RATE-SENSITIVE DIFFERENCES BETWEEN MODAL AND NON-MODAL VOWELS IN SAN MARTIN PERAS MIXTEC

Ben Eischens

University of California, Los Angeles
beischens@ucla.edu

ABSTRACT

The difference between laryngeal consonants and non-modal phonation is often difficult to determine based solely on phonetic data. This pilot study demonstrates a novel source of phonetic evidence for an analysis of [Vʔ] sequences in San Martín Peras Mixtec as non-modal vowels that are phased as modal-then-glottalized. In slow speech, the non-modal portion of these vowels lengthens significantly more than the portion made up of modal voicing. This pattern differs from sequences of a modal vowel and a consonant, supporting an analysis of [Vʔ] sequences as glottalized vowels. The study also investigates the behavior of [h], a sound whose (supra-)segmental status is not as clear, finding inter-speaker variation that merits further study.

Keywords: Mixtec, speech rate, non-modal phonation

1. INTRODUCTION

Cross-linguistically, the representational distinction between laryngeal consonants and non-modal phonation is not always apparent from investigations of their acoustic correlates [1]. For example, Davidson [2] found that the phonemic glottal stop consonant in Hawaiian is typically realized as a period of creaky voice, especially intervocalically, similarly to the phonetic realization of contrastively creaky vowels in Jalapa Mazatec [3]. Additionally, laryngealized vowels in many languages are often associated with glottal closure [4, 5, 6]. Because of this, analyses of whether laryngeal gestures are best analyzed as consonants or instantiations of phonation types are usually made based on phonological evidence. This is the case in Mixtec languages, which are commonly analyzed as contrasting modal and 'glottalized' vowels [7, 8, 9, 10, 11, 12, 13], though see also [14, 15, 16]. In these languages, glottalized vowels are usually realized as a sequence of modal voicing followed by creaky voicing or glottal closure. This contrast is found in San Martín Peras Mixtec (SMP Mixtec; ISO: jmx; glottocode: west2643), where the motivation for an analysis of [Vʔ] sequences as glottalized vowels comes from phonotactic evidence [17]. The present study analyzes the characteristics of modal and glottalized vowels across speech rates and provides additional support for this analysis.

In SMP Mixtec, the proportion of a glottalized vowel that is made up of modal voicing decreases in slow speech. This happens regardless of whether the vowel is followed by a consonant or by another vowel. For example, in both words in Figure 1, the period of glottalization increases in duration disproportionately to the preceding modal voicing.

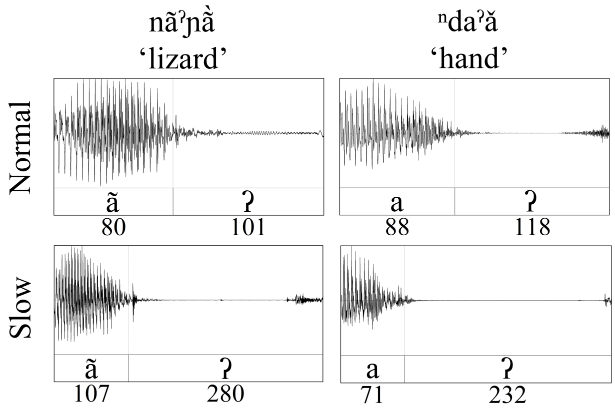

Figure 1: Waveforms and duration (ms) of the glottalized [a^ʔ] vowel in [nã^{ʔ}ɲà̃] ('lizard') and [^{n}da^{ʔ}ǎ] ('hand') at normal and slow speech rates

Crucially, it is not the case that modal vowels never lengthen much in slow speech. The first vowel in the words in Figure 2 lengthens proportionately to the following consonant in slow speech.

The lengthening pattern in Figure 1 has implications for the decision to analyze phonetic [Vʔ] sequences as glottalized vowels or as sequences of a vowel and a consonant. If they constitute glottalized vowels, the disproportionate reduction in duration of the modal vowel in slow speech in Figure 1 can be said to result from a production strategy in which the glottalized portion of a vowel is disproportionately lengthened in slow speech. However, an analysis of [ʔ] as a consonant has trouble accounting for the pattern

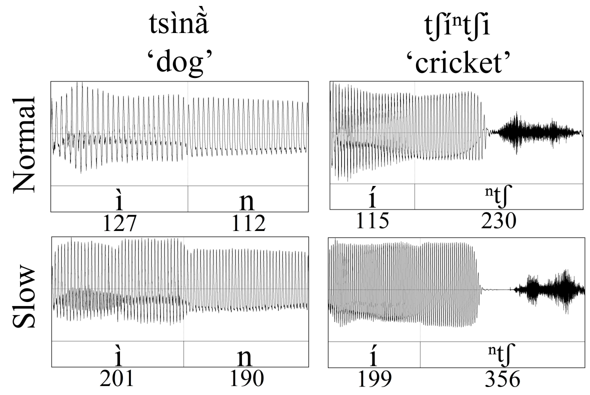

Figure 2: Waveforms and duration (ms) of the modal vowel [i] and the following consonant from [tsìnã̀] ('dog') and [tʃí^{n}tʃi] ('cricket') at normal and slow speech rates

because of how [ʔ] would be syllabified: It should be syllabified as a coda in [nã^{ʔ}ɲã̀] ('lizard') and as an onset in [^{n}da^{ʔ}ă] ('hand'). Importantly, there is no evidence of exceptional coda syllabification in CVCV words, and [ʔC] is not a possible onset.

When syllabified as a coda in [nã^{ʔ}ɲã̀] ('lizard'), the disproportionate lengthening of [ʔ] can be understood as lengthening of the initial rime via selective lengthening of the coda (see [18] for similar asymmetries in Cantonese, and [19, 20] for onset-coda lengthening asymmetries in English and Dutch). However, when syllabified as an onset in [^{n}da^{ʔ}ă] ('hand'), the [ʔ] no longer forms a rime with the preceding vowel. The preceding vowel nucleus should lengthen more than in [nã^{ʔ}ɲã̀] ('lizard') because the coda is no longer available to be lengthened. However, this prediction is not borne out—the modal vowel decreases disproportionately to the following [ʔ] in slow speech in both cases. The fact that [ʔ] lengthens uniformly relative to the preceding vowel regardless of whether it would be a coda or an onset points toward an analysis of [Vʔ] sequences in SMP Mixtec as glottalized vowels, and away from an analysis of [ʔ] as a consonant.

This paper reports the results of a pilot study demonstrating this lengthening pattern for glottalized vowels. It also investigates the lengthening patterns of words with contrastive and non-contrastive [h] (referred to as breathiness and pre-aspiration throughout, respectively), which share many phonotactic properties with [ʔ] but cannot be as conclusively analyzed as non-modal phonation on the basis of phonotactic evidence. Ultimately, the findings of this study provide convergent evidence in favor of an analysis of [Vʔ] sequences as glottalized vowels in SMP Mixtec, and they do so through a novel method of speech rate manipulation. Though the results must still be replicated with a larger participant pool, they present a novel method for probing representational distinctions between laryngeal consonants and non-modal phonation that may also prove useful outside of SMP Mixtec.

2. LANGUAGE BACKGROUND

SMP Mixtec is an Otomanguean language spoken by about 10,000 people in western Oaxaca, Mexico [21] and by diaspora communities throughout Mexico and the US. As is the case across Mixtec [13, 22], it has a complex tonal system [17, 23], and roots are minimally (and usually maximally) bimoraic. There are no coda consonants. The first vowel in a root may be modal (V), glottalized (V^{ʔ}), or followed by [h] (V^{h}). [h] predictably precedes all root-medial, voiceless consonants, but is contrastive before vowels and voiced consonants. Examples of roots of these types can be seen in Table 1.

3. METHODS

3.1. Participants

Two middle-aged, female speakers of SMP Mixtec who live and work in Watsonville, CA participated separately in this production task. Both grew up in the municipality of San Martín Peras, are bilingual in SMP Mixtec and Spanish, and speak Mixtec on a daily basis. The task was administered in Spanish.

3.2. Materials and procedure

Each participant (henceforth P1 and P2) produced target words in the carrier sentence in (1). They were asked to produce the first utterance at a normal rate, the second more slowly, and the third very slowly.

(1) kã̂^{ʔ}=ĩ̀ __ βi^{h}tsĩ
POT.say=1SG __ now
'I will say __ now.'

To avoid priming effects, and because the language currently has no standardized orthography, the prompt sentences were read out loud in Spanish by the author. If a participant produced a target word other than the desired one, the author then specified the desired word in SMP Mixtec, and the participant produced the three utterances using that word. Target words varied in whether the initial vowel was glottalized, breathy, modal, or was followed by a pre-nasalized or pre-aspirated consonant (Table 1). Glottalized and breathy words varied in whether or not they included a root-medial consonant, with a roughly equal number of items in each condition.

				V1+P-V1		
				V1	P-V1	
Glottal	C2	nã^{ʔ}ñã̀ 'lizard'	n	ã	ʔ	ñã̀
	No C2	^{n}da^{ʔ}ă 'hand'	^{n}d	a	ʔ	ă
Breathy	C2	nũ̀^{h}nĩ̌ 'corn'	n	ũ̀	h	nĩ̌
	No C2	nĩ̀^{h}ĩ̌ 'blood'	n	ĩ̀	h	ĩ̌
Preasp		tʃú^{h}tu 'cat'	tʃ	ú	h	tu
Prenas		tʃí^{n}tʃi 'cricket'	tʃ	í	n	tʃi
Modal		tsìnã̀ 'dog'	ts	ì	n	ã̀

Table 1: Relevant portions of target words by condition. 'P-V1' stands for 'Post-V1.'

Words with a medial, pre-nasalized consonant were included for purposes of comparison because pre-nasalization occurs in the same local context as glottalization and breathiness but is unambiguously a feature of the medial consonant [17].

P1 produced the sentence in (1) at three different speech rates for 151 target words (41 glottalized + 40 breathy + 21 modal + 28 pre-aspirated + 21 pre-nasalized) for a total of 453 productions. P2 produced this three-sentence sequence for 153 target words (42 glottalized + 43 breathy + 22 modal + 22 pre-aspirated + 22 pre-nasalized) for a total of 459 productions. Few target words were prompted more than once, except in the pre-nasalized condition because pre-nasalization of root-medial consonants is rather uncommon in the language.

3.3. Data analysis

Each token was spliced and annotated by hand in Praat [24]. A Praat script extracted the total duration of each carrier sentence and each segment in the target words. Sentence duration spanned from the utterance-initial stop burst to the end of periodic voicing on the utterance-final vowel. In the target word, vowel duration was measured from the offset of the preceding consonant until either (1) the onset of the following consonant, or (2) the point at which periodic voicing ceased and silence or aperiodic noise began. When present, consonant off-glides were included in the duration of the following vowel because many such vowels never reached a steady state. [ʔ] duration was measured from the end of periodic vocal fold vibration on the preceding vowel until the onset of periodic vocal fold vibration in the following consonant or vowel. [h] duration was measured from the onset of frication noise, excluding any periodic, breathy voicing, (c.f. [25]) to the onset of the following consonant or vowel. The relevant portions of the tokens whose duration was measured are given in Table 1.

The dependent variable was the ratio of V1 to V1 + Post-V1, which is the ratio of the modal vowel's duration to the duration of the modal vowel plus whatever immediately follows it in the same word. Raw duration was not used because the phenomenon of interest is how much the modal vowel lengthens in slow speech relative to what follows, whether that is [ʔ], [h], the [n] of pre-nasalization, or a full consonant in the case of the modal words. To satisfy the linear regression's assumption of unbounded variables, each token's ratio of V1 to V1 + Post-V1, which was necessarily bounded at [0, 1], was centered and scaled. To obtain a continuous measure of speech rate, rate was quantified as moras per second (μ/sec) for each token's carrier sentence, all of which had 6 moras. P1's average μ/sec values for the three prompted rates from normal to slowest were 4.19 (SE = .04), 2.77 (SE = .03), and 1.79 (SE = .02). P2's rate change was smaller, with average μ/sec values for the three prompted rates being 4.66 (SE = .01), 3.7 (SE = .03), and 3.41 (SE = .03). Because μ/sec is bounded at 0, each token's value was centered and scaled for the linear regression.

3.4. Statistical models

All statistical tests were carried out in R [26]. The dependent variable was the centered and scaled value for each token's ratio of V1 to V1 + Post-V1, and the independent variables were speech rate (centered and scaled μ/sec), phonation type, and their interaction. A random effect of item was also included. To establish a baseline for comparison, Pearson's correlation coefficient was calculated to check for correlation between speech rate (μ/sec) and the ratio of V1 to V1 + Post-V1 for roots with a medial, pre-nasalized consonant. The test did not reach significance for either participant (P1: t = -.55, p = .58; P2: t = -1.32, p = .19). Because this ratio did not reliably correlate to speech rate, the pre-nasalized condition was used as the baseline for each participant. A linear mixed effects model was run using the lme4() package [27]. Because there were only two participants in this task, separate statistical models were run for each participant, with the results subjected to Bonferroni correction for multiple comparisons (α=0.05/2).

4. RESULTS

The models' results for P1 and P2 are given in Table 2. The residuals for each model were normally distributed (R = .991 for P1, .973 for P2). Model criticism was carried out using the drop1() function in the lmerTest package [28]. The full models were compared with simpler models omitting the phonation by μ/sec interaction. These comparisons were both significant ($p < .001$) using Satterthwaite's method, indicating that this interaction should not be excluded from the models. No further model simplification was possible.

Both models found negative main effects of glottalized and breathy phonation type, meaning the ratio of V1 to V1 + Post-V1 was lower across the board in these conditions than in pre-nasalized words. There was no main effect of pre-aspiration or modal phonation for either participant. For P1, there were significant positive interactions between rate (μ/sec) and the ratio of V1 to V1 + Post-V1 for glottalized and breathy words. This interaction held regardless of whether or not there was a medial consonant, and also for pre-aspirated words, which always have a medial consonant. This means that the modal portion of the vowel decreased disproportionately relative to [ʔ] or [h] as rate slowed down. This interaction was not seen for modal words, suggesting that the ratio of V1 to V1 + Post-V1 is consistent across speech rates for modal words for P1.

For P2, the positive interaction between rate and pre-aspirated roots was significant, and the interaction of rate with glottalized phonation was significant regardless of whether or not there was a medial consonant. However, the interaction between rate and breathy phonation was significant only for words with a medial consonant. That is, modal vowel duration decreased disproportionately relative to [ʔ] as rate slowed down whether or not there was a following consonant, but the modal vowel only decreased disproportionately relative to [h] when there was a following consonant. There was no interaction between rate and modal phonation.

5. DISCUSSION

For both participants, modal vowels are shorter before [ʔ] than in the baseline case, and their proportion relative to [ʔ] shrinks as speech rate slows regardless of whether or not there is a consonant following [ʔ]. As discussed in §1, this finding is straightforwardly explained by an analysis of [Vʔ] sequences in SMP Mixtec as glottalized vowels, but not by an analysis of [ʔ] as a consonant.

Predictor	β (SE) P1	β (SE) P2	|t| P1	|t| P2	p-value P1	p-value P2
Intercept	.9 (.18)	.93 (.15)	5.1	6.1	***	
μ/sec	-.04 (.05)	-.07 (.04)	-.84	-1.7	.4	.09
Glottal (w/ C2)	-1.28 (.2)	-1.62 (.19)	-6.3	-8.7	***	
Glottal (no C2)	-1.99 (.2)	-2.05 (.19)	-9.8	-11.06	***	
Breathy (w/ C2)	-1.31 (.21)	-1.25 (.2)	-6.2	-6.26	***	
Breathy (no C2)	-1.91 (.2)	-1.58 (.19)	-9.41	-8.45	***	
Preasp	-.09 (.2)	.11 (.15)	-.49	.72	.63	.47
Modal	-.17 (.2)	.003 (.18)	-.82	.02	.41	.98
μ/sec x Glottal (w/ C2)	.56 (.07)	.33 (.06)	7.54	5.54	***	
μ/sec x Glottal (no C2)	.45 (.08)	.26 (.06)	5.92	3.96	***	
μ/sec x Breathy (w/ C2)	.21 (.08)	.29 (.06)	2.73	4.95	*	***
μ/sec x Breathy (no C2)	.3 (.08)	.09 (.06)	3.89	1.58	***	.12
μ/sec x Pre-asp	.36 (.07)	.28 (.06)	5.19	4.46	***	
μ/sec x Modal	.06 (.07)	.03 (.06)	.77	.55	.44	.59

Table 2: Results of the mixed effects models with Bonferroni-corrected α codes: * = < .025, ** < .005, *** = < .0005

The data regarding [h] are less clear. For P1, [h] behaves the same as [ʔ]. However, as speech rate slows for P2, [h] lengthens disproportionately only when it would be syllabified as a coda, not an onset. This is consistent with the consonantal analysis. Interestingly, the phonotactic evidence for [h] as the realization of a phonation type in SMP Mixtec is not quite as strong as it is for [ʔ], despite their phonotactic similarity: [ʔ] is never root-initial, but one lexeme, the demonstrative [hã̰ã́] ('that'), begins with [h]. With only two participants, it is not possible to tell if this inter-speaker variation reflects a principled distinction between groups of speakers. That said, the behavior of [h] does clear up one confound: The cases in which modal vowels decrease disproportionately as rate slows down involve the voiceless consonants [ʔ] and [h] (c.f. [1]).[1] The other cases involve sonorants and pre-nasalized obstruents, which are voiced. However, P2's modal vowels do not always decrease disproportionately to [h], obviating this confound.

6. ACKNOWLEDGEMENTS

I am very grateful to Natalia Gracida Cruz, Roselia Durán Cruz, Juan Gracida Ortiz, and other members of the San Martín Peras community in Oaxaca and in California for graciously sharing their language with me.

7. REFERENCES

[1] M. Garellek, Y. Chai, Y. Huang, and M. Van Doren, "Voicing of glottal consonants and non-modal vowels," *Journal of the International Phonetic Association*, 2021.
[2] L. Davidson, "Effects of word position and flanking vowel on the implementation of glottal stop: Evidence from Hawaiian," *Journal of Phonetics*, vol. 88, 2021.
[3] D. Silverman, B. Blankenship, P. Kirk, and P. Ladefoged, "Phonetic structures in Jalapa Mazatec," *Anthropological Linguistics*, pp. 70–88, 1995.
[4] D. Silverman, "Laryngeal complexity in Otomanguean vowels," *Phonology*, vol. 14, no. 2, pp. 235–261, 1997.
[5] C. Gerfen and K. Baker, "The production and perception of laryngealized vowels in Coatzospan Mixtec," *Journal of Phonetics*, vol. 33, no. 3, pp. 311–334, 2005.
[6] H. Avelino, "Acoustic and electroglottographic analyses of nonpathological, nonmodal phonation," *Journal of Voice*, vol. 24, no. 3, pp. 270–280, 2010.
[7] M. Macaulay and J. C. Salmons, "The phonology of glottalization in Mixtec," *International Journal of American Linguistics*, vol. 61, no. 1, pp. 38–61, 1995.
[8] C. Gerfen, *Phonology and phonetics in Coatzospan Mixtec*. Springer Science & Business Media, 2013, vol. 48.
[9] I. McKendry, "Tonal association, prominence and prosodic structure in South-Eastern Nochixtlán Mixtec," 2013.
[10] B. Becerra Roldán, "Un estudio fonológico del mixteco de Santo Domingo Huendio, Oaxaca," Master's Thesis, 2015.
[11] J. Mendoza Ruiz, "Fonología segmental y patrones tonales del Tu'un Savi de Alcozauca de Guerrero," Master's Thesis, 2016.
[12] O. León Vásquez, "Sandhi tonal en el mixteco de Yucuquimi de Ocampo," *MA thesis, Centro de Investigaciones y Estudios en Antropología Social*, 2017.
[13] K. Penner, "Prosodic structure in Ixtayutla Mixtec: Evidence for the foot," Ph.D. dissertation, 2019.
[14] C. Mak, "The tonal system of a third Mixtec dialect," *International Journal of American Linguistics*, vol. 24, no. 1, pp. 61–70, 1958.
[15] K. L. Pike, "Analysis of a Mixteco text," *International Journal of American Linguistics*, vol. 10, no. 4, pp. 113–138, 1944.
[16] L. Pankratz and E. V. Pike, "Phonology and morphotonemics of Ayutla Mixtec," *International Journal of American Linguistics*, vol. 33, no. 4, pp. 287–299, 1967.
[17] B. Eischens, "Tone, phonation, and the phonology-phonetics interface in San Martín Peras Mixtec," Ph.D. dissertation, University of California, Santa Cruz, 2022.
[18] E. Zee, "The effect of speech rate on the temporal organization of syllable production in Cantonese," in *Speech Prosody 2002, International Conference*, 2002.
[19] A. E. Turk and J. R. Sawusch, "The domain of accentual lengthening in American English," *Journal of Phonetics*, vol. 25, no. 1, pp. 25–41, 1997.
[20] T. Cambier-Langeveld and A. E. Turk, "A cross-linguistic study of accentual lengthening: Dutch vs. English," *Journal of Phonetics*, vol. 27, no. 3, pp. 255–280, 1999.
[21] Instituto Nacional de Estadística y Geografía, "Censo de población y vivienda," 2010.
[22] C. DiCanio and R. Bennett, "Prosody in Mesoamerican languages," *The Oxford Handbook of Language Prosody*, 2018.
[23] S. L. Peters, "The inventory and distribution of tone in Tù'un Ndá'vi, the Mixtec of Piedra Azul (San Martín Peras), Oaxaca," Master's Thesis, 2018.
[24] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer (version 6.1.27)," 2020.
[25] J. Kingston, "The coordination of glottal with oral articulations in Icelandic and Hindi," *The Journal of the Acoustical Society of America*, vol. 82, no. S1, pp. S114–S114, 1987.
[26] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013. [Online]. Available: http://www.R-project.org/
[27] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
[28] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest Package: Tests in Linear Mixed Effects Models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[1] The phonotactic properties of SMP Mixtec, in which all root-medial voiceless consonants are preceded by [h], do not allow for direct comparison between [Vh/Vʔ] sequences and [VC] sequences where the C is voiceless.