Perception of rearticulated and checked phonations in Sierra Norte Zapotec: the effect of glottalization position and duration

Yuan Chai

Department of Linguistics, University of Washington, Seattle, 98105, USA

(Dated: 31 October 2024)

Yateé Zapotec is a Zapotec variety spoken in the Sierra Norte region of Oaxaca. It features contrastive glottalized phonations: rearticulated phonation and checked phonation. Rearticulated phonation features glottalization in the middle of vowel, whereas checked phonation features glottalization at the end. However, the exact range of "middle" and "end" remains unclear. This study for the first time investigates the effect of the position of glottalization and duration in perceiving two contrastive glottalized phonations in Zapotec. The results show that as long as there is a portion of modal voice before and after the glottalization, rearticulated vowels is more likely to be elicited. Conversely, checked vowels requires glottalization to be in vowel-final position with no following modal voicing. Duration also casts an effect on phonation perception in Zapotec: shortening the duration increases the probability of eliciting checked phonation, while lengthening the duration elicits more rearticulated phonation. Overall, glottalization position is a more effective perceptual cue than duration for distinguishing phonation types in Yateé Zapotec.

I. INTRODUCTION

Yateé Zapotec is a variety of Northern Core Zapotec, spoken in San Francisco Yateé, Oaxaca, Mexico, and by diaspora community in Los Angeles, US. According to a census conducted by the local clinic in 2017, there are 480 people in the village. Yateé Zapotec features two contrastive glottalized phonations: rearticulated phonation (Vˀ V) and checked phonation (Vˀ). These contrastive glottalized phonations have also been found in other varieties of Zapotec, such as Teotitlá del Valle (Uchihara and Gutiérrez, 2019, 2020), Isthmus (Pickett *et al.*, 2010), Choapan (Lyman and Lyman, 1977; Oliva-Juarez *et al.*, 2014), Yalálag (Avelino, 2004, 2016), Betaza (Crowhurst *et al.*, 2016; Teodocio Olivares, 2009), Texmelucan (Speck, 1978a,b, 1984), Guienagati (Benn, 2016, 2021), Zoogocho (Sonnenschein, 2004), Tabaa (Earl, 2011), and Mitla (Stubblefield and Hollenbach, 1991), and San Pablo Macuiltianguis Zapotec (Barzilai and Riestenberg, 2021). The phonetic difference between rearticulated and checked vowels in these varieties of Zapotec are mainly in two dimensions: the position of glottalization and duration. Regarding the position of glottalization, rearticulated vowels have glottalization in the middle of vowels, whereas checked vowels have glottalization at the end. However, the phonetic realization of glottalization position is known to vary. For example, Crowhurst *et al.* (2016) reported that, in non-phrase-final positions, for rearticulated vowels, glottalization can occur in the first third, first half, and first two thirds of the vowels; for checked vowels, glottalization has been found in the beginning, middle, and the end of the vowel. In Yateé Zapotec, we observed similar variability of glottalization position. We found rearticulated vowels with glottalization in the first half

(Figure 1a), middle (Figure 1b), and latter half (Figure 1c) of the vowel; and checked vowels with glottalization in the last two thirds (Figure 1d) and at the end (Figure 1e) of the vowel.

Thus, while we describe rearticulated and checked vowels as having mid-phased and late-phased glottalization, the actual phonetic realization of the "mid" and "late" phases actually occurs across a range. This raises a perceptual question: if we move the glottalization on the vowel from the beginning to the end as a continuum, at what point do listeners perceive a rearticulated vowel, and at what point do listeners perceive a checked vowel? We have not found studies that systematically repositioned glottalization along the time continuum of a vowels and tested its effect on the perception of phonation. However, some studies have involved stimuli with glottalization at different positions within the vowel, illustrating its effects in tone perception. In Vietnamese, the C1 (Chao numeral 312) and C2 (325) tones resemble the rearticulated phonation in Zapotec, with glottalization occurring in the middle of the vowel; while the B2 tone resembles the checked phonation in Zapotec, with glottalization occurring at the end of the vowel (Brunelle, 2009; Kirby, 2011). Brunelle (2009) used words with B2 and C1 tones as the base stimuli tokens and manipulated their f0. They found that, C1 and C2 tones were mostly elicited by stimuli with mid-glottalization (C1 base), while the B2 tone was elicited by stimuli with final glottalization (B2 base). Another example comes from Mandarin Chinese. Mandarin has four tones. When being produced in isolation, Tone 2 is a rising tone (15) that has the lowest f0 at the beginning of the tone, while Tone 3 (214) frequently has the lowest f0 in the middle when produced in isolation, resembling the phonetics of rearticulated phonation in Zapotec (Tseng, 1982; Xu, 1997). Huang (2018) added glottalization to the beginning of Tone 2 and to the middle of Tone 3.

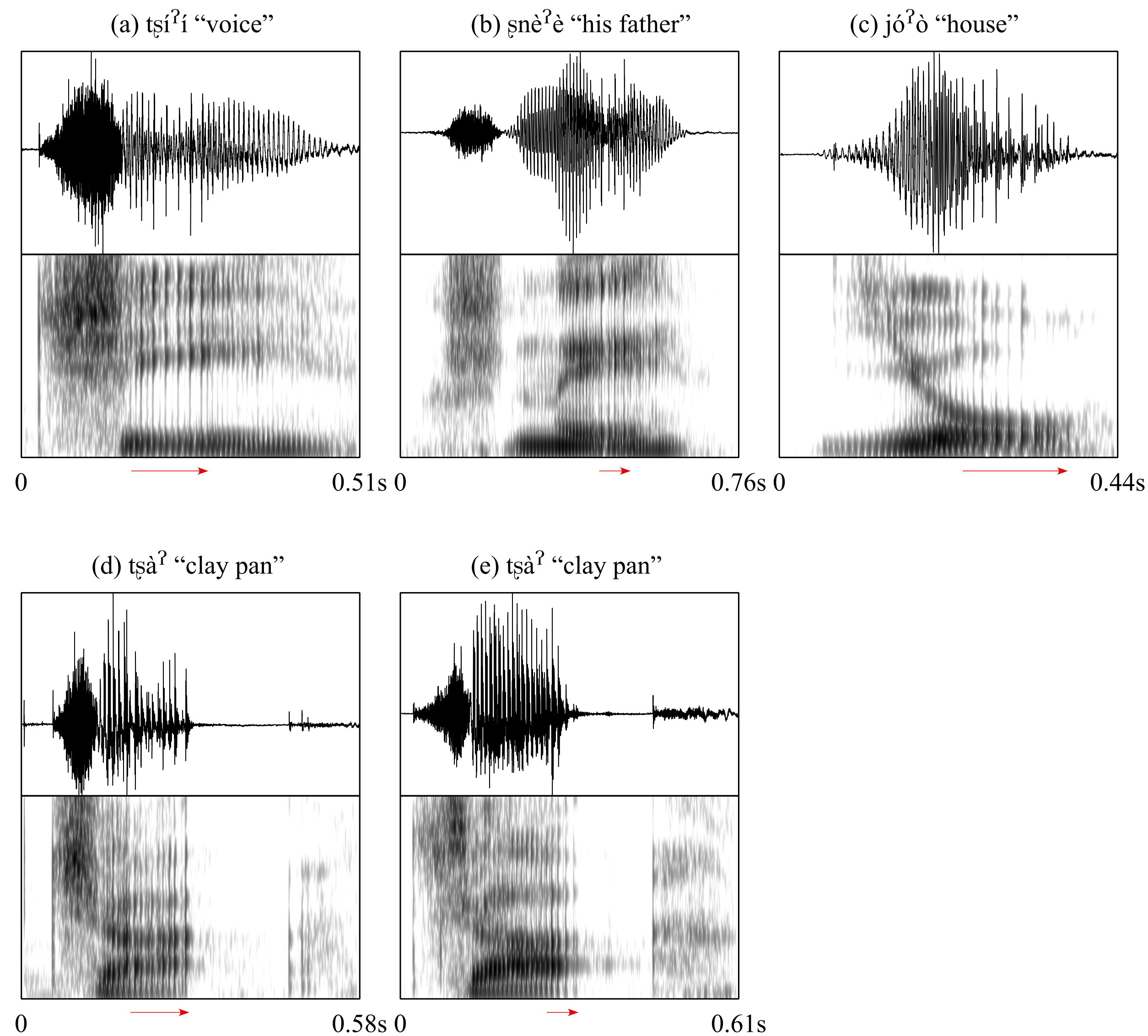

FIG. 1. Examples of words with rearticulated and checked vowels, showing varied positions of glottalization. Red arrows highlight the glottalization portion in the vowel. (a) Early glottalization in rearticulated vowel [t íˀí] “voice”; (b) Mid glottalization in rearticulated vowel [nèˀè] “his father”; (c) Late glottalization in rearticulated vowel [jóˀò] “house”; (d) Last two thirds glottalization in checked vowel [t àˀ] “clay pan”; (e) Late glottalization in checked vowel [t àˀ] “clay pan”

They found that adding glottalization decreased the identification reaction time for Tone 2 and increased the identification accuracy for Tone 3, indicating that adding glottalization to the position where the tone has the lowest f0 facilitated the perception of that specific tone.

In terms of duration, the difference between rearticulated vowel and checked vowel is fairly consistent in Zapotec. Checked vowels have been reported to be shorter compared to rearticulated and modal vowels in Yalálag (Avelino, 2004), Betaza (Teodocio Olivares, 2009), and Yateé (Chai *et al.*, 2023) Zapotec. While previous studies have established the duration differences among these three phonation types in production, this study aims to explore the perceptual function of duration. Specifically, our second research question asks: Is duration an effective cue in differentiating rearticulated phonation from checked phonation? If duration and the position of glottalization jointly distinguish rearticulated vowels from checked vowels in Zapotec, do listeners rely more on one cue than the other?

Several studies have examined the role of duration in the perception of rearticulated-like and checked-like phonetic elements. For instance, Mandarin's rearticulated-like tone, dipping Tone 3 (214), has a longer duration than the other three Mandarin lexical tones (Jongman *et al.*, 2006; Liu and Samuel, 2004; Moore and Jongman, 1997). Liu and Samuel (2004) masked the f0 cues of the four Mandarin tones by using whispered speech, and found that listeners still had above-average accuracy in identifying the original tone. Specifically, duration was highly correlated with the listeners' responses of Tone 3, such that longer durations predicted a higher likelihood of Tone 3 response. In terms of checked phonation perception, the "creaky" tone (-m) in White Hmong (Garellek *et al.*, 2013), the "glottalized" tone in Sgaw Karen (Brunelle and Finkeldey, 2011), the mid-registered checked Tone 3 in Taiwanese Min (Zhang and Lu, 2023), and the high- and the low-checked tones in Xiapu Min (Chai, 2022) share phonetic properties with the checked phonation in Zapotec. In these languages, the aforementioned perception studies have reported that shortening vowel duration significantly elicited more of these checked-like tones. Among these studies, (Garellek *et al.*, 2013) and (Chai, 2022) discussed the relative weighting of duration and glottalization as cues in tone perception: Garellek *et al.* (2013) found that in White Hmong, glottalization is a redundant cue, while duration is an effective cue for perceiving the "creaky" tone; whereas Chai (2022) suggested that while both glottalization and duration serve as effective cues for checked tone perception, duration is the more reliable cue in predicting a checked tone response.

In summary, this study aims to address two key questions: 1)In Yateé Zapotec, which part of the vowel needs to be glottalized for the listeners to perceive a rearticulated vowel, and which part for a checked vowel; 2) How does duration help differentiate rearticulated and checked vowels, and are listeners more sensitive to glottalization or duration when identifying the phonation? To answer these two questions, we created resynthesized stimuli by systematically manipulating the position of glottalization within the vowel and the vowel's duration in steps. We then conducted a word-identification experiment with native listeners of Yatee Zapotec.

II. METHOD

Yateé Zapotec has four tones—high, low, rising, and falling—and three contrastive phonations: modal, rearticulated, and checked (Chai *et al.*, 2023). Our identification task focuses on phonation identification, meaning that, ideally, the word options available to participants in the identification task would be identical in segments and tones, differing only in phonation. However, we were unable to find a minimal pair that contrasts phonation in all three types (modal, rearticulated, and checked) while maintaining identical tone and segmental structure. The closest three-way phonation contrasts we identified in Yateé Zapotec are represented by the six words listed in Table I, with their waveform and spectrogram shown in Figure 2. These six words share the segmental structure [ja] but differ in both phonation and tone: modal with falling and rising tones; rearticulated with low, rising, and falling tones; and checked with a high tone. We measured the f0 of three repetitions[1] of each word in natural production in isolation by a male speaker (see Table II), and plotted the f0 tracks over time, normalized into nine equal intervals (see Figure 3). Because the response choices in the identification task vary in tone, we needed to create an f0 contour that is ambiguous across different tones. We chose to make the f0 contour ambiguous between the rising tone (94 to 125 Hz) and the high tone (103 to 101 Hz).[2] The f0 contour that we used in the base token for the stimuli resynthesis begins at 100 Hz and ends at 115 Hz.

TABLE I. Options for identification experiment

Transcription	Tone	Phonation	Orthography	English/Spanish
[jâ]	falling	modal	ya	“reed”/“carrizo”
[jǎ]	rising	modal	yaa	“metal”
[jàˀ͡à]	low	rearticulated	ya’a	“mountain”/“cerro”
[jàˀ͡á]	rising	rearticulated	ya’a	“market place”/“plaza”
[jáˀ͡à]	falling	rearticulated	ya’a	“green”/“verde”
[jáˀ]	high	checked	ya’	“San Andres Yaa” (village name)

TABLE II. Average f0 and duration of three tokens for each word in the identification options. 1/9, 2/9, ..., 9/9 means the time interval in the vowel.

	1/9	2/9	3/9	4/9	5/9	6/9	7/9	8/9	9/9	Duration
reed	114	116	112	109	105	101	97	93	89	157 ms
metal	95	96	94	94	95	101	111	121	126	213 ms
mountain	94	97	93	82	73	73	84	85	76	268 ms
market place	92	95	93	82	84	90	106	121	123	297 ms
green	103	112	113	109	100	97	97	102	104	249 ms
San Andres Yaa	103	102	101	99	99	99	100	102	101	146 ms

A. Stimuli creation

We used a modal token [jǎ] “metal” produced by a male speaker of Yateé Zapotec as the base token of resynthesis and resynthesized it in three steps. The first step is to modify the duration of the base tokens. We manipuated the duration tier of the sound file in Praat to modify the base token into three durations: 150 ms, 225 ms, and 300 ms. The 150 ms and 300

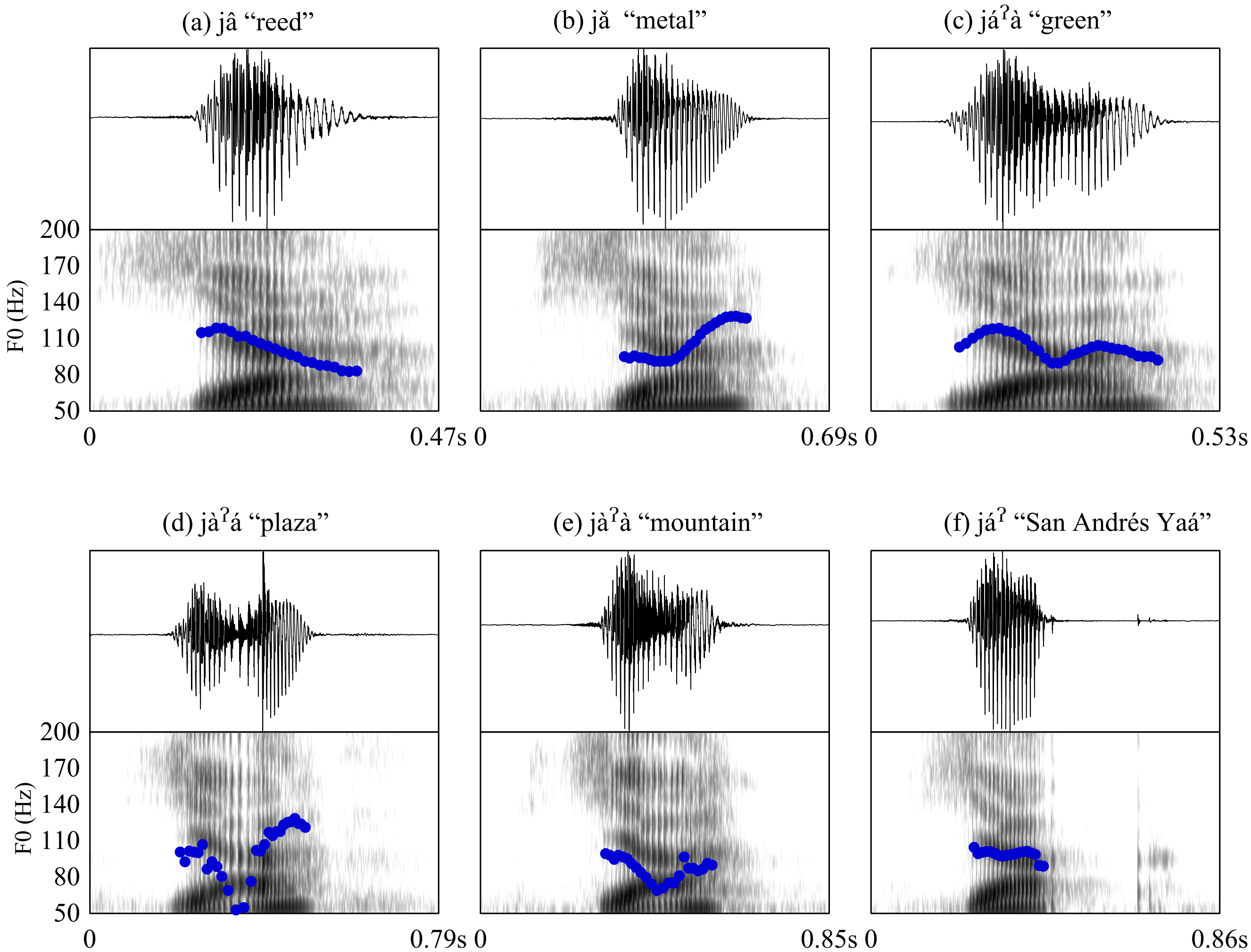

FIG. 2. Spectrograms of natural production of the options in the identification task.

ms durations are in reference to the shortest (146 ms; [jáˀ] “San Andres Yaa.”) and longest (297 ms; [jàˀá] “plaza.”) average duration (146 ms) among the six words in the identification task (Table II). The 225 ms is in the middle of the 150 ms and 300 ms conditions, and is also approximating the mean duration (213 ms) of the modal token “metal.” We selected these conditions to ensure covering the extreme short and long conditions among the three phonations in Yateé Zapotec. The second step is to modify the f0 track of the token. We

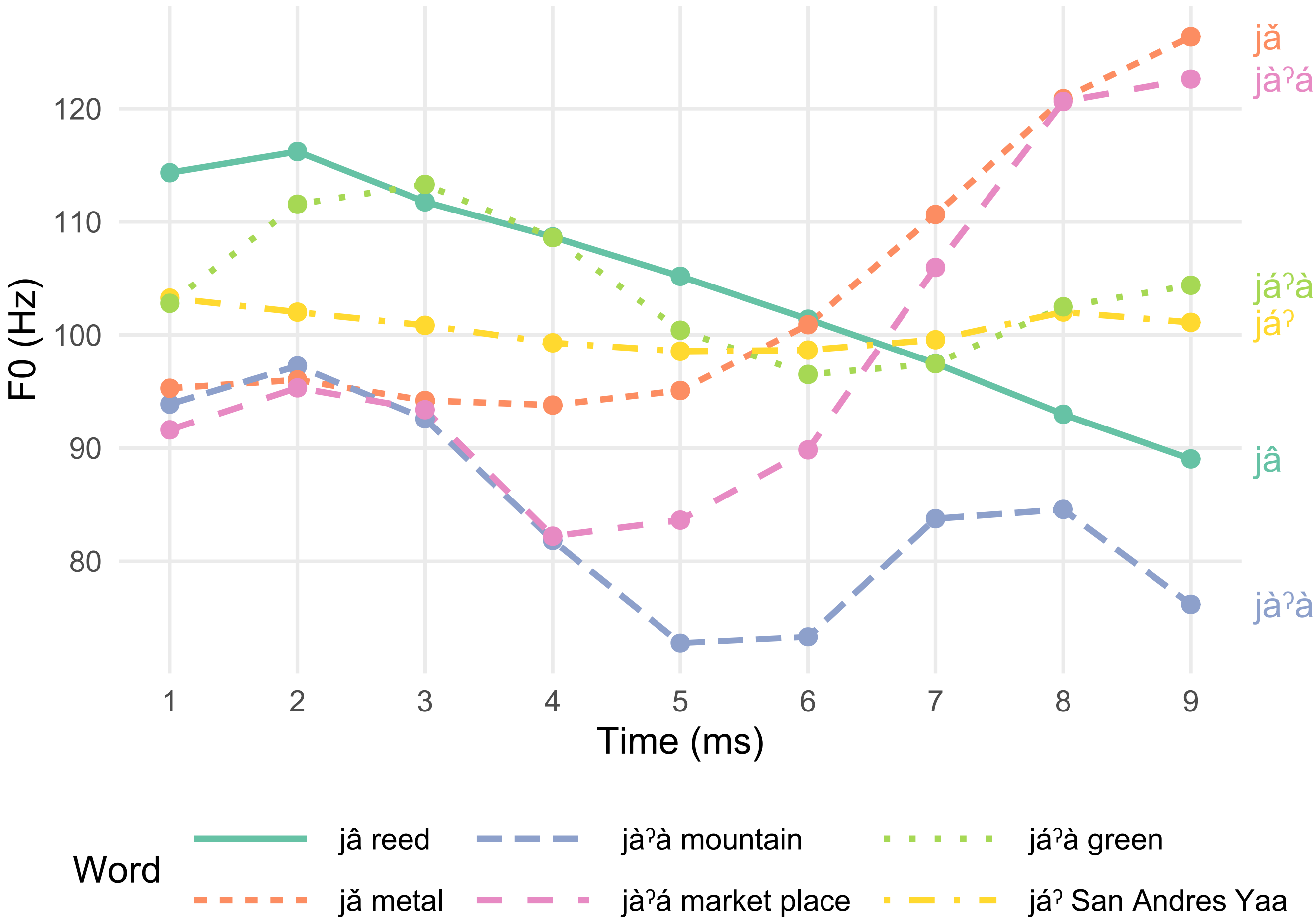

FIG. 3. Pitch track of natural productions of the word options in the identification task. The durations are normalized into nine equal-timed intervals.

used PSOLA algorithm in Praat to modify the f0 track of the tokens as starting at 100 Hz, and ending at 115 Hz, and evenly interpolate other pitch points in between the middle point of each pulse.

The third step is to create glottalization at different positions of the vowel. Each base vowel is evenly divided into five intervals. In order to create a glottalized percept, we lowered and jittered the f0, and also lowered the amplitude. Because we observed full glottal stop release in the production of checked phonation, we also synthesized full glottal stop closure and release, along with a token with vowel-final glottalization plus glottal stop. The three conditions of glottalization at 5/5 of the vowel, glottal stop, and final glottalization plus glottal stop represent three degrees of glottalization, from weak to strong. Previous studies have suggested that the degree of glottalization could be correlated with the likelihood of perceiving a glottalized phonation. Yucatec Maya has glottalized tone where there is glottalization in the middle of the vowel (Frazier, 2016). Frazier (2016) synthesized stimuli varying the degree of glottalization: weak glottalization with only one pitch point of extra-low f0; creaky voice with two pitch points of extra-low f0 and lower intensity during the extra-low f0; and full glottal stop, finding that as the degree of glottalization increases, the likelihood of the listeners selecting a glottalized tone increases. Therefore, with the stimuli varying in the degree of glottalization, we will be able to examine if the observation in Frazier (2016) is replicable in Yateé Zapotec. In total, we created 24 conditions—3 durations (150, 225, 300 ms) * 8 glottalization positions (no glottalization; 1/5, 2/5, 3/5, 4/5, 5/5 glottalization; glottal stop; 5/5 glottalization + glottal stop). The waveform and spectrogram of the resynthesized stimuli for stimuli with a 300 ms duration are in Figure 4.

B. Participants and procedure

Twenty-four individuals participated in the experiment (14 women, 10 men; average age: 43). All participants identified Zapotec as their primary language and were bilingual in Zapotec and Spanish. The identification task consisted of three parts: listening to the natural productions of the six words in the response options, listening to resynthesized stimuli, and producing the words from the identification options. The first and third parts of the

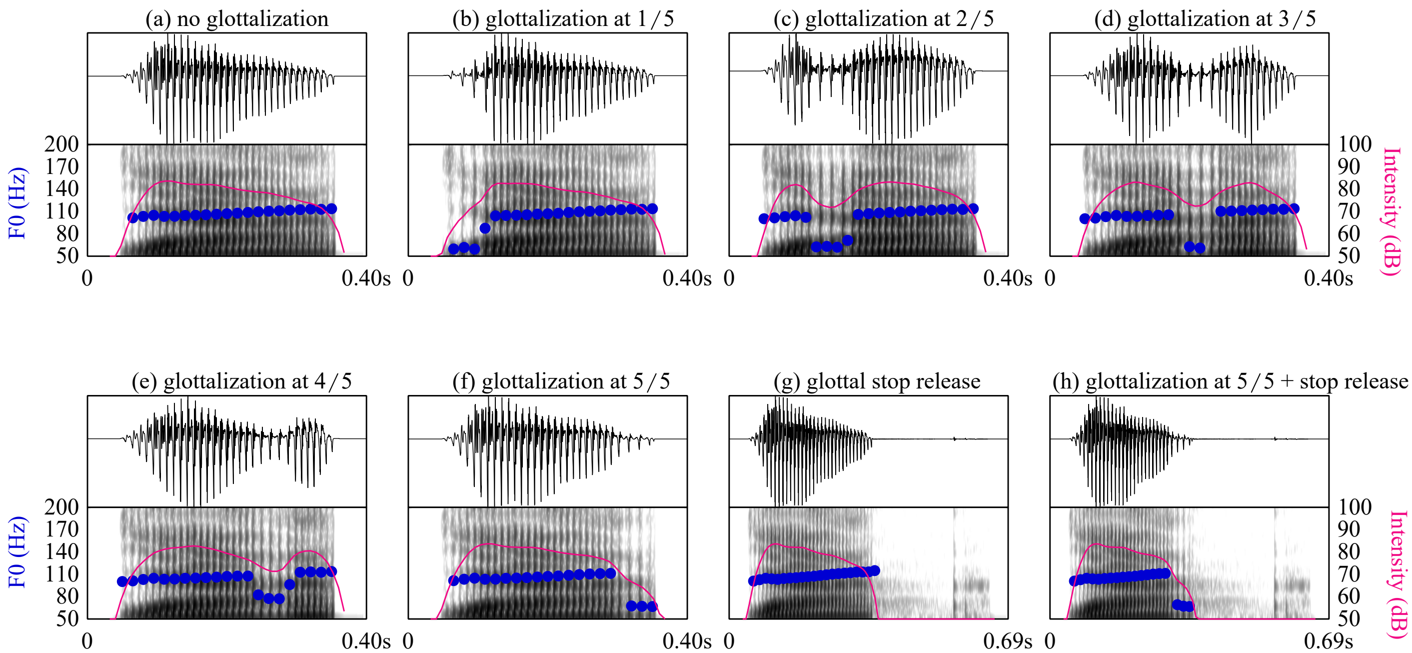

FIG. 4. Waveforms and spectrograms of resynthesized stimuli with 300 ms duration, and eight different glottalization positions. Blue dots represent f0; pink lines represents intensity.

task served as screening steps to determine participant eligibility for analysis. During our field research, we realized that there is notable variability in tone and phonation production across speakers. Thus, Part I was used to assess whether participants shared the same understanding of phonation and tone for each word as intended in our stimuli. For example, if a participant correctly identified the word “mountain” when listening to the natural production of “mountain [jà͡ˀà],” we could assume that, in subsequent tasks, their selection of “mountain” likely indicates a perception of rearticulated phonation. In contrast, if a participant selected “metal [jǎ]” in response to the natural production of “mountain [jà͡ˀà],” it suggests that they might not be aware of the phonation difference between “mountain” and “metal” in Zapotec. As a result, we cannot assume that their selection of “mountain” in later tasks reflects the intended rearticulated phonation. In Part 1, nine out of twenty-four participants correctly identified the phonation for all natural stimuli. However, a “wrong” selection in this part did not necessarily indicate a lack of phonation awareness; it might reflect that the natural token presented was not prototypical for some listeners. To further confirm participants’ understanding, we used the third part, a production task. Here, the participants were instructed to produce each word in the identification task three times. For words incorrectly identified in Part 1, we checked if the participants produced them with the phonation that we expected in the production task. Based on this criterion, ten additional participants who made incorrect selections in Part 1 perception test nonetheless produced the correct phonation in the production test. In total, nineteen participants (10 women, 9 men; average age: 44) were included in the final analysis. Among the five excluded participants, three were younger speakers (average age: 27) who appeared to exhibit a less robust distinction between phonation and tone. The remaining participant (age: 79) had a different vocabulary item for the word “reed” and was excluded from the analysis.

Part II contains all the test trials for the identification task. The participants listen to the test stimuli. Each word in the test stimuli is presented in the orthography of Zapotec and its Spanish translation. Each word is also represented with a image, because some participants were not literate in Zapotec orthography. Part II was split into two sub-sections. The 24 stimuli tokens were played to the participants once in each section in a random order. The listeners can listen to each token as many times as they desire by pressing the ”Replay (Reproducir)” button. Figure 5 shows the page display of a question in Part II. In total, we elicited 888 responses (48 questions * 18 participants + 24 questions * 1 participant). We have to exclude the first sub-section of one participant because they did not fully understand the task in the first section.

FIG. 5. A sample page for the Part II test trials.

III. RESULTS

We summarized the percentage of each condition in Table III, illustrating the general trends in phonation elicitation by glottalization position and duration. Checked phonation is elicited predominantly by glottalization at the end of the vowel, by vowel-final glottal stop, and by glottalization followed by a glottal stop. Additionally, checked phonation is elicited by shorter vowel durations. In contrast, rearticulated phonation is more likely to be elicited when glottalization occurs between the second fifth and fourth fifth of the vowel and is associated with longer vowel durations. Modal phonation is most commonly elicited in conditions without glottalization.

To reveal the more detailed interactions between specific glottalization and duration combinations, we visualized the response percentages for each condition in a heatmap in Figure 6. In the heatmap, darker colors indicating higher percentage of eliciting a specific phonation type within that specific combination of glottalization position and duration. We chose to analyze conditions with a probability higher than one-third for each phonation type, as this threshold represents an above-average probability, given that there are three different phonations to choose from in this experiment.

In Figure 6, we observe several glottalization positions that consistently elicit a specific phonation type response with a probability exceeding one-third, regardless of the duration condition. For rearticulated phonation, glottalization positions at the 2/5, 3/5, and 4/5 of the vowel consistently elicit responses with a probability over one-third across all durations. Checked phonation responses exceed one-third probability for conditions with glottalization at the 5/5 of the vowel, with glottal stop, and with the combination of 5/5 glottalization plus glottal stop, independent of duration. In conditions without glottalization, modal phonation consistently receives a probability larger than one-third, regardless of duration. These findings highlight the glottalization positions that favor each phonation type when considered across all durations.

Other specific combinations between glottalization position and duration also elicit responses with greater than one-third probability for specific phonation. When the glottalization is at 1/5 of vowel, 150 ms predominantly elicits checked phonation; 225 ms modal; 300

ms rearticulated. In addition, the no-glottalization condition with durations of 150 ms and 300 ms yields probability over one-third for checked phonation. Modal phonation responses exceed one-third probability with 4/5 glottalization at 150 ms. These observations suggest that duration, combined with glottalization position, plays a role in phonation perception. A more detailed exploration of these response patterns and their potential causes will be discussed in the Section IV.

TABLE III. Percentage of checked, rearticulated, and modal responses by fixed effects

	glottalization								Duration		
	no gl	1/5	2/5	3/5	4/5	5/5	gl release	5/5+gl release	150	225	300
Checked	36.04	34.23	17.12	14.41	7.21	59.46	63.96	75.68	50.34	35.47	29.73
Rearticulated	14.41	38.74	65.77	75.68	72.07	18.92	18.02	10.81	26.01	41.22	50.68
Modal	49.55	27.03	17.12	9.91	20.72	21.62	18.02	13.51	23.65	23.31	19.59

To complement our observations in the descriptive data, we conducted a statistical test to determine, for each condition of glottalization position and duration, which phonation response has a significantly higher probability of elicitation than the other phonations. For this purpose, we fit a multinomial mixed-effects model with the selected phonation as the dependent variable, glottalization position and duration as the predictors, and a random intercept for each participant. The model was fit using a Bayesian approach with the *brms* package (Bürkner, 2021) in R.

In the model, the priors for all the slopes have a normal distribution with mean of 0 and standard deviation of 10. This prior centers the slope at 0, assuming no strong initial

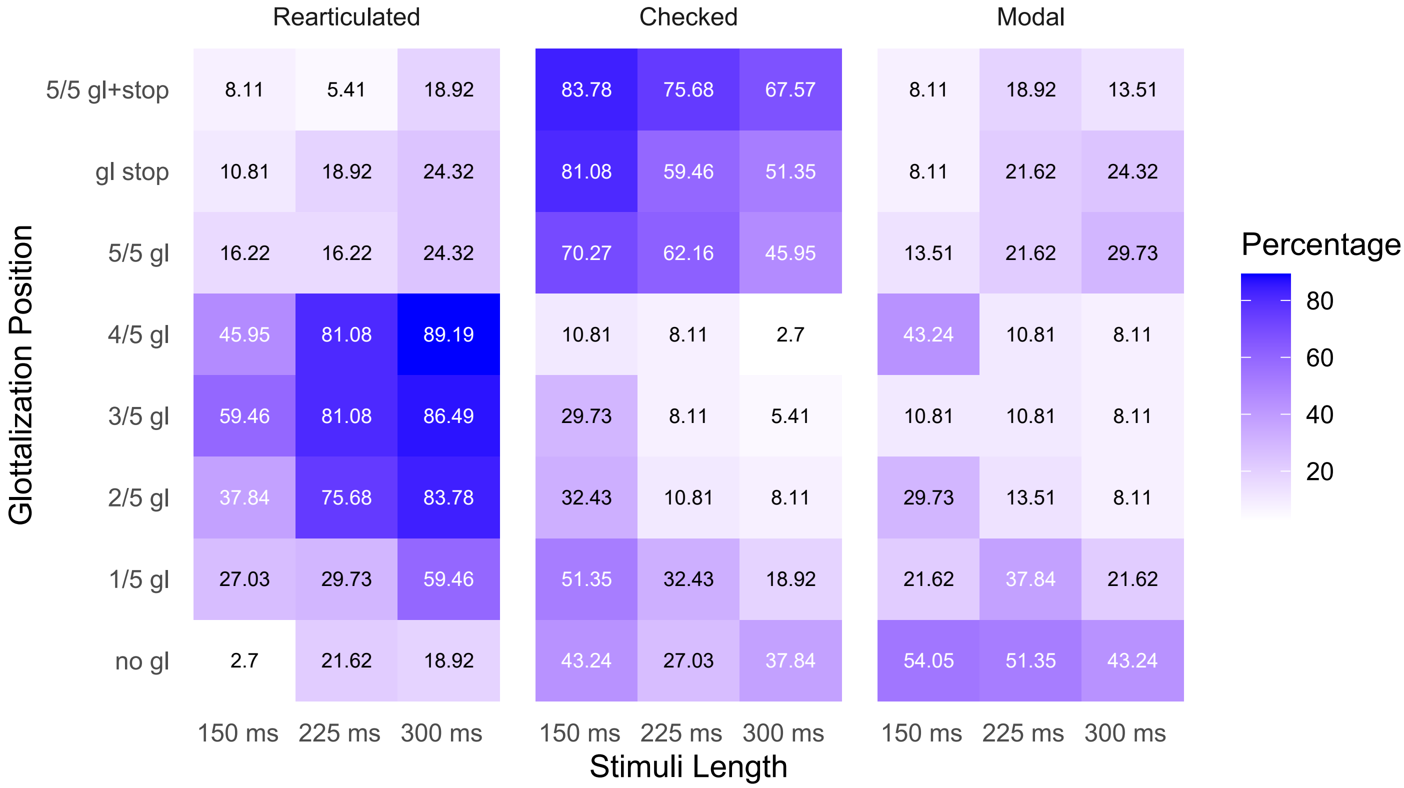

FIG. 6. Percentage of responses of rearticulated, checked, and modal vowel by stimuli condition. The number in each cell represent the percentage of the specific response in the specific condition of the cell (i.e. Number 2.7 in the bottom left corner represents in the condition of 150 ms duration and no glottalization, among all the responses in that condition, 2.7% of the responses has checked phonation.). Percentages higher than 34% is marked with white color. The darkness of the background color in each condition is correlated with how large the percentage is. The higher the percentage, the darker the color.

bias in either direction, while a standard deviation of 10 provides enough flexibility to cover a wide range of effect sizes. All the variables are coded in dummy coding. The baseline condition is glottalization position at 5/5 and duration of 150 ms. This condition has a mean probability of around 0.5 (Figure 6). Thus, the standard deviation of 10 will be able to capture probabilities across the full 0 to 1 range, making the priors to be weakly informative for the slopes[3]. The prior for the random intercept is the default setting in the *brms* package—a half-Student' s t-distribution prior, which is also a weakly informative prior (Bürkner, 2017). As there is no prior research directly addressing how glottalization phasing and vowel duration affect phonation perception, these weakly informative priors were selected to minimize the influence of prior assumptions on posterior predictions. The model was fit with 4 chains, each running for 10,000 iterations (2,000 for warm-up), as recommended in Vasishth *et al.* (2018). Convergence was assessed via R-hat values, all of which equaled to 1. Effective sample sizes for each parameter were sufficiently large ($>$ 1000), indicating reliable parameter estimation.

Because our goal is to compare the probability of the checked, rearticulated, and modal responses in each condition, we drew 4000 posterior predictions for each of the 456 unique observations in the data (456 = 8 glottalization positions * 3 durations * 19 participants) using the *posterior_epred()* function in the *brms* package (Bürkner, 2017) in R. Each prediction provided estimation of the probability of each phonation response for each specific observation. We calculated the mean of the probability for each phonation in each condition, and the 95% credible interval by getting the 2.5% and 97.5% quantile of all the predicted probability. These probabilities represent marginal effects, illustrating the likelihood of each phonation at each glottalization position (or duration), averaged over the other factors (participants and either duration or glottalization position, respectively).

In Figure 7, for each level of each predictor, we plotted the distribution of the predicted probability, alongside the mean and 95% confidence interval. When two response categories do not show overlapping confidence intervals, we interpret them as differing significantly in their predicted probabilities. Using this criterion, for glottalization position, when there is no glottalization, the predicted probabilities for checked and modal responses are significantly higher than for rearticulated phonation. At the 1/5 position, the predicted probabilities for all three phonation types do not differ significantly. At the 2/5, 3/5, and 4/5 positions, the predicted probability of eliciting rearticulated phonation is higher than the other phonations. In addition, in the 4/5 position, the predicted probability of a modal response is significantly higher than for checked phonation. When glottalization occurs at 5/5, with a glottal stop, or as a combination of 5/5 glottalization plus glottal stop, the predicted probability of eliciting checked phonation is higher than the other two phonations.

For duration, the results show that in the 150 ms condition, checked responses have a higher predicted probability than modal and rearticulated responses. In the 225 ms condition, both checked and rearticulated responses are predicted to be more probable than modal responses. In the 300 ms condition, rearticulated responses have a higher probability than checked responses, and checked responses are more probable than modal responses.

By examining the descriptive data, we observe that glottalization position appears to be a stronger predictor of phonation perception than duration. Specifically, certain glottalization positions consistently elicit a dominant phonation response (over 1/3 probability) across all durations. In contrast, no single duration condition elicits a dominant phonation response across all glottalization positions. This suggests that glottalization position may play a more definitive role in influencing phonation perception. To statistically evaluate this observation, we used a random forest model to calculate importance scores for glottalization position and duration. We used the *cforest()* function in the *randomForest* package (Breiman, 2001) in R. The model grew 500 trees in total (ntree = 500). Two predictors (i.e. both the glottalization

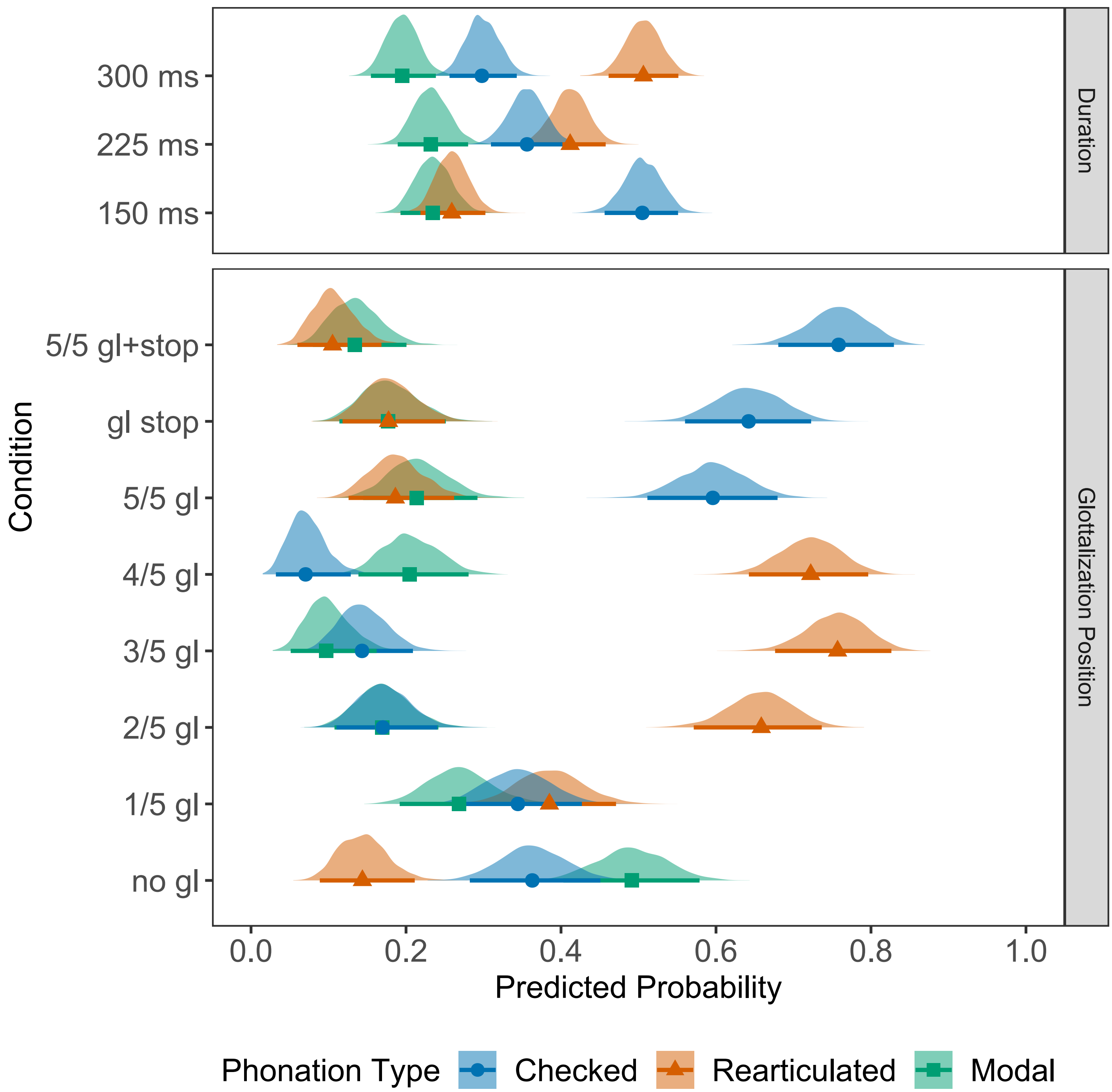

FIG. 7. Posterior prediction of the possibility of the phonation response at eight different glottalization position levels and three duration levels. The density plots show the distributions of the probability for each specific phonation response among the 4000 iterations. The error bar represent the 2.5% to 97.5% quantile (i.e. 95% confidence interval) of the 4000 iterations over 456 observations in the data.

position and the duration predictors) were sampled at each node (mtry = 2). The dataset was divided into an 80% training set and a 20% test set, with the selected phonation type as the dependent variable and glottalization position and duration as predictors. The resulting importance scores were 0.22 for glottalization position and 0.023 for duration, indicating that glottalization position is more influential in predicting phonation perception. We tested the random forest model on the test data. The classification accuracy is 0.591 (chance level = 0.392; $p < 0.001$), suggesting that the random forest model is effective in making predictions for unseen data.

While Random Forest models calculate the weighting among the predictors in the model, it does not directly demonstrate the relationship between the predictors and the responses. In order to more directly demonstrate what conditions lead to what phonation responses, and how the predictor of glottalizatoin position is more dominant than the predictor of duration in predicting the phonation responses, we constructed a classification tree using the same training and test sets as the random forest model. The classification tree was created with ten-fold cross-validation and a tune length of 100, implemented using the *rpart* package (Therneau *et al.*, 2023) in R. Based on the best tuning results, we selected a complexity parameter (cp) value of 0.002. We set a minimum split and bucket size of 12, slightly above the chance level of 11 based on the category frequencies in the training data. This threshold helps capture splits that represent decisions with a higher than chance probability. The resulting decision tree, shown in Figure 8, illustrates that glottalization position predominantly determines phonation type: glottalization in the 2/5, 3/5, and 4/5 segments of the vowel tends to elicit rearticulated responses; glottalization in the 5/5 position, glottal stop release, and the combination of 5/5 glottalization with glottal stop leads to checked responses; and the absence of glottalization generally results in modal responses. Duration only affects phonation perception when the glottalization position is less definitive. At the 1/5 glottalization position, shorter durations (150 ms) lead to checked phonation, mid-range durations (225 ms) result in modal phonation, and longer durations (300 ms) elicit rearticulated phonation. The decision tree clearly demonstrates glottalization is more effective in determining the phonation response. Glottalization position alone decided 87% of the responses; whereas duration decided only 12% of the responses.

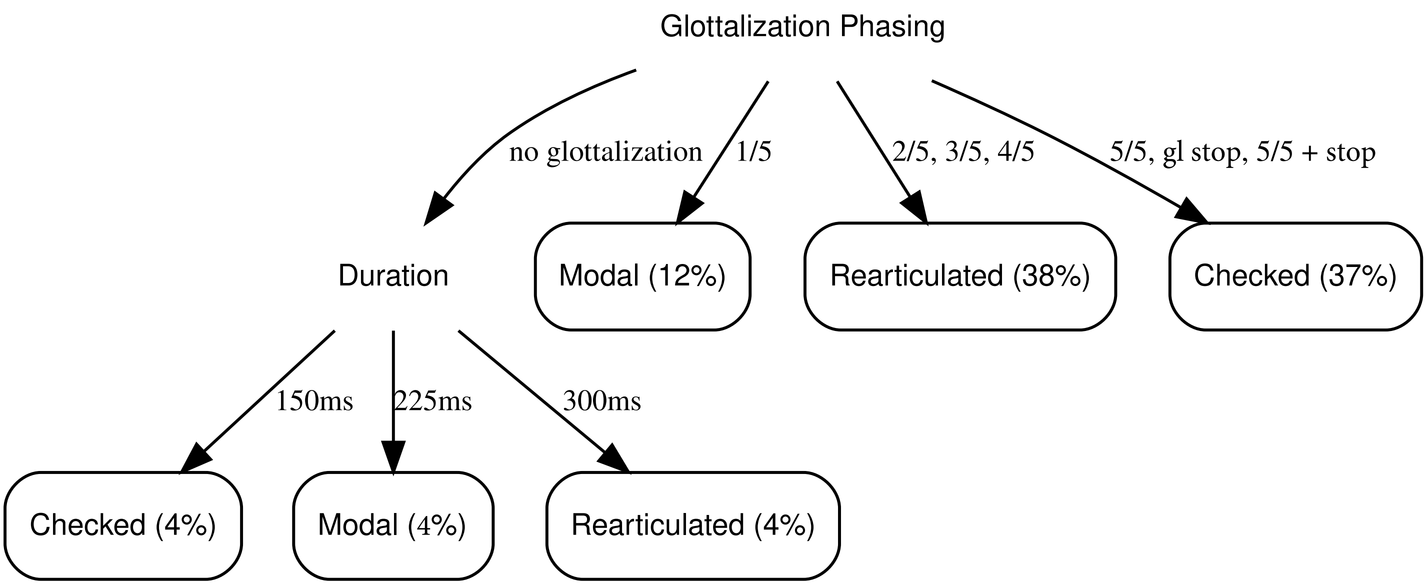

FIG. 8. Classification tree of the relation between the cue and the perceived phonation.

IV. DISCUSSION

Our study addresses the following questions: (1) Which part of the vowel needs to be glottalized for listeners to perceive a rearticulated vowel? (2) Does vowel duration play a role in phonation differentiation, and if so, do listeners rely more on duration or glottalization cues? By resynthesizing glottalization at different positions of the vowel and eliciting listeners' identification of vowel phonation, we observed that the absence of glottalization leads to a modal phonation percept, middle-position glottalization (2/5, 3/5, and 4/5) elicits a rearticulated percept, and final-position glottalization (5/5, glottal stop, and 5/5 plus glottal stop) results in a checked phonation percept. These findings reflect that the requirements for eliciting a rearticulated phonation percept are relatively flexible: the glottalization may occur in various parts of the vowel's middle section, whether early-middle, middle, or late-middle. As long as there is a modal portion before and after the glottalization, a rearticulated percept is likely. In contrast, the glottalization position for checked vowels is more restricted, requiring glottalization to occur at the very end of the vowel with no modal portion following. Glottalization at the 1/5 position creates an ambiguous percept, eliciting modal, checked, and rearticulated responses at chance levels. This ambiguity is consistent with production patterns in Yateé Zapotec, as no phonation consistently shows glottalization only at the beginning of the vowel in natural productions.

The degree of glottalization also impacts perception. While vowel-final glottalization generally leads to a high probability of a checked phonation percept, stronger degrees of glottalization increase the likelihood of this response. For instance, the predicted probability of checked phonation ranks glottalization $<$ glottal stop $<$ glottal stop + glottalization. The non-overlapping credible intervals between the glottalization and glottal stop + glottalization conditions suggest a significant difference in checked phonation elicitation between these categories. This finding aligns with previous work in Yucatec Maya (Frazier, 2016), suggesting that listeners use the degree of glottalization as a cue to enhance the glottalized phonation perception.

Our data reveal two notable patterns regarding modal phonation responses: (1) modal phonation is most likely to be elicited in conditions with no glottalization, but its probability remains relatively low even in the most likely conditions; and (2) modal responses appear unexpectedly in certain conditions, particularly in the 150 ms and 4/5 glottalization condition, where rearticulated phonation would generally be expected.

For the first pattern, we propose two explanations. First, in Yateé Zapotec, modal vowels in open syllables in utterance-final positions often feature a breathy quality. This could mean that participants needed a breathy phonation to consistently select the "modal" response. Second, the f0 contour used in our stimuli is not the prototypical f0 of naturally produced modal words in this language, potentially causing perceptual ambiguity. In our experiment, the modal phonation word [jǎ] has an f0 contour starting at 95 Hz and ending in 126 Hz. The f0 of the stimuli used in the current experiment is between 100 to 115 Hz, which may affect the listeners to be less inclined to select the modal word.

The second trend—the relatively high percentage of modal responses in the condition of 4/5 glottalization with 150 ms duration—is probably due to the briefness of the modal portion after the glottalizaiton. The overall duration of 150 ms is short. When the glottalization is at 4/5 of the vowel, the modal portion after the glottalization is only 30 ms (compared with glottalization at 3/5 with modal portion of 60 ms; see Figure 9). Since rearticulated vowel favors long duration, stimuli in this condition are not stereotypical tokens for rearticulated vowel, reducing the probability of eliciting a rearticulated phonation, creating ambiguity of the phonation type. Since the checked phonation percept strongly disfavor any modal portion after the glottalization, the ambiguity has to between the rearticulated phonation and modal phonation, leading the probability of modal phonation reponse to be relatively high in this condition. Future studies can test stimuli with even shorter modal duration after the glottalization to see whether listeners consistently perceive modal phonation for short vowels with glottalization in late-medial position.

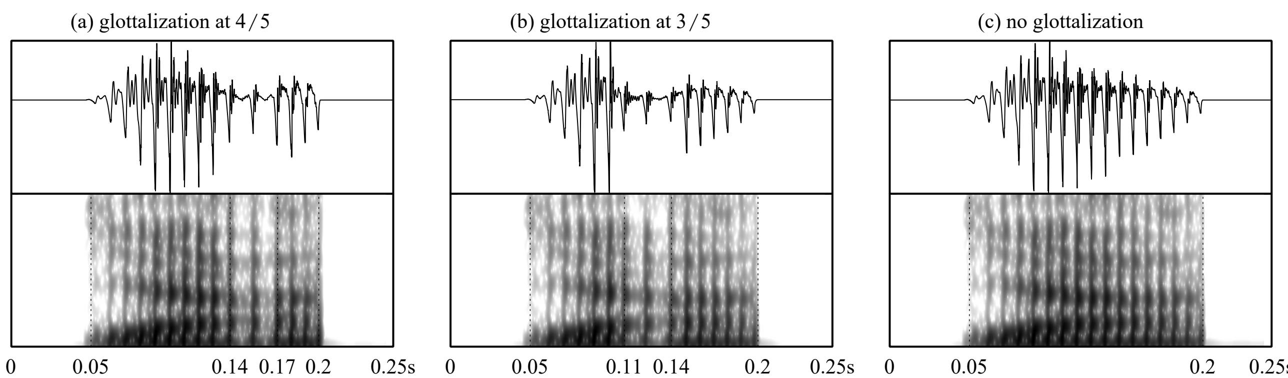

FIG. 9. Waveform and spectrogram for stimuli with (a) 150 ms and 4/5 glottalization; (b) 150 ms and 5/5 glottalization; (c) 150 ms without glottalization

Our findings indicate that duration also influences phonation perception. The shortest duration condition (150 ms) leads to more checked responses, while the longest duration (300 ms) elicits more rearticulated responses. Across durations, the confidence intervals for checked responses rank as 150 > 225 > 300 ms, whereas rearticulated responses follow the opposite ranking, supporting duration as an additional cue in phonation differentiation.

The random forest model and the classification tree analyses further support the importance of glottalization position over duration. Random Forest models show higher importance scores for glottalization positioning, and the decision tree analysis reveals that glottalization predominantly determines phonation type, with duration only contributing when glottalization is ambiguous (e.g., at vowel-initial positions).

When comparing Yateé Zapotec to other languages reviewed in Section I, we find its similarities with Vietnamese, where glottalization positioning influences rearticulated and checked phonation perception, and with Mandarin, Sgaw Karen, and Taiwanese Min, where duration also plays a role. In contrast, Yateé Zapotec differs from White Hmong and Xiapu Min, where listeners prioritize duration over glottalization in perceiving low creaky tones.

Future research can explore more levels in the duration predictor. In the current experiment, as vowel duration increases, the glottalization duration is proportionally stretched. It remains unclear whether the observed duration effect is due to the duration of the modal portion, the glottalization portion, or a combination. Future studies could isolate these factors by fixing glottalization duration while varying the modal portion or vice versa to dissect these components further. Future research could also examine the role of f0 in phonation perception. While this study used an ambiguous f0 contour, future studies can create stimuli that vary in f0 and glottalization position independently. This design can test when two words differ in both tone and phonation, whether the listeners will prioritize tone or phonation in word identification.

[1]One repetition for the word “cerro” and one repetition for the word “market place” were excluded from the analysis because of failure of pitch tracking in the glottalization portions of these vowels.

[2]Checked phonation occurs only with the high tone in our stimuli options, so we first aimed to make the f0 ambiguous between high and another tone. We then needed a tone present in both rearticulated and

modal phonations, which limited our choices to the rising and falling tones. The rising tone was chosen due to its similarity in f0 shape and height between rearticulated and modal phonations, whereas the falling tone showed more contour differences between these phonations. To ensure ambiguity across phonations, we therefore created an f0 contour that is ambiguous between high and rising tones.

[3]With normal distribution normal(0,10), there is 95% probability that the slope's value falls between -20 to 20. The slope represents the difference in log odds between the target level and the reference level. The reference level has a probability around 0.5 and a log odds around 1. If the log odds of the target level is larger than the base level by 20, its probability is almost equal to 1; if the log odds of the target level is lower than the base level by 20, its probability is almost equal to 0. Thus, with the normal(0,10) prior for the slopes, the model should be able to capture all the possible probabilities between 0 to 1.

Avelino, H. (**2004**). "Topics in Yalálag Zapotec, with particular reference to its phonetic structures," Dissertation, University of California, Los Angeles, Los Angeles, CA.

Avelino, H. (**2016**). "Phonetics in Phonology: A Cross-Linguistic Study of Laryngeal Contrast," in *The Phonetics and Phonology of Laryngeal Features in Native American Languages*, edited by H. Avelino, M. Coler, and W. L. Wetzels (Brill, Leiden/Boston), pp. 157–179.

Barzilai, M. L., and Riestenberg, K. J. (**2021**). "Context-dependent phonetic enhancement of a phonation contrast in San Pablo Macuiltianguis Zapotec," Glossa: a journal of general linguistics **6**(1), https://www.glossa-journal.org/article/id/5398/, doi: 10.5334/gjgl.959.

Benn, J. (**2016**). "Consonant-Tone-Phonation Interactions in Guienagati Zapotec," in *5th International Symposium on Tonal Aspects of Languages (TAL 2016)*, ISCA, Buffalo, New

York, pp. 125–128, doi: 10.21437/TAL.2016-27.

Benn, J. (**2021**). "The phonetics, phonology, and historical development of Guienagati Zapotec," PhD diss., State University of New York at Buffalo, Buffalo, New York.

Breiman, L. (**2001**). "Random Forests," Machine Learning **45**(1), 5–32, http://link.springer.com/10.1023/A:1010933404324, doi: 10.1023/A:1010933404324.

Brunelle, M. (**2009**). "Tone perception in Northern and Southern Vietnamese," Journal of Phonetics **37**(1), 79–96, doi: 10.1016/j.wocn.2008.09.003.

Brunelle, M., and Finkeldey, J. (**2011**). "Tone perception in Sgaw Karen," in *Proceedings of the ICPhS XVII 2011*, Hong Kong, pp. 372–375.

Bürkner, P.-C. (**2017**). "**brms** : An *R* Package for Bayesian Multilevel Models Using *Stan*," Journal of Statistical Software **80**(1), http://www.jstatsoft.org/v80/i01/, doi: 10.18637/jss.v080.i01.

Bürkner, P.-C. (**2021**). "Bayesian Item Response Modeling in R with brms and Stan," Journal of Statistical Software **100**, 1–54, doi: 10.18637/jss.v100.i05.

Chai, Y. (**2022**). "Phonetics and phonology of checked phonation, syllables, and tones," PhD diss., University of California San Diego, San Diego, CA.

Chai, Y., Fernández, A., and Mendez, B. (**2023**). "Phonetics of glottalized phonations in Yateé Zapotec," in *Proceedings of the 20th International Congress of Phonetic Sciences*, edited by R. Skarnitzl and J. Volín, Guarant International, Prague, Czech, pp. 1751–1755.

Crowhurst, M. J., Kelly, N. E., and Teodocio Olivares, A. (**2016**). "The influence of vowel laryngealisation and duration on the rhythmic grouping preferences of Zapotec speakers," Journal of Phonetics **58**, 48–70, doi: 10.1016/j.wocn.2016.06.001.

Earl, R. (**2011**). "Gramática del Zapoteco de Tabaa," https://www.sil.org/resources/archives/40770.

Frazier, M. (**2016**). "Pitch and Glottalization as Cues to Contrast in Yucatec Maya," in *The phonetics and phonology of laryngeal features in Native American languages*, edited by H. Avelino, M. Coler, and W. L. Wetzels (Brill, Leiden/Boston), pp. 203–234.

Garellek, M., Keating, P., Esposito, C. M., and Kreiman, J. (**2013**). "Voice quality and tone identification in White Hmong," The Journal of the Acoustical Society of America **133**(2), 1078–1089, doi: 10.1121/1.4773259.

Huang, Y. (**2018**). "Tones in Zhangzhou: Pitch and Beyond," Ph.D. thesis, Australian National University, Canberra, Australia, publisher: The Australian National University.

Jongman, A., Wang, Y., Moore, C. B., and Sereno, J. A. (**2006**). "Perception and production of Mandarin Chinese tones," in *The Handbook of East Asian Psycholinguistics*, edited by P. Li, L. H. Tan, E. Bates, and O. J. L. Tzeng (Cambridge University Press, Cambridge), pp. 209–217, https://www.cambridge.org/core/product/identifier/CBO9780511550751A031/type/book_part, doi: 10.1017/CBO9780511550751.020.

Kirby, J. (**2011**). "Vietnamese (Hanoi Vietnamese)," Journal of the International Phonetic Association **41**, 381–392, doi: 10.1017/S0025100311000181.

Liu, S., and Samuel, A. G. (**2004**). "Perception of Mandarin lexical tones when F0 information is neutralized," Language and Speech **47**, 109–138.

Lyman, L., and Lyman, R. (**1977**). "Choapan Zapotec phonology," in *Studies in Otomanguean phonology*, edited by W. R. Merrifield, number 54 in SIL International Publications in Linguistics (Summer Institute of Linguistics and the University of Texas at

Arlington, Mexico), pp. 137–161, https://www.sil.org/resources/archives/8836.

Moore, C. B., and Jongman, A. (**1997**). "Speaker normalization in the perception of Mandarin Chinese tones," The Journal of the Acoustical Society of America **102**(3), 1864–1877, https://pubs.aip.org/jasa/article/102/3/1864/557637/Speaker-normalization-in-the-perception-of, doi: 10.1121/1.420092.

Oliva-Juarez, G., Martinez-Licona, F., Martinez-Licona, A., and Goddard-Close, J. (**2014**). "Identification of Vowel Sounds of the Choapan Variant of Zapotec Language," in *Nature-Inspired Computation and Machine Learning*, edited by A. Gelbukh, F. C. Espinoza, and S. N. Galicia-Haro, Springer International Publishing, Cham, pp. 252–262.

Pickett, V. B., Villalobos, M. V., and Marlett, S. A. (**2010**). "Isthmus (Juchitán) Zapotec," Journal of the International Phonetic Association **40**(3), 365–372, doi: 10.1017/S0025100310000174.

Sonnenschein, A. H. (**2004**). "A descriptive grammar of San Bartolomé Zoogocho Zapotec," PhD diss., University of Southern California, Los Angeles, California.

Speck, C. H. (**1978**a). "The Phonology of Texmelucan Zapotec Verb Irregularity," Master's thesis, University of North Dakota, Grand Forks, ND, https://commons.und.edu/theses/2660.

Speck, C. H. (**1978**b). "Texmelucan Zapotec suprasegmental phonology," Work Papers of the Summer Institute of Linguistics, University of North Dakota Session **22**(1), doi: 10.31356/silwp.vol22.09.

Speck, C. H. (**1984**). "The Phonology of the Texmelucan Zapotec Verb," International Journal of American Linguistics **50**(2), 139–164, https://www.jstor.org/stable/1265602 publisher: University of Chicago Press.

Stubblefield, M., and Hollenbach, B. E. (**1991**). *Gramática zapoteca: Zapoteca de Mitla, Oaxaca* (Instituto Lingüístico de Verano, A.C.), https://www.sil.org/resources/archives/88999.

Teodocio Olivares, A. (**2009**). "Betaza Zapotec Phonology: Segmental and Suprasegmental Features," Master's thesis, University of Texas at Austin, Austin, TX, https://repositories.lib.utexas.edu/handle/2152/19162.

Therneau, T., Atkinson, B., and Ripley, B. (**2023**). "rpart: Recursive Partitioning and Regression Trees" https://cran.r-project.org/package=rpart.

Tseng, C.-y. (**1982**). "An acoustic phonetic study on tones in Mandarin Chinese," Ph.D. Dissertation, Brown University.

Uchihara, H., and Gutiérrez, A. (**2019**). "El texto Don Crescencio: ilustración del sistema tonal del zapoteco de Teotitlán del Valle," Tlalocan **24**, 127–155, doi: 10.19130/iifl.tlalocan.2019.487.

Uchihara, H., and Gutiérrez, A. (**2020**). "Subject and agentivity in Teotitlán Zapotec," Studies in Language **44**(3), 548–605, doi: 10.1075/sl.18025.uch.

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., and Kong, E. J. (**2018**). "Bayesian data analysis in the phonetic sciences: A tutorial introduction," Journal of Phonetics **71**, 147–161.

Xu, Y. (**1997**). "Contextual tonal variations in Mandarin," Journal of Phonetics **25**, 61–83.

Zhang, W., and Lu, Y.-A. (**2023**). "The role of duration in the perception of checked versus unchecked tones in Taiwanese Southern Min," in *Proceedings of 20th International Congress* *of Phonetics Science*, edited by R. Skarnitzl and J. Volín, GUARANT International spol. s r.o., Prague, Czech, pp. 226–230.