



ARTICLE

Interactions between tone and voice quality in San Sebastián del Monte Mixtec modal and rearticulated vowels

Jae Sidika Weller^{1*} , Jeremy Steffman² , Félix Cortés³ and Iara Mantenuto⁴ 

¹University of Wisconsin – Madison, ²The University of Edinburgh, ³Proyecto Club Alma Mixteca, and

⁴California State University Dominguez Hills

*Corresponding author. Email: jdweller@wisc.edu

(Received 9 April 2024; revised 21 April 2025; accepted 1 May 2025)

Abstract

San Sebastián del Monte Mixtec (henceforth SSM), also known as Tò'on Ndá'vi, is a language of the Mixtecan family, Otomanguean stock. SSM has lexical tones that are orthogonal to rearticulation on vowels. The aim of this production study is to examine both long modal and rearticulated vowels to gain insight into the SSM tonal system, contrastive voice quality, and any potential interactions between voice quality and f₀. Rearticulated vowels are described as having a glottal gesture between two vowels of the same quality (\bar{V}^iV), while modal vowels have no such gesture (VV). To this end, we examined the phonetic realization of the lexical tones in long modal vowels in terms of f₀. All tones are distinguished by f₀; f₀ patterns largely as expected given previously ascribed labels, with minor deviations. Secondly, the phasing and degree of glottalization in rearticulated vowels was measured using 'strength of excitation' (SoE); generally the glottal gesture was vowel medial with a dip in SoE at the beginning of the glottal gesture and a rise in SoE following the glottal gesture. However, there was a large degree of interspeaker variation in the production of rearticulated vowels. Additionally, lexical tone category was found to have an impact on the phasing and degree of glottal gesture in rearticulated vowels, and on voice quality in long modal vowels. This supports the idea that voice quality is an additional correlate of lexical tone in SSM.

Keywords: Otomanguean Language; Mixtec; Lexical Tone; Contrastive Voice Quality; Phonetics

1. Introduction

This study investigates the phonetic interaction between lexical tones and glottalization in San Sebastián del Monte Mixtec (henceforth SSM, ISO: 639 code mks), also known as Tò'on Ndá'vi. SSM is a language of the Mixtecan family, Otomanguean stock (Rensch 1976, Cortés et al. 2023), spoken in the Mixteca Baja region of Oaxaca (Josserand 1983:107). A phonetic description of SSM was written by Cortés et al. (2023). In that study, the authors provided some qualitative assessment of glottalization in so-called 'rearticulated' vowels (described below). However, that study, and work on glottal phenomena in other Mixtec varieties has focused on just a few tones (e.g., Gerfen & Baker 2005), leaving open the

question of how glottalization phenomena are manifest across the tonal inventory. In this study we build on that preceding literature by testing a near-exhaustive combination of tone and re-articulation in the language. We further complement past research in using several time-series modeling approaches (Generalized Additive Mixed Models, and Functional Principle Components Analysis) to provide a quantified and dynamic characterization of the data, elicited from 14 speakers of SSM.

1.1 Background on glottalization and tone

The precise manner in which contrastive glottalization is realized in speech has been the subject of much phonetics study (see Blankenship 2002, Gerfen & Baker 2005, Garellek & Keating 2011, DiCanio 2012, Esposito & Khan 2020, Garellek 2022, and others). This research has shown that contrastive glottalization is associated with multiple phonetic features (i.e., duration, f_0 , spectral tilt) and that languages vary in their exact realization of these phonetic attributes.

Garellek *et al.* (2021) investigated the acoustics of various glottal phenomena across a number of the world's languages using speech data collected from illustrations of the IPA. Their analysis found that phasing and degree of glottal constriction in glottalized vowels varied, based on both the type of glottal gesture and its position within the word. Garellek *et al.* (2021) found that checked and rearticulated vowels, unlike creaky vowels, were characterized by dynamic voice quality over the course of the vowel. That is, voice quality was not constant over the vowel, but changed over time. Additionally, rearticulated vowels were found to have an earlier phasing of the glottal gesture as compared to checked vowels. Both rearticulated and checked vowels had a lower rise in amplitude following the glottal gesture when in word-final position as compared to word-medial position.

During non-modal phonation, f_0 may lower or become highly irregular (Silverman 1997a, Silverman 1997b, Gerfen & Baker 2005, Garellek & Keating 2011, DiCanio 2012). This leads to an interesting problem for languages in which contrastive voice quality and lexical tone can co-occur on the same vowel. Such languages are labeled 'laryngeally complex' in Silverman (1997a). Laryngeally complex languages, Silverman argues, are able to maintain tone contrasts by sequencing non-modal phonation with respect to modal phonation such that tone information can occur on the modal portions of the vowel. In addition, tonal contrasts can be enhanced with changes in voice quality (Esposito & Khan 2020, Garellek 2022). In languages with dense tonal systems, cuing lexical tone category with voice quality differences may aid the listener in distinguishing lexical tones (see Garellek 2022 for further discussion). For instance, Black Miao has a dense tonal system with five level tones; both production and perception studies have shown that voice quality is an important cue of tone in the language (Kuang 2013a, Kuang 2013b, Kuang 2017). In addition, Andruski & Ratliff (2000) and Huang (2020) present evidence for voice quality as a tonal correlate and cue in Green Mong and Mandarin Chinese, respectively. Andruski & Ratliff (2000) showed, via a production study of six speakers, that three of the seven contrastive tones in Green Mong are best predicted by H1–H2 values. Huang (2020) performed two perception experiments in which Mandarin speakers identified tones with f_0 and voice quality manipulations, with and without noise. They found that voice quality differences were more important for tone identification when f_0 was partially obscured by noise.

Chai (2022) further expands on the concept of laryngeal complexity as laid out in Silverman (1997a). Languages which include both lexical tone and contrastive glottalization are further categorized into six different types based on to the extent that lexical tone and contrastive glottalization are orthogonal to each other. These six types range from fully crossed (completely orthogonal) to no overlap (in which modal and glottal vowels have no

tones in common).¹ This study concerns a fully crossed language, as lexical tone and voice quality are orthogonal in SSM.

1.2 Voice quality and tone in Otomanguean languages

There is substantial work describing the phonetics of tone and voice quality in Otomanguean languages, including several studies testing interactions between them. Relevant aspects are detailed briefly below.

Mazatec is an Otomanguean language of the Popolocan branch (Garellek & Keating 2011). San Felipe Jalapa de Díaz (shortened to Jalapa) Mazatec has a lexical tone system with three tones: high, mid, and low, which are fully orthogonal to the three-way voice quality contrast (Garellek & Keating 2011). Garellek & Keating (2011) investigated acoustic correlates of voice quality contrasts (spectral tilt measurements) and lexical tone in Jalapa Mazatec, including their interactions. The study analyzed audio recordings of 14 speakers of Jalapa Mazatec. Certain acoustic correlates were found to be more important for differentiating phonation type in certain tones than others. For example, high tone tokens did not have a single acoustic correlate which differentiated breathy, modal, and laryngealized vowels from each other, whereas for mid tone tokens, H1*–A2* was significantly different for all three phonation types. This indicates that voice quality contrasts in laryngeally complex languages (like SSM) may not be consistent across all tones: it is useful and perhaps necessary to investigate voice quality differences between modal and non-modal vowels for each individual tone.

San Lucas Quiavini Zapotec is an Otomanguean language of the Zapotec family (Chávez Peón 2001). The language has four tones: high, low, rising, and falling. In addition, there are four phonation types: modal, breathy, creaky, and interrupted. Interrupted vowels are realized as either checked (for high tone vowels) or rearticulated (for low and falling tones). The rising tone is only ever present on modal vowels. Chávez Peón (2001), as part of larger work surveying the phonetics and phonology of the language, examined phonetic tone and non-modal phonation using recordings of two speakers of the language. In order to quantify voice quality, the measure of jitter was used at the midpoint of the vowel. When comparing between creaky vowel tokens, high tone creaky vowels were found to have the lowest amount of jitter (corresponding to less creakiness) when compared to low and falling tone creaky vowels. This indicates that the implementation of a voice quality contrast in laryngeally complex languages like SSM may be influenced by lexical tone.

Itunyoso Trique is an Oto-Manguean language of the Mixtecan branch (DiCanio 2010). DiCanio (2012) studied the phasing of f₀ perturbation and non-modal phonation in vowels adjacent to the glottal consonants [h] and [ʔ] in Itunyoso Trique, across 12 different lexical tone combinations. Analysis of audio recordings taken from eight speakers indicated that a more gradual phasing of the glottal gesture with a greater degree of coarticulation led to a dip in f₀ on the adjacent vowel. The effect was not the same across all tones; low tones did not show any f₀ perturbation. Again, we see that tonal category is important: it can affect the interaction between glottal consonants and f₀. In the case of SSM, this indicates that examining the interactions of contrastive glottalization and f₀ across tones is crucial to understanding the whole picture.

Gerfen & Baker (2005) examined both production and perception of what they refer to as laryngealized vowels in Coatzacoapan Mixtec, analogous to what we refer to as re-articulated vowels in SSM. In the production study, modal and laryngealized vowels of eight speakers

¹ While Chai (2020) primarily focuses on checked vowels, this framework can also be applied to rearticulated vowels.

were compared in terms of f_0 , amplitude, H1–H2, and duration. Speakers were highly variable in their degree of creak during the production of laryngealized vowels. The most common correlate of laryngealization was dip in amplitude, which was sometimes accompanied by a drop in f_0 or creaky voicing. In terms of duration, laryngealized vowels were found to be shorter than modal vowels. While there was interspeaker variation in the voice quality of laryngealized vowels, overall, H1–H2 was found to be significantly lower for laryngealized vowels as compared to modal vowels. Lastly, several perception studies examined judgments of 31 speakers. These studies found that small f_0 and amplitude dips, individually, were enough to cue the presence of a laryngealized vowel, and that combining the two cues led to listeners detecting a laryngeal vowel with an even smaller dip in both. Neither the production nor the perception studies examined all tones, however: the production study included only HH and LH tones, and the perception studies included only the HH tone. This leaves open the question of whether these findings would hold for a wider variety of tones. As mentioned above, tone has been shown to have an impact on voice quality contrasts and on the interaction between glottal contrasts and f_0 in other Otomanguean languages.

In light of the prior research on voice quality and tone, we seek to answer the following questions as they pertain to SSM:

- How is lexical tone realized in terms of f_0 (as a replication of Cortés *et al.* 2023)?
- How is contrastive glottalization (in rearticulated vowels) realized over time?
- Does lexical tone in modal vowels have any impact on voice quality, or in other words, could voice quality be a non- f_0 correlate of lexical tone in SSM? Such a finding would be commensurate with previous research showing that lexical tone contrasts incorporate voice quality (Andruski & Ratliff 2000; Huang 2020).
- Does lexical tone impact the magnitude and phasing of the contrastive glottal gesture in rearticulated vowels?

1.3 Background: San Sebastián del Monte Mixtec

San Sebastián del Monte Mixtec is part of the Mixtecan language family, which consists of Mixtec, Cuicatec and Trique, though Mixtec and Cuicatec are part of the same subgroup, also called Mixtecan (Josserand 1983:99–101). San Sebastián del Monte is a town in the Santo Domingo Tonalá municipality of Oaxaca State, Mexico, in the district of Huajuapán de León. The town is 45 km southwest of Huajuapán de León, with a population of approximately 2,000 people (latitude: 17.677778, longitude: –98.021944). The majority of people between the age of 19 and 80 are bilingual Mixtec and Spanish speakers and specifically younger generations (below the age of 19) are mostly monolingual Spanish speakers; though in the town there is an ongoing effort to revitalize the language (Cortés *et al.* 2023).

The consonantal inventory of SSM consists of 21 contrastive sounds /p, b, t, tʰ, k, kʷ, mʰb, nʰd, nʰdʲ, tʃ, m, n, ɲ, r, v, s, ʃ, ʒ, x, ɣ, l/, there are five contrastive oral vowels /a, e, i, o, u/, and four contrastive nasal vowels /ã, ĕ, ĭ, õ/ (Cortés *et al.* 2023). There are long and short vowels, though vowel length is not contrastive in the language. Vowel length is predictable based on the shape of the morpheme. For example, morphemes that present a long vowel in the language are only CVV or VV. Lexical words are minimally and generally formed by two moras (Pike 1948, Longacre 1957), and they can also have three or four moras; however, only function words can be monomoraic (Cortés *et al.* 2023).

Glottalization in SSM is employed contrastively in vowels, and it is evident in four types of bimoraic word: $\bar{V}^{\text{h}}V$, $C\bar{V}^{\text{h}}V$, $\bar{V}^{\text{h}}CV$ and $C\bar{V}^{\text{h}}CV$. The vowels flanking the glottal event in $\bar{V}^{\text{h}}V$ and $C\bar{V}^{\text{h}}V$ are always of the same quality and are called ‘rearticulated vowels’ (eg, *n̄dīi*

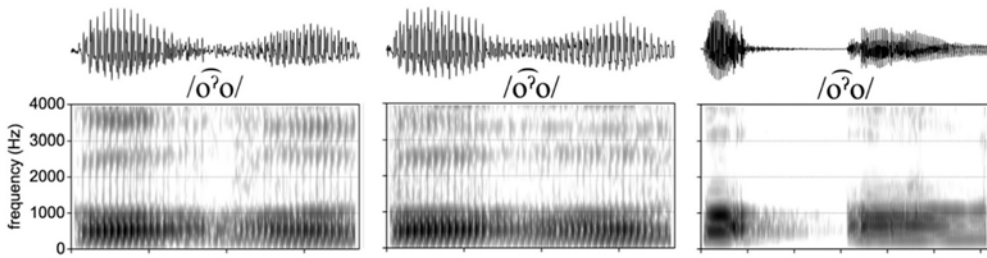


Figure 1. Time-aligned waveforms and spectrograms exemplifying three realizations of SSM rearticulated $/\text{o̞}^{\text{F}}\text{o}/$. Ticks on the x axis are placed at 100 ms intervals. Because speakers were recorded in situ (see the following section), there may be some reverberation present in the recordings, as is visible in the vowel medial portion of the rightmost spectrogram.

‘pimple’ and *loʔo* ‘little’). In the $\text{V}^{\text{F}}\text{CV}$ and $\text{CV}^{\text{F}}\text{CV}$ shapes, glottalization occurs at the end of a vowel preceding a consonant and the two vowels in the syllable can be of different quality (Cortés et al. 2023). While the glottal gesture in rearticulated and checked vowels is commonly transcribed as a glottal stop $/\text{ʔ}/$ for other languages, this study will transcribe it as a tie with a superscript glottal stop, as in $/\text{V}^{\text{F}}\text{V}/$. This convention has been adopted to describe Mixtecan languages due to a lack of full glottal stop closure in most productions of checked and rearticulated vowels (Garellek et al. 2021, Cortés et al. 2023). Figure 1 shows an example of three rearticulated vowel productions, adopted from Cortés et al. (2023). As is already apparent from the figure, the production of rearticulation can vary substantially, in some cases showing full closure and silence in between flanking vowels, and in other cases only a dip in amplitude.

There are three tones in SSM, a high $\langle \text{á} \rangle$, a mid tone $\langle \text{a} \rangle$, and a low tone $\langle \text{à} \rangle$. As in other Mixtec varieties, each mora carries a separate tone. In word shapes that are analyzed as bimoraic (C)V̄V (a long vowel) or (C)V̄^FV (a rearticulated vowel), a tone associates with each mora to form the following tonal combinations (Cortés et al. 2023): high level (HH), mid-level (MM), mid-falling (ML), low level (conventionally labeled as LM), low rising (LH), high falling (HL), low falling (LL). As stated earlier, SSM Mixtec is a fully crossed language with respect to modal and rearticulated vowels; all tones which occur on modal vowels can also occur on rearticulated vowels, and vice versa.

2. Methods

2.1 Speakers, stimuli, and recording

Two bimoraic word types (henceforth referred to as word shapes) were examined in this study. One modal word shape was included in the corpus: long modal vowel CVV, and one word shape which includes a glottal gesture: rearticulated vowel $\text{CV}^{\text{F}}\text{V}$.

As noted above, there are seven tonal contours which occur on CVV and $\text{CV}^{\text{F}}\text{V}$ words: HH, MM, LL, ML, LM, HL, and LH (Mantenuto 2020, Cortés et al. 2023). These were the seven tonal combinations examined across both word shapes in this study. No tokens of HL tones were included in the CVV data. This tone pattern can occur on CVV words, but only in certain aspects of verbs, which were judged too difficult to control in the elicitation.² All five SSM vowels $/i\ e\ a\ o\ u/$ were represented in the corpus.

² One difference between the data obtained for Cortés et al. (2023) and this set of data is related to the HL tone pattern, which was present with one token in the former but not in the latter, the word that we used was *jáa* ‘hot’. However in the present study it was determined that this may be ambiguous as a verb or adjective to speakers, which leads to a tonal ambiguity. For this reason, we excluded it from the present corpus.

This study included 14 speakers of SSM (nine female, five male) who acquired the language from birth and use the language on a daily basis. The speakers were recorded in situ in San Sebastián del Monte. Target words were presented to the speakers in a randomized list of 36 unique words. Target words varied in vowel quality, tone, and word shape, with some lexical gaps. Speakers produced target words in the following carrier sentence:

- (1) Tiáái __ vîtî.
 /tʰáái __ vîtî/
 Tiáái=i __ vîtî.
 write.POT=1SG __ now
 ‘I am going to write __ now.’

The word list was randomized and read twice, with the second reading in reverse order. Each token was repeated twice in both readings of the list, resulting in a total of four repetitions of each unique word. Tokens with background noise or speech errors were not included in analysis, this ultimately being 12% of the recorded data. The tokens which were processed for data analysis included 919 $\text{CV}^{\text{T}}\text{V}$ productions and 944 CVV productions. Additionally, data for CVCV and $\text{CV}^{\text{T}}\text{CV}$ were collected; however, these words were not included in the data presented here, as they were ultimately outside of the scope of this study. See Appendix A for the full word list.

2.2 Measurements

Recordings were annotated by hand using Praat (Boersma & Weenink 2022). STRENGTH OF EXCITATION (SoE) and f_0 measurements were taken to evaluate voice quality and pitch, respectively.

SoE is an amplitude measure which captures relative loudness of voicing without other noise in the recording (Murty & Yegnanarayana 2008; Mittal *et al.* 2014). In recent years, it has been used to analyze non-modal phonation in a variety of languages (see Garellek *et al.* 2021). Unlike other voice quality measures which utilize spectral tilt – such as H1–H2 and Cepstral Peak Prominence – SoE does not require clear harmonic structure, making it appropriate to quantify voicing which may change to very low amplitude, or silence, during a glottal gesture. Because the corpus contained a large number of tokens with heavy glottalization, this feature of SoE was crucial for examining voice quality in the present study. In addition, large amounts of both inter- and intra-speaker variation in the production of the glottal gesture would have meant that segmenting the rearticulated vowel tokens into more modal voiced portions would have made analyzing the corpus in any unified way quite complex.

SoE and f_0 were computed in Voicesauce; f_0 was measured using STRAIGHT (Shue *et al.* 2011). Twenty time-normalized measurements were taken for both SoE and f_0 across the vowels in $\text{CV}^{\text{T}}\text{V}$ and CVV productions, which were segmented manually.

For SoE we present data in log-transformed and range-normalized SoE, following Garellek *et al.* (2021). We log-normalized the SoE measurements because SoE is expected to be log-normal with a right tailed distribution. Following Garellek *et al.* (2021), we also excluded outliers (z -scored larger than |3|). We then range-normalized these measures for the purpose of visual illustration in the figure. This was done by taking the maximum and minimum SoE values produced by a given speaker, and then using the following formula where SoE_i is the (logged) SoE measurement at a given time point.

$$(2) \quad (\text{SoE}_i - \min(\text{SoE})) / (\max(\text{SoE}) - \min(\text{SoE}))$$

With this procedure, a value of one represents the maximum SoE produced by a given speaker, while a value of (very near) zero represents voicelessness. As described in Dhananjaya & Yegnanarayana (2009), SoE values which are less than 1% of the maximum SoE can be considered voiceless.

f₀ measurements were transformed from Hz to Equivalent Regular Bandwidth (ERB) and scaled within speaker to account for between-speaker variation in overall f₀ height, and f₀ range.

2.3 Statistical modeling

GENERALIZED ADDITIVE MIXED MODELS (GAMMs), throughout the paper, are implemented using R packages *mgcv* and *itsadug* (Wood 2011, 2015, 2017; van Rij et al. 2020). To reduce residual autocorrelation, the GAMMs were set up as AR1 error models, from which the rho value was obtained by a previous run of the model without the AR1 structure. In recent years, GAMMs have played a growing role in phonetics research, as a means of examining the dynamics of acoustic and articulatory properties of speech over time (e.g., Carignan et al. 2020, Sóskuthy 2021). We use them here as a means of assessing statistical differences in acoustics measures, as they unfold over time, with 20 time-normalized measures taken from each interval of interest, e.g., VV in CVV and $\bar{V}^T V$ in $CV^T V$. The default number of knots was used in all models, and each model was inspected and found to be adequate using the *gam.check* function.

We structured several different models to address different research questions. To address the question of how f₀ varies as a function of lexical tone in CVV (Section 3.1), we set the GAMM to predict f₀ as a function of lexical tone. We predicted f₀, which was represented as ERB and scaled within speaker as a function of the tone variable, using both a parametric term for tone (capturing overall f₀ height by tone) and smooth term, capturing the (potentially) non-linear dynamics of f₀ over the course of the vowel. Random effects were specified for vowel quality, as control on possible vowel-quality-related modulation in f₀, using bs = 're' as a term in the model. We also include random effects analogous to random effects for subject and by-subject random slopes by tone, using bs = 'fs', and the reference-difference smooth technique described in Sóskuthy (2021), with the *m* parameter set to the value of 1 (Baayen et al. 2016; Sóskuthy 2017). In subsequent analyses we modeled SoE, which was represented in the models as log SoE values, scaled by speaker. In Section 3.2, we were interested in the overall effect of rearticulation on voice quality, measured with SoE. To this end, the GAMM predicted the SoE measure, over normalized time, as a function of word shape, a two level variable with levels CVV (no rearticulation) and $CV^T V$. The random effect structure was the same in this model, with the only difference being that the fixed effect of word shape was built as a by speaker random slope equivalent using the reference difference smooth approach. In this analysis, the parametric term measures the overall difference between word shapes in SoE (across the entire interval of normalized time), and is thus interpretable in a similar fashion to a coefficient in a linear regression model of SoE as a function of word shape.

Finally, we were interested in the interaction between tone and glottalization and their joint influence on voice quality (Section 3.3). To this end, we fit a model to predict SoE as a function of a combined variable which was the combination of each word shape with each tone for which data existed for that word shape. This model does not orthogonally represent tone and word shape, but rather examines their combination as a predictor of SoE. The advantage of this technique is that it allows us to capture (potentially) non-linear and dynamic differences between tones, as a function of word shape, and simultaneously,

between word shapes as a function of tones. Fitting the data in this way, and then examining comparisons of interest thus allow for a full picture of the relationship between tone and rearticulation, and also allows us to examine tone-based differences within each word shape.

In presenting the GAMM results, we primarily use the ‘visual inspection’ technique to assess significance in the models (see Sóskuthy 2017, 2021, for discussion), with reference to computed difference smooths. Difference smooths capture the difference between two fit smooths of interest, and also have 95% confidence intervals (CI). When 95% CI exclude the value of 0, this indicates high confidence in a non-zero difference between smooths (i.e., a significant difference between smooths) at that particular point in normalized time. In what follows, we show both fit smooths and difference smooths for comparisons of interest.

To further examine the possible interactive influence between tone and rearticulation on SoE, we complemented our GAMM modeling of this data with FUNCTIONAL PRINCIPLE COMPONENT ANALYSIS (FPCA; Ramsay & Silverman, 2005). FPCA is a data-driven approach of modeling how a given parameter (here, SoE) changes over time. FPCA has fruitfully been applied to the study of dynamic f0 and formant frequencies (e.g., Gubian *et al.* 2015; Asano & Gubian 2018; Baltazani *et al.* 2022; Arvaniti *et al.* 2024), and offers a way to capture and quantify modes of variation in dynamic contours, which can be submitted to statistical modeling to examine the influence of categories of interest; in our case tone and rearticulation. The present paper provides only a very brief and conceptual introduction to FPCA below; for a more in-depth explanation, we refer readers to Gubian *et al.* (2015).

In our application of the method, sampled SoE curves are represented as continuous functions over time. Independent modes of variation in the data are extracted, and these PRINCIPAL COMPONENTS (PCs) are each associated with a numerical score. A particular contour in the data is modeled as a function of time, and as the sum of functional components, including the mean of the contours in the data $\mu(t)$. In addition to $\mu(t)$, PC scores are multiplied by each PC curve and summed, as shown in the equation below.

$$(3) \quad f(t) \approx \mu(t) + s_1 \times PC1(t) + s_2 \times PC2(t) + s_3 \times PC3(t) + \dots$$

As given in (3) above, $\mu(t)$ is modified by the sum of other PC contours, which together determine the shape of the modeled curve. Each contour in the data set will be associated with a particular value of a PC score, which determines how variation associated with that score is reflected in the contour. Importantly, PC curves are not actual curves in the data, but can be used to reconstruct curves on the basis of the summed PC curves, each with a corresponding PC score. FPCA is ‘data driven’ in the sense that the assignment of PC scores to contours in the data is not made on the basis of a-priori categories in the data (e.g., here, tone and word shape). Nevertheless, it allows for the incorporation of hypothesis tests about linguistic categories and distinctions in a secondary stage where PC scores are passed to a statistical model. Linear modeling can at this stage be used to examine how PC scores, which map to modes of variation in the time-series data, vary as a function of tone and rearticulation.

One benefit of FPCA as a complement to GAMM modeling is that it allows for more direct testing of the relationship and interaction between tone and rearticulation in SSM. In particular, our GAMM modeling fits the model to a combined tone + rearticulation variable, which represents how each trajectory for this combined variable changes over time. Assessment of interactions between tone and rearticulation, made on the basis of visual inspection, are more qualitative in the sense that we can observe, e.g., if particular tones vary in how they are impacted by rearticulation. This is validated statistically by the inspection of difference smooths, which may be different in the scope and location of significant differences. However, the GAMM assessment still lacks a more quantified assessment of the

presence of an interaction between tone and rearticulation. In modeling PC scores with a linear mixed effect model, we can test if both tones and words shapes differ in that score, and if there are statistical interactions between these factors in the model.

The modeling approach we took for the resulting PC scores was as follows: the first three PC scores were submitted to a mixed effects model, implemented in the Bayesian framework using brms (Bürkner 2017). In order to account for possible co-variance among the PC scores, we opted to model the data using a multivariate approach: a single model was thus used to predict how each of the dependent variables (PC1, PC2, and PC3) vary, while also accounting for the possible relationship between the PC scores in the data set. We also ran individual models, one for each PC score, and found that the results were quite comparable across these two approaches, both of which are included on the open access repository. The model predicted a given PC score as a function of tone and word shape, with both of these variables contrast coded. The interaction between the two fixed effects was included as well, to address the central question of how tone and word shape (may) jointly impact SoE. Random effects in the model were by-speaker random intercepts with random slopes for both fixed effects and their interaction. A random intercept for vowel category was also included, as was the case for GAMM modeling. The priors in the model were set to be weakly informative student-t priors, centered on zero with a standard deviation of 1, and 3 df, for both the intercept and fixed effects in the model (encoding no prior expectation of an effect of tone or word shape on each PC score). The models were set to draw 4,000 samples in each of four Markov chains from the posterior distribution over parameter values. The first 1,000 samples from each chain were discarded, retaining 75% of samples for inference.

In reporting results from the model, our key interest is the overall (marginal) effect of word shape on each PC (that is, the effect of word shape across tones), and presence of any interactions between tone and word shape on PC scores. To examine marginal effects and interactions of interest we used emmeans (Lenth 2023) to compute marginal contrasts for word shape, tone, and the effect of word shape within each tone category (detailed in the results below). In reporting effects of interest, we provide the median posterior estimate for an effect, 95% Credible intervals (CrIs) for the estimate. CrIs refer to the range of an estimated posterior distribution in which 95% of the estimates fall. When estimates for an effect are reliably estimated to be non-zero, this indicates that an effect is reliable, or ‘credible’. A common convention for considering which effects are credible is thus to assess if the 95% CrI for a given effect INCLUDES OR EXCLUDES the value of zero. If the value of zero is included in the interval, this would suggest a lack of certainty that the effect is not zero, or a lack of certainty that the effect has a consistent directionality (i.e. 95% CrI spanning zero indicates substantial variation in the estimated directionality of an effect). Conversely, 95% CrI which EXCLUDE the value of zero can be taken to represent a consistently estimated effect. We additionally report the percentage of a distribution for a given effect which shows a consistently estimated directionality, as computed with bayestestR (Makowski et al. 2019). This ‘probability of direction’ (pd) metric can be taken to represent graded evidence for effect existence; a posterior estimate centered precisely on the value of zero will have a pd value of 50 (50% of the distribution on either side of zero). On the other hand, a distribution with a consistently estimated effect will have a pd value approaching 100. A pd value in excess of 97.5 corresponds precisely to an estimate with 95% CrIs just excluding zero; however, those values approaching this threshold can also be taken as graded evidence for an effect (i.e., the probability that an effect exists with a particular directionality).

For both of the data analysis approaches described above, the numerical model output is not particularly insightful for interpreting effects of interest. The full models, in

addition to the data and scripts for analysis are thus contained online, hosted on the OSF at: <https://osf.io/crx2q/>. We also present marginal effect estimates for comparisons of interest in Appendix B of the paper.

3. Results

3.1 *f0 as a function of tone in CVV*

First, we investigate the manner in which lexical tone is realized with *f0* in long modal vowels, as a basic characterization of the tonal contrasts when not influenced by re-articulation. To this end, we use GAMMs to test how *f0* varies across the vowel in CVV word shapes. The *f0* results for rearticulated tokens will not be examined here, as the glottalized vowel-medial portion had irregular *f0*, which led to a high number of erroneous *f0* readings.³

Figure 2 shows findings for this section. Panel A contains fit smooths for each of the six tonal melodies elicited for the CVV tokens (coloration indicating tone). The tone smooths in Panel A are generally as expected given the lexical tone labels, with a couple of exceptions. The non-contour tones (according to phonological tone labels) HH, MM, and LL have a mostly constant *f0* and have highest to lowest *f0* in that order. The LL (low falling) tone drops slightly over the course of the vowel. LM and LH both have rising *f0* over the duration of the vowel, with LH rising farther than LM. ML starts with a higher *f0* than expected given the standard tone label: at the beginning of the vowel, it has a similar *f0* to HH. Unfortunately, as mentioned earlier, we were not able to collect any HL long modal vowel tokens for comparison. Panel B of Figure 2 shows the difference smooth for each pair of tonal melodies. Highlighted yellow sections mark portions of the vowel for which *f0* is significantly different between tonal melodies. Generally, tone pairs are significantly different in terms of *f0* for all or most of the vowel, with the exception of LL and LM, which only show small differences in *f0* over relatively short portions of the beginning and end of the vowel.

Cortés et al. (2023) also presented *f0* data for tones in SSM Mixtec CVV tokens. Findings were similar to those in this study for all tones except LM. Cortés et al. (2023) characterizes LM as a low steady tone, citing their result that the *f0* of LM was mostly stable and did not show a substantial rise in pitch in the latter part of the vowel. However, the LM data in this experiment is not consistent with that finding. The LM tone *f0* smooth here dips down in the beginning of the vowel and then rises slightly in the latter two thirds of the vowel. Cortés et al. (2023) had a similar result for ML tokens (which we found had slightly higher than expected at the beginning of the vowel given the tone label). In their study, HL tokens were included, and HL was differentiated from ML by a higher peak in *f0* during the first half of the vowel and a lower dip in *f0* during the last third of the vowel. Cortés et al. (2023) found that the HL *f0* smooth dipped down to approximately match the level of the LL *f0* smooth near the end of the vowel, while the ML *f0* smooth stayed higher. While our ML tone findings include a slightly higher *f0* rise at the beginning of the vowel, our results are very similar to Cortés et al. (2023) in that the *f0* of ML does not dip down to meet the LL tone at the end of the vowel.

³ Over half of the *f0* tracks for rearticulated vowels were identified as containing *f0*-tracking errors, with some speakers being more affected than others. We also measured *f0* from manually annotated regions of modal voicing flanking the glottal gesture, though demarcation of this region was somewhat difficult to assess by manual labeling. Potential shortcomings of both of these approaches to measuring *f0* in the rearticulated vowels led us to opt not to report these measures in the present paper.

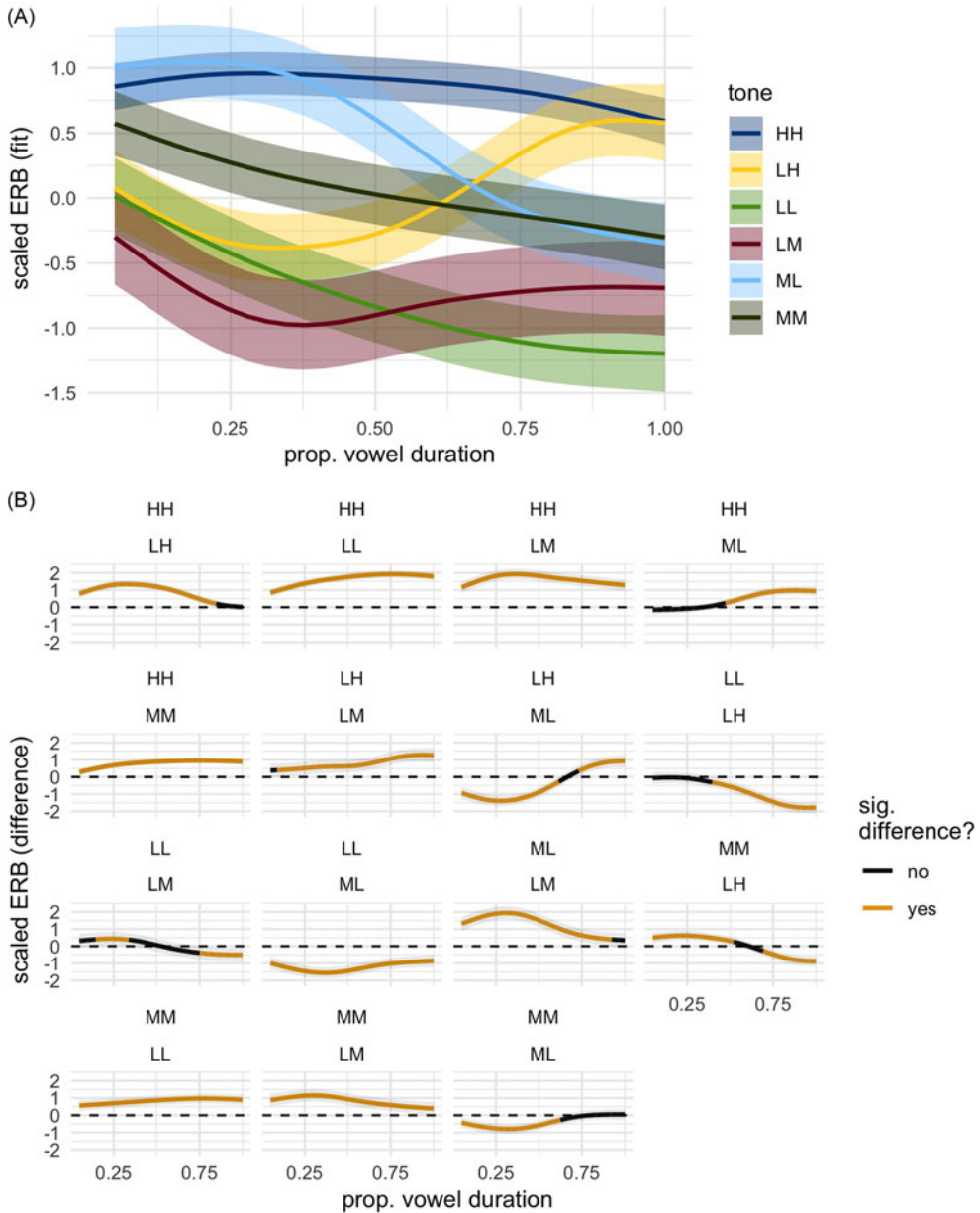


Figure 2. Panel A (top) shows GAMM fits for f_0 (in scaled ERB) over vowel duration for each tone in CVV tokens. Panel B (bottom) shows difference smooths for all tone pairs in Panel A; yellow marks parts of the vowel duration where there is a significant difference in f_0 between tones.

3.2 Strength of excitation by word shape

In this section, we examine how rearticulated vowels are produced, focusing on the dynamics of strength of excitation (SoE) over the vowel. As in the previous section, we use GAMMs, this time to test how SoE varies across the vowel in both rearticulated and CVV word shapes. Two separate GAMM structures are described in this section.

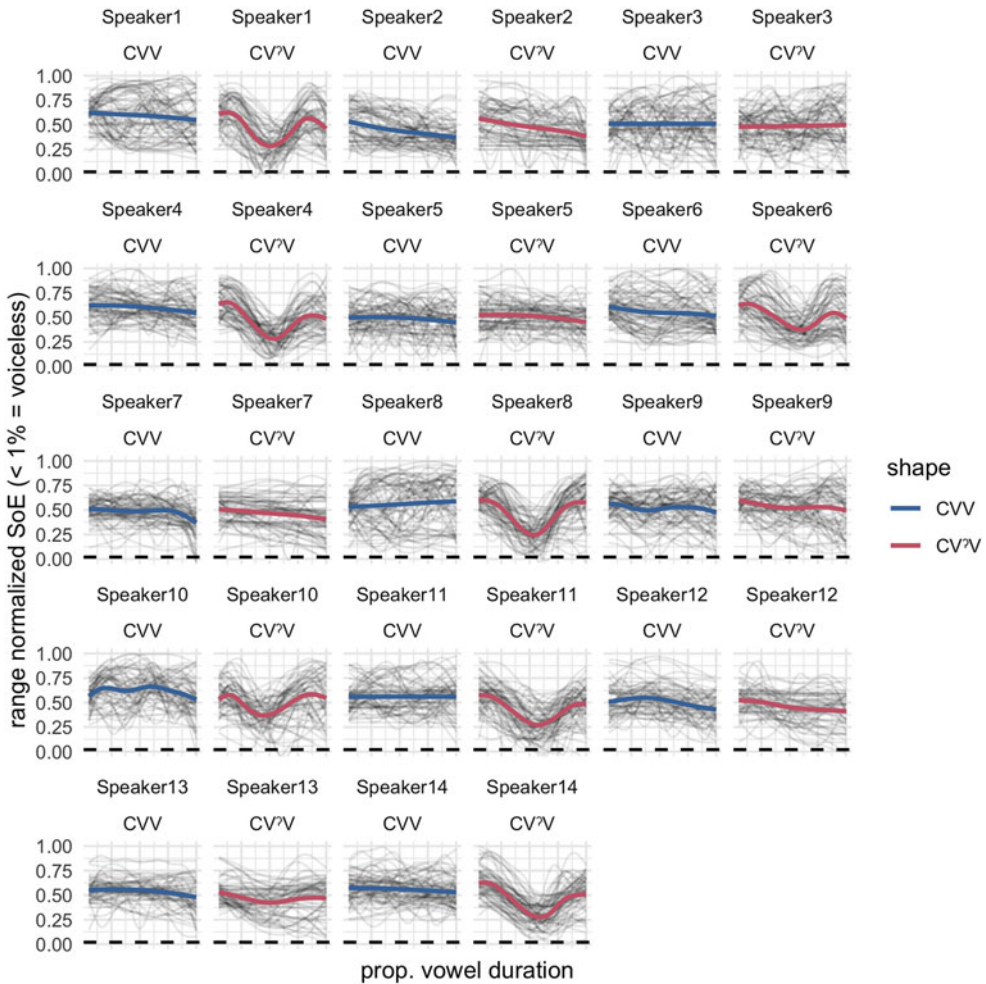


Figure 3. SoE tracks over normalized vowel duration, split by speaker. For each speaker, a plot for CVV words is shown on the left and a plot for CV[^]V words is shown on the right. Thin lines represent individual SoE tracks, and thick colored lines show means.

Prior to proceeding to the analysis, we present the data in Figure 3, which shows by-speaker measures of SoE over time in both rearticulated and CVV word shapes. Figure 3 shows these measures plotted for both the individual productions, and speaker means as smoothed functions (GAMMs) over time, to visually simplify and factor out micro-fluctuations in SoE. The thin black lines in the figure represent individual SoE trajectories and the thicker colored lines represent SoE mean.

All speakers show relatively stable mean smooths for SoE over the course of long modal vowels in CVV tokens. While SoE means are stable, individual SoE trajectories can be dynamic over time (for example, see speakers 1, 3, and 11). However, the timing of the SoE dip within the vowel differs across tokens, leading to a relatively flat SoE mean trajectory. This is not the case for rearticulated vowel tokens. For seven speakers (1, 4, 6, 8, 10, 11, and 14), there is a pronounced dip in mean SoE about halfway through rearticulated vowel tokens. This dip corresponds with the expected glottal gesture in the middle of rearticulated vowels. Other speakers who do not show this regularly phased glottal

gesture in rearticulated vowels may be producing contrastive glottalization with other phonetic cues. We examine this inter-speaker variation in quantified terms using FPCA in Section 3.4.

In our first analysis of rearticulated vowels, we focus on the difference between the two word shapes overall by fitting a GAMM to model SoE over time. There are two essential questions examined here: first, whether rearticulation leads to a dip in SoE, as predicted based on findings from Garellek et al. (2021); and secondly, how this dip is phased in the vowel overall. To this end, we fit the GAMM to model log SoE, scaled within each speaker, as a function of word shape. The model was fit with a parametric term for word shape. Parametric terms are analogous to a fixed effect in a linear model. This term will capture the overall value of SoE, across the interval as a whole, as a function of word shape. The model was additionally fit with smooth terms for word shape over time, which will allow us to capture the dynamics of (potential) differences in SoE as a function of word shape. Random effects in the model were a random intercept for vowel ($bs = 're'$) and a random intercept for speaker with by-speaker slopes for shape, using the reference-difference smooth technique described in Sóskuthy (2021). These random effects thus account for vowel-based

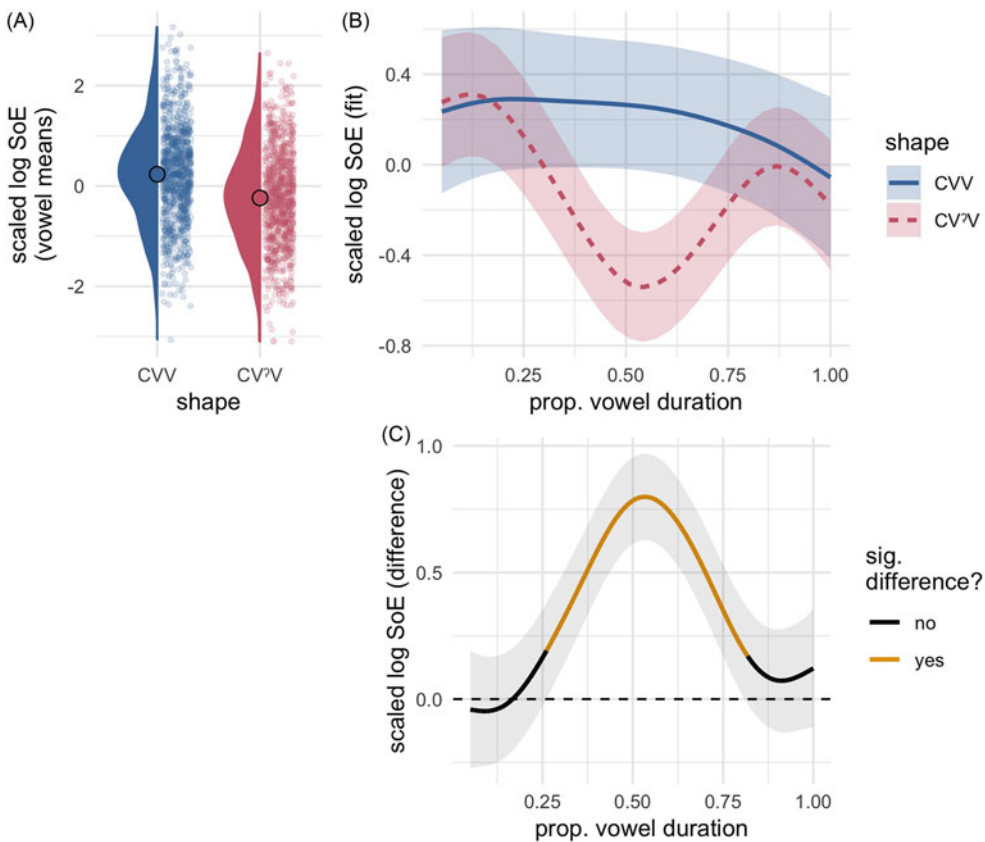


Figure 4. Panel A (top left) shows the overall mean SoE for long modal and rearticulated vowels (larger points), with half-violin plots showing the distribution of the data and individual data points shown at the right of each half-violin. Panel B (top right) shows smooths for SoE in long modal (blue) and rearticulated vowels (red). Shading shows 95% confidence intervals. Finally, panel C (bottom) shows the difference smooth corresponding to the pair of smooths in panel B.

variation in SoE, where we may expect more sonorous segments to have higher SoE (e.g., Chong *et al.* 2020), potentially as a function of e.g., vowel openness. The random effects also account for speaker-based variation in overall SoE, and in the effects of word shape for each speaker. The overall effects of word shape on SoE are shown in Figure 4.

First, there was a main effect of shape on SoE, whereby with CV^TV as the reference level on the model, the parametric term capturing the differences between shapes was significant, showing that, across the interval as a whole, CVV shapes have higher SoE ($\beta = 0.31$, $t = 4.19$, $p < .001$). This effect is shown in Figure 4, Panel A, which presents the empirical scaled mean SoE measure across word shapes. This overall lowering of SoE is consistent with what we expect in terms of glottalization lowering voicing amplitude overall. To inspect the dynamics of the effect, we plot the fit smooths for each word shape (Figure 4, Panel B), and the difference smooth for each word shape (Figure 4, Panel C). The difference smooth in a GAMM model allows for a statistical assessment of where (in time) two trajectories are different from one another. As can be seen by coloring in Panel C, the region of significant difference is in the middle portion of the vowel, and in this region only. In other words, in assessing the differences in voicing amplitude across word shapes overall, rearticulation leads to a dip in SoE in the middle of the vowel, which is otherwise not significantly different from CVV at the beginning or end of the vowel's duration. This corresponds to the expectation that rearticulation is phased in the middle of the vowel. Interestingly, we found that SoE in rearticulated vowels, as a whole, rose more after the initial dip than was reported in Garellek *et al.* (2021), where results indicated that word-final rearticulated vowels had a dip in SoE, after which SoE did not rise.

3.3 Interactions between glottalization and tone

3.3.1 GAMM modeling

Our second analysis sought to examine the influence of lexical tone on the dynamics of SoE. Our interest here is twofold. First, are there reliable differences in SoE based on tone in CVV and CV^TV shapes? Descriptively, we are interested in testing for non-f₀ correlates that may distinguish lexical tones in SSM. Secondly, how does the previously described phasing pattern, evident in the aggregated modeling in Section 3.2, vary by tone? That is, we are interested to test if lexical tone mediates the magnitude and phasing of SoE differences across word shapes. To this end, we fit smooths with a combined tone+wordshape variable, which was also included as a parametric term and a random effect using the reference difference smooth method employed in the previous GAMM. The same random effects for vowel were additionally used. Crucially, for our purposes, this technique allows us to model non-linear interactions between word shape and tone. Figure 5 shows these GAMM fits, grouped by word shape in Panel A, and grouped by tone in Panel B.

Panel A shows that long vowel tokens have a more stable in SoE over time, whereas rearticulated vowel tokens are more dynamic with a large dip in SoE corresponding to the contrastive glottal gesture in that word shape (as shown in the aggregated analysis in Figure 4).

In Panel B, blue shading indicates portions of the vowel during which the long vowel smooths and rearticulated vowel smooths are significantly different. For all tones, there was some vowel-medial (and/or end) portion in the rearticulated vowel tokens during which SoE dipped significantly lower than the CVV smooth. Crucially, the timing, duration, and degree of this difference in SoE shapes varied by tone. The SoE of rearticulated LH and LM tone tokens only briefly and very slightly dips below the 95% CI for long modal vowels. Comparing the modal and rearticulated vowel smooths for the HH and MM tones, on the other hand, shows that the rearticulated vowels have a more robust dip in SoE

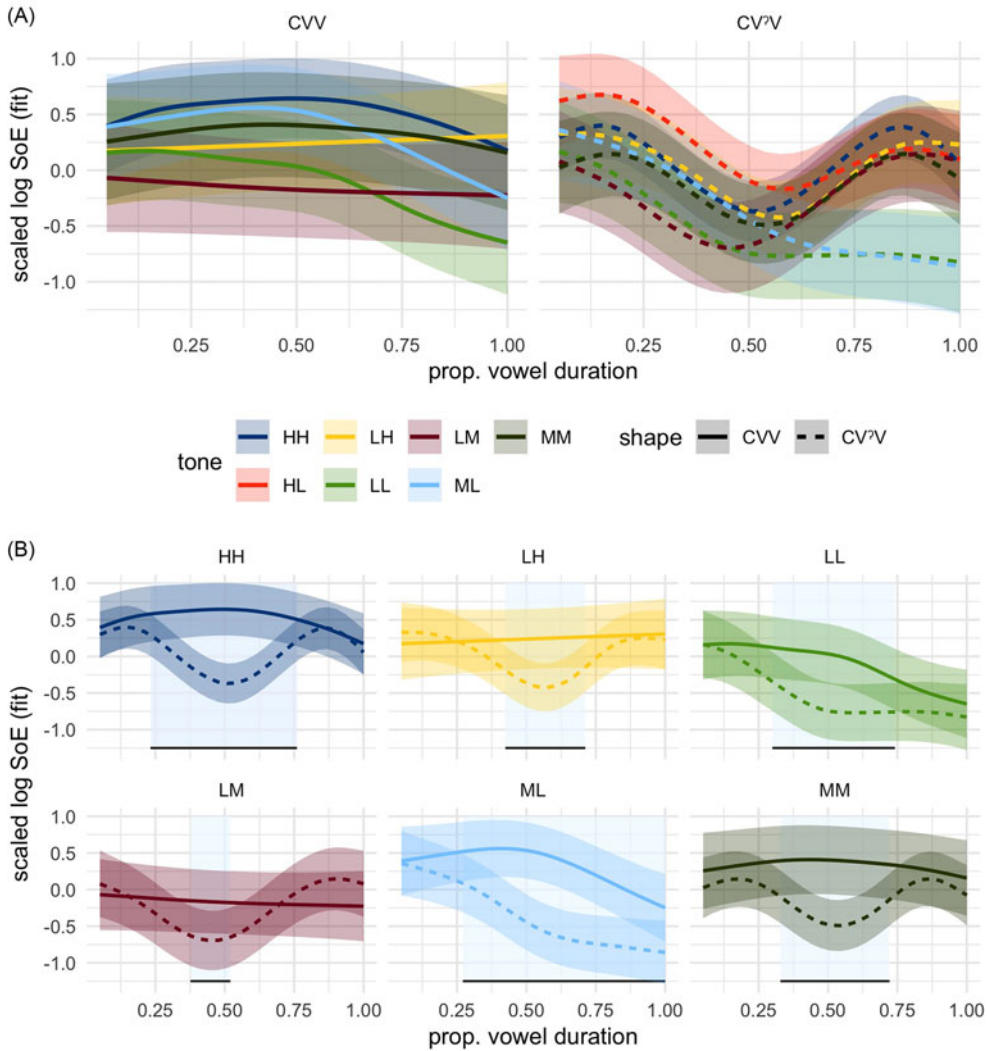


Figure 5. Panel A (top) shows by-tone SoE smooths for long modal and rearticulated vowels. Panel B (bottom) shows smooths which compare the SoE for long modal and rearticulated of the same tone.

vowel-medially as compared to modal vowels. Lastly, the ML and LL rearticulated data show a different pattern from the others: rather than SoE falling and rising, SoE falls after the first quarter of vowel duration and does not rise.

In the smooths in Panel A of Figure 5 above, it appears that SoE smooths for each tone differ from one another within word shape (comparing tones within long modal and rearticulated vowels). This is further confirmed by inspection of pairwise difference smooths, which show clearly in aggregate that the scope and location of differences varies between tones and word shapes. Most tone pairs within word shape, eight of 15 tonal pairs for long modal vowels and 17 of 21 tonal pairs for rearticulated vowels, had a significantly different SoE for some portion of the vowel. These difference smooths are included in Appendix C below. This trend in the data will be explored more fully in the following FPCA section.

3.3.2 FPCA modeling

Figure 6 shows the FPCA modeling results, where Panel A gives a visual representation of the impact of variation in PC1 (left) PC2 (middle) and PC3 (right); see the figure caption for

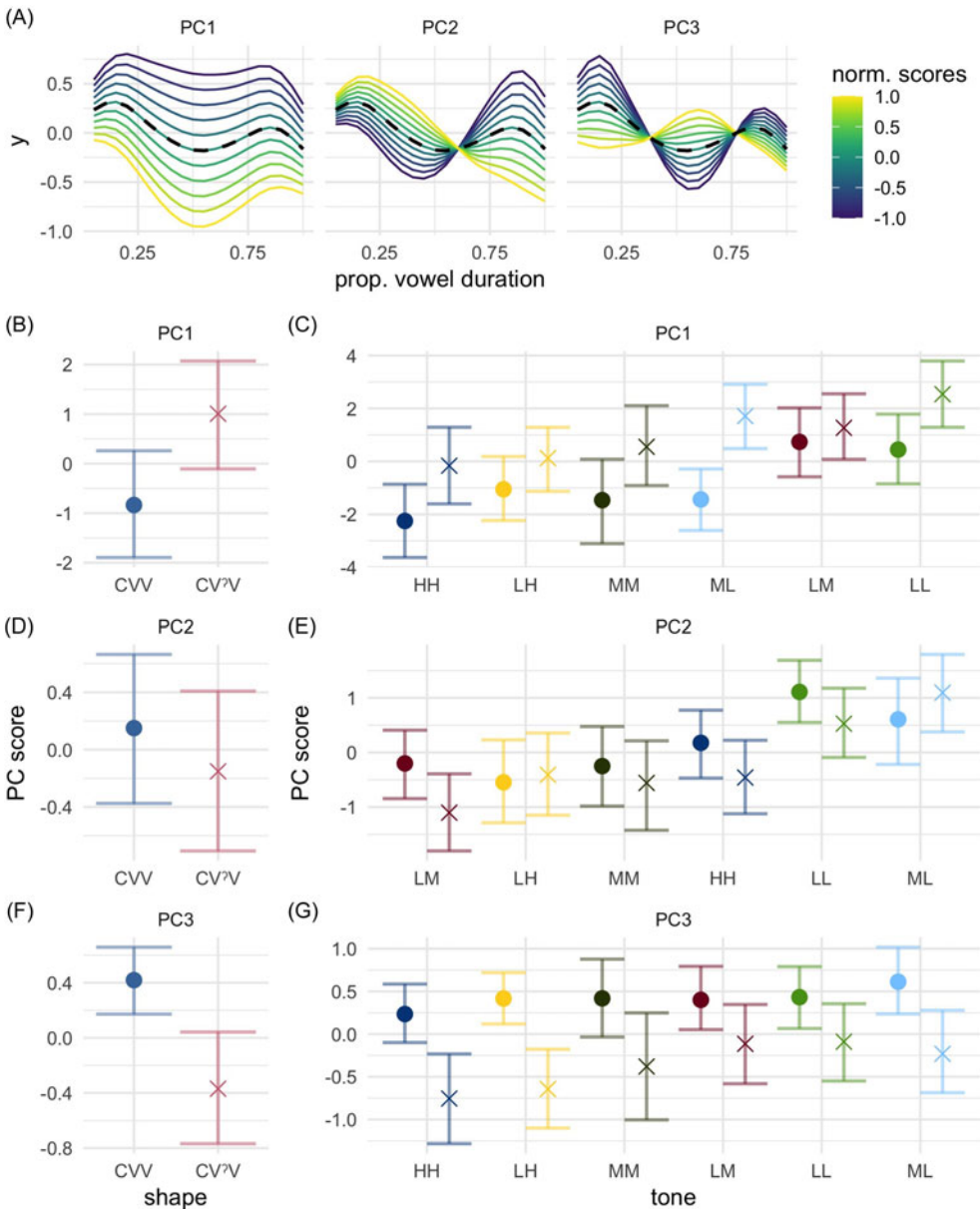


Figure 6. Panel A (top): visual representation of the impact of each PC, represented by how a particular PC score results in deformation of the mean curve in the data (dashed line, same in all panels). These were generated by adding a particular PC curve (PC1 at left, PC2 at middle, PC3 at right) multiplied by a particular PC score (shown by coloration and indexed at right) to the mean curve. The scores are represented as SD to normalize across PCs visually, ranging from -1 SD to $+1$ SD. Panels B–G show the model estimates for each PC as a function of word shape (indicated by point shape, left) and tone and word shape (right). Points are median estimates, error bars show 95% CrI.

Table 1. This includes the qualitative characteristics of the curve corresponding to higher and lower PC scores for each calculated PC value

PC#	PC Score Value	Interpretation – associated modification of mean curve
PC1	High	lower SoE over all; more dipping
	Low	higher SoE over all; less dipping
PC2	High	SoE is falling
	Low	SoE is more dipping and rising
PC3	High	More stable SoE throughout (and perhaps rising in the middle)
	Low	SoE rising in the beginning portion, dipping in the middle, and then rising and falling at the end

details. The first three PCs in the data set captured 55%, 18%, and 12% of the variance in the data for PC1, PC2, and PC3, respectively, 85% in aggregate. This indicates that the first three PCs collectively explain a large proportion of variance present in the data set. In presenting the results which follow, our goal is to describe how a particular PC score corresponds to variation in contour shape in the data set. With that variation in mind, we observe how the PC score varies as both a function of tone and word shape. We explain each of the three PCs in succession in what follows.

Table 1 above contains a brief description for each PC score, i.e. what a higher or lower value corresponds to in deformation of the mean curve. We will step through each in detail here. PC1, broadly speaking, appears to capture two different properties simultaneously. The first is contour height, whereby SoE contours with higher PC1 scores are LOWER in SoE– numerically higher (normalized) scores in PC1 result in lower-valued contours (i.e., weaker voicing). PC1 additionally varies in the extent to which the contours dips in the middle. Numerically higher PC1 values map to a more-dipping contour shape. This observation that PC1 maps to SoE height and propensity to dip leads us to expect that there will be a difference in this PC score as a function of word shape: rearticulated vowels, having been shown to have lower and more-dipping SoE contours, should have numerically HIGHER PC1 scores. The statistical model of PC1 scores confirms this observation. The marginal effect for word shape in the model shows that rearticulated vowels have credibly higher PC1 scores ($\beta = 1.84$, 95% CrI = [0.97, 2.81]; $pd = 100$), shown in Figure 6 Panel B. This essentially confirms the previous GAMM modeling analysis based on word shape: overall, PC1 captures SoE lowering and dipping which is present in rearticulated vowels. Marginal effects for tone also indicate some differences in PC1. Given the many possible pairwise comparisons of tones with one another, we just focus on several comparisons of interest, which can also be seen in Figure 6, Panel C. The points in the panel are sorted from lowest to highest PC1 score by tone (left to right). Table 2 in Appendix B contains marginal estimates for all pairwise tone differences (for all three PCs). One clear difference in PC1 is that HH tones have credibly lower PC1 scores as compared to three other tones: LM, LL, and ML ($pds > 99$). This suggests that these latter tones are effectively lower in SoE (and/or more dipping), also confirming the observations from the fit GAMM smooths. Like HH, LH is also credibly lower in PC1 as compared to LM and LL ($pds > 99$). This clearly shows that tone AND word shape are jointly structuring SoE variation associated with PC1, such that the information is jointly encoded in the trajectories' PC1 score.

The model also indicates the presence of credible interactions between tone and word shape ($pds > 99$), leading us to inspect the effect of shape within each tone, shown in Figure 6, Panel C. Comparison of marginal shape effects within tone show that for all but one tone, there is clear evidence for a difference based on shape in line with the marginal shape effect ($pds > 97$). For LM however, there was no credible evidence for a difference in PC1 score as a function of word shape ($pd = 84$), though the directionality of the effect was the same as the marginal effect for word shape. The estimates also reveal clear variation in the *MAGNITUDE* of the shape effect within each tone. For example, ML shows a larger difference in PC1 score as a function of word shape. In summary, PC1, which varies credibly as a function of both tone and word shape, shows that these dimensions are intricately linked in their impact on SoE variation captured in this particular PC. Statistical interactions further confirm that tone and word shape are jointly encoded in PC1, and that the effects of rearticulation are not uniform in this dimension.

Looking at Panel A in Figure 6, we can note that PC2 variation is primarily associated with variation between monotonically falling shapes (high PC2 score) or scooped and rising shapes (low PC2 score). Lexical tone varied clearly in this dimension, with many tones showing credible pairwise differences from one another (see Table 2 in Appendix B), as evident from the variation in PC2 score left to right as a function of lexical tone (ordered low to high in Panel E). The evidence was weaker for a difference in PC2 based on the marginal effect of word shape, as shown in Figure 6, Panel D. The estimates are strongly skewed one way, however 95% CrI do narrowly include the value of zero ($\beta = -0.30$, 95% CrI = $[-0.60, 0.01]$; $pd = 97$). The presence of several credible interactions between tone and word shape in the model indicate possible asymmetries of this effect across tones, which become apparent when we consider the estimates for these two factors in interaction in Panel E. MM and LH tones show no credible difference in this PC, though the remaining tones show some evidence of credible difference in PC2 score as a function of word shape ($pds > 99$, save for ML where $pd = 95$). For all but one, this aligns with the marginal effect of word shape, though interestingly ML tones show a reversal of this effect whereby rearticulated vowels tend to have higher PC2 scores. Observing the GAMM fits for this tone shown in Figure 5, we can recall that rearticulation is expressed differently for this tone: instead of an SoE dip which falls and rises, SoE for ML toned rearticulated vowels falls monotonically, aligning more closely with numerically higher PC2 score contributions to contour shape. In this sense, the statistical interaction observed in this PC confirms the tone-dependent realization of rearticulation over the time course of the vowel.

Finally, PC3 can be characterized as encoding the presence or absence of a dip in SoE in the middle of the vowel (PC3 in Panel A). In this sense, it is similar to PC1. However, it does not encode as much global height variation, and instead is mostly an indication of dipping in the middle of the vowel. In accordance with this qualitative observation, we have the clear expectation that rearticulation should cause variation in the PC. This is confirmed by the model, (Figure 6, Panel F) which shows credibly lower PC3 scores for rearticulated vowels ($\beta = -0.79$, 95% CrI = $[-1.17, -0.42]$; $pd = 100$), where lower scores map to more of an SoE dip. This difference was credible for all tones ($pds > 99$, save for LL where $pd = 97$), evidencing a clear tone-general manifestation of word shape in this PC. Comparisons across lexical tone also show that PC3 variation is less driven by tone as compared to the preceding 2 PCs, though there are several credible differences where $HH < \{LL, LM, ML\}$ and $LH < \{ML, LL\}$ in terms of PC3 score.

In sum, the FPCA analysis confirms the GAMM modeling in showing that modes of variation in the SoE trajectories are reflective of both tone and rearticulation. Moreover, even for each of the first three PCs analyzed here, credible interactions between tone and word shape indicated that a particular PC was impacted by both of these factors. The patterns

overall comported with what was observed more qualitatively in the GAMM modeling analysis; rearticulation was associated with overall lower SoE (higher PC1 score) and more of an SoE dip in the analyzed interval (lower PC2 and PC3 scores, though PC2 variation was smaller as a function of rearticulation). In addition to overall variation in PC scores based on tone, notable tone-based interactions emerged. For example, ML tone showed evidence for the opposite of the main effect for rearticulation in PC2 scores. In observing the actual contours fit by the GAMMs in Figure 5, it became apparent that this difference arises from the tone-specific realization of rearticulation, wherein ML rearticulated vowels were realized more as falling instead of dipping (reflected in a higher PC2 score overall). Patterns such as this show clearly that both dimensions of contrast – tone and rearticulation – are intimately tied up in voice quality dynamics, as measured with SoE.

3.4 Quantifying speaker variability

Given the group-level results presented above, one point of interest is the extent to which individual speakers align with these results, as already noted in the substantial variation across speakers that was identified in Section 3.2. In this section, we briefly quantify and examine individual variation, focusing on the overall word shape differences. As previously noted in Figure 3, speakers do certainly vary in the extent they produce SoE modulations associated with re-articulation, leading us to examine these patterns in a bit more detail here.

To this end, we modified two of the previous analyses to be speaker specific. For a numerical and objective quantification of speaker differences, we modified the FPCA analysis to include Speaker as a fixed effect. While individual speakers are certainly conceptually in line with random effects, fitting them as a fixed effect in this case allows us to estimate speaker-wise difference in PC scores as a function of rearticulation, with, critically, the same PC scores modeled as in the group-level analysis. With this structure, we can estimate marginal contrasts for word shape and for each speaker, asking, for example, does Speaker 1 show a credible difference in PC1 across word shapes? Looking at the constellation of PC scores and whether or not there are credible differences across speakers allows us to get a bird's-eye, quantified view of individual variation. To provide a more qualitative assessment of individual variability, not unlike Figure 3, we fit a GAMM to each speaker and word shape combination, mirroring the analysis of overall word shape effects. This allows us to get a more specific view of the actual SoE trajectories produced under these conditions, and can be related to the PC score evaluation as a way of understanding how PC score variation maps to actual trajectory shapes. (For a summary of how each PC score corresponds to a different trajectory shape, see Table 1.)

Figure 7 below shows both of these results. In Panel A, the estimated contrast for the shape variable is plotted in three dimensions, one for each PC. The x axis is the word shape contrast for PC1 (see Table 1 for how varying each PC corresponds to deviations from the mean trajectory). Note that this is the estimated DIFFERENCE in PC1 scores between rearticulated and modal vowels, when the difference is positive, rearticulated vowels have a higher PC1 score (a more dipping shape as shown in Figure 5 Panel A). The error bars along the x axis are 95% CrI for the estimated contrast, when they exclude the value of zero, we conclude there is credible evidence for a difference in PC1 score based on shape. As a specific illustration, we can note that Speaker 9 has an estimated difference in PC1 score of approximately zero: no credible evidence for an effect of word shape on PC1 for this speaker. Conversely, Speaker 4 has the largest estimated difference, and a clearly credible (non-zero) effect. The word shape contrast for PC3 is shown in the y axis, chosen because it was more clearly related to word shape differences. Finally, the word shape contrast for

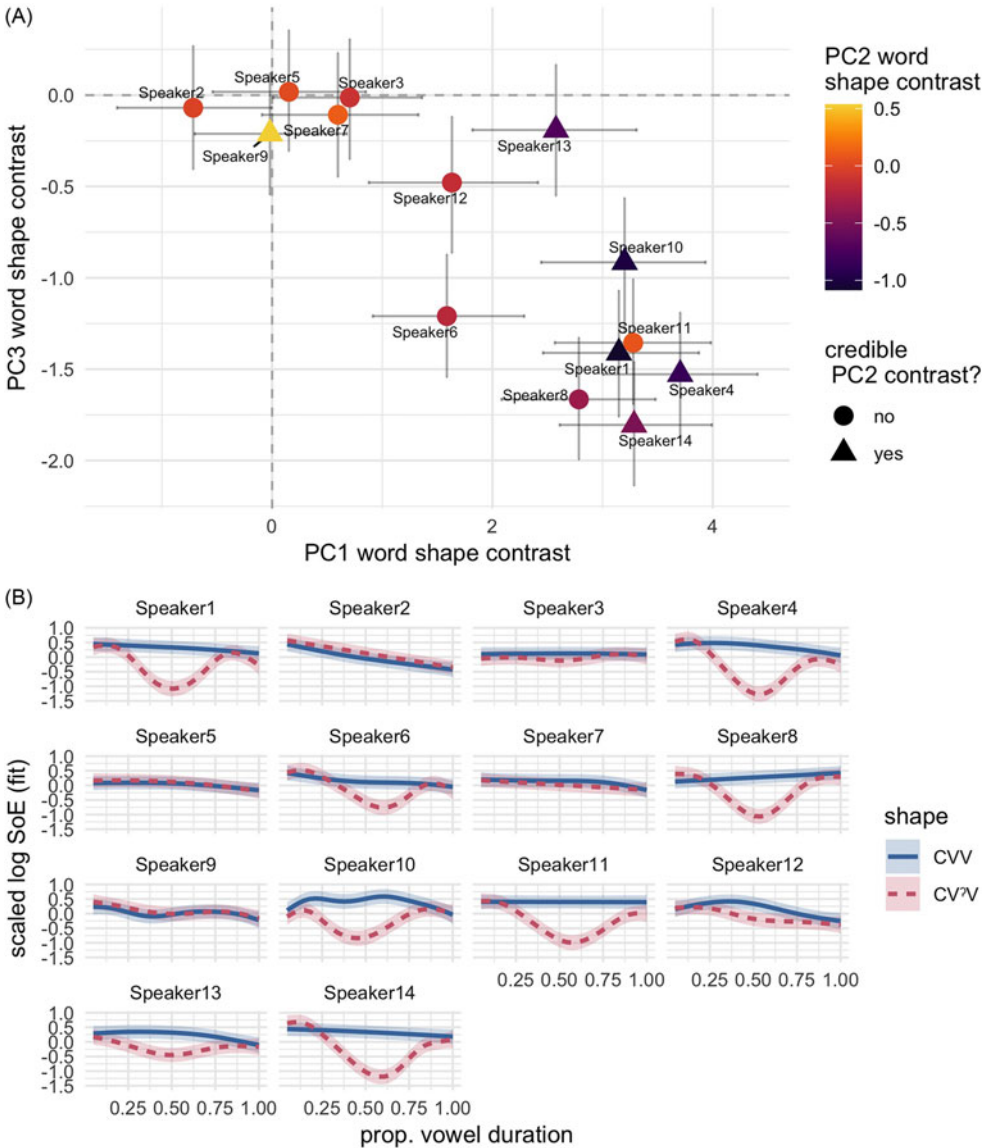


Figure 7. Panel A (top): PC score contrasts for word shape, computed for each individual speaker, from the individual analysis showing PC1 on the x axis, PC3 on the y axis, and PC2 in terms of point coloration, shown in the legend on the right. Panel B (bottom): By-speaker smooths for modal (blue solid line) and rearticulated (red dashed line) vowels.

PC2 is shown in terms of the coloration of each speaker's point, represented at right on the color scale. When a PC2 contrast is credibly non-zero, the shape is a triangle, otherwise a circle. In aggregate, this three dimensional representation shows if individual speakers have credible differences in each of the PC components, if and how these different aspects of SoE variation co-vary across speakers. Panel B below shows the GAMM smooths fit to each speaker and word shape, plotted together.

We first consider some basic observations of this representation of inter-speaker variability. There are a set of speakers who show credible differences in all three PCs: Speakers

1, 4, 10 and 14. These speakers, as can be seen in Panel B, produce rearticulated vowels with pronounced dips in SoE which involve a clear lowering, and centralized dipping. This shape evidently combined aspects of each of the three PCs to the extent that they all are differentiated for these speakers. These four speakers can thus be considered maximalists in terms of their production of rearticulation in SoE: they produced large dips which involve lowering, central dipping and the early dip-then-rise shape associated with negative PC2 scores.

Speakers 8 and 11 represent a similar qualitative dip, as shown in Panel B, though differ from the aforementioned four in not having a credible difference in PC2 scores as a function of rearticulation. Recall that in the group-level analysis PC2 was the least associated with re-articulation. While both of the PC2 scores for these two speakers lean in the direction that is expected (a negative estimate, meaning rearticulated vowels have more negative PC2 scores, the lack of a credible difference means that component of variation is not consistent or large enough for these two speakers to constitute a credible effect. Speaker 11 in particular shows a smaller (less-negative) contrast, which we can hypothesize is related to the fact that their productions of rearticulated vowels do not rise quite as much at the end, though do not contain as much variation in the vein of PC2. For each of the six aforementioned speakers though we can conclude they are producing large and robust dips in SoE, evident in credible differences in PC1 and PC3 in all cases.

Next, there is a group of speakers who show credible differences for PC1. Speaker 3 shows a very slight credible difference for PC1 alone. This is as expected given Speaker 3's extremely slight difference between rearticulated and model vowels in Panel B (with rearticulated vowels having a very slight dip). Speakers 6 and 12 show credible differences for PC1 and PC3, which matches the qualitative patterns shown in panel B – both speakers have rearticulated vowels which dip slightly in SoE word medially. Interestingly, Speaker 13 appears to be a slight outlier. This speaker shows credible differences in PC1 and PC2, but not PC3.

Finally, we have a group of speakers who did not show credible differences in any PC score. This group includes five speakers out of our 14 speakers, which all cluster around the origin in the plot in Panel A. Speaker 7 showed very weak differences in PC1 which were not credible. Lastly, Speakers 2, 5, and 9 showed no credible PC score differences at all. Speaker 2 shows a slight difference in PC1 which is not credible, but interestingly this difference is in the opposite direction to other speakers with a PC1 difference.

4. Discussion and conclusions

This study provided an examination of lexical tone and contrastive glottalization in San Sebastián del Monte Mixtec rearticulated vowels. The language, being laryngeally complex in the terminology of Silverman (1997a) provides a good opportunity to study both lexical tone and contrastive glottalization, as well as interactions between the two. The main focus of the study was twofold: first, to better understand the phonetic realization of lexical tone and rearticulated vowels in San Sebastián del Monte Mixtec, and second, to examine any potential interactions between lexical tone and voice quality.

In terms of f0, lexical tone patterns as expected given tonal labels, largely corroborating the findings of Cortés et al. (2023). Strength of Excitation (SoE) was used to study the degree and phasing of the contrastive glottal gesture in rearticulated vowels. Garellek et al. (2021) found that a dip in SoE is a robust acoustic correlate of contrastive glottal gestures across a number of the world's languages, making it a good measure to quantify rearticulation in our data. Our findings broadly support the results of Garellek et al. (2021) in that SoE is an effective measure of the glottal gesture in these rearticulated vowels. Our findings in SSM

indicate that the phasing of the glottal gesture may be affected by lexical tone; rearticulated ML and LL tokens have a dip in SoE which does not rise, as reported in Garellek *et al.* (2021). Other tones, however, have a dip in SoE which rises towards the end of the vowel. Garellek *et al.* (2021) did not report tone information, so it is unclear whether that pattern was also present in their data.

In addition, voice quality results for rearticulated vowels in SSM are generally consistent with those reported in Gerfen & Baker (2005) for laryngealized vowels in Coatzospan Mixtec. Like Gerfen & Baker (2005), we observed interspeaker variation in the production of non-modal vowels. By comparing PC scores, we found that different speakers make different adjustments in SoE when producing rearticulated vowels. In addition, Gerfen & Baker found that Coatzospan Mixtec laryngealized vowels were produced with a dip in amplitude in the middle of the vowel, as well as lower H1–H2 than their modal counterparts. This aligns with our finding that there is a dip in SoE for rearticulated vowels, as compared to modal vowels. Lastly, we sought to expand on the production study of Gerfen & Baker (2005), which examined two tones (HH and LH), by including seven tones, and we found that phasing and degree of the glottal gesture in non-modal vowels did vary by tone. This is significant because it indicates that the production of a glottal gesture is affected by tonal category.

We found that lexical tone impacts SoE, which we presume is linked to non-modal phonation (*cf.* Garellek *et al.* 2021). This impact includes long modal vowels (Figure 5). Difference smooths (Figure 9, in the Appendix) show that SoE between 8 of fifteen tonal pairs examined for these long vowels was different in some region of the SoE contour. Since there is no contrastive glottal gesture in long modal vowels, differences in f_0 across tones can only be explained through some influence of tone on voice quality. In addition, SoE was significantly different across most tones for rearticulated vowel tokens, with 17 of 21 tonal pairs being significantly different for some portion of the vowel, according to the GAMM analysis (Figure 8 in the Appendix). Moreover, the FPCA modeling of the SoE contours for rearticulated vowel tokens show clearly that SoE variation is credibly impacted by tone, most clearly for the first two PCs. In sum, this suggests that voice quality serves as a non- f_0 correlate of lexical tone in SSM.

In general, our findings on the nature of interaction between lexical tone, voice quality, and contrastive glottalization corroborate Silverman (1997a, 1997b) in that contrastive voice quality is necessarily sequenced with more modal phonation, which carries tonal information. Rearticulated vowels consist of more modal vowel portions at the beginning and end of the vowel, with a glottalized vowel-medial portion which contains the glottal gesture – with the exception of the ML and LL tones, which did not show a rise in SoE following the glottal gesture. While lexical tone does impact voice quality (as we observed in CVV data), the degree of glottalization is far smaller and less dynamic than the contrastive glottal gesture in rearticulated vowels. Therefore, while voice quality may be a correlate of lexical tone in SSM, tone-based voice quality differences remain distinct from contrastive voice quality differences in rearticulated vowels.

One interesting finding of this study (mentioned in the preceding paragraph) is that SoE in rearticulated vowels was not the same across tones. So, tone-based variation in voice quality is not entirely constrained by the presence of a contrastive glottal gesture. In this sense the results contribute to our understanding of how glottal events are manifest in languages with lexical tone, and complement previous research in Otomanguean, in particular in evidencing intricately linked realization of both dimensions. Previous studies, like this one, have found various interrelations between voice quality and tone in Otomanguean. For instance, Garellek & Keating (2011) found that voice quality contrasts in Jalapa Mazatec can be measured using different acoustic qualities depending on tone. In addition, Chávez Peón (2001) found that creaky vowels in San Lucas Quiaviní Zapotec had more or less jitter at the

midpoint of the vowel depending on tone. Lastly, Dicanio (2010) found that tonal category was important for the phasing of glottal consonants in Itunyoso Trique. In the broadest terms, our results align with these in showing that tone and rearticulation jointly influence dynamics of SoE across SSM vowels.

Our findings suggest that ideas around laryngeal complexity may be extended – not only are tone and glottal gestures sequenced with one another, tone category itself may exert a mediating influence into this phasing equation. In sum, the results contribute to our understanding of how glottal events are manifest in languages with lexical tone, and complement previous research in Otomanguean, in particular in evidencing intricately linked realization of both dimensions.

4.1 Limitations and future directions

There is much more interesting phonetic research left to be explored in SSM Mixtec. For instance, SSM Mixtec has checked vowels in addition to rearticulated vowels. A future study is needed to examine the glottal gesture in checked vowels and its interactions with lexical tone. An additional limitation of this study was that no f_0 data was examined from rearticulated vowel tokens, due to substantial f_0 tracking errors. Future research could include f_0 data for rearticulated vowels, to determine whether the glottal gesture might have any impact on f_0 across lexical tones.

Another limitation of the present study is that our target words, while produced in an utterance as compared to isolation, were elicited in relatively careful speech, and in the context of a phrase which did not vary. This leaves open the question of how the phasing and strength of the glottal gesture in rearticulated vowels will play out in more variable contexts, and naturalistic speech. One possible future avenue is thus to examine these effects in more naturalistic speech contexts, in for example, a communicative task like the Map Task (e.g., Brown et al. 1983). What we can assume is that the patterns we see here are most representative of careful speech, in which we can presume the effects of rearticulation, and intersection with tone, will be most clearly produced by speakers. Testing if and how these effects are reduced or modulated by contextual factors will be another important piece of the puzzle.

One aim of this study was identifying possible non- f_0 correlates of lexical tone in SSM Mixtec. We have identified SoE as an acoustic correlate of lexical tone production in SSM, which we take to constitute a measure of voice quality changes, as also reflected in rearticulated vowels. It remains to be seen whether and how voice quality variation plays a role as a cue in speech perception. Whereas some research on Mixtec has examined perception of re-articulation (Gerfen & Baker 2005), we are not aware of existing research on the perception of lexical tone in Mixtec varieties. Given the production patterns documented in this study, SSM represents a good extension along these lines. A fundamental question to address is the relative perceptual import of f_0 and non- f_0 cues to lexical tone, which could be manipulated along continua as in many cue-weighting studies, or investigated by resynthesizing f_0 across base files of differing lexical tones, examining how perceptual processing of these correlates operates when they are in conflict. Examining how this works for particular pairs of tones that share relatively similar SoE profiles (e.g., relatively flat SoE contours for modal LH and LM tones), or are more distinct, would present another comparison of interest. Though modal vowels seem a logical starting point to address this question, further exploring if/how patterns differ in rearticulated tone perception would constitute an important step. Each of these possible studies would help provide insight into how the acoustic differences shown here translate perceptually.

To answer a different but related question, another valuable extension would be to study the cues relevant to the perception of rearticulation in SSM, in similar fashion to Gerfen & Baker (2005). This question becomes particularly interesting in light of the tone-dependent manifestation of rearticulation seen in this study. A test of how cue weighting for perception of rearticulation may vary across tonal context strikes us as another intriguing perception-oriented question to address.

Extending the study in these ways would provide a fuller understanding of lexical tone and glottalization in SSM, and in doing so, contribute to the overall knowledge of laryngeal contrasts cross-linguistically.

Acknowledgements First and foremost, we want to thank the community of San Sebastián del Monte for their invaluable help with this project. Without their assistance, this project simply would not be possible. Secondly, many thanks to the University of Wisconsin – Madison Language Sciences program for providing the funding for this work.

Appendix A

Word list

Table 2. List of all words elicited in the study

SSM	IPA	English
ndì'i	[ⁿ di [?] i]	pimple
jì'í	[xi [?] í]	female
kee	[kee]	lay (an egg)
tíí	[tíí]	to be cold
sàá	[sàá]	new
và'a	[và ^a a]	well (adj)
tú'ú	[tú [?] ú]	strange
nà'án	[nà [?] á]	long time
ki'i	[ki [?] i]	wear (verb)
yúú	[ʒúú]	hard
ní'in	[ni [?] í]	strong noise
yùù	[ʒùù]	stone
jìì	[xìì]	happy
yò'ó	[ʒò [?] ó]	you
tò'on	[tò [?] õ]	word
kò'ò	[kò [?] ò]	plate
sà'a	[sà [?] à]	foot
yàà	[ʒàà]	ash
kèè	[kèè]	empty (verb, future tense)
tìín	[tí [?] í]	mouse
lo'ò	[lo [?] ò]	little
ndii	[ⁿ di]	deceased

Table 2. Continued

SSM	IPA	English
kaa	[kaa]	pile (verb, future tense)
ve'e	[ve'e]	house
ñòó	[ɲòó]	night
kàa	[kàa]	metal
xe'è	[xe'e]	ring (noun)
too	[too]	frequently
yí'í	[ʒí'í]	raw
ka'à	[ka'a]	bottom
kaà	[kaà]	there
ñí'in	[ɲí'í]	sweat lodge
yóó	[ʒóó]	we
leè	[leè]	baby
xoò	[foò]	hen
yó'ò	[ʒó'ò]	here

Appendix B

PC scores for FPCA analysis

Table 3. Summary of marginal estimates for tone pairs, for each PC, from the FPCA analysis. The contrast column shows the tone pair compared with median, lower and upper CrI, and pd

	Contrast	Estimate	L-95% CrI	U-95% CrI	pd
PC1	HH - LH	-0.74	-1.67	0.21	94
	HH - LL	-2.69	-3.76	-1.61	100
	HH - LM	-2.2	-3.39	-1.05	100
	HH - ML	-1.33	-2.34	-0.33	99
	HH - MM	-0.74	-2.21	0.8	85
	LH - LL	-1.96	-2.66	-1.25	100
	LH - LM	-1.46	-2.24	-0.72	100
	LH - ML	-0.6	-1.24	0.02	97
	LH - MM	0	-1.01	1	50
	LL - LM	0.49	-0.18	1.17	93
	LL - ML	1.36	0.57	2.12	100
	LL - MM	1.95	0.79	3.21	100
	LM - ML	0.86	0.03	1.67	98
	LM - MM	1.46	0.21	2.74	99
	ML - MM	0.6	-0.41	1.59	89

Table 3. Continued

	Contrast	Estimate	L-95% CrI	U-95% CrI	pd
PC2	HH - LH	0.34	-0.26	0.94	87
	HH - LL	-0.96	-1.41	-0.53	100
	HH - LM	0.51	0.09	0.94	99
	HH - ML	-0.99	-1.57	-0.39	100
	HH - MM	0.27	-0.34	0.88	83
	LH - LL	-1.3	-1.84	-0.72	100
	LH - LM	0.17	-0.44	0.75	73
	LH - ML	-1.33	-2.13	-0.52	100
	LH - MM	-0.07	-0.88	0.77	58
	LL - LM	1.47	1.05	1.89	100
	LL - ML	-0.03	-0.55	0.49	55
	LL - MM	1.23	0.65	1.83	100
	LM - ML	-1.5	-2.06	-0.91	100
	LM - MM	-0.24	-0.86	0.36	81
	ML - MM	1.26	0.54	1.98	100
PC3	HH - LH	-0.15	-0.39	0.11	89
	HH - LL	-0.43	-0.72	-0.14	100
	HH - LM	-0.4	-0.8	-0.01	98
	HH - ML	-0.45	-0.77	-0.13	100
	HH - MM	-0.28	-0.73	0.13	91
	LH - LL	-0.28	-0.5	-0.06	99
	LH - LM	-0.25	-0.57	0.07	94
	LH - ML	-0.3	-0.56	-0.02	98
	LH - MM	-0.13	-0.47	0.25	77
	LL - LM	0.03	-0.31	0.36	57
	LL - ML	-0.02	-0.35	0.31	55
	LL - MM	0.15	-0.24	0.56	79
	LM - ML	-0.05	-0.44	0.34	60
	LM - MM	0.13	-0.39	0.67	69
	ML - MM	0.17	-0.27	0.64	79

Table 4. Summary of marginal estimates for the effect of word shape for each tone, and for each PC, from the FPCA analysis. The estimates are shown with the median, lower and upper CrI, and pd

	Tone	Contrast for shape	L-95% CrI	U-95% CrI	pd
PC1	HH	2.09	0.83	3.36	100
	LH	1.18	-0.04	2.42	97
	LL	2.08	0.77	3.36	100
	LM	0.54	-0.56	1.66	84
	ML	3.16	2.08	4.22	100
	MM	2.02	0.4	3.61	99
PC2	HH	-0.63	-1.08	-0.21	100
	LH	0.14	-0.37	0.65	71
	LL	-0.58	-1.02	-0.14	99
	LM	-0.9	-1.46	-0.36	100
	ML	0.49	-0.09	1.09	95
	MM	-0.3	-1.01	0.36	82
PC3	HH	-0.99	-1.49	-0.49	100
	LH	-1.06	-1.54	-0.56	100
	LL	-0.52	-1.07	0.03	97
	LM	-0.52	-0.95	-0.09	99
	ML	-0.85	-1.29	-0.37	100
	MM	-0.8	-1.44	-0.13	99

Appendix C

Pairwise difference smooths for SoE modeling

For rearticulated vowel tokens (Figure 8), all tone pairs had some portion of the vowel during which the smooths were significantly different, except for HH & LH, HH & MM, LL & ML, and MM & LH. The MM & LM pair of smooths are only significantly different for a brief period toward the beginning of the vowel. The same difference smooths were calculated for long modal vowel tokens (Figure 9). For CVV smooths, the tonal pairs which showed a statistically significant difference for some portion of the vowel were HH & LL, LH & LM, LL & LH, LL & ML, ML & LM, MM & LL, and MM & LM. The HH & LH pair also showed a significant difference, but only briefly at the middle of the vowel. Overall, the pairwise inspection of difference smooths indicates that there are significant differences in SoE between lexical tones in both word shapes. These results indicate that voice quality may enhance tonal contrasts in both rearticulated and long modal vowels. For rearticulated vowels, lexical tone may affect the phasing and degree of glottalization in the contrastive glottal gesture.

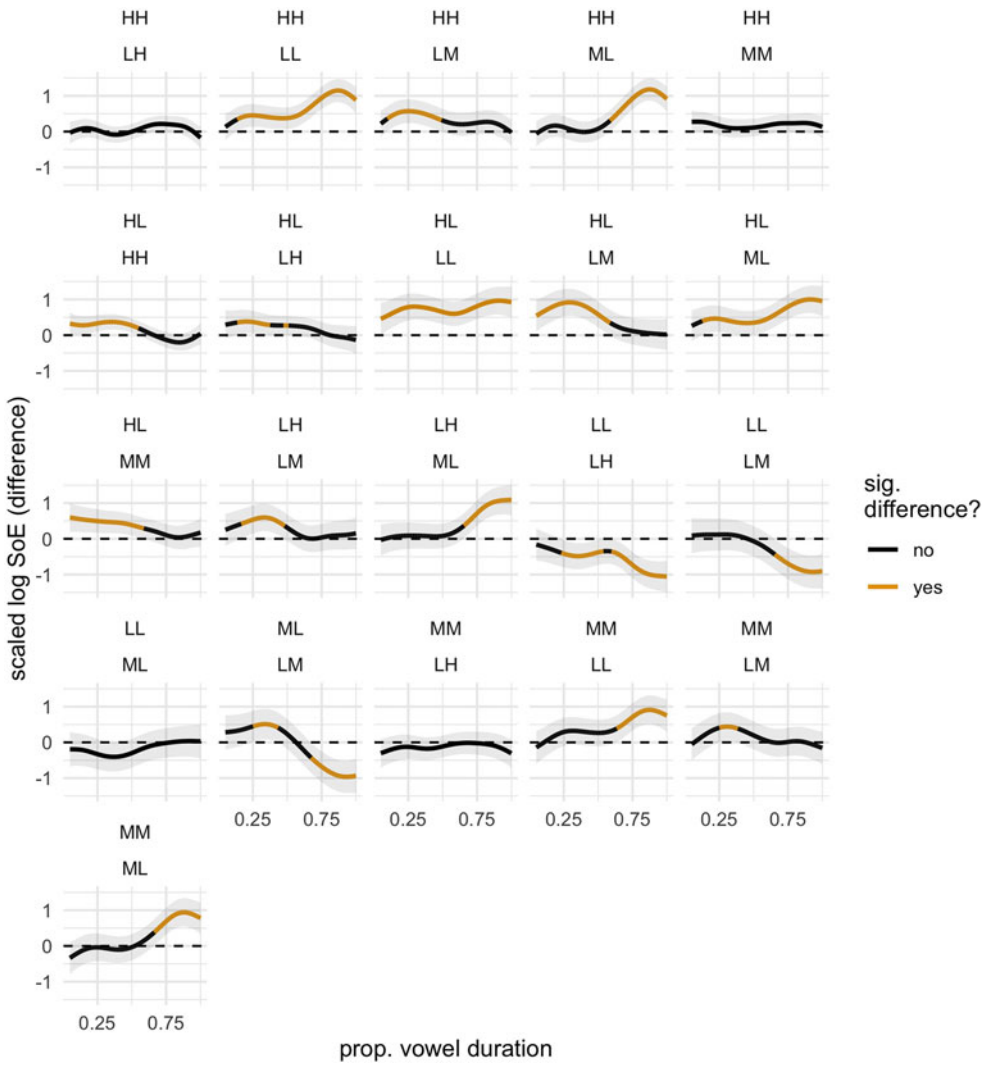


Figure 8. Difference smooths for SoE in rearticulated vowels, each plot shows the smooth for one tone pair.

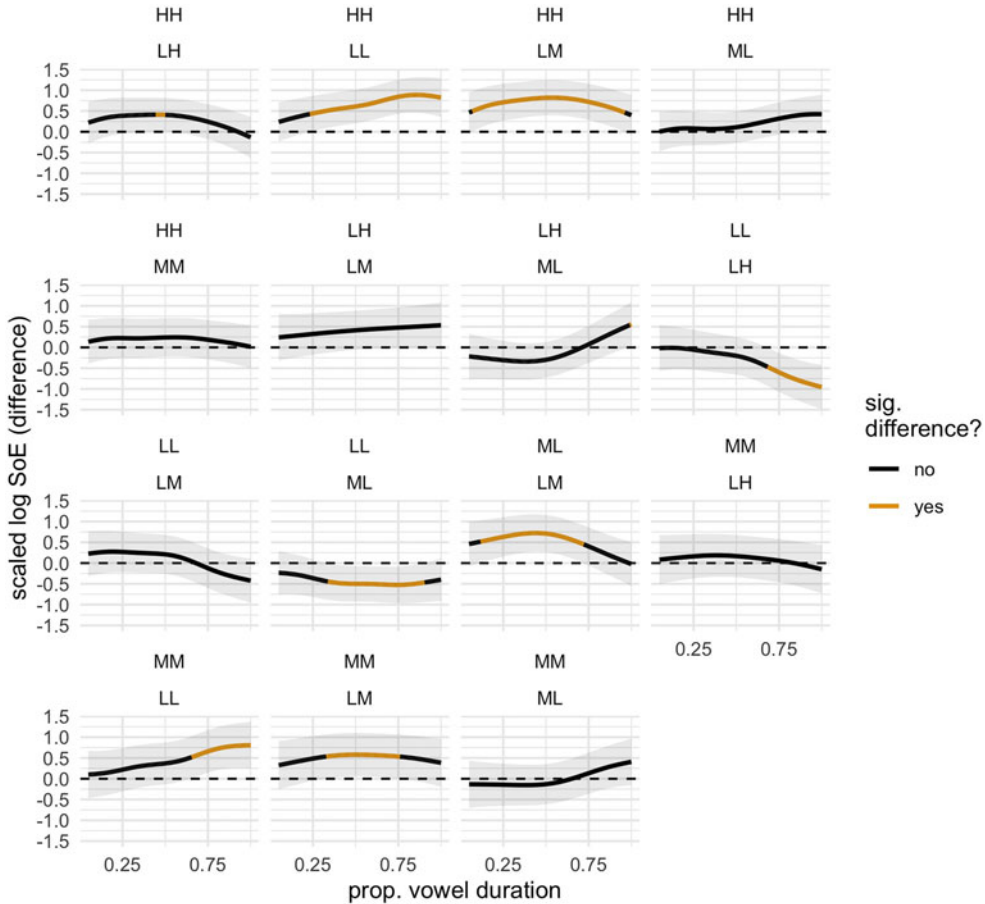


Figure 9. Difference smooths for SoE in CVV vowels, each plot shows the smooth for one tone pair.

References

- Andruski, Jean E., & Martha Ratliff. (2000). Phonation types in production of phonological tone: The case of Green Mong. *Journal of the International Phonetic Association* 30, 37–61. <https://doi.org/10.1017/S0025100300006654>
- Arvaniti, A., A. Katsika, & N. Hu. (2024). Variability, overlap, and cue trading in intonation. *Language* 100(2), 265–307.
- Asano, Yuki, & Michele Gubian. (2018). “Excuse meeee!”: (Mis) coordination of lexical and paralinguistic prosody in L2 hyperarticulation. *Speech Communication* 99, 183–200.
- Baayen, R. H., Jacolien van Rij, Cecile de Cat, & Simon N. Wood (2016). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models. *Mixed-effects regression models in linguistics*, 49–69.
- Baltazani, Mary, John Coleman, Elisa Passoni, & Joanna Przedlacka. (2022). Echoes of past contact: Venetian influence on Cretan Greek intonation. *Language and Speech*, 00238309221091939.
- Blankenship, Barbara. (2002). The timing of nonmodal phonation in vowels. *Journal of Phonetics* 30, 163–191. <https://doi.org/10.1006/jpho.2001.0155>
- Boersma, Paul & David Weenink (2022). *Praat: doing phonetics by computer* [Computer program]. Version 6.3.03, retrieved 17 December 2022 from <http://www.praat.org/>
- Brown, Gillian, Anne Anderson, George Yule, & Richard Shillcock. (1983). *Teaching talk*. Cambridge: Cambridge University Press.
- Bürkner, Paul C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80, 1–28.

- Carignan, Christopher, Phil Hoole, Esther Kuna, Marianne Pouplier, Arun Joseph, Dirk Voit, Jens Frahm, & Jonathan Harrington. (2020). Analyzing speech in both time and space: Generalized additive mixed models can uncover systematic patterns of variation in vocal tract shape in real-time MRI. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1).
- Chai, Yuan. (2022). *Phonetics and phonology of checked phonation, syllables, and tones*. Dissertation, University of California, San Diego.
- Chávez-Peón, Mario (2001). *The interaction of metrical structure, tone and phonation types in Quiaviní Zapotec*. Dissertation, The University of British Columbia, Canada.
- Chong, Adam J., Megan Risdal, Ann Aly, Jesse Zymet, & Patricia Keating. (2020). Effects of consonantal constrictions on voice quality. *The Journal of the Acoustical Society of America* 148(1), EL65–EL71.
- Cortés, Félix, Iara Mantenuto, & Jeremy Steffman. (2023). San Sebastián del Monte Mixtec. *Journal of the International Phonetic Association*, 1–22. <https://doi.org/10.1017/S0025100322000226>
- Dhananjaya, N., & B. Yegnanarayana. (2009). Voiced/nonvoiced detection based on robustness of voiced epochs. *IEEE Signal Processing Letters* 17(3), 273–276.
- DiCanio, Christian T. (2010). Itunyoso Trique. *Journal of the International Phonetic Association* 40(2), 227–238. <http://www.jstor.org/stable/44527019>
- DiCanio, Christian T. (2012). Coarticulation between tone and glottal consonants in Itunyoso Trique. *Journal of Phonetics*, 40(1), 162–176.
- Espósito, Christina M., & Sameer D. Khan (2020). The cross-linguistic patterns of phonation types. *Language and Linguistics Compass*, 14, e12392. <https://doi.org/10.1111/lnc3.12392>
- Garellek, Marc. (2022). Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality. *Journal of Phonetics* 94, 101155.
- Garellek, Marc, Yuan Chai, Yaqian Huang, & Maxine Van Doren (2021). Voicing of glottal consonants and non-modal vowels. *Journal of the International Phonetic Association* 53(2), 305–332.
- Garellek, Marc, & Patricia Keating. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association* 41, 185–205. <https://doi.org/10.1017/S0025100311000193>
- Gerfen, Chip & Kirk Baker (2005). The production and perception of laryngealized vowels in Coatzacoapan Mixtec. *Journal of Phonetics* 33(3), 311–334.
- Gubian, Michele, Francisco Torreira, & Lou Boves. 2015. Using functional data analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics* 49, 16–40.
- Huang, Yaqian. (2020). Different attributes of creaky voice distinctly affect Mandarin tonal perception. *Journal of the Acoustical Society of America* 147, 1441–1458. <https://doi.org/10.1121/10.0000721>
- Josserand, Judy K. (1983). *Mixtec dialect history*. Doctoral dissertation, Tulane University.
- Kuang, Jianjing. (2013a). The tonal space of contrastive five level tones. *Phonetica* 70, 1–23. <https://doi.org/10.1159/000353853>
- Kuang, Jianjing. (2013b). *Phonation in tonal contrasts*. Ph.D. thesis, UCLA.
- Kuang, Jianjing. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *Journal of the Acoustical Society of America* 142, 1693–1706. <https://doi.org/10.1121/1.5003649>
- Lenth, Russell V. (2023). emmeans: Estimated Marginal Means, aka Least-Squares Means. *R package version 1.8.9*. <https://CRAN.R-project.org/package=emmeans>
- Longacre, Robert E. (1957). Proto-Mixtecan. *Indiana University Research Center in Anthropology, Folklore, and Linguistics* 5.
- Makowski, D., Mattan S. Ben-Shachar, & D. Lüdtke. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian Framework. *Journal of Open Source Software* 4(40), 1541. <https://doi.org/10.21105/joss.01541>
- Mantenuto, Iara. (2020). *'Other' and 'more' in SSM and beyond*. Dissertation, UCLA.
- Mittal, Vinay K., B. Yegnanarayana, & Bhaskararao P. (2014). Study of the effects of vocal tract constriction on glottal vibration. *The Journal of the Acoustical Society of America* 136(4), 1932–1941.
- Murty, K. Sri Rama & B. Yegnanarayana. (2008). Epoch extraction from speech signals. *IEEE Transactions on Audio, Speech, and Language Processing* 16, 1602–1613.
- Pike, Kenneth Lee. (1948). *Tone languages: A technique for determining the number and type of pitch contrasts in a language, with studies in tonemic substitution and fusion*. Ann Arbor: University of Michigan Press.
- Ramsay, J. & B. Silverman. 2005. *Functional data analysis*. New York, NY: Springer. Online. https://doi.org/10.1007/0-387-22751-2_8
- Rensch, Calvin R. (1976). *Comparative Otomanguean phonology*. Bloomington, IN: Indiana University.
- Shue, Yen-Liang, Patricia Keating, Chad Vicenik, & Kristine Yu. (2011). VoiceSauce: A program for voice analysis. *Proceedings of the ICPHS XVII*, 1846–1849.
- Silverman, Daniel. (1997a). Laryngeal complexity in Otomanguean vowels. *Phonology* 14(2), 235–261.

- Silverman, Daniel. (1997b). Phasing and recoverability. *Outstanding dissertations in linguistics*. London: Routledge.
- Sóskuthy, Maérton. (2017). Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. arXiv preprint arXiv:[1703.05339](https://arxiv.org/abs/1703.05339).
- Sóskuthy, Maérton. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics* 84, 101017.
- van Rij, Jacolien, Martijn Wieling, R. Baayen, & Hedderik van Rijn. (2020). itsadug: Interpreting time series and autocorrelated data using GAMMs. R package, version 2.4.
- Wood, Simon N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73(1), 3–36.
- Wood, Simon N. (2015). Package ‘mgcv’. *R package version 1*(29), 729.
- Wood, Simon N. (2017). *Generalized additive models: An introduction with R* (2nd edition). New York: Chapman and Hall/CRC.