

The Phonetic Realization of Devoiced Vowels in the Southern Ute Language

Stacey Oberly^a Viktor Kharlamov^b

^aUniversity of Arizona, Tucson, Ariz., and ^bFlorida Atlantic University, Boca Raton, Fla., USA

Abstract

In this article, we analyze the phonetic realizations of devoiced vowels from 8 fluent speakers of Southern Ute, a severely endangered Southern Numic Uto-Aztecan language spoken in Southwestern Colorado. Devoiced vowels are considered to be an important aspect of the phonology of Southern Ute, yet very little is known about the pronunciation of such segments. Our findings indicate that devoiced vowels are realized phonetically in three ways: (i) fully voiceless, (ii) partially devoiced, and (iii) fully reduced with concurrent lengthening, lower intensity and greater voicelessness of the preceding consonant. Such variable and noncategorical devoicing is seen for both high and nonhigh vowels and all consonants regardless of their manner of articulation.

© 2015 S. Karger AG, Basel

1 Introduction

Vowel devoicing is considered to be an important areal feature of the American Indian languages of the Rockies and the Plains (Ladefoged and Maddieson, 1996). It has been attested in Chemehuevi (Press, 1979), Comanche (Canonge, 1957; Armagost, 1985; Charney, 1993), Hopi (Whorf, 1934; Manaster-Ramer, 1986), Shoshone (Armagost and Miller, 2000), Southern Paiute (Sapir, 1930; Harms, 1966), Southern Ute (Givon, 1979; Charney, 1996; Oberly, 2008), and Tohono O'odham (Miyashita, 2000). In these languages, vowels are known to be either voiced or devoiced, as in the case of the Southern Ute word [tʰɪ'kwa] 'deep' that has a devoiced vowel [ɪ̥] in the initial unstressed syllable and a voiced vowel [a] in the second stressed syllable (Givon, 1979, p. 191).

However, limited to no acoustic data are available for most of the indigenous languages with vowel devoicing, so it is not clear how such vowels are actually realized at the phonetic level. They may be pronounced as 'whispered vowels' that have no visible voicing bar and at best limited formant structure (Ladefoged and Maddieson, 1996). Alternatively, they may show gradation in devoicing similar to what has been reported for Korean, Japanese, Montreal French and Cypriot Greek where devoiced vowels

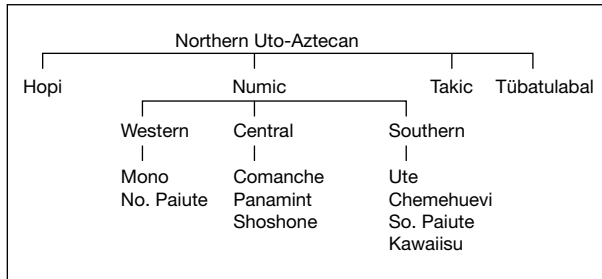


Fig. 1. Northern Uto-Aztecan branch (Lewis, 2009).

appear to range from fully voiced to partially voiced to fully voiceless (Cedergren and Simoneau, 1985; Jun and Beckman, 1993; Beckman, 1996; Eftychiou, 2010). With devoiced vowels lacking perceptual distinctness (Gordon, 1998), devoicing may also result in complete reduction of the vowel when the segment is no longer being pronounced, as seen for Montreal French (Cedergren and Simoneau, 1985). For Native American languages with vowel devoicing, this could mean that previously unattested consonantal sequences are now being produced by speakers, which would in turn necessitate reanalyses of the phonotactic constraints of these languages.

The current study examines the phonetic realization of devoiced vowels in Southern Ute, a Uto-Aztecan language spoken in Colorado. We describe the surface forms of devoiced vowels based on data from 8 fluent speakers and show that such segments are not necessarily realized as separate vowels. Instead, the presence of these syllabic elements in the mental representations of the words can be signaled at the phonetic level by the acoustic properties of the preceding consonants.

In the next section, we present an overview of the Southern Ute language, its language family, and the sound inventory. This is followed by examples and description of vowel devoicing in Southern Ute.

2 Language Information

Southern Ute is a member of the large Uto-Aztecan language family (Lewis, 2009). It belongs to the Southern Numic branch, which also includes Southern Paiute, Chemehuevi and Kawaiisu, as shown in figure 1. It is spoken primarily on the Southern Ute reservation in Southwestern Colorado, which includes the town of Ignacio and surrounding areas.

2.1 *The Sound System of Southern Ute*

The consonantal inventory of Southern Ute is shown in figure 2. It includes 17 phonologically distinct consonants ([p, t, tʃ, k, q, x, χ, ʎ, ʁ, ʟ, v, r, s, m, n, w, j]), a voiced velar stop [g] (an allophone of the voiced velar fricative [ɣ]; listed in parentheses), and the sequence of [qχ] that we analyze as an affricate (due to single release in our data).

The sound system of Southern Ute also includes 7 vowels ([i, u, e, ε, ø, o, a]) that are shown in figure 3. All these vowels can be long, which is represented in the orthography as a sequence of two identical vowel letters (e.g., [i:] written as ‘ii’; Givon, 2013). The vowels can also be stressed or unstressed, with stress usually falling on the second vowel of the word if the first vowel is short or the first vowel if it is long (Givon, 2013).

	Bilabial	Labio-dental	Alveolar	Palatal	Velar	Uvular	Glottal
Plosive	p		t		k (g)	q	ʔ
Fricative		v	s		x ɣ	χ	ʁ
Affricate				tʃ		(qʒ)	
Nasal	m		n				
Flap			r				
Approximant	w			j			

Fig. 2. Southern Ute consonants (adapted from Givon, 2013).

	Front		Back	
	-Round	+Round	-Round	+Round
High	i		ɯ	u
Mid	ɛ	ø		o
Low			ɑ	

Fig. 3. Southern Ute vowels (adapted from Givon, 2013).

2.2 Vowel Devoicing in Southern Ute

An important feature of Southern Ute is the devoicing of vowels. As shown in (1) below, it affects all 7 vowels of the language and is also found in the case of a vowel whose underlying identity cannot be reconstructed on the basis of currently available data (transcribed as [H]; according to Givon, 2013, this may be a high central vowel).

(1) Devoiced vowels in Southern Ute (marked with [◌̥] and bolded; based on Charney, 1996):

- a. [i̥] ʃi̥ʃi̥ɛ
'to be hard'
- b. [ɯ̥] tɯ̥kwa
'to be deep'
- c. [y̥] ky̥penapɯ̥
'bat, drum stick'
- d. [ø̥] kø̥pokiti:
'to cause to break'
- e. [ø̥] tø̥təna
'to hit repeatedly'
- f. [ɑ̥] pi:pɑ̥koʔwa
'to burp'
- g. [ɛ̥] ʃi̥ʃi̥ɣkatumɯ̥
'hard/difficult'
- h. [H] ʃi̥:wHʃapɯ̥
'beaded belt'

The sampling in (1) illustrates that devoicing tends to affect short unstressed vowels adjacent to voiceless consonants. In (1a–e), the devoiced vowels occur in the first unstressed syllable of the word when the vowel is both preceded and followed by voiceless consonants. In (1f), the devoiced [ɑ̥] occurs in the second unstressed syllable, which is unstressed because the first vowel is phonologically long, and is located in between

the voiceless stops [p] and [k]. The examples in (1g) and (1h) show that devoicing can occur when only the following consonant is voiceless. In (1g), the devoiced unstressed front unrounded mid vowel [ɛ̥] is preceded by the voiced fricative [ɣ] and followed by the voiceless stop [k]. In (1h), the unidentified devoiced short unstressed vowel [H] is preceded by the voiced glide [w] and followed by the voiceless affricate [tʃ].

The words in (1c), (1g) and (1h) also show devoicing of the word-final vowel [ʉ] that is preceded by either the voiceless stop [p] (1c and 1h) or the voiced nasal [m] (1g). This type of devoicing is not necessarily predictable by a phonological rule as it interacts with various grammatical processes, such as suffixation, compounding, and case marking (Givon, 1979, 1980, 2013; Charney, 1996). For example, for nouns, the word-final vowel is often part of a suffix that consists of a consonant and a vowel, such as [-vi, -ʃi, -pu, -tu] (Goss, 1972; Givon, 2013). When a noun is the subject of a sentence, the word-final vowel is devoiced. When the same word is an object, the final vowel is voiced. This results in minimal pairs such as [tukuavɪ̥] ‘meat (subj.)’ and [tukuavi] ‘meat (obj.)’ (Givon, 2013, p. 48).

Voicing alternations are also attested in the case of number marking and case marking. For example, the high front vowel is devoiced in the singular form [mamaʃɪ̥] ‘woman’ but not when the same word is suffixed with the plural suffix [-u] and is pronounced as [mamaʃiu] ‘women’ (Givon, 1979, p. 20). Similarly, the final vowel is devoiced in the noun [tukaʔnapu] ‘table’ but the same vowel becomes voiced when the postpositional suffix [-vwan] ‘on’ is added onto the word, which results in the surface form [tukaʔnapuvwan] ‘on the table’ (Givon, 1979, p. 23). The noun [puku] ‘horse’ ends in a devoiced high back rounded [ʉ̥] but when the first-person possessive suffix [-n] is added onto the word, the [u] is voiced and the word is pronounced as [pukun] ‘my horse’ (Givon, 1979, p. 7). Alternations of this type seem to involve mostly high vowels [i, u, ʉ], which is likely due to the fact that these vowels are at the end of the most common noun suffixes (e.g., [-ʃi, -pi, -vu, -pu, -ru, -tu]) (Oberly, 2008). However, the exact nature of such alternations and the environments in which vowel devoicing does and does not apply in Southern Ute are not yet fully understood.

Even less is known about the phonetic realization of devoiced vowels in Southern Ute. According to Givon (1979), such segments may not be pronounced by some speakers. Other speakers produce ‘silenced’ segments, when the vowel is not audible but its presence ‘can to some extent be told by the lip articulation’ (Givon, 1979, p. 7). Oberly (2008) provides descriptive phonetic data from 4 speakers of Southern Ute and notes the shorter durations of devoiced vowels compared to their voiced counterparts as well as a devoicing effect that such segments have on the preceding consonants. No other acoustic data have previously been reported for the language.

The current study addresses the gap in phonetic knowledge on the realization of devoiced vowels in Southern Ute. We aim to provide general phonetic descriptions of such segments and their effect on the preceding consonants, including making statistical comparisons between voiced versus devoiced vowels, as well as compare the phonetics of Southern Ute devoiced vowels to how similar segments are realized in other languages. Given the many unanswered questions related to the phonological aspects of the devoicing process in Southern Ute, we analyze both predictable and unpredictable cases of devoicing together, in the hope that future research will be able to determine the extent to which the patterns described below are representative of the different types of devoicing.

The following section provides a detailed description of the methods of data collection and analysis used in the current study.

Table 1. Southern Ute speaker information

Female	Age	L2	Literate in S. Ute	Male	Age	L2	Literate in S. Ute
F1	72	EN, SP	N	M1	68	EN	N
F2	72	EN	N	M2	78	EN	N
F3	72	EN	Y	M3	79	EN, SP	N
F4	84	EN	N	M4	90	EN, SP	N

EN = English; SP = Spanish; N = no; Y = yes.

3 Methods

3.1 Speakers

Our data came from 8 fluent Southern Ute speakers, 4 males and 4 females. Their ages ranged from 68 to 90, with an average age of 76.9. All speakers learned Ute as a first language and later learned English. Three speakers (F1, M3, M4) also spoke Spanish. Only 1 speaker (F3) was literate in Southern Ute. Speaker characteristics are summarized in table 1.

3.2 Recordings

We elicited speech tokens using a translation task (English to Southern Ute) with a list of 202 Southern Ute words and phrases from Oberly (2008). The translation task made it possible to obtain data from all speakers regardless of whether or not they were literate in Southern Ute. The word list contained all CV and VC sequences that are possible in the language. The speakers said the words both in isolation and in a carrier sentence [*Maas tani ____ maiku*] ‘He always says ____’. It is important to note that Southern Ute is an agglutinative language and speakers occasionally produced words with additional suffixes added onto the stems. They also sometimes produced alternate forms with related meanings. All these items were included in the analyses, which resulted in a larger dataset that could be analyzed using not only descriptive but also inferential statistics.

The recordings were done indoors on the Southern Ute reservation in Colorado using a Marantz PMD671 portable solid state digital recorder and a Countryman Isomax E60P5T omnidirectional head-mounted microphone with a frequency response of 30 Hz to 20 kHz. The audio was recorded in .wav format at a rate of 16 bits and with a sampling frequency of 44,100 Hz.

3.3 Measurements and Analyses

Phonetic analysis was done with Praat (Boersma and Weenink, 2014). Three acoustic measurements were calculated for the vowels and the preceding consonants: (1) segmental duration, (2) intensity of the segment, and (3) the voicing ratio. Segmental duration was measured in milliseconds, using segmental boundaries that were determined on the basis of waveforms and spectrograms (using such properties as the presence of periodic cycles in the waveform, presence of a voicing bar in the spectrogram, onset and offset of frication noise, presence and changes in formant structure, etc.). For vowels and preceding fricatives, nasals, flaps and glides, the measurement was done over the entire speech sound. For plosives and affricates, release duration (a single measurement including both burst duration and release/aspiration duration) and closure duration were measured separately, with closures measured for non-word-initial segments only.

Sound intensity, which is a correlate of loudness, was measured in decibel (dB) using high-pass-filtered data (400 Hz; Hann) and a window set at 70% of the measured segment duration. For plosives and affricates, intensity was measured during the release portion of the consonant only, with the window centered around the midpoint of release. The data were not normalized as the recording settings and microphone distance were kept constant both within and across the recording sessions.

The voicing ratio was measured as the percentage of the total duration of the segment that had phonetic voicing in it as indicated by the concurrent presence of a voicing bar and formant structure

Table 2. Phonetic realization of devoiced vowels

Phonetic form	n	%
Completely devoiced	15	1.42
Partially devoiced	178	16.87
Fully reduced	862	81.71
Total	1,055	100

in the spectrogram and periodic vibration in the waveform, even if it was of extremely low intensity. Closure voicing and release voicing were measured separately.

The data were averaged across items (i.e., averaging across the tokens from the same category produced by the same speaker) and analyzed using repeated measures ANOVAs and paired-samples t tests. Acoustic measurements were used as the dependent variables. When the devoiced vowel was produced as a separate segment, the analyses compared (i) devoiced vowels to (ii) voiced short unstressed vowels. When the vowel was not produced as a separate segment, comparisons were made between consonants that preceded the two types of vowels. The data for preceding plosives, fricatives and affricates were analyzed separately as Southern Ute lacks tokens for some places and manners of articulation (e.g., no labial affricates). Only the role of the preceding consonant could be investigated due to the prevalence of word-final vowels in the dataset. The Gender factor was not included in the statistical design because pilot testing showed no effect of speaker's gender on the degree of vowel devoicing for any of the acoustic parameters (all F values <2.0, all probabilities >0.1). The exact type of statistical treatment and the list of independent predictors for different stimulus types are described in more detail in the Results section.

The following section presents the results of acoustic analyses, starting with a general typology of the phonetic realizations of devoiced vowels in the dataset. This is followed by detailed descriptions of the observed types of devoiced vowels.

4 Results

4.1 General Summary

The Southern Ute dataset was found to contain a total of 1,055 instances of devoiced vowels. Three different types of phonetic realizations were observed: (i) fully voiceless, (ii) partially devoiced and (iii) fully reduced with concurrent lengthening, lower intensity and greater voicelessness of the preceding consonant. This is shown in table 2.

As can be seen from the table, 18.29% of all devoiced vowels were produced as separate speech sounds that could be segmented out from the adjacent consonants. This includes instances of devoiced [j̥] (n = 51), [ɥ̥] (n = 20), [ɥ̥ɹ] (n = 57), [ɔ̥] (n = 3), [ḁ] (n = 50), and [H] (n = 12). Out of these tokens, 1.42% were realized as fully voiceless and 16.87% were either partially or fully voiced. The remaining 81.71% of devoiced vowels show effects on the acoustic properties of the preceding consonants without being represented by a vowel-like segment at the phonetic level. The different realizations of devoiced vowels are described in sections 4.2 (fully voiceless and partially devoiced) and 4.3 (fully reduced with a concurrent effect on the preceding consonant).

4.2 Fully Voiceless and Partially Devoiced Vowels

Figure 4 provides examples for devoiced vowels produced as fully voiceless and partially devoiced. In figure 4a, notice that the fully voiceless [j̥] shows irregularity of the waveform, absence of a voicing bar, and a faint formant structure in

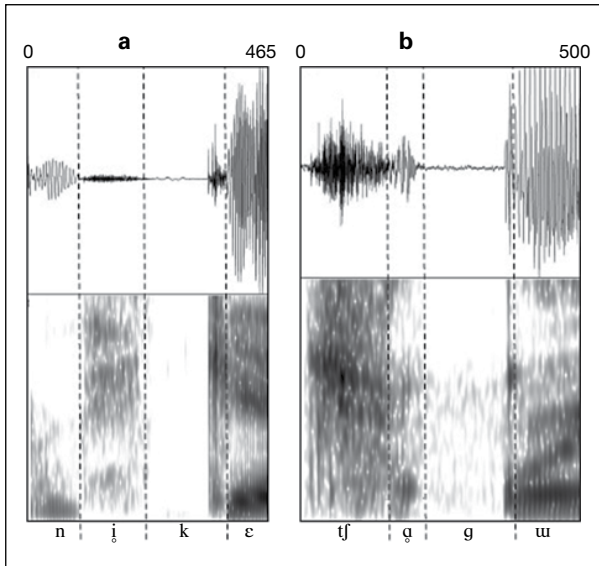


Fig. 4. Fully voiceless (a) and partially devoiced (b) vowels. Data from speakers F3 (a) and M3 (b).

Table 3. Voiced vowels versus devoiced vowels; the vowel [H]

Type	Measurement	Voiced	Devoiced	Difference	F	p
Vowels (excl. [H])	duration, ms	<i>110 (13.5)</i>	<i>82 (17.4)</i>	<i>28</i>	<i>25.75</i>	<i>0.002</i>
	intensity, dB	<i>66.7 (4.6)</i>	<i>59.8 (4.1)</i>	<i>6.9</i>	<i>21.48</i>	<i>0.004</i>
	voicing ratio, %	<i>91.2 (4.1)</i>	<i>63.7 (23.1)</i>	<i>27.5</i>	<i>13.90</i>	<i>0.01</i>
[H]	duration, ms	–	77	–	–	–
	intensity, dB	–	54.8	–	–	–
	voicing ratio, %	–	58.5	–	–	–

Means and standard deviations (in parentheses) of vowel duration, intensity and voicing ratio. Statistically significant differences (voiced vs. devoiced) are in italics.

the spectrogram. In figure 4b, the partially devoiced vowel [q] shows periodicity of the waveform, presence of a voicing bar, and clearly visible formants in the spectrogram.

Table 3 lists the means of duration, intensity and the voicing ratio of voiced vowels, devoiced vowels (averaged across both fully and partially devoiced vowels and excluding [H]), and the vowel [H] specifically. Significance of the observed differences was tested in separate repeated measures ANOVAs for each acoustic measurement, with factors Voicing (voiced, voiceless) and Vowel Height (high, nonhigh) used as independent predictors. The 2×2 ANOVAs were run on the data of 7 (out of 8) speakers who produced sufficient numbers of tokens in all voicing and height

Table 4. Phonetic realization of partially devoiced vowels

Voicing ratio	1–50% voiced	51–99% voiced	100% voiced	Total
Token counts, %	5.40 (n = 57)	4.83 (n = 51)	6.64 (n = 70)	16.87 (n = 178)

categories. (Speaker F1 did not produce any tokens of devoiced nonhigh vowels.) Only descriptive statistics are provided for the vowel [H] that does not have a phonologically voiced counterpart in Southern Ute and does not have a clear specification for height.

As can be seen from table 3, devoiced vowels were found to be significantly shorter than voiced short unstressed vowels (28 ms difference). Devoiced vowels also show significantly lower average intensity (6.9 dB difference), which indicates that they are less loud than their voiced counterparts, and have significantly fewer voicing cycles (27.5% difference), which demonstrates that the rate of phonetic devoicing is higher in devoiced than voiced vowels. At the same time, the voicing ratio rate of 63.7% observed in the case of devoiced vowels reveals that phonetic voicing often continued throughout more than half of the devoiced vowel.

Among the 16.87% of devoiced vowels that were either partially or fully voiced, 5.40% (n = 57) of tokens had phonetic voicing continuing for less than 50% of the vowel, 4.83% (n = 51) of tokens had voicing continuing for more than 50% but less than 100%, and 6.64% (n = 70) of tokens were fully voiced. This is shown in table 4.

As confirmed by the absence of an interaction between voicing and vowel height for any of the three acoustic parameters ($F < 2$, $p > 0.1$), the observed differences between voiced and devoiced vowels are representative (at the group level) of both high and nonhigh vowels alike. They are also highly in line with the differences observed in the case of the vowel [H], which is also shorter (33 ms difference), less loud (11.9 dB difference) and less voiced (32.7% difference) than voiced vowels. The [H] vowel also tends to be either partially or fully voiced (n = 11) rather than fully voiceless (n = 1), with phonetic voicing usually continuing throughout more than 50% of its duration. This demonstrates that, when pronounced, all devoiced vowels likely follow the same general pattern of shorter duration, lower intensity, and limited phonetic voicing.

The data also showed a general effect of vowel height on intensity, with lower values observed in the case of high vowels [$\text{mean}_{\text{high}} = 59.5$ dB, $\text{SD} = 4.1$; $\text{mean}_{\text{nonhigh}} = 66.8$ dB, $\text{SD} = 4.6$; $F(1, 6) = 29.67$, $p = 0.002$]. These differences are the result of the phonetic properties of the vocal tract (Lehiste and Peterson, 1959) and they did not interact with voicing ($F < 1$, $p > 0.1$).

4.3 Fully Reduced Vowels: Effect on the Preceding Consonants

The 81.71% of tokens with devoiced vowels that were not realized phonetically as separate segments showed consistent effects of voicing on the acoustic properties of the preceding consonants. In the sections that follow, these effects are described separately for different types of consonants, starting with plosives.

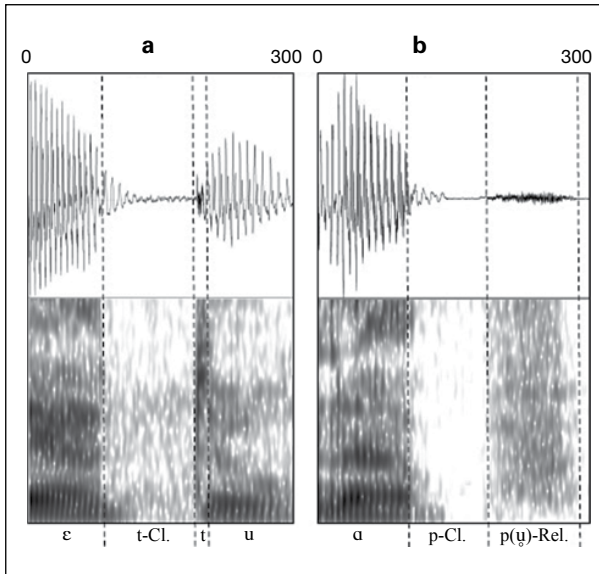


Fig. 5. Plosives followed by a voiced vowel (**a**) vs. a fully reduced vowel (**b**). Data from speaker M3.

Table 5. Plosives not followed versus followed by fully reduced vowels

Type	Measurement	Reduced V		Difference	F	p	
		N	Y				
Plosives	duration, ms	closure	110 (18.3)	108 (24.8)	2	0.14	0.717
		release	<i>31 (7.2)</i>	<i>102 (25.3)</i>	<i>-71</i>	<i>73.3</i>	<i><0.001</i>
	intensity, dB	release	<i>57.1 (3.4)</i>	<i>52.3 (3.4)</i>	<i>4.8</i>	<i>41.78</i>	<i><0.001</i>
	voicing ratio, %	closure	44.2 (19.2)	38.8 (12.0)	5.4	2.32	0.179
		release	<i>35.0 (9.3)</i>	<i>13.0 (8.7)</i>	<i>22.0</i>	<i>18.82</i>	<i>0.003</i>

Means and standard deviations (in parentheses) of closure duration, release duration, release intensity, closure voicing ratio and release voicing ratio. Statistically significant differences are in italics.

4.3.1 Plosives

Figure 5 demonstrates plosives followed by a voiced vowel versus a fully reduced vowel. In the waveform and spectrogram in figure 5a, the voiceless plosive [t] consists of a closure and a short release and is then followed by a voiced [u] with a voicing bar and formant structure. In figure 5b, the voiceless bilabial plosive [p] is followed by a fully reduced vowel. Notice that there is no vowel or vowel-like segment in the waveform or the spectrogram in figure 5b. Instead, the long low amplitude release of the [p] is what signals the syllabic element.

Table 5 lists the means of closure duration, release duration, release intensity, closure voicing ratio and release voicing ratio of plosives that are not followed versus

followed by fully reduced vowels. Significance of the differences was investigated in separate repeated measures ANOVAs for each acoustic measurement. Independent predictors included (i) adjacency to a fully reduced vowel (no, yes) and (ii) consonantal place of articulation (labial, coronal, dorsal). The 2×3 ANOVAs were run on the data from 8 speakers for releases and 7 speakers for closures. (Speaker M2 did not produce a sufficient number of non-word-initial tokens.) Closure duration and closure voicing ratio were measured on the basis of word-medial and word-final plosives only ($n = 224$).

As reflected in table 5, plosives followed by devoiced vowels demonstrated longer release durations (71 ms difference), lower intensity (4.8 dB difference), and fewer voicing cycles during releases (22.0% difference). Durations of closures and voicing ratios during closures were not affected. This shows that the acoustic properties of releases reveal the presence of a syllabic element at the phonological level.

The data also showed a general effect of consonantal place of articulation on closure duration [$F(2, 12) = 11.09, p = 0.003$] and closure voicing ratio [$F(2, 12) = 5.79, p = 0.041$] but there was no interaction between consonantal place and the presence of a fully reduced vowel ($F < 3.5, p > 0.05$). This suggests that the differences shown in table 5 are representative of all Southern Ute plosives regardless of their place of articulation.

4.3.2 Affricates

Figure 6 demonstrates the effect of a fully reduced vowel on the properties of the preceding affricate. The figure shows affricates followed by a voiced vowel versus a fully reduced vowel. Notice the difference in release durations and amplitudes of the two affricates preceding the high vowel /i/, with longer duration and lower amplitude observed in the case of the fully reduced vowel in figure 6b.

Table 6 compares the means of closure duration, release duration, release intensity, closure voicing ratio and release voicing ratio of affricates not followed versus followed by fully reduced vowels ($n = 274$). The means are based on the data from 8 speakers. Significance of the differences was tested in separate paired-samples *t* tests for each acoustic measurement. The potential role of consonantal place of articulation could not be investigated for affricates because the affricate [tʃ] represents 98% of cases of affricates followed by phonologically voiceless vowels. Closure-related measurements are based on word-medial and word-final affricates only ($n = 232$).

As shown in table 6, tokens with fully reduced vowels demonstrate significant lengthening of releases (80 ms difference), lower intensity of releases (2.4 dB difference), and fewer voicing cycles during releases (7.1% difference). Closure duration and voicing during closure are not affected, which fully parallels the findings for plosives.

4.3.3 Fricatives

Figure 7 shows examples of fricatives followed by a voiced vowel versus a fully reduced vowel. Notice that the fricative is longer and has lower amplitude in figure 7b. There is also no voice bar or formant structure in the spectrogram, and there is no periodicity in the waveform.

Table 7 lists the means of frication duration, intensity and the voicing ratio of fricatives not followed versus followed by fully reduced vowels ($n = 151$). The means are based on the data from all 8 speakers. Significance of the differences was tested in separate paired-samples *t* tests for each acoustic measurement. The potential role of

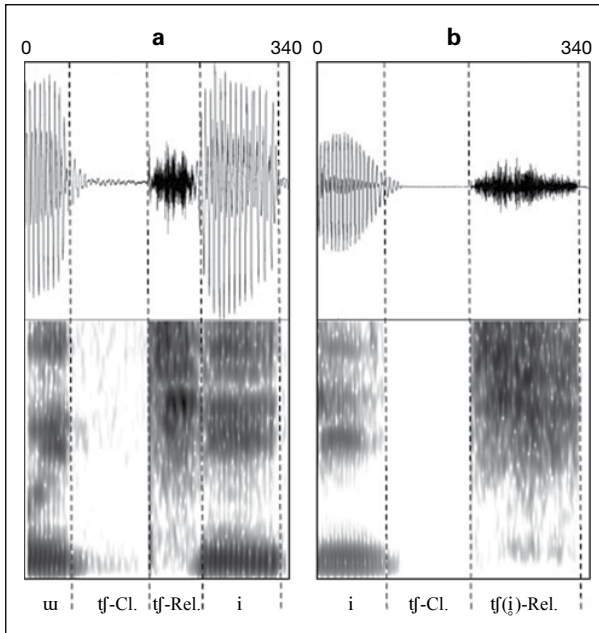


Fig. 6. Affricates followed by voiced vowel (**a**) vs. a fully reduced vowel (**b**). Data from speaker F3.

Table 6. Affricates not followed versus followed by fully reduced vowels

Type	Measurement		Reduced V		Difference	t	p
			N	Y			
Affricates	duration, ms	closure	83 (1.4)	84 (1.7)	-1	0.32	0.755
		release	77 (9.8)	157 (22.3)	-80	9.77	<0.001
	intensity, dB	release	61.8 (3.6)	59.4 (3.7)	2.4	4.51	0.003
voicing ratio, %	closure		50.6 (12.3)	47.7 (16.2)	2.9	1.15	0.287
		release	15.9 (3.9)	8.8 (6.2)	7.1	4.16	0.004

Means and standard deviations (in parentheses) of closure duration, release duration, release intensity, closure voicing ratio, and release voicing ratio. Statistically significant differences are in italics.

consonantal place of articulation could not be investigated for fricatives because of the insufficient numbers of tokens for some speakers/categories.

The data in table 7 demonstrate that, as in the case of plosives and affricates, fricatives preceding fully reduced vowels show significant lengthening (56 ms difference), lower intensity (6.6 dB difference), and fewer voicing cycles (32.3% difference). This suggests that all obstruents become longer, less loud and more voiceless when they are followed by a fully reduced vowel.

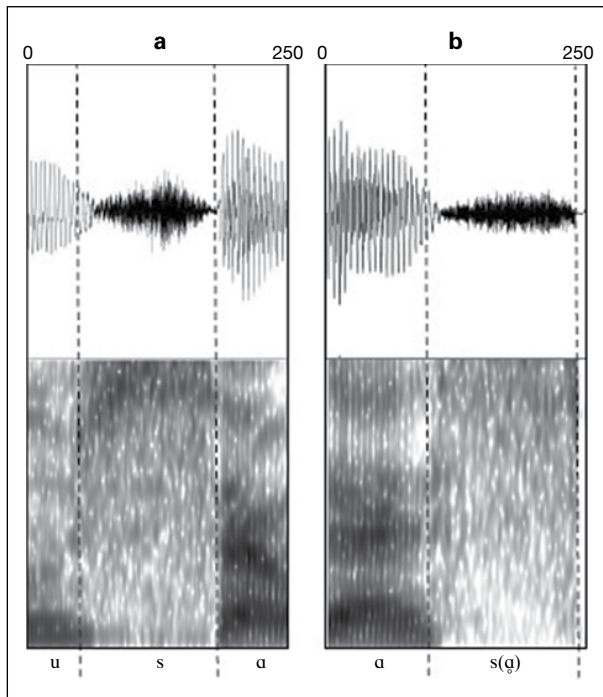


Fig. 7. Fricatives followed by a voiced vowel (**a**) vs. a fully reduced vowel (**b**). Data from speaker F3.

Table 7. Fricatives not followed versus followed by fully reduced vowels

Type	Measurement	Reduced V		Difference	t	p
		N	Y			
Fricatives	duration, ms	108 (32.9)	164 (18.9)	-56	5.24	0.001
	intensity, dB	57.4 (5.6)	50.8 (5.2)	6.6	2.73	0.029
	voicing ratio, %	65.6 (12.3)	33.3 (12.5)	32.3	5.15	0.001

Means and standard deviations (in parentheses) of frication duration, intensity, and the voicing ratio. All differences are statistically significant.

4.3.4 Nasals

Figure 8 demonstrates the surface forms of nasals followed by a voiced vowel versus a fully reduced vowel. Notice that the nasal is longer and has lower amplitude in figure 8b than 8a.

Table 8 gives the means of segmental duration, intensity and the voicing ratio across nasals not followed versus followed by fully reduced vowels ($n = 94$). Significance of the observed differences was tested in separate repeated measures ANOVAs for each acoustic measurement. Independent predictors included (i) adjacency to a fully reduced vowel (no, yes) and (ii) consonantal place of articulation (labial, coronal). The 2×2

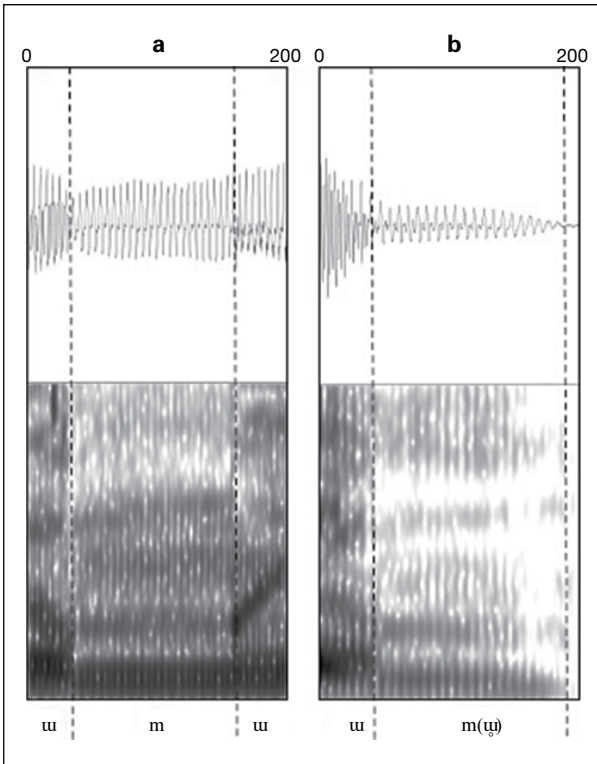


Fig. 8. Nasals followed by a voiced vowel (**a**) vs. a fully reduced vowel (**b**). Data from speaker F3.

Table 8. Nasals not followed versus followed by fully reduced vowels

Type	Measurement	Reduced V		Difference	F	p
		N	Y			
Nasals	duration, ms	<i>93 (9.8)</i>	<i>127 (10.8)</i>	<i>-34</i>	<i>26.01</i>	<i>0.002</i>
	intensity, dB	<i>57.5 (5.7)</i>	<i>51.5 (4.7)</i>	<i>6.0</i>	<i>52.90</i>	<i><0.001</i>
	voicing ratio, %	94.8 (2.7)	87.1 (10.3)	7.7	4.60	0.076

Means and standard deviations (in parentheses) of segmental duration, intensity, and the voicing ratio. Statistically significant differences are in italics.

ANOVAs were run on the data of 7 (out of 8) speakers who produced sufficient numbers of tokens in all categories. (Speaker M2 did not produce any coronal nasals followed by fully reduced vowels.)

As reflected in table 8, nasals showed significantly longer durations (34 ms difference) and lower intensity (6.0 dB difference) when followed by a fully reduced vowel. The difference in the voicing ratio went in the expected direction (7.7% less voicing

compared to nasals followed by other segments) but it was not statistically significant. These results largely parallel the findings for obstruents.

Nasal consonants did not show any significant differences across labial and coronal places of articulation for any of the acoustic parameters ($F < 3$, $p > 0.1$), and no significant interactions were found between consonantal place of articulation and adjacency to a fully reduced vowel ($F < 3.6$, $p > 0.1$). This indicates that the differences shown in table 8 are representative of both labial and coronal nasals.

4.3.5 Flaps

Examples and results for flaps are provided in figure 9 and table 9. Figure 9 shows flaps followed by a voiced vowel versus a fully reduced vowel. Note the longer duration, lower amplitude and greater voicelessness of the flap in figure 9b, which resembles a fricativized trill more than a prototypical flap.

Table 9 compares the durations, intensity values and the voicing ratios of the flaps not followed versus followed by fully reduced vowels ($n = 51$). Significance of the differences was tested in separate paired-samples *t* tests for each acoustic measurement, using the data from all 8 speakers.

The values listed in table 9 reveal that, when followed by a fully reduced vowel, coronal flaps show significant lengthening (75 ms difference), lower intensity (6.4 dB difference), and fewer voicing cycles during the flap (53.3% difference). This is fully in line with the effects observed for obstruents and nasals.

4.3.6 Glides

Figure 10 shows glides followed by a voiced vowel versus a fully reduced vowel. Notice that the glide [w] has lower amplitude, irregular voicing, and a less clear spectral pattern in figure 10b than 10a.

Table 10 compares the durations, intensity values and the voicing ratios of the two categories of glides. Only descriptive statistics are provided here due to the small number of tokens in the fully reduced vowel category ($n = 5$). As can be seen in table 10, glides followed by fully reduced vowels show longer durations (35 ms difference), lower intensity (6.4 dB difference), and fewer voicing cycles during the glide (46.6% difference). These results appear to be consistent with the patterns observed for other types of consonants; however, more tokens would be needed to determine if these effects are statistically significant.

5 Summary and Discussion

The results presented above show that the devoiced vowels of Southern Ute are realized phonetically in three different ways. The most common type of phonetic realization (81.71%) is a fully reduced vowel with concurrent lengthening, decreased intensity and/or increased voicelessness of the preceding consonant. The second most common type (16.87%) is a partially devoiced vowel with visible formant structure and a voicing bar. The least common type (1.42%) is a fully voiceless vowel that has only limited formant structure and no voicing bar in the spectrogram. This demonstrates that devoiced vowels are seldom realized as prototypical ‘whispered vowels’. However, even without the presence of a vowel-like segment in phonetic outputs, the acoustic signal still contains cues to their underlying syllabic structure.

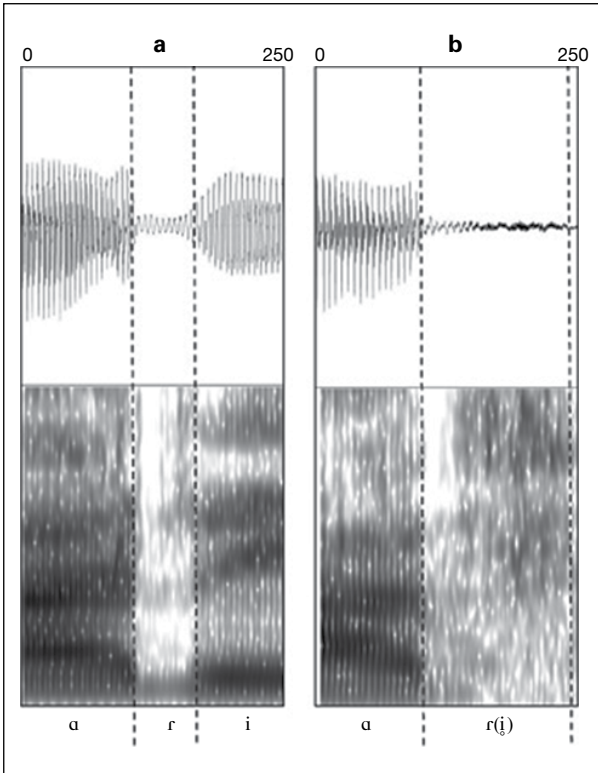


Fig. 9. Flaps followed by a voiced vowel (**a**) vs. a fully reduced vowel (**b**). Data from speaker F3.

Table 9. Flaps not followed versus followed by fully reduced vowels

Type	Measurement	Reduced V		Difference	t	p
		N	Y			
Flaps	duration, ms	65 (13.5)	140 (47.5)	-75	16.10	0.005
	intensity, dB	58.9 (7.3)	52.5 (3.7)	6.4	8.61	0.022
	voicing ratio, %	94.2 (5.7)	40.9 (22.1)	53.3	5.79	0.001

Means and standard deviations (in parentheses) of segmental duration, intensity, and the voicing ratio. All differences are statistically significant.

The varied realizations of devoiced vowels in Southern Ute are highly similar to the phonetic forms observed in such languages with vowel devoicing as Montreal French, Korean, Japanese, and Cypriot Greek (Cedergren and Simoneau, 1985; Jun and Beckman, 1993; Beckman and Shoji, 1994; Beckman, 1996; Tsuchida, 1997, 2001; Ogasawara, 2007; Eftychiou, 2010). Cedergren and Simoneau (1985), for example, reported that the devoiced vowels of Montreal French can be fully voiceless, partially

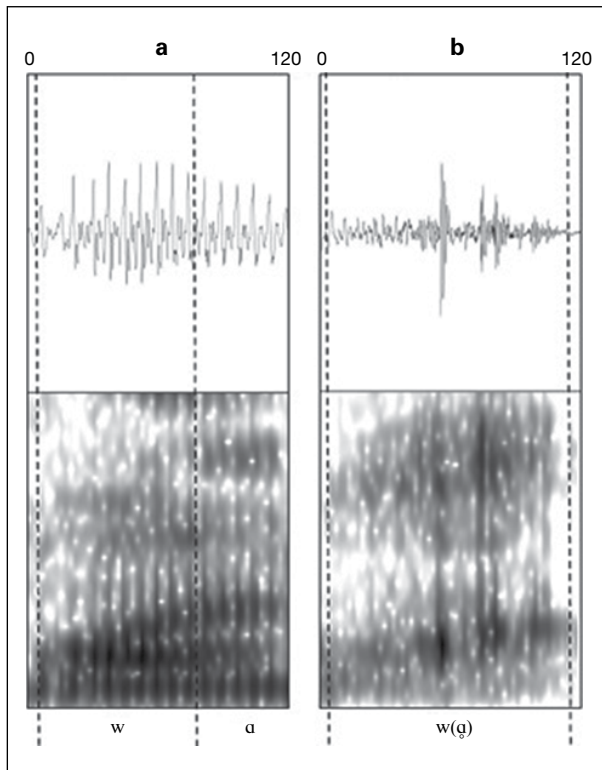


Fig. 10. Glides followed by a voiced vowel (**a**) vs. a fully reduced vowel (**b**). Data from speaker F4.

Table 10. Glides not followed versus followed by fully reduced vowels

Type	Measurement	Reduced V		Difference
		N	Y	
Glides	duration, ms	75	110	-35
	intensity, dB	66.7	60.3	6.4
	voicing ratio, %	93.4	46.8	46.6

Means of segmental duration, intensity, and the voicing ratio.

devoiced, and deleted. Jun and Beckman (1993) found that Korean speakers produce devoiced vowels as fully voiceless, partially devoiced, and voiced. Similar variability and gradation in devoicing has also been attested in the case of vowel devoicing in Japanese (Beckman, 1996; Tsuchida, 1997, 2001; Ogasawara, 2007) and Cypriot Greek (Eftychiou, 2010). It is also in line with the findings for vowel reduction in languages such as English and German where acoustic outputs also appear to contain variable remnants of reduced vowels (Jun and Beckman, 1993; Jannedy, 1994; Beckman, 1996).

The observed modification of consonants adjacent to fully reduced vowels is also consistent with the crosslinguistic findings on the effect of vowel elision on the preceding consonant, when the acoustic properties of consonantal segments signal the presence of vowels at the phonological level (Ostreicher and Sharf, 1976; Martin, 1987; Davidson, 2006; Mo, 2007; Ogasawara, 2007; Eftychiou, 2010). For example, Davidson (2006) and Mo (2007) show that in languages such as English and Korean, syllables with devoiced or elided vowels have the same or longer durations than identical syllables with voiced vowels. Southern Ute shows a similar effect on the duration of the preceding consonant, with the segment becoming longer when the devoiced vowel is not pronounced as a separate segment. The consonant also becomes less loud and more voiceless compared to how the same segment is realized when not followed by a devoiced vowel. Such noncategorical behavior can be attributed to gestural overlap between the gesture of the vowel and the gesture of the adjacent consonant that can result in one of the two segments being seemingly deleted in phonetic outputs (Browman and Goldstein, 1990; Kröger, 1993; Beckman and Shoji, 1994; Beckman, 1996).

The modified pronunciation of adjacent consonants that is seen in Southern Ute likely ensures the perceptibility of the vowel even in the absence of a vowel-like segment at the acoustic level. Listeners are known to be very good at identifying plosive-vowel combinations even when they have very little acoustic information, such as when listening to the stimuli containing a single glottal pulse (Blumstein and Stevens, 1980). Southern Ute speakers can therefore be expected to benefit even from limited and indirect cues to the presence of devoiced vowels, such as durational differences between outputs with and without devoiced vowels. This seems especially plausible considering that the language has unaspirated plosives, which likely makes lengthened releases perceptually salient for the speakers. Furthermore, although formant values could not be measured reliably for the devoiced vowels, visual inspection of the spectrograms revealed that the second formant transitions associated with the consonant may also signal the identity of devoiced vowels. Thus, phonetic outputs likely contain even more acoustic cues to their syllabic structure, which can be expected to further facilitate correct decoding of the word forms during lexical access.

Both direct and indirect preservation of devoiced vowels at the phonetic level is essential for Southern Ute as it helps satisfy the phonotactic constraints of the language. Southern Ute has highly restrictive CV(V)(C) syllable structure (Givon, 2013). If fully reduced vowels were no longer being pronounced, Southern Ute speakers would encounter and produce words with previously unattested CC sequences, such as [pkɥ] ‘horse’ (cf. [pɥkɥ]). As a result, the language would be reanalyzed as allowing complex onsets and codas.

Since all vowels, not just high vowels, are devoiced in Southern Ute and not all of the devoiced vowels are adjacent to voiceless consonants, our data also make an important contribution to the question of whether or not vowel devoicing is restricted to high/nonlow vowels and requires adjacency to a voiceless segment (among others, Eftychiou, 2010). Namely, devoiced vowels have been attested in a variety of languages from different families, including Awadhi, Campa, Chatino, Cypriot Greek, Dagur, Huichol, Japanese, Korean, Montreal French, Serbo-Croatian, Tunica, and Uzbek (Greenberg, 1969; Jun and Beckman, 1993; Beckman and Shoji, 1994; Beckman, 1996; Tsuchida, 1997, 2001; Gordon, 1998; Imai, 2004; Mo, 2007; Ogasawara, 2007). In the majority of these languages, devoicing usually targets only high vowels adjacent

to voiceless consonants and almost never affects low vowels (Greenberg, 1969; Jaeger, 1978; Gordon, 1998). The apparent effect of vowel height is traditionally attributed to the higher oral pressure associated with the production of close vowels that interferes with the pressure drop across the glottis needed for the production of voicing (Ohala, 1975), and the effect of adjacency to a voiceless consonant has been described as the result of the glottal closing gesture of the vowel overlapping with or being delayed by the glottal opening gesture for the voiceless consonant (Jun and Beckman, 1993; Beckman and Shoji, 1994; Beckman, 1996; Tsuchida, 2001; Imai, 2004; Mo, 2007; Ogasawara, 2007).

In the data we present here, vowel height is not found to be a significant feature in the devoicing process, which shows that high, mid and low vowels can all demonstrate similar patterns of devoicing behavior. One likely explanation for this is that devoicing in Southern Ute may have started as the result of gestural overlap between high vowels and adjacent voiceless consonants but it has since been lexicalized, with devoicing being extended to both high and nonhigh vowels alike. Additional support for such lexicalization comes from the fact that, as noted in the section on the Southern Ute language, the devoicing environment is not always predictable and devoicing can even occur when vowels are followed by voiced segments (150 such cases in our dataset). Our results also show that, in Southern Ute, devoicing behavior is the same for different types of consonants. This contrasts with the findings for Korean, a language with greater rates of devoicing observed with fricatives than plosives (Mo, 2007). This suggests that vowel height and adjacency to voiceless consonants are not the only triggers for vowel devoicing in languages such as Southern Ute where it seems to no longer be a purely phonetically motivated phenomenon.

Finally, our data also show that vowels located next to a strong prosodic boundary can undergo devoicing, which is different from what has been reported for Japanese where devoiced vowels are unlikely to occur before a pause (Imai 2004). We attribute this difference to the fact that, in Southern Ute, devoicing of word-final vowels can play a grammatical role (e.g., subject-marking), which appears to override the effects of a prosodic boundary.

6 Conclusion

Southern Ute, a Native American language spoken in Southwestern Colorado, has long been known for the presence of devoiced vowels in its sound inventory. However, very little is known about the way devoiced vowels are realized at the phonetic level, including the question of whether or not such segments are actually being produced by the speakers. The current study shows that there are three different types of phonetic realizations of devoiced vowels: partially devoiced, fully voiceless, and fully reduced with concurrent lengthening, lower amplitude, and decreased voicing of the preceding consonant. This confirms the continued presence of devoiced vowels in the language. The findings also hold true for all vowels regardless of their height and all consonants regardless of their manner of articulation, which suggests that devoicing is highly lexicalized in Southern Ute.

References

- Armagost J (1985): On predicting voiceless vowels in Comanche. *Kansas Working Papers Linguist: Stud Native Am Lang* IV:1–15.
- Armagost J, Miller W (2000): Are voiceless vowels phonemic in Comanche? In Casad E, Willet T (eds): *Uto-Zatecan: Structural, Temporal, and Geographic Perspectives*. Hermosillo, Universidad de Sonora, pp 57–82.
- Beckman ME (1996): When is a syllable not a syllable? In Otake T, Cutler A (eds): *Phonological Structure and Language Processing: Cross-Linguistic Studies*. Berlin, Mouton de Gruyter.
- Beckman ME, Shoji A (1994): Spectral and perceptual evidence for CV coarticulation in Japanese devoiced /si/ and /syu/. *Phonetica* 41:61–71.
- Blumstein SE, Steven KN (1980): Perceptual invariance and onset spectra for stop consonants in different vowel environments. *J Acoust Soc Am* 67:648–662.
- Boersma P, Weenink D (2014): Praat: doing phonetics by computer (computer program); version 5.3.59. <http://www.praat.org/> (accessed November 25, 2013).
- Browman CP, Goldstein L (1990): Gestural specification using dynamically defined articulatory structures. *J Phon* 18:299–320.
- Canonge E (1957): Voiceless vowels in Comanche. *Int J Am Linguist* 23:63–67.
- Cedergren HJ, Simoneau L (1985): La chute des voyelles hautes en Français de Montreal: as-tu entendu la belle syncope? In Lemieux M, Cedergren HJ (eds): *Les tendances dynamiques de Français parlé à Montreal*. Montreal, Office de la Langue Française, pp 57–144.
- Charney J (1993): *A Grammar of Comanche*. Lincoln, University of Nebraska Press.
- Charney J (1996): *Ute Dictionary*. Ignacio, Southern Ute Tribe.
- Davidson L (2006): Schwa elision in fast speech: segmental deletion or gestural overlap? *Phonetica* 63:79–112.
- Efthychiou E (2010): Routes to lenition: an acoustic study. *PLoS One* 5:1–13.
- Givon T (1979): *Ute Dictionary*. Ignacio, Southern Ute Tribe.
- Givon T (1980): *Ute Reference Grammar*. Ignacio, Southern Ute Tribe.
- Givon T (2013): *Ute Reference Grammar*. Philadelphia, Benjamins.
- Gordon M (1998): The phonetics and phonology of non-modal vowels: a cross-linguistic perspective. *Berkeley Linguist Soc* 24:93–105.
- Goss J (1972): *Ute Lexical and Phonological Patterns*; PhD diss University of Chicago.
- Greenberg J (1969): Some methods of dynamic comparison in linguistics; in Puhvel J (ed): *Substance and Structure of Languages*. Los Angeles, Center for Research in Languages and Linguistics, pp 147–204.
- Harms R (1966): Stress, voice, and length in Southern Paiute. *Int J Am Linguist* 32:228–235.
- Imai T (2004): *Vowel Devoicing in Tokyo Japanese: A Variationist Approach*; PhD diss Michigan State University.
- Jaeger A (1978): Speech aerodynamics and phonological universals; in Proc 4th Annu Meet Berkeley Linguist Soc, pp 311–329.
- Jannedy S (1994): Rate effects on German unstressed syllables. *Working Papers Linguist, Ohio State University* 44:105–124.
- Jun S-A, Beckman ME (1993): A gestural-overlap analysis of vowel devoicing in Japanese and Korean. 1993 Annu Meet Linguist Soc Am, Los Angeles.
- Kröger B (1993): A gestural production model and its application to reduction in German. *Phonetica* 501:213–233.
- Ladefoged P, Maddieson I (1996): *Sounds of the World's Languages*. Malden, Wiley Blackwell.
- Lehiste I, Peterson G (1959): Vowel amplitude and phonemic stress in American English. *J Acoust Soc Am* 31:428–435.
- Lewis PM (ed) (2009): *Ethnologue: Languages of the World*, ed 16. Dallas, SIL International.
- Manaster-Ramer A (1986): Genesis of Hopi tones. *Int J Am Linguist* 52:154–160.
- Martin SE (1987): *The Japanese Language through Time*. New Haven, Yale University Press.
- Miyashita M (2000): Less stress, less pressure, less voice. Proc Southwest Workshop on Optimality Theory Conf 4, pp 43–55.
- Mo Y (2007): Temporal, spectral evidence of devoiced vowels in Korean. Proc 16th Int Congr Phon Sci, ID 1597, pp 445–448.
- Oberly S (2008): *A Phonetic Analysis of Southern Ute with a Discussion of Southern Ute Language Policies and Revitalization*; PhD diss University of Arizona.
- Ogasawara N (2007): *Processing of Speech Variability: Vowel Reduction in Japanese*; PhD diss University of Arizona.
- Ohala JJ (1975): A mathematical model of speech aerodynamics; in Fant G (ed): *Speech Communication: Speech Production and Synthesis by Rule*, 2. Stockholm, Almqvist & Wiksell, pp 65–72.
- Ostreicher HJ, Sharf DJ (1976): Effects of coarticulation of the identification of deleted consonant and vowels. *J Phon* 4:285–301.
- Press M (1979): *Chemehuevi: A Grammar and Lexicon*. Berkeley, University of California Press.
- Sapir E (1930): The Southern Paiute Language. *Proc Am Acad Arts Sci* 65:1–296.
- Tsuchida A (1997): *Phonetics and Phonology of Japanese Vowel Devoicing*; PhD diss Cornell University, Ithaca.
- Tsuchida A (2001): Japanese vowel devoicing: cases of consecutive devoicing environments. *J East Asian Linguist* 10:225–245.
- Whorf B (1934): *Hopi Verb Classes*. Ms No 1570. Philadelphia, Library of the American Philosophical Society.