlate to articulatory constriction (Gordon et al. 2002). I first segmented the release periods of all aspirated stops using the visual waveform and wide-band spectrogram in Praat, and then used automated scripts to compile measurements. Following Cho and Ladefoged 1999, VOT was measured from the release of the stop burst to the beginning of periodicity of the following vowel. In the case of multiple bursts, VOT was measured from the first burst, and 25 statistical outliers with a VOT greater than 200ms were removed. CoG was measured with a power spectrum weighting of two using the release period as the window of measurement. Following Sundara 2005, a 200Hz highpass filter was applied with a 100Hz smoother in order to remove effects of voicing, and a 12,000Hz low-pass filter to remove higher frequencies irrelevant for audible cueing (Thomas 2011). A total of 71 tokens were removed due to background noise.

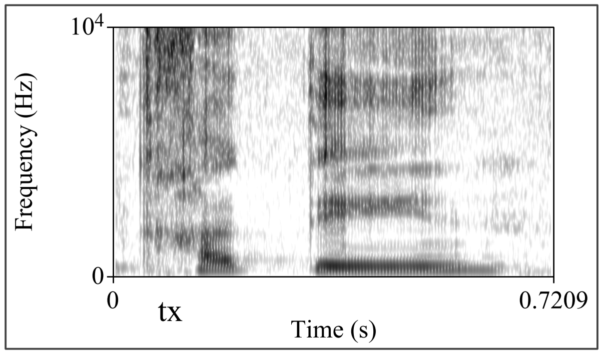

Figure 1: Spectrogram of *tádiin* from Annie, a 75-year-old woman. VOT = 88ms; CoG = 5016Hz.

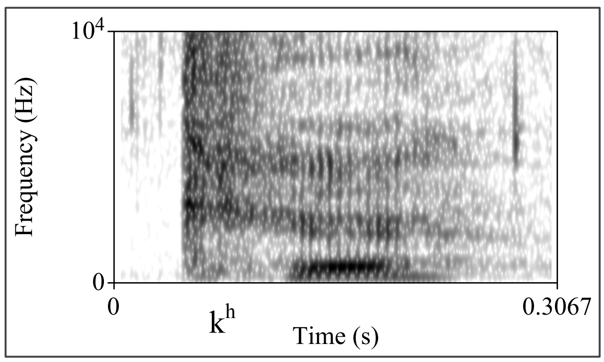

Figure 2: Spectrogram of *ké* from Lemuel, a 46-year-old man. VOT = 71ms; CoG = 2495Hz.

3.3 Statistical Methods

After compiling the acoustic measurements, I coded each observation for relevant linguistic and social predictors then fit mixed-effects linear regression models to values of VOT and CoG using the *lme4* package for *R* (Bates et al. 2015). Models included random intercepts for SPEAKER and WORD and the independent predictors in Table 3 as tested fixed effects. Following Zuur et al. 2009, the fixed effects structure was determined by starting with a maximal model containing all main effects and interactions calculated with maximum likelihood estimation and applying a backwards model selection process in which non-significant interactions and predictors were individually removed.[2] Final models were computed with restricted maximum likelihood estimation and additional analysis conducted with post-hoc Tukey tests using the *multcomp* package (Hothorn et al. 2008).

4 Results

4.1 Observed Results

As displayed in Figure 3, participants in this study have a longer VOT for /tx/ (mean = 117ms; sd = 26) than for /k^h/ (mean = 91ms; sd = 28), a result that contradicts typological patterns and results of the earlier study (McDonough and Ladefoged 1993). However, unlike the present data, previous

---

[2]Only the VOT model included PHONETIC ENVIRONMENT, while only the CoG model included FOLLOWING VOWEL LENGTH. The maximal models were otherwise the same.

measurements mainly targeted intervocalic segments. To facilitate accurate real-time comparison, VOT means and standard deviations are displayed by phonetic environment in Table 4 alongside earlier data. This comparison reveals that the VOT of intervocalic /k^h/ has shortened by 51ms, while the VOT of /tx/ has shortened by 5ms.

Independent Predictors	Levels	VOT n=	CoG n=
SEGMENT	k	865	827
	t	1258	1280
FOLLOWING VOWEL	a	424	418
	e	578	595
	i	490	474
	o	631	620
PHONETIC ENVIRON-MENT	Word-initial	741	708
	V_V	810	792
	C_	572	607
FOLLOWING VOWEL LENGTH[3]	long	960	971
	short	1163	1129
AGE GROUP	Young (18-38)	595	565
	Mid (39-58)	892	900
	Older (59-78)	636	642
GENDER	Men	1297	1271
	Women	826	836
REGION	East	930	934
	West	1193	1173

Table 3: Distribution of observations by independent predictors.

	/k^h/	n=	/tx/	n=
Word-initial	83 (26)	351	109 (23)	390
C_	90 (23)	279	110 (24)	293
V_V	**103 (31)**	**235**	**125 (26)**	**575**
V_V (1993)	**154 (43)**		**130 (29)**	

Table 4: Observed VOT by Phonetic Environment.

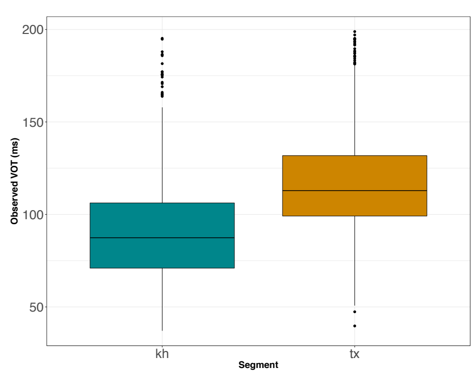

Figure 3: Observed VOT by Segment.

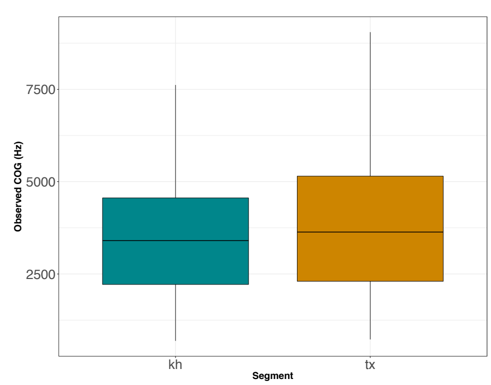

Figure 4: Observed CoG by Segment.

[3]This refers to phonemic vowel length.

Figure 4 shows that /tx/ has a higher CoG (mean=3452Hz; sd=1594) than /k^h/ (mean=3327Hz; sd=1425). Extensive co-articulation with the following vowel results in a wide range of CoG values.

4.2 Linear Regression Results

The model fit to VOT *(R^2=29%; cR^2=64%)* included significant interactions: SEGMENT:AGE, SEGMENT:GENDER, PHONETIC ENVIRONMENT:GENDER, AGE:FOLLOWING VOWEL, and GENDER:FOLLOWING VOWEL. The model fit to CoG *(R^2= 57%; cR^2= 68%)* included significant interactions: SEGMENT:AGE, SEGMENT:GENDER, AGE:FOLLOWING VOWEL, REGION:FOLLOWING VOWEL, AGE:REGION, AGE:GENDER, and a main effect of FOLLOWING VOWEL LENGTH. Except for GENDER, all fixed effects in the interactions were significant as individual main effects. Select interactions will be discussed thematically below.

4.2.1 Results by Age Group

Figures 5 and 6 display model predictions given the significant interaction between SEGMENT and AGE. With regards to the VOT of /k^h/ shown in Figure 5, young speakers are predicted to have the lowest values, while the VOT of /tx/ is fairly stable across age groups. The CoG values in Figure 6 reflect a change-in-apparent-time effect of increasing CoG for /tx/ and decreasing CoG for /k^h/. In both acoustic parameters, the segments are increasingly distinct from each other by age group.

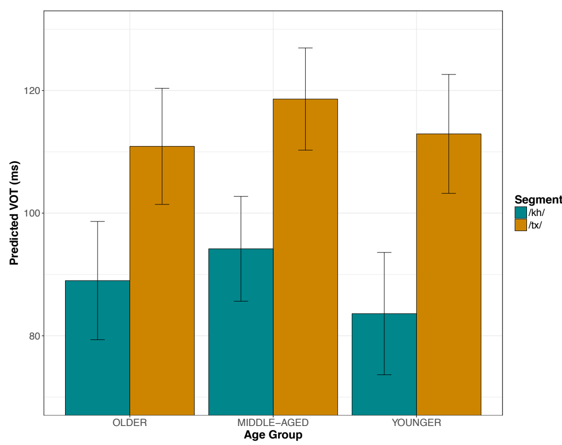

Figure 5: VOT by Segment and Age.

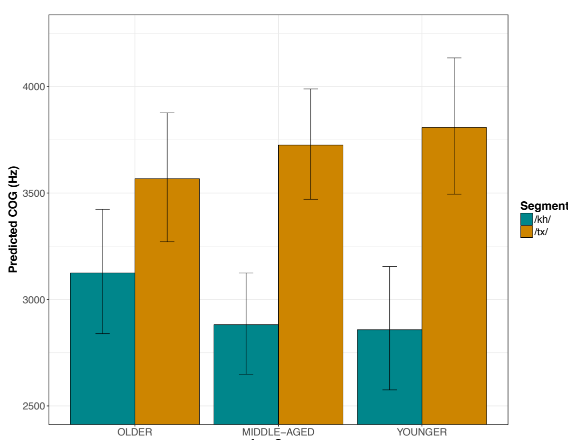

Figure 6: CoG by Segment and Age.

4.2.2 Results by Gender

Figures 7 and 8 display model predictions given the significant interaction between SEGMENT and GENDER. As shown in Figure 7, women have a shorter VOT for /k^h/ than men, while VOT values for /tx/ are similar across genders. In the CoG model output shown in Figure 8, women have much higher CoG values for /tx/ than men and slightly higher CoG values for /k^h/. In the CoG model, there is an additional significant interaction between GENDER and AGE, in which older women have a much higher CoG than men regardless of segment. Gender differences within other age groups are minimal. These results may partly reflect physiological differences in vocal tract sizes (Thomas 2011), but that explanation fails to adequately account for the generational patterning where only older women appear to have a more anterior constriction in their releases than men.

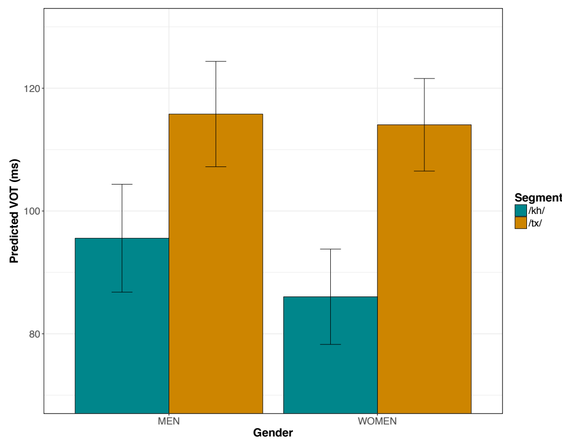

Figure 7: VOT by Segment and Gender.

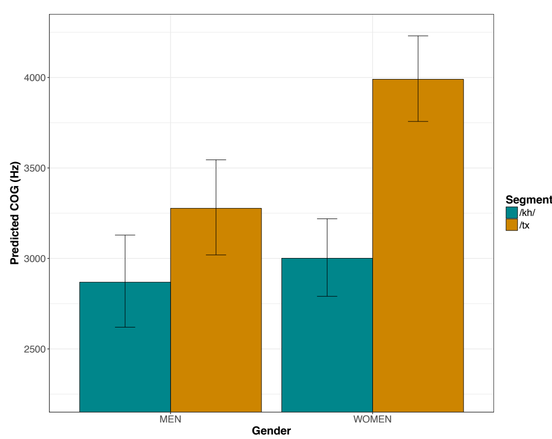

Figure 8: CoG by Segment and Gender.

4.2.3 Results by Region

Figure 9 displays model predictions given a significant interaction between AGE and REGION in the CoG model. REGION was not included in the final model fit to VOT.

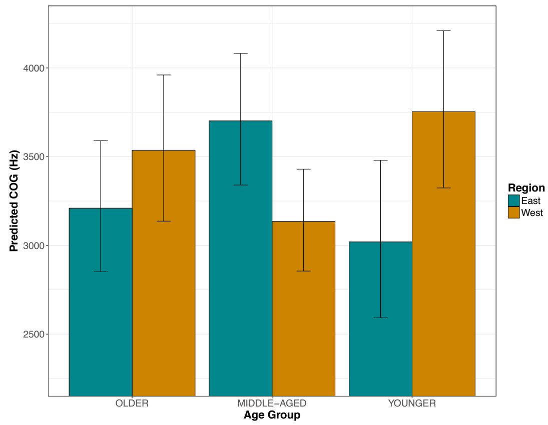

Figure 9: CoG by Age and Region.

Model results show that Western speakers have a higher CoG among older and younger speakers. Middle-aged speakers show the opposite pattern, possibly because Navajo language teachers are overly represented amongst Eastern middle-aged speakers. Language teachers may be more prone to self-monitor for long releases or to hyperarticulate. Similarity between younger and older speakers, but not middle-aged speakers, may also stem from widespread language transmission patterns; most of these young speakers learned Diné bizaad through close relationships with their grandparents, perhaps leading to more shared spectral characteristics amongst these age groups than the middle generation.

5 Discussion

Together these results contribute to answering the questions posed earlier, the first of which concerns the degree to which targeted measurements of aspiration pattern synchronically with social variables. Results indicate that phonetic measures of aspiration do vary significantly by age, gender, and region. For the segment /k^h/, young speakers have the shortest release periods and the lowest CoG, indicative of a more posterior articulatory constriction. At the same time, young speakers also have the highest CoG values for /tx/.

Regional patterns in these data show that Diné speakers from the Western/Arizona region have higher overall CoG values, especially for /tx/. These spectral differences could relate to the earlier documented perception that Western speakers have "stronger" aspiration (Reichard 1945). In a section of the sociolinguistic interview focused on discussing perceptions of differences between speakers, many participants expressed the opinion that the Western part of the reservation was associated with "better" or "more traditional" Diné language usage. These attitudes were present even among participants who had limited personal experience speaking with people from the Western area. Similar attitudes were noted in Jacobsen 2017. A more anterior release of /tx/ is just one feature amongst many variables present in speech, but due to its patterning with Western speakers, for some speakers this feature may index a "correct" Diné accent.

Women, especially older speakers, have much higher CoG values for /tx/ than men, making /tx/ and /k^h/ more distinct in CoG among women. This same progression may be observed as a change-in-apparent time effect amongst all speakers, evidence that women are leading a subphonemic change towards a more anterior constriction in the release of /tx/. These data thus expand the typology of the female-led changes found in many communities including those that speak indigenous minority languages (Ravindranath 2009). Earlier documentation also raises the possibility of stylistic variation in velar fricatives and plosive frication among men in particular: "big orators such as Chee Dodge would make the softer h sound while an 'old time man' would make the harder x sound" (Peery 2012:118). Such stylistic variation may account for lower CoG values among older men, if they are imitating the softer [h] in the style of prestigious Diné speakers.

The second question addresses whether the famously long release periods are changing given increased contact with English. I propose that a subphonemic change of timing has taken place in real time for /k^h/, but not /tx/. As is found in many studies of bilingual speech production, the perception of /k^h/ as similar to English /k/ may cause speakers to shorten their releases to align more closely with the English phonetic target. Even the oldest group of speakers, aged 59–78, have much shorter values for /k^h/ than values reported in earlier data, and the youngest group of speakers, aged 18-38 and most exposed to English, produce stops with the shortest VOT.

This result raises an additional question of why a similar timing change has not taken place with /tx/. I suggest that the salience, taken here to be "the degree to which something stands out relative to other, neighboring items" (Drager and Kirtley 2016:12), of the affrication of /tx/ is sufficient to distinguish it from English /t/ and precludes any gradient convergence towards an English-like phonetic target. Evidence of the salience of this affrication comes from the many sources and language learning materials that consistently describe a more strongly fricated release of /tx/ than /k^h/. As mentioned earlier, Young and Morgan (1987) write /tx/ as an affricate distinct from /k^h/, and McDonough and Wood (2008) also note that several Dené languages have a less fricated release for /k^h/ than /t^h/. In the present data, when compared with /k$^{h/}$, /tx/ has a longer release, a higher CoG, and a positively correlated and higher spectral standard deviation, indicative of dispersed frication energy. Given these acoustic characteristics, it is more likely that speakers will attend to the affrication of /tx/ than /k^h/, even if it remains below the level of conscious awareness.

Though these results, especially the shortening of the release of /k^h/, are explainable by appealing to effects of contact with English, there are also endogenous motivations for the observed subphonemic changes. A possible internal explanation is that there has been a conflation in timing of the release periods for ejective /k'/ and aspirated /k^h/; a similar change has been observed amongst bilingual Spanish/Yucatec Maya speakers (Michnowicz and Carpenter 2013). Synchronic measurements from the ejectives have not been analyzed, but Table 5 presents VOT values for aspirated stops compared with earlier data for ejectives at each place of articulation (McDonough and Ladefoged 1993). When compared with recorded values from 1993, the present VOT values of the aspirated stops are more similar, especially the velars, to the VOT of ejective stops. The ejective measurements also have a longer VOT for /t'/ than /k'/, mirroring current results for /tx/ and /k^h/.

Furthermore, ejectives are highly salient in many languages including Diné bizaad: "the lengthy pause that follows the release of the ejective stop is a salient aspect of Navajo" (Cho and Ladefoged 1999:223), and the contrast between Diné aspirated stops and ejectives does not depend on VOT (McDonough 2003). Given this salience, the shortening of the release periods of aspirated stops to match the duration of their ejective counterparts is a compelling possibility.

	Alveolar	**Velar**
Aspirated stops (1993)	130 (29)	154 (43)
Aspirated stops (present study)	117 (26)	91 (28)
Ejective stops (1993)	108 (31)	94 (21)

Table 5: VOT of ejective and aspirated stops.

6 Conclusion

In summary, results show that with respect to phonetic measurements of aspiration, Diné bizaad is not simply converging with English or developing random idiosyncrasies, as sometimes occurs in languages experiencing intense contact (Cook 1989). The relevance of social factors as significant predictors of variation suggests that phonetic variation continues to carry social meaning in the speech community. The shortening of the release of /k^h/ in real time, along with a gradient change-in-apparent time effect of increasing center of gravity for /tx/, constitute subphonemic changes that ultimately "may have minimal impact on the native structure of the language" (Babel 2009:24). Further, these data support the importance of perceptual salience in linguistic change, as the salience of Diné affricated /tx/ is argued to prevent its convergence with English /t/. Finally, ongoing phonetic divergence of Diné /tx/ from /k^h/ provides an example of how languages may continue to develop innovative distinctions despite the threat of intergenerational shift.

References

Babel, Molly. 2009. The phonetic and phonological effects of obsolescence in Northern Paiute. In *Variation in Indigenous Minority Languages*, ed. J. Stanford and D. Preston, 23–46. Philadelphia, PA: John Benjamins Publishing Company.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48.

United States Census Bureau. 2010. *Native North American Languages Spoken at Home in the United States and Puerto Rico: 2006-2010.*

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

Cook, Eung-Do. 1989. Is phonology going haywire in dying languages? Phonological variations in Chipewyan and Sarcee. *Language in Society* 18:235–255.

Drager, Katie and M. Joelle Kirtley. 2016. Awareness, salience, and stereotypes in exemplar-based models of speech production and perception. In *Awareness and Control in Sociolinguistic Research.* ed. A. Babel, 1–24. Cambridge, England: Cambridge University Press.

Fowler, Carol A., Valery Sramko, David J. Ostry, Sarah A. Rowland, and Pierre Hallé. 2008. Cross language phonetic influences on the speech of French-English bilinguals. *Journal of Phonology* 36:649–663.

Flege, James E. 1991. Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America* 89:395–411.

Gordon, Matthew, Paul Barthmaier, and Kathy Sands. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32:141–174.

Heselwood, Barry and Louise McChrystal. 1999. The effect of age-group and place of L1 acquisition on the realisation of Panjabi stop consonants in Bradford: An acoustic sociophonetic study. In *Leeds Working Papers in Linguistics and Phonetics* 7, ed. Paul Foulkes, 49–69.

Hothorn, Torsten, Frank Bretz, and Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50:346–363.

House, Deborah. 2002. *Language Shift Among the Navajos*. Tucson, Arizona: University of Arizona Press.

Jacobsen, Kristina M. 2017. *The Sound of Navajo Country: Music, Language and Diné Belonging*. Chapel Hill, NC: University of North Carolina Press.

Lisker, Leigh and Arthur S. Abramson. 1964. Cross-language study of voicing in initial stops: acoustical measurements. *Word* 20:384–422.

McDonough, Joyce and Peter Ladefoged. 1993. Navajo Stops. In *UCLA Working Papers in Phonetics 84,* ed. P. Ladefoged and I. Maddieson, 151–164.

McDonough, Joyce and Valerie Wood. 2008. The Stop Contrasts of the Athabaskan languages. *Journal of Phonetics* 36:427–449.

McDonough, Joyce. 2003. *The Navajo Sound System*. Dordrecht, The Netherlands: Kluwer Academic Press.

Michnowicz, Jim and Lindsey Carpenter. 2013. Voiceless stop aspiration in Yucatan Spanish: A sociolinguistic analysis. *Spanish in Context* 10:410–437.

Nagy, Naomi and Alexei Kochetov. 2013. Voice onset time across the generations. In *Multilingualism and Language Diversity in Urban Areas. Acquisition, identities, space, education,* ed. P. Siemund, I. Gogolin, M. Schulz, and J. Davydova, 19–38. Philadelphia, PA: John Benjamins Publishing Company.

Newlin-Łukowicz, Luiza. 2014. From interference to transfer in language contact: Variation in voice onset time. *Language Variation and Change* 26:359–385.

Peery, Char. 2012. New Deal Navajo linguistics: Language ideology and political transformation. *Language and Communication* 32:114–123.

Ravindranath, Maya. 2009. Language shift and the speech community: Sociolinguistic change in a Garifuna community in Belize. Doctoral dissertation, University of Pennsylvania.

Reichard, Gladys. 1945. Linguistic diversity amongst the Navaho Indians. *International Journal of American Linguistics* 11:156–168.

Reichard, Gladys A. 1951. *Navaho grammar*. New York: Augustin.

Saville-Troike, Muriel and Lynn A. McCreedy. 1980. Synchronic variation in Navajo: Regional, social, and developmental evidence from child language. *Final Project Report NSF 1979-1980*.

Sundara, Megha. 2005. Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *Journal of the Acoustical Society of America* 118:1026–1037.

Thomas, Erik R. 2011. *Sociophonetics: An Introduction*. Basingstoke, UK/New York: Palgrave.

Yao, Yao. 2007. Closure duration and VOT of word-initial voiceless plosives in English in spontaneous connected speech. *UC Berkeley Phonology Lab Annual Report* 3:183–225.

Young, Robert and William Morgan. 1987. *The Navajo language: A Grammar and Colloquial Dictionary*. Albuquerque: UNM Press.

Zuur, Alain F., Elena N. Ieno, Neil Walker, Anatoly A. Saveliev, and Graham M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. New York: Springer Science and Business Media.

Department of Linguistics
University of California, Santa Barbara
3432 University Drive
Santa Barbara, CA 93111
kaylapalakurthy @gmail.com