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

Yuan Chai

Department of Psychiatry, University of San Diego, La Jolla, 92037, USA

(Dated: 1 November 2025)

Yateé Zapotec, spoken in the Sierra Norte region of Oaxaca, contrasts two glottalized phonations: rearticulated (glottalization in the middle of the vowel) and checked (glottalization at the end of the vowel), though their precise temporal boundaries have not been clearly defined. This study is the first to systematically test how glottalization position, proportion, and degree, together with vowel duration, shape the perception of these phonations. In Experiment I, nineteen listeners listened to stimuli differing in glottalization position and vowel duration, and identified words out of a six-way near-minimal set contrasting in phonation and tone. The results show that, as long as there is a portion of modal voice before and after the glottalization, listeners are more likely to identify the word as having a rearticulated vowel. Conversely, identifying a word with a checked vowels requires glottalization to be in vowel-final position, with no following modal voicing. Duration also has 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. Experiments II–IV used strict minimal pairs between each two phonations to validate the results from Experiment I. Experiment II confirmed that glottalization position is the strongest cue determining the perception of phonation, with duration playing a secondary role. Experiment III focuses on the ambiguity found in stimuli with vowel-initial glottalization in Experiment I, and found that increasing the proportion of vowel-initial glottalization (to at least one third of the vowel) consistently yields rearticulated responses. Experiment IV, together with Experiment I, demonstrated that stronger glottalization increases glottalized responses. These findings show that listeners are highly sensitive to subsegmental timing and amplitude cues in Yateé Zapotec, and that glottalization position is the most influential cue in perceiving contrastive phonations.

I. INTRODUCTION

Yateé Zapotec is a variety of Northern Core Zapotec, spoken in San Francisco Yateé, Oaxaca, Mexico, and by the 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 phonation types with glottalization: rearticulated phonation ($\widehat{\mathrm{V^{ʔ}V}}$) and checked phonation ($\mathrm{V^{ʔ}}$). 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 is mainly described in terms of the position of glottalization and duration. As for 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 Betaza Zapotec, 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 primarily describe rearticulated and checked vowels as having mid-phased and late-phased glottalization, the actual phonetic realization of the "mid" and "late" phases varies substantially. 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? To my knowledge, no previous study has directly tested the effect of the position of glottalization on the perception of rearticulated vs. glottalized phonation types. 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, which has four tones. When being produced in isolation, Tone 2 is a rising tone (35) that has the lowest f0 at the beginning of the tone, while Tone 3 (214) frequently has the lowest f0 (and concomitant glottalization) in the middle when produced in isolation, resembling the phonetics of rearticulated phonation in

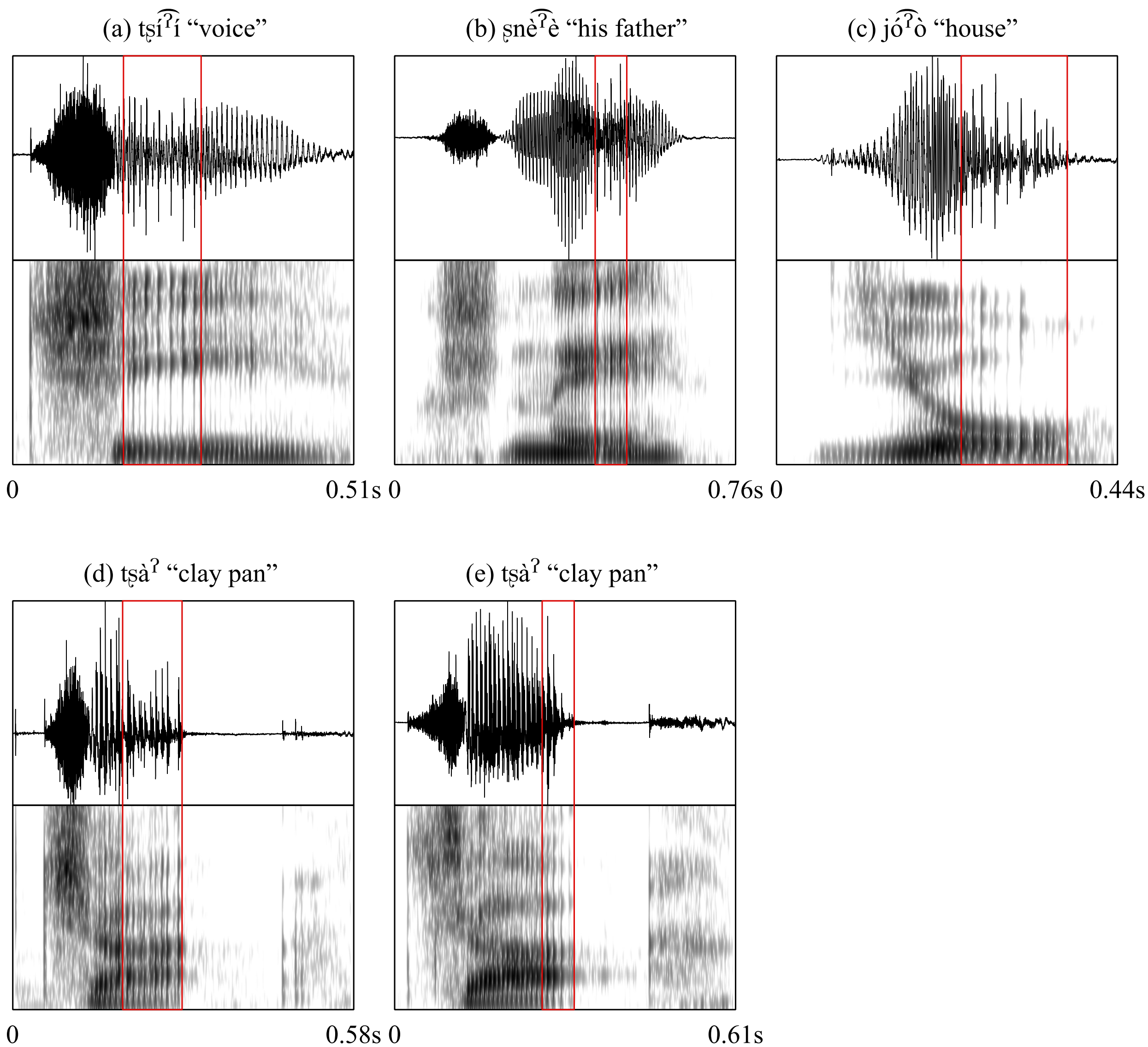

FIG. 1. Examples of words with rearticulated and checked vowels, showing varied positions of glottalization. Red boxes highlight the glottalization portion in the vowel. (a) Early glottalization in the rearticulated vowel of [tʂ̩í͡ˀí] “voice”; (b) Mid glottalization in the rearticulated vowel of [nè͡ˀè] “his father”; (c) Late glottalization in the rearticulated vowel of [jó͡ˀò] “house”; (d) Glottalization during the final two-thirds in the checked vowel of [tʂ̩àˀ] “clay pan”; (e) Late glottalization in the checked vowel of [tʂ̩àˀ] “clay pan”

Zapotec (Tseng, 1982; Xu, 1997). Huang (2018) added glottalization to the beginning of Tone 2 and to the middle of Tone 3. 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 cooccur with the tone's 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 when in isolation (Jongman *et al.*, 2006; Liu and Samuel, 2004; Moore and Jongman, 1997; Xu, 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. In contrast, Chai suggested that in Xiapu Min, both glottalization and duration serve as effective cues for checked tone perception, but duration is the more reliable cue.

In summary, this study aims to address three questions: 1) In Yateé Zapotec, which part of the vowel needs to be glottalized for the listeners to perceive a rearticulated vs. a checked vowel? 2) How does duration help differentiate rearticulated and checked vowels? and 3) Are listeners more sensitive to glottalization or duration when identifying the phonation? To answer these 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. EXPERIMENT I

A. Method

Yateé Zapotec has four tones—high, low, rising, and falling—and three contrastive phonations: modal, rearticulated, and checked. Phonation and tone are independent of each other (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 set 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] and differ in phonation and/or 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). To tackle the issue that the tone is not identical in all word options, we made the f0 contour ambiguous between the rising tone (94 to 125 Hz) and the high tone (103 to 101 Hz) to motivate the listeners to pay less attention to tone information and make judgment based on phonation[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”
[jà͡ˀà]	low	rearticulated	ya’a	“mountain”/“cerro”
[jà͡ˀá]	rising	rearticulated	ya’a	“market place”/“plaza”
[já͡ˀà]	falling	rearticulated	ya’a	“green”/“verde”
[jáˀ]	high	checked	ya’	“San Andres Yaa” (village name)

TABLE II. Average f0 (Hz) 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
[jâ] reed	114	116	112	109	105	101	97	93	89	157 ms
[jǎ] metal	95	96	94	94	95	101	111	121	126	213 ms
[jà͡ˀà] mountain	94	97	93	82	73	73	84	85	76	268 ms
[jà͡ˀá] market place	92	95	93	82	84	90	106	121	123	297 ms
[já͡ˀà] green	103	112	113	109	100	97	97	102	104	249 ms
[jáˀ] San Andres Yaa	103	102	101	99	99	99	100	102	101	146 ms

1. *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 two steps. The first step is to modify the duration of the base tokens. We manipuated the duration tier of the sound file in Praat (Boersma and Weenink, 2023) to modify the base token into three durations: 150 ms, 225

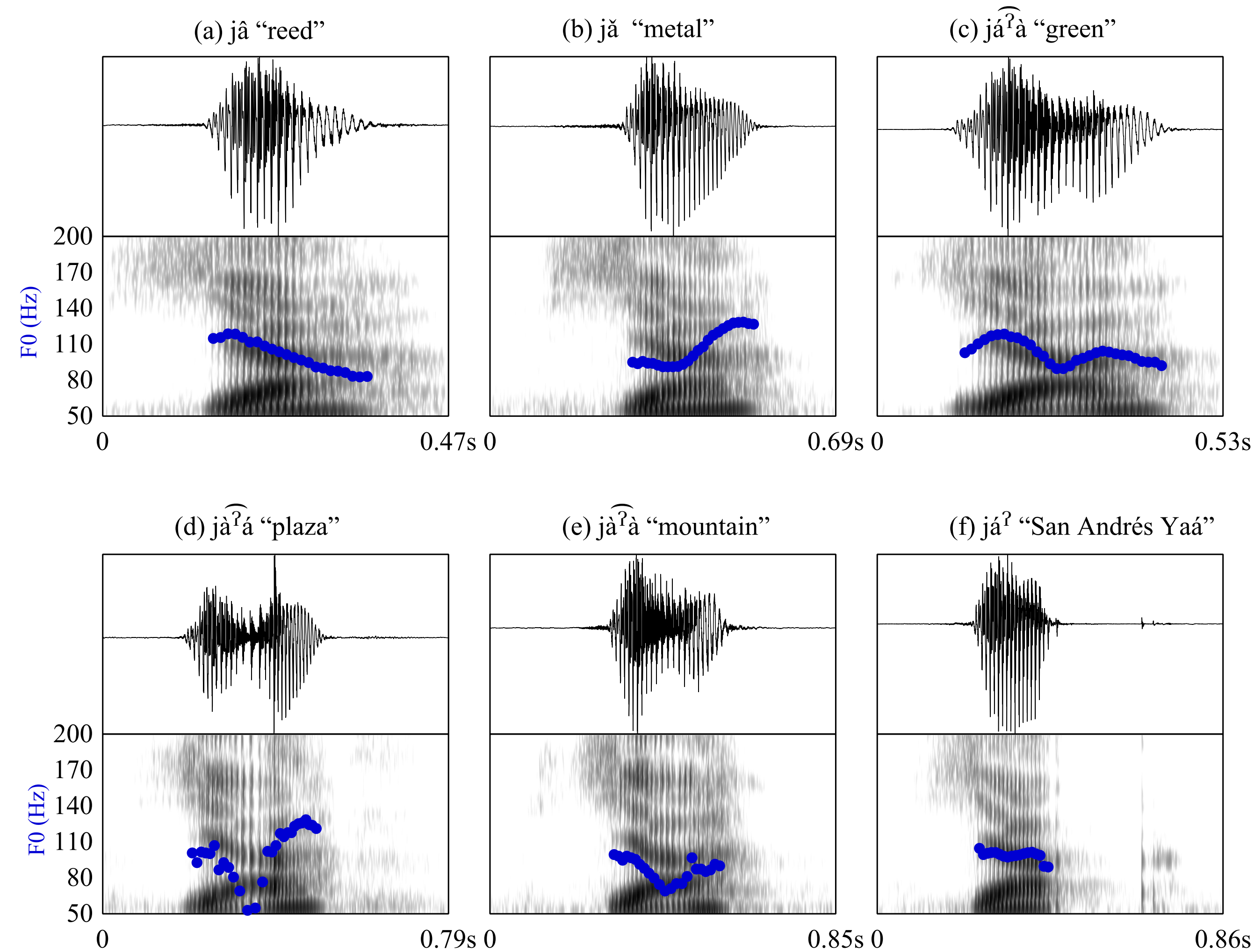

FIG. 2. Waveforms and spectrograms (showing 0-5000 Hz) of natural production of the options in the identification task. The blue contours overlaid on the spectrogram refer to the f0 contour. The y-axis of the spectrogram is showing the f0 range.

ms, and 300 ms. The 150 ms and 300 ms durations are in reference to the shortest (146 ms; [jáʔ] “San Andres Yaa.”) and longest (297 ms; [jà͡ʔá] “plaza.”) average duration among the six words in the identification task (Table II). The 225 ms falls within the middle of the 150 ms and 300 ms conditions, and is also approximating the mean duration (213 ms) of the

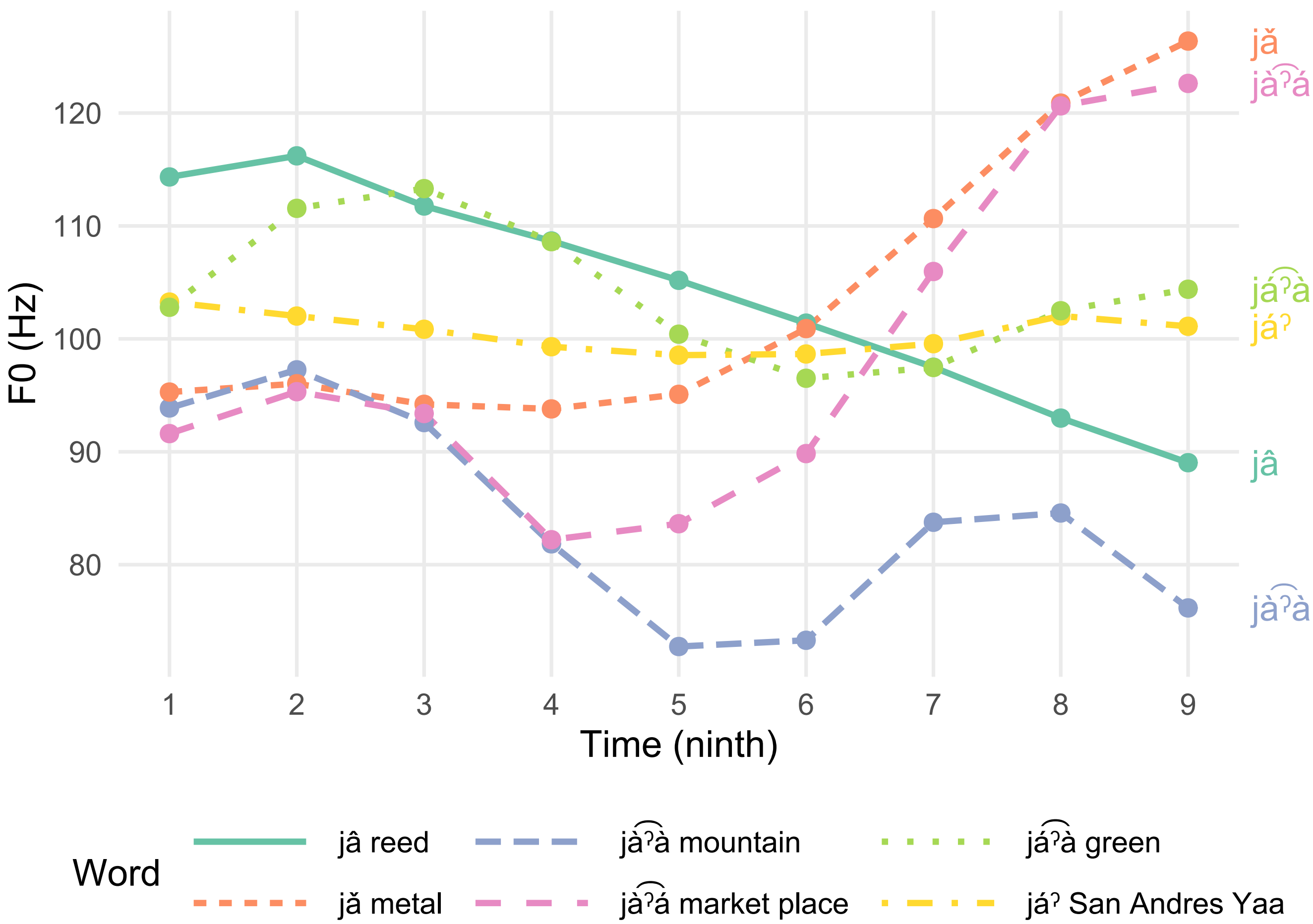

FIG. 3. Pitch track of natural productions of the word options in the identification task. The durations are normalized into nine equal-timed intervals.

modal token [jǎ] "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 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 product after the first steps result in three tokens for the stimuli for the "no glottalization" condtions with 150 ms, 225 ms, and 300 ms. Based on the three "no glottalization" tokens, we then 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 aim to make the stimuli to sound natural, the exact value of pitch and intensity adjustment differ slightly for each condition. We plotted the original and adjusted f0 and intensity values for all stimuli with a duration of 300 ms in Figure 4. The Praat scripts used for creating glottalization are available in the Supplementary Material at https://doi.org/10.17605/OSF.IO/SA2TD

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. We used a cross-splicing method, combining the first part of a vowel and the second part of a glottal stop into one stimuli token. For the no glottalization plus glottal stop condition, the first part of the token is the stimuli with no glottalization created earlier. For the 5/5 glottalization plus glottal stop condition, the first part is the stimuli with 5/5 glottalization created in the earlier step. The glottal stop release burst is extracted from a natural token of [jáˀ] "San Andres Yaa," produced by the same speaker who produced the base token of the stimuli. We found that in that natural token [jáˀ], the amplitude of the glottal stop release is half of the voicing portion in the word. Thus, we adjusted the glottal stop release amplitude to the half of its first part (i.e. stimuli with no glottalization and stimuli with 5/5 glottalization), then concatenated the glottal stop release to the end of the first part.

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 5. The audio of the stimuli are available in the Supplementary Materials at https://doi.org/10.17605/OSF.IO/SA2TD.

2. Participants and procedure

Twenty-four individuals (14 women, 10 men; average age: 43) participated in the experiment in San Francisco Yateé, Oaxaca, Mexico. 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 task serve to determine participant eligibility for

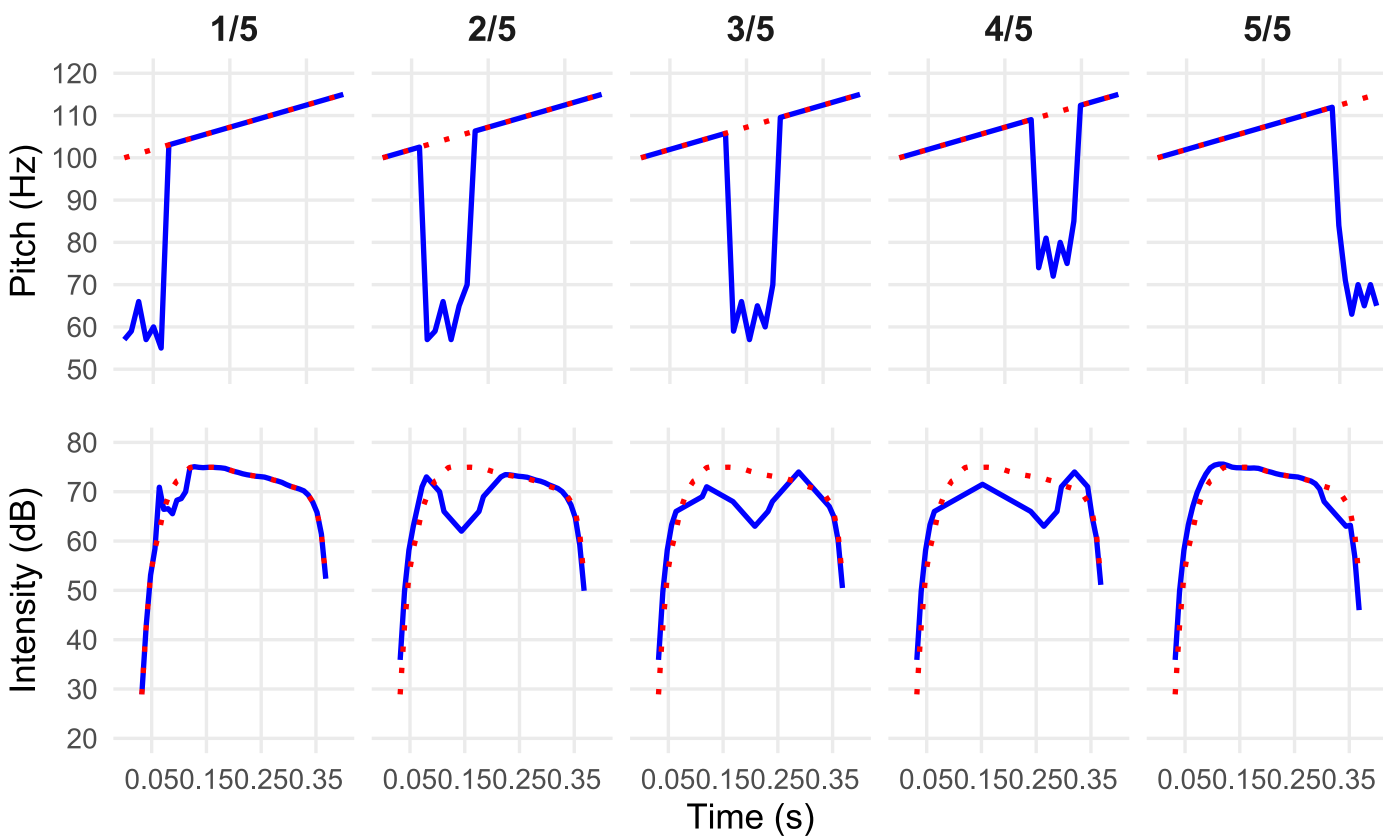

FIG. 4. Pitch track and intensity track that are superimposed on the base modal token with 300 ms duration in Step 3 of stimuli resynthesis. The blue lines represent the manipulated values. The red dotted lines represent the pitch and intensity of the base modal token.

analysis. During our field research, we realized that there is notable variability in tone and phonation production across speakers, especially among younger speakers. To ensure that we analyze data from participants who speak the same variety of Zapotec as the elderly speakers of Yateé Zapotec, we used Part I and III of the experiment for a screening purpose. For example, in Part I, 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.

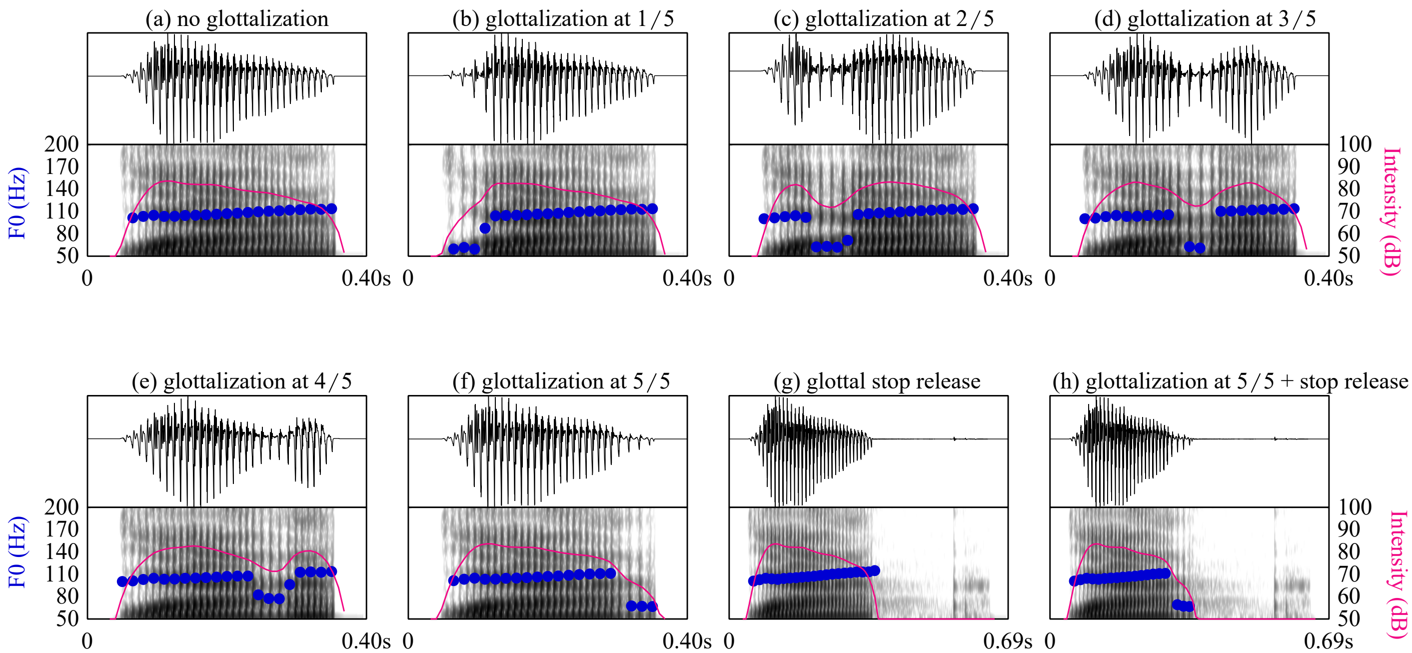

FIG. 5. Waveforms and spectrograms of resynthesized stimuli with 300 ms duration, and eight different glottalization positions. Blue dots represent f0; pink lines represents intensity.

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 I, 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 I, we checked if the participants produced them with the "correct" phonation that we expected. Based on this criterion, ten additional participants who made incorrect selections in Part I perception test nonetheless produced the correct phonation in the production test and were included in the analysis. 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" [jâ] 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. The experiment can be accessed at https://yuanchaiyc.github.io/zapotecperception/. 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. The data and the scripts for data analysis are available in the Supplementary Materials at https://doi.org/10.17605/OSF.IO/SA2TD.

B. Results

We summarized the percentage of phonation identified in each condition in Table III, illustrating the general trends in phonation elicitation by glottalization position and duration. Checked phonation responses are elicited predominantly by stimuli with glottalization at the end of the vowel, by those ending in a vowel-final glottal stop, and by those with vowel-final glottalization followed by a glottal stop. Additionally, checked phonation responses are elicited by stimuli with 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 stimuli 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 highlighted the phonation responses that received the highest percentage in each condition and analyze the pattern of under what conditions each phonation become the most popular choice.

In Figure 6, we observe several glottalization positions where a specific phonation type consistently receives the highest probability, regardless of the duration condition. Glottalization at the 2/5, 3/5, and 4/5 of the vowel consistently elicits rearticulated responses as the majority. Vowels with a 5/5 glottalization, ending in a glottal stop, and with the combination of 5/5 glottalization plus glottal stop predominantly elicit checked responses. In stimuli without glottalization, modal responses consistently receives a the highest probability among the three phonation responses.

When the glottalization is at 1/5 of vowel, the majority phonation response changes by duration. For stimuli that have the shortest duration (150 ms), the responses are largely "checked." When the duration is longer, the responses are largely "modal." For stimuli with the longest duration (300 ms), the majority of the responses are "rearticulated."

Additionally, there is one outlier response, which is the modal phonation response in the condition of 4/5 glottalization with 150 ms. While the rearticulated response has the highest probability among the three phonations, modal response also received a relatively high probability (43.24%). More analyses about the conditions that lead to this modal phonation response will be discussed in Section VII.

TABLE III. Percentage (%) of checked, rearticulated, and modal responses by fixed effects. The majority response is bolded in each condition.

	glottalization								Duration		
	no gl	1/5	2/5	3/5	4/5	5/5	gl stop	5/5+gl stop	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

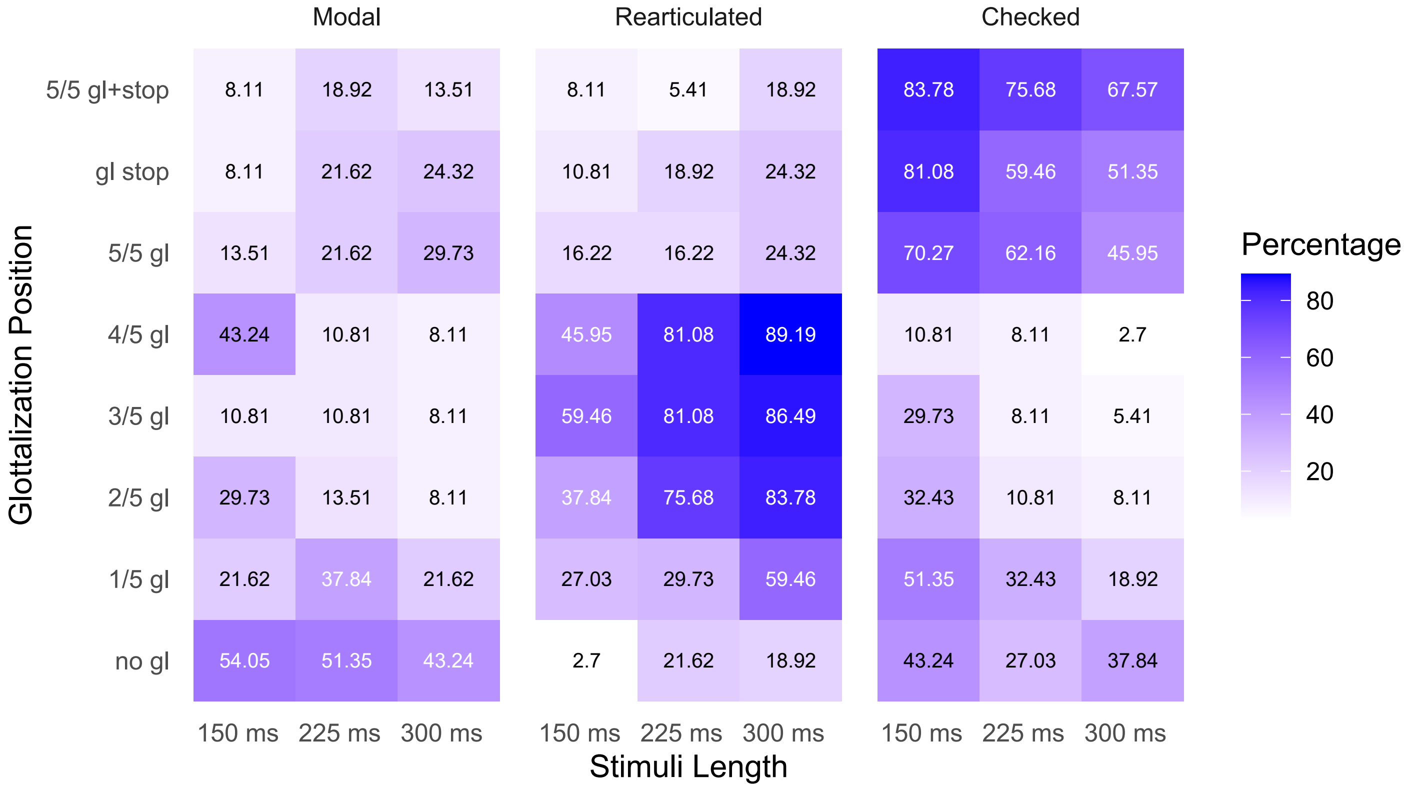

FIG. 6. Percentage of responses of rearticulated, checked, and modal vowel by stimuli condition. The number in each cell represent the percentage of the specific response in the specific condition of the cell (i.e. Number 2.7 in the bottom left corner represents in the condition of 150 ms duration and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

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 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 dummy-coded. 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 few research directly addressing how glottalization phasing and vowel duration affect modal vs. rearticulated vs. checked phonation identification, 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 responses. At the 1/5 position, the predicted probabilities for all three phonation responses do not differ significantly. At the 2/5, 3/5, and 4/5 positions, the predicted probability of eliciting rearticulated responses is higher than responses with the other phonations. In addition, in the 4/5 position, the predicted probability of modal responses is significantly higher than for checked responses. For vowels with 5/5 glottalization, with a vowel-final glottal stop, or as a combination of 5/5 glottalization plus vowel-final glottal stop, the predicted probability of eliciting checked responses is higher than responses with 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. The results suggest that modal responses are not strongly influenced by duration; checked responses increase as duration decreases; and rearticulated responses increase as duration increases.

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 position and the duration predictors) were sampled at each node (mtry = 2). The dataset was divided into an 80% training set (712 observations) and a 20% test (176 observations) 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,

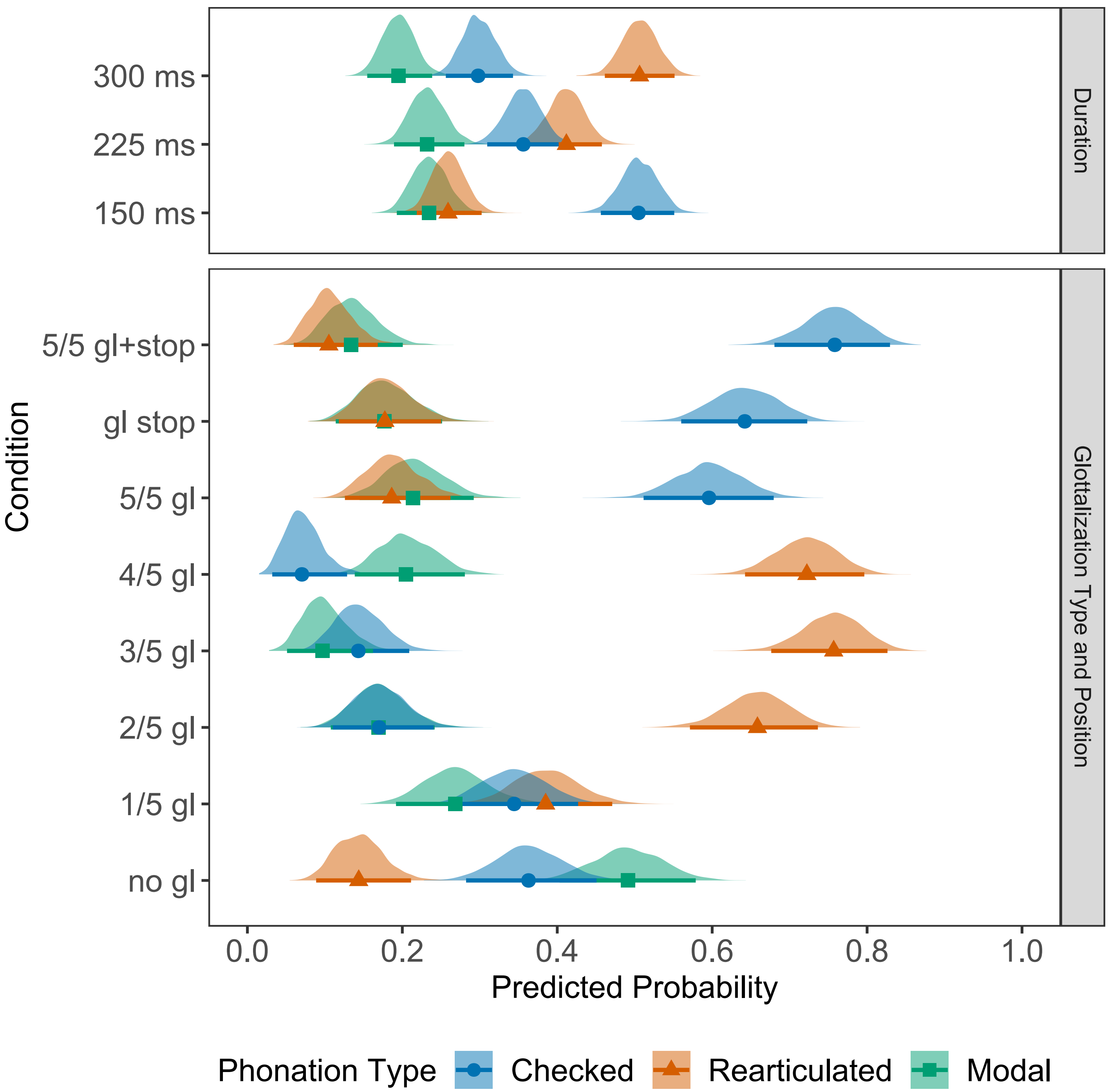

FIG. 7. Posterior prediction of the possibility of the phonation response at eight different glottalization position levels and three duration levels. The density plots show the distributions of the probability for each specific phonation response among the 4000 iterations. The error bar represent the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

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 to capture splits that represent the majority decision in each given condition[4] The resulting classification tree, shown in Figure 8, illustrates that glottalization position predominantly determines the phonation type of the responses in all conditions except the 1/5 position condition. At the 1/5 glottalization position, shortest duration (150 ms) leads to checked responses, mid-range duration (225 ms) results in modal responses, and longest duration (300 ms) elicits rearticulated responses. The classification tree demonstrates that glottalization is more effective than duration in determining the phonation of the responses: glottalization position alone decided 88% of the responses; whereas duration decided only 12% of the responses.

C. Discussion

Experiment I 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 middle-position glottalization (2/5,

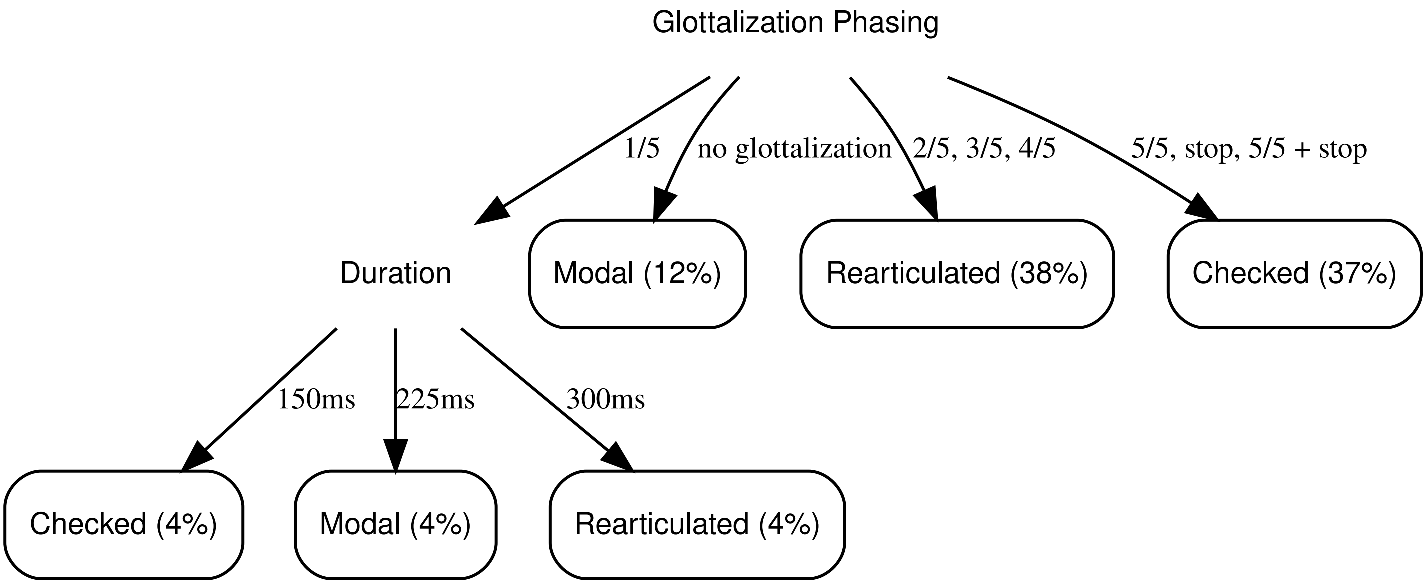

FIG. 8. Classification tree of the relation between the cue and the perceived phonation.

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. In Experiment III, we tested whether when increasing the proportion of glottalization in vowel at the beginning of the vowel would bias listeners towards a rearticulated phonation and clear the ambiguity.

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 mean predicted probability of checked phonation ranks (voiced) glottalization $<$ glottal stop $<$ (voiced) glottalization + glottal stop (Figure 7). The non-overlapping credible intervals between the conditions of (voiced) glottalization vs. (voiced) glottalization + glottal stop suggest a significant difference in checked response elicitation between these categories. These differences suggest that higher degree of vowel-final glottalization enhances the percept of checked phonation. Our findings complement the previous work in Yucatec Maya by Frazier (2016), where they found that a higher degree of glottalization in vowel-medial position yields higher probability for listeners to perceive a glottalized phonation in Yucatec Maya. In Experiment IV, we tested the effect of glottalization magnitude in vowel-medial position in Yateé Zapotec.

Our findings indicate that duration also influences phonation perception. The shorter duration condition (150 ms) leads to more checked responses, while the longer duration (300 ms) elicits more rearticulated responses. Duration is not used as a reliable cue for listeners to perceive modal phonation.

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).

Experiment I employed stimuli that were not strictly minimal pairs, as the tokens also differed in tone. In addition, listeners were asked to choose among six response options, making the identification task relatively demanding. To address these limitations, Experiment II used two-way minimal pairs contrasting phonation types while controlling for tone, thereby reducing the task to a two-alternative forced-choice design. Three phonation contrasts were tested: modal vs. checked, modal vs. rearticulated, and checked vs. rearticulated.

III. EXPERIMENT II

Experiment II aims to the effect of duration and the position of glottalization found in Experiment I, using minimal pairs between modal and checked phonations; modal and rearticulated phonations; and checked and rearticulated phonations. For the modal vs. checked contrast, stimuli were created at three vowel durations (150, 225, and 300 ms) and two glottalization conditions: modal phonation and vowel-final glottalization occupying one-fifth of the vowel's duration. For the modal vs. rearticulated contrast, the same three durations were fully crossed with two glottalization conditions: modal phonation and vowel-medial glottalization occupying one-fifth of the vowel's duration. These identification tasks were designed to address the question: **Do listeners rely more on vowel duration or on the presence of glottalization when distinguishing modal from non-modal phonations?** For the checked vs. rearticulated contrast, we again used three vowel durations (150, 225, and 300 ms), this time fully crossed with five glottalization positions: at first fifth (1/5), second fifth (2/5), third fifth (3/5), fourth fifth (4/5), and last fifth (5/5) of the vowel. This experiment address the question: **Are listeners more sensitive to vowel duration, or to the phase of glottalization within the vowel when distinguishing between checked and rearticulated phonation?**.

A. Method

1. Stimuli

The target words used in the perception experiment are listed in Table IV. The contrast between modal and checked phonation is represented by [gǎ] “nine” and [gǎ͡ˀ] “basket”; the contrast between modal and rearticulated phonation by [chǐ] “ten” and [ch‘i͡ˀ’i] “her/his voice”; and the contrast between checked and rearticulated phonation by [z’i͡ˀ] “pain” and [z’i͡ˀ’i] “heavy”. We added an additional modal‑rearticulated pair [jǎ] “metal” vs. [jà͡ˀá]“plaza” starting from the 18th participant we ran, because several listeners reported that the word [ch‘i͡ˀ’i] (“her/his voice”) was unfamiliar to them. Figure 9 illustrates the waveforms and spectrograms of the natural productions of all target words listed in Table IV.

To test whether listeners rely more on vowel duration or on the presence of glottalization when distinguishing modal from non-modal phonations,we resynthesized stimuli with three duration levels crossed with phonation type: modal phonation, mid-vowel glottalization (for the modal‑rearticulated contrast), and vowel-final glottalization (for the modal‑checked contrast). The corresponding waveforms and spectrograms are shown in Figure 10. Glottal-

TABLE IV. Options for Experiment II-IV identification tasks.

Contrast between	Transcription	Tone	Phonation	Orthography	English/Spanish
M vs. C	[gǎ]	rising	modal	ga	“nine”/“nueve”
	[gǎ͡ˀ]	rising	modal	ga’	“basket”/“canasta”
M vs. R	[ʂǐ]	rising	modal	chi	“ten”/“diez”
	[tʂì͡ˀí]	rising	rearticulated	chi’i	“her/his voice”/“su voz”
M vs. R	[tjà͡ˀá]	rising	rearticulated	ya’a	“market place”/“plaza”
	[jǎ]	rising	modal	yaa	“metal”
C vs. R	[zí͡ˀ]	high	checked	zi’	“pain”/“duele”
	[zí͡ˀí]	high	checked	zi’i	“heavy”/“pesado”

ization was resynthesized using the PSOLA algorithm in Praat by lowering and jittering the f0, and lowering the intensity of the glottalized portion via modification of the vowel’s intensity tier. The Praat scripts used for stimulus generation are provided in the Supplementary Materials.

To test whether listeners are more sensitive to vowel duration, or to the phase of glottalization within the vowel when distinguishing between checked and rearticulated phonation, we resynthesized stimuli with three durations crossed with glottalization at the first (1/5), second (2/5), third (3/5), fourth (4/5), and last (5/5) fifth of the vowel. Example waveforms and spectrograms of the stimuli of 300 ms are shown in Figure 11. The stimuli of 150 ms and 225 ms are in Supplementary Materials.

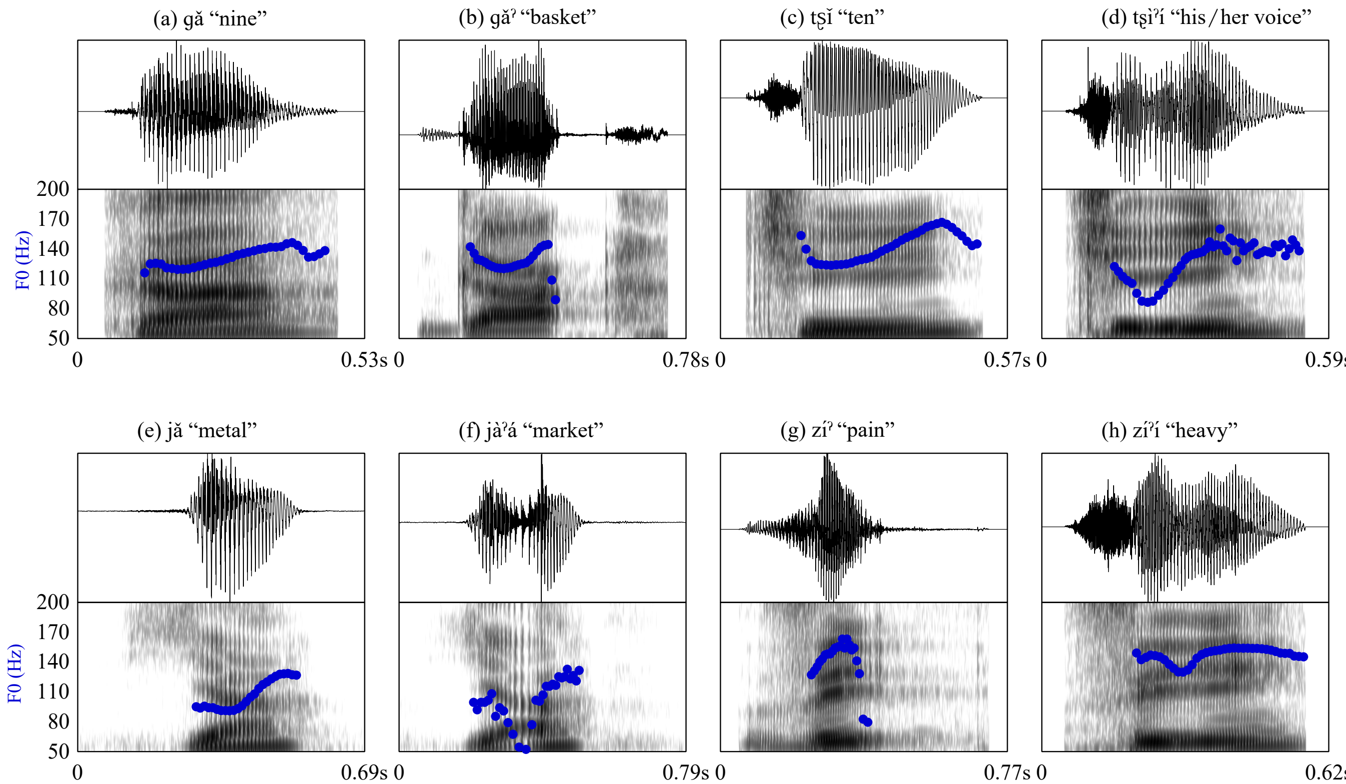

FIG. 9. Waveforms and spectrograms (showing 0-5000 Hz) of natural production of the options in the Experiment II-IV identification task. The blue contours overlaid on the spectrogram refer to the f0 contour. The y-axis of the spectrogram is showing the f0 range. Options (a) and (b) are modal vs. checked contrast; (c) and (d), (e) and (f) are modal vs. rearticulated contrast; (g) and (h) are checked vs. rearticulated contrast.

2. Participants

Twenty-seven native speakers of Yateé Zapotec (16 women, 11 men; mean age = 48 years) participated in the experiment. All participants reported being bilingual in Zapotec and Spanish, with Zapotec as their first-acquired language, learned in the village of San Francisco Yate'e. Eleven participants also took part in Experiment I. Participants completed a two-alternative forced-choice identification task. The contrasts tested were: [gǎ] vs. [gǎ͡ʔ];

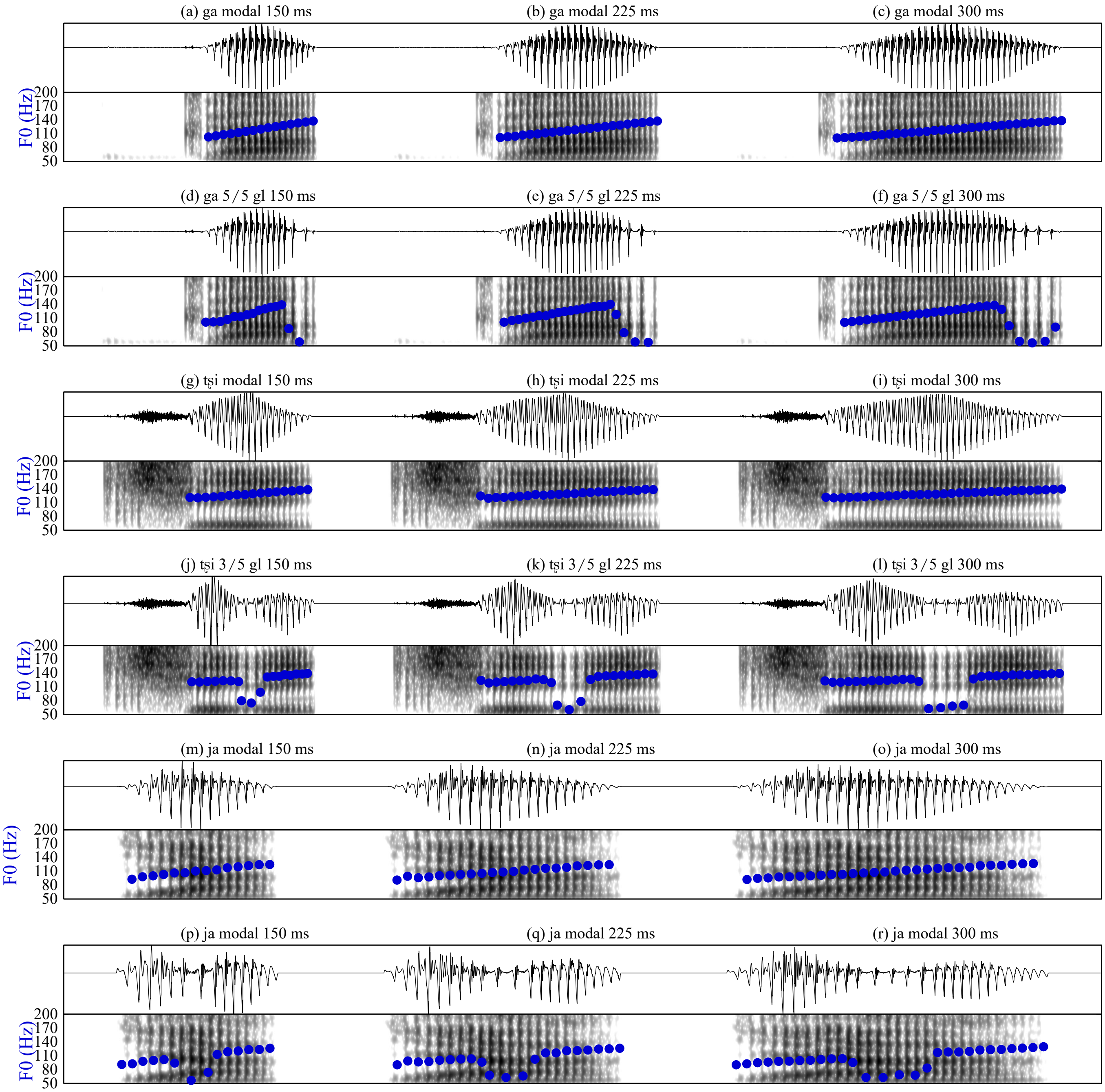

FIG. 10. Waveforms and spectrograms (showing 0-5000 Hz) of resynthesized stimuli for modal phonation vs. mid glottalization and modal phonation vs. final glottalization. Figures (a)-(f) are stimuli for the identification of [gǎ] vs. [gǎ͡ˀ], for testing the effect of duration and glottalization on modal vs. checked phonations identification. Figures (g)-(l) are stimuli for the identification of [ts̩ǐ] vs. [ts̩i͡ˀí]. Figures (m)-(r) are stimuli for the identification of [jǎ] vs. [jà͡ˀá], both for testing the effect of duration and glottalization on modal vs. rearticulated phonations identification.

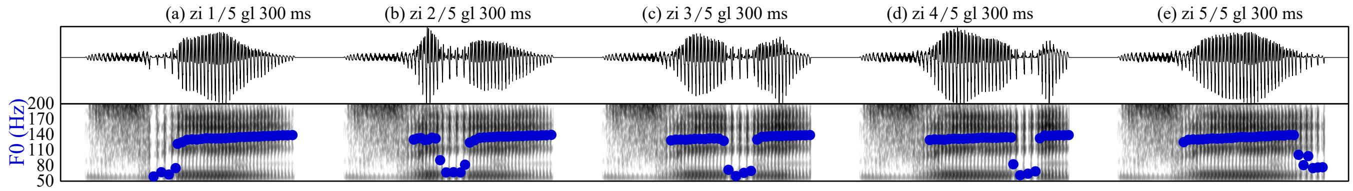

FIG. 11. Waveforms and spectrograms (showing 0-5000 Hz) of resynthesized stimuli for five positions of glottalization. Figures (a)-(e) display stimuli with glottalization inserted at each of the five temporal positions within the vowel with 300 ms duration.

[tʂ̩ĭ] vs. [tʂ̩ʻi͡ˀʼi]; [jă] vs. [jʻa͡ˀʼa], and [zʼi͡ˀ] vs. [zʼi͡ˀʼi]. All written instructions were provided in the Zapotec orthography, and the experimenter explained the task in Zapotec. The experiments were run on the same laptop computer as in Experiment I. The listeners listened to the stimuli played over the same headphone as they did in Experiment I. On each trial, listeners heard a target stimulus and viewed two images representing the response options, with the corresponding word written in Zapotec orthography displayed beneath each image. Participants could replay the stimulus as many times as desired by pressing a play button. They responded by clicking on the image corresponding to the perceived word, after which the next trial began. Figure 12 is an example page of the experiment.

The identification task consisted of two parts. In the first part, participants listened to the natural productions of the target words by a native speaker of Yateé Zapotec. This part served as a screening step to verify that listeners could reliably perceive each phonation contrast. Only responses from participants who correctly identified the natural tokens for a given contrast were included in the analysis. In the second part, they completed the identification task using resynthesized stimuli, with each stimulus presented three times

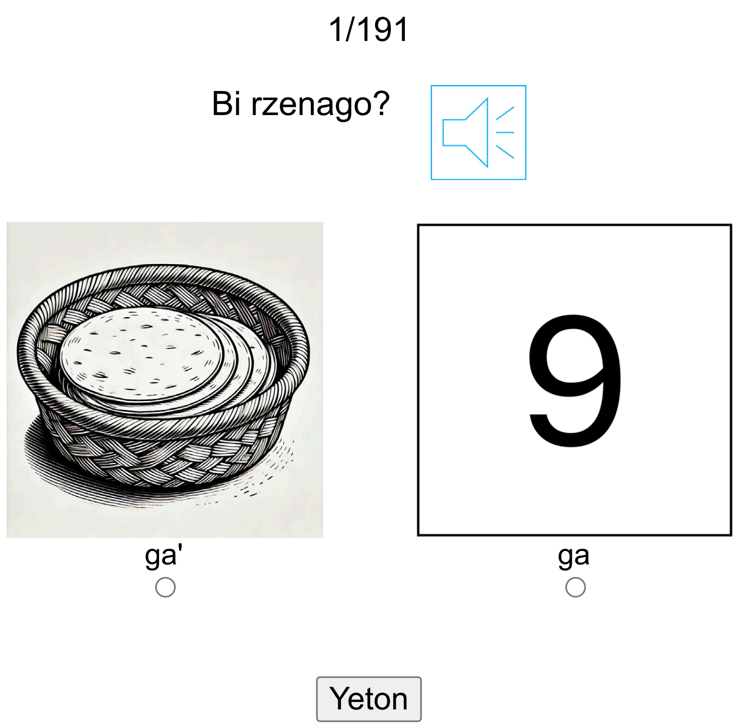

FIG. 12. Example page of Experiment II trials.

across three randomized blocks[5]. For the [z’i͡ˀ] vs. [z’i͡ˀ’i] contrast, 26 participants correctly identified the natural tokens of that contrast, yielding 1164 responses (3 durations × 5 glottalization positions × 3 repetitions × 26 participants − 6 missing trials[6]). For the [gǎ] vs. [gǎ͡ˀ] contrast, 17 participants passed the screening, yielding 304 responses (3 durations × 2 glottalization positions × 3 repetitions × 17 participants − 2 missing trials). For the [tʂ̯ǐ] vs. [tʂ̯‘i͡ˀ’i] contrast, 22 participants passed the screening, producing 319 responses (3 durations × 2 glottalization positions × 3 repetitions × 16 participants − 5 missing trials + 3 durations × 2 glottalization positions × 1 repetition × 6 participants). For the [jǎ] vs. [j‘a͡ˀ’a] contrast, 8 of 10 participants passed, yielding 96 responses (3 durations × 2 glottalization positions × 2 repetitions × 8 participants). Because the [tʂ̯ǐ] - [tʂ̯‘i͡ˀ’i] and [j‘a͡ˀ’a] - [jǎ] pairs both tested the modal-rearticulated contrast, their data were combined

(319 + 96 = 415 responses), and word pair was included as a random intercept in statistical modeling to control for potential lexical effects.

B. Results

The results are shown in Figure 13. As illustrated in Figure 13a, for the modal vs. checked contrast, listeners consistently selected the modal response for stimuli with modal phonation and the checked response for stimuli with vowel-final glottalization, with this pattern holding across durations. Similarly, for the modal vs. rearticulated contrast (Figure 13b), modal responses were predominantly chosen for modal stimuli, whereas rearticulated responses were more likely when glottalization occurred in the middle of the vowel, again regardless of duration. To statistically evaluate these patterns, we fit logistic regression models with perceived phonation as the dependent variable and duration, glottalization condition, and their interaction as fixed effects. A random intercept was included for each participant, and for the modal vs. rearticulated contrast, an additional random intercept was included for the word pair ([tʂ̩ǐ] vs. [tʂ̩ʻi͡ʔʼi] or [jʻa͡ʔʼa] vs. [jǎ]) to account for lexical differences.

To test the effect of duration and glottalization position for the discrimination between modal and the other two glottalized phonations, we fitted two generalized linear mixed-effects model (modal vs. checked and modal vs. rearticulated), with the phonation of the response as the dependent variable, and the duration, the glottalization condition, and their interaction as the independent variable. Each participant was assigned a random intercept. For both models, we compared the full model with a simpler model without the interaction between duration and glottalization, and found that the interactions did not

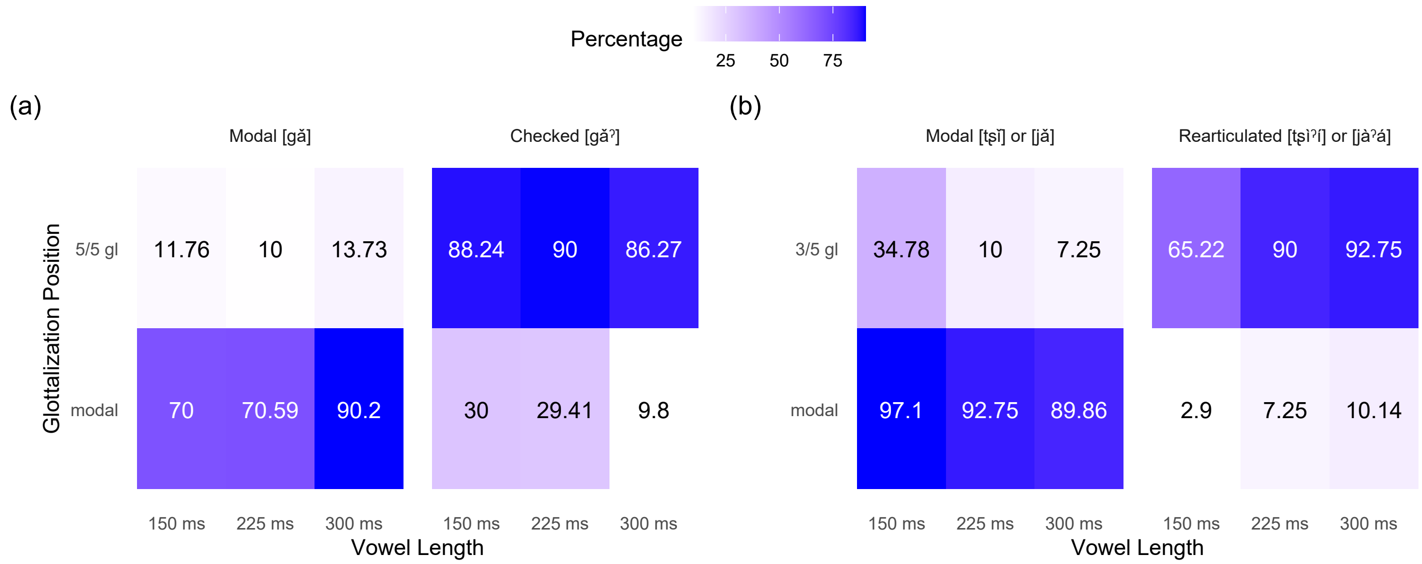

FIG. 13. Percentage (%) of modal vs. checked (a), and modal vs. rearticulated (b) responses by duration and position of glottalization. The majority response is marked with white text in each condition.

significantly improve either model ($p = 0.285$ for modal vs. checked; $p = 0.755$ for modal vs. rearticulated), thus the interactions were dropped from both model.

The statistical results for the modal vs. checked phonation contrast are shown in Table V. Vowel-final glottalization elicited significantly more checked responses than modal responses ($b = 3.419$, $p < 0.001$). Shorter durations (150 ms and 225 ms) also led to a higher likelihood of checked responses compared to modal responses ($b = 0.856$, $p = 0.013$). The coefficients of the independent effects reveal that glottalization has a stronger influence than duration in predicting the phonation response. The coefficient for glottalization (*gl5vsmodal*) was 3.26 on the logit scale, corresponding to an increase in predicted probability of checked responses from 22% for modal stimuli to 88% for vowel-final glottalization stimuli, an increase of 66%. In contrast, the coefficient for duration (*length150_225vs300*) was 1.14, reflecting an increase in predicted probability of checked responses from 40% at 300 ms to 68% at the average of

150 ms and 225 ms, a difference of 28%. Together, these results indicate that vowel-final glottalization has a substantially stronger effect on the perception of checked phonation than vowel shortening.

TABLE V. Fixed-effect estimates from the mixed-effects logistic regression model predicting modal vs. checked response. Reference level: Modal phonation = 0; Checked phonation = 1.

Predictor	Estimate	Std. Error	z value	$\Pr(>\|z\|)$	
(Intercept)	0.422	0.188	2.237	0.025	*
length150_225vs300	0.856	0.345	2.483	0.013	*
length150vs225	-0.045	0.382	-0.119	0.905	
gl5vsmodal	3.419	0.357	9.590	<0.001	***

The statistical results for the modal vs. rearticulated phonation contrast are presented in Table VI. Vowel-medial glottalization elicited significantly more rearticulated responses than modal vowels ($b = 4.933$, $p < 0.001$). Durations of 225 ms and 300 ms elicited more rearticulated responses than 150 ms ($b = 1.735$, $p < 0.001$). Examining the magnitude of the coefficients, glottalization had a stronger influence than duration on the phonation response. The coefficient for glottalization (*gl3vsmodal*) was 4.933 on the logit scale, corresponding to an increase in predicted probability from approximately 5% for modal vowels to 87% for vowels with medial glottalization, a difference of 82%. In contrast, the coefficient for duration (*length225_300vs150*) was 1.735, corresponding to an increase in predicted probability from 24% at 150 ms to 50% at the mean of 225 ms and 300 ms, a difference of 26%. Together, these findings indicate that vowel-medial glottalization has a stronger effect on the perception of rearticulated phonation than vowel lengthening.

TABLE VI. Fixed-effect estimates from the mixed-effects logistic regression model predicting modal vs. rearticulated response. Reference level: Modal phonation = 0; Rearticulated phonation = 1.

Predictor	Estimate	Std. Error	z value	Pr(> $\|z\|$)	
(Intercept)	-0.569	0.219	-2.595	0.009	**
length225_300vs150	1.735	0.377	4.600	<0.001	***
length300_225	0.349	0.426	0.818	0.413	
gl3vsmodal	4.933	0.449	10.986	<0.001	***

The results for the checked vs. rearticulated vowels are shown in Figure 14. To examine the effects of vowel duration and the relative position of glottalization on the discrimination between checked and rearticulated phonation, we fitted a generalized linear mixed-effects model with phonation response as the dependent variable and duration and glottalization position, as well as their interaction, as independent variables. A random intercept was included for each participant. Statistical results are provided in Table VII. No significant interaction between duration and glottalization position was found, and the interaction terms were removed from the model ($p = 1$). Non-vowel-final glottalization (1/5, 2/5, 3/5, 4/5) elicited significantly more rearticulated responses than vowel-final glottalization (5/5) (*glpos1_2_3_4vs5*, $b = 5.715$, $p < 0.001$). Vowel-medial glottalization (2/5, 3/5, 4/5) elicited significantly more rearticulated responses than vowel-initial glottalization (1/5) (*glpos2_3_4vs1*, $b = 1.356$, $p < 0.001$). Early-to-mid glottalization (2/5, 3/5) elicited significantly more rearticulated responses than late-medial glottalization (4/5) (*glpos2_3vs4*, $b = 1.117$, $p < 0.001$). In terms of duration, 225 ms and 300 ms on average elicited more rearticulated responses than 150 ms (`length225_300vs150`, $b = 2.075$, $p < 0.001$), and

300 ms elicited more rearticulated responses than 225 ms (*length300vs225*, b = 0.657, p = 0.036). Looking at the coefficients of the independent effects, glottalization position has a much stronger influence on the phonation response than duration. The coefficient for glottalization (*glpos1_2_3_4vs5*) was 5.72 on the logit scale, corresponding to an increase in predicted probability of a rearticulated phonation response from approximately 6% for vowel-final glottalization (5/5) to 95% for non-final glottalization (1/5–4/5), an 89% increase. In contrast, the coefficient for duration (*length225_300vs150*) was 2.08, indicating an increase in predicted probability of a rearticulated phonation response from 60% at 150 ms to 92% at the average of 225 ms and 300 ms, a 33% increase. These results show that non-final glottalization is more effective than longer duration in eliciting rearticulated phonation responses.

TABLE VII. Fixed-effect estimates from the mixed-effects logistic regression model predicting checked vs. rearticualted phonation response. Reference level: Checked phonation = 0; Rearticulated phonation = 1.

Predictor	Estimate	Std. Error	z value	Pr(> $\|z\|$)	
(Intercept)	1.769	0.215	8.223	<0.001	***
length225_300vs150	2.075	0.246	8.429	<0.001	***
length300vs225	0.657	0.314	2.095	0.036	*
glpos1_2_3_4vs5	5.715	0.350	16.350	<0.001	***
glpos2_3_4vs1	1.356	0.253	5.369	<0.001	***
glpos2_3vs4	1.117	0.308	3.620	<0.001	***
glpos2vs3	0.261	0.435	0.600	0.548	

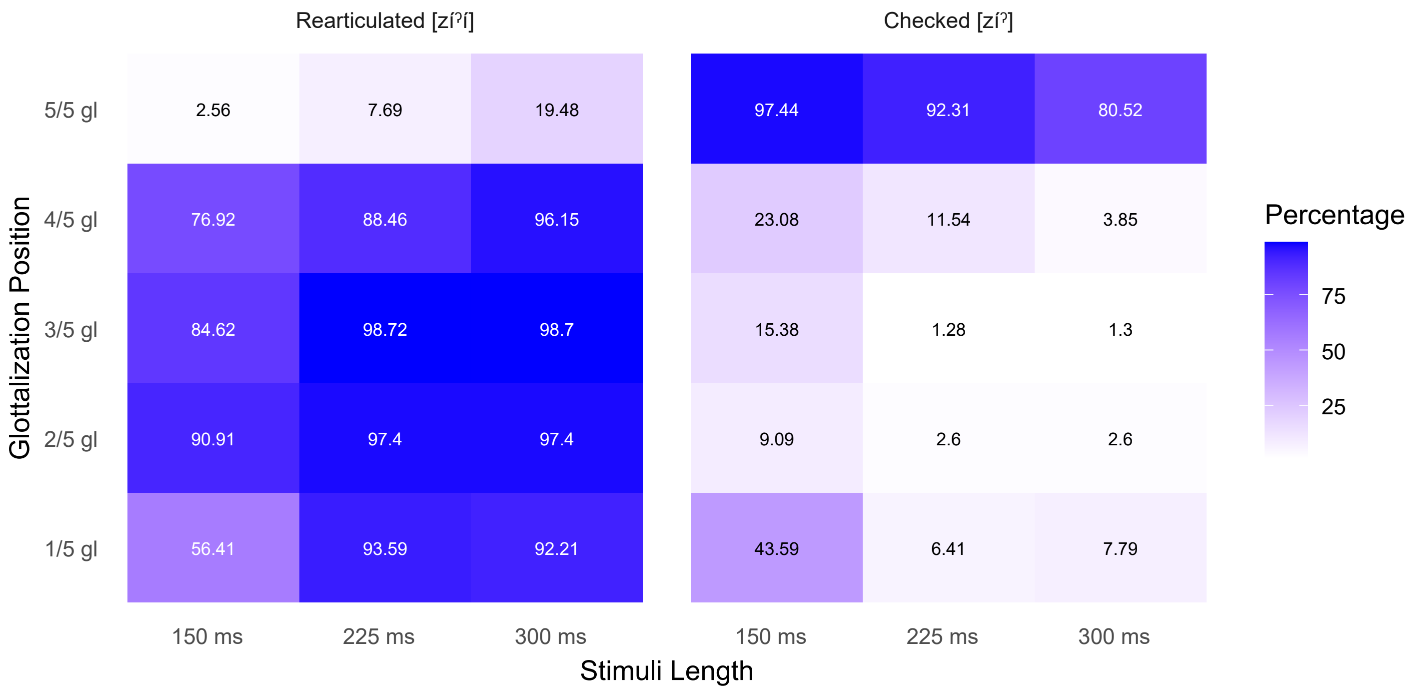

FIG. 14. Percentage (%) of checked vs. rearticulated responses by duration and glottalization position. The majority response in each condition is in white text.

IV. DISCUSSION

The results from Experiment II reinforce the findings from Experiment I, now with stronger evidence due to the use of strict minimal pairs in the identification tasks. Experiment II confirms that both vowel duration and glottalization shape the perception of phonation categories in Yate e Zapotec, but listeners do not treat these cues equally. For checked phonation, vowel-final glottalization was a much stronger cue than short duration. In both the modal vs. checked and rearticulated vs. checked contrasts, stimuli with vowel-final glottalization consistently led listeners to choose the checked category, regardless of duration. Short duration also promoted checked responses, but its influence was smaller. For rearticulated phonation, glottalization earlier in the vowel was the most reliable cue, with longer duration also increasing rearticulated responses. In both the modal vs. rearticulated and rearticulated vs. checked contrasts, early or medial glottalization reliably guided listeners toward rearticulated responses, and duration played a secondary role. Finally, no particular duration favored modal responses. Instead, the absence of glottalization was the key factor for eliciting modal perception across both modal vs. checked and modal vs. rearticulated contrasts. Overall, these results show that the relative timing of glottalization within the vowel is a more important perceptual cue than duration in distinguishing the three phonation categories in Yate e Zapotec.

V. EXPERIMENT III

In Experiment I (Figure III), we observed ambiguity when glottalization appeared at the beginning of the vowel. For first-fifth glottalization, the proportions of modal, rearticulated, and checked responses were 21.62%, 27.03%, and 51.35% at 150 ms; 37.84%, 29.73%, and 32.43% at 225 ms; and 21.62%, 59.46%, and 18.92% at 300 ms. In none of these duration conditions did one category clearly dominate, indicating that listeners did not consistently map initial glottalization to a single phonation category. However, in natural production, vowel-initial glottalization is attested, and when it occurs, it typically spans a much larger portion of the vowel, often close to half its duration (Figure 1). The ambiguity observed in vowel-initial glottalization in Experiments I and II may stem from the glottalized interval being too short to clearly elicit a rearticulated phonation percept. To address this question, Experiment III systematically manipulated the proportion of vowel-initial glottalization. We tested whether lengthening the glottalized portion biases perception toward rearticulated phonation. The experiment examined both the modal vs. rearticulated contrast and the rearticulated vs. checked contrast by creating stimuli with first-fifth, first-third, and first-half of the vowel duration glottalized at the beginning of the vowel.

As shown in Figure 1c–d, we also observe cases in natural speech where glottalization spans the latter half of the vowel in both rearticulated (c) and checked (d) phonations. This motivates a related question: if we lengthen the glottalized portion at the end of the vowel, will listeners shift toward a rearticulated percept, under the hypothesis that a longer glottalized interval favors rearticulated phonation percept, or will they instead categorize the vowel more consistently as checked, if extended glottalization at the end of the vowel strengthens a checked phonation percept? To test between these possibilities, we manipulated the duration of vowel-final glottalization in the rearticulated vs. checked contrast, examining whether increasing glottalization at the end of the vowel biases perception toward rearticulated phonation.

A. Stimuli

Figure 15 shows the waveform and spectrogram for stimuli in which the proportion of vowel-initial and vowel-final glottalization was varied, illustrated here for the 300 ms condition. Stimuli with 150 ms and 225 ms durations can be found in the Supplementary Materials. For the stimuli shown in Figure 15a–f, listeners judged between [z’i͡ˀ] “pain” and [zí͡ˀí] “heavy”. For the stimuli shown in Figure 15g–i, listeners judged between [ʂ̦ǐ] “ten” and [tʂ̦ì͡ˀí] “her/his voice”. For the stimuli shown in Figure 15j–l, listeners judged between [jǎ] “metal” vs. [j‘a͡ˀ’a] “plaza”.

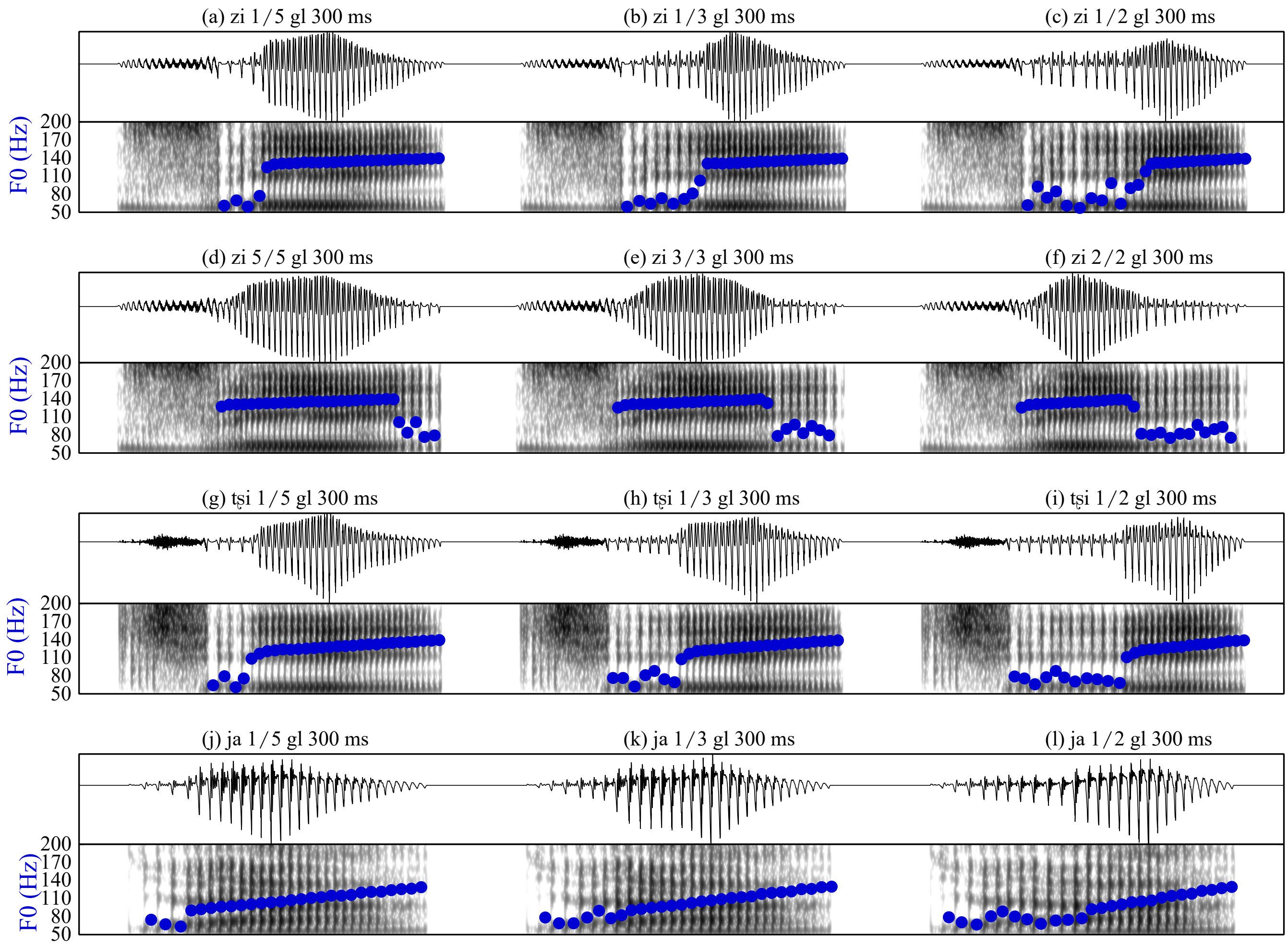

FIG. 15. Resynthesized stimuli with glottalization at first fifth, third, and half of the vowel (a)-(c); and at last fifth, third, and half of the vowel (d)-(e), for the contrast between rearticulated and checked phonations ([zí͡ʔ] vs. [zí͡ʔí]). Resynthesized Stimuli with glottalization at first fifth, third, and half of the vowel, for the contrast between modal and rearticulated phonations [ʂ̩ǐ] vs. [tʂ̩ì͡ʔí] (g)-(i) and between [jǎ] vs. [jʻa͡ʔʼa] (j) - (l).

B. Participants and procedure

The participants and procedure is the same as Experiment II. The data exclusion process is the same as in Experiment II. Each stimulus were presented to the listeners three times in three blocks in a randomized order. For [zí͡ʔ] "pain" and [zí͡ʔí] "heavy" pair, we elicited

697 responses for the proportion variation in the beginning of the vowel (3 durations * 3 proportions * 3 repetitions * 26 participants - 5 missing responses due to one participant requested to end experiment early). And we elicited 698 responses for the proportion variations at the end of the vowel (3 durations * 3 proportions * 3 repetitions * 26 participants - 4 missing responses). For the [ʂǐ] and [tʂì͡ˀí] pairs, we elicited 481 responses (3 durations * 3 proportions * 3 repetitions * 16 participants - 5 missing responses + 3 durations * 3 proprotions * 1 repetition * 6 participants). For the [jà͡ˀá] vs. [jǎ] pair, we elicited 144 responses (3 durations * 3 proportions * 2 repetitions * 8 participants). The responses for [ʂǐ] and [tʂì͡ˀí] and [jà͡ˀá] vs. [jǎ] were collapsed together, resulting in 625 responses in total for the modal vs. rearticulated pair.

C. Results

The descriptive results for the distribution of rearticulated and modal responses across durations and vowel-initial glottalization proportions are shown in Figure 16a. We fit a generalized linear mixed-effects model with phonation response as the dependent variable, and the proportion of vowel-initial glottalization and duration, along with their interaction, as fixed effects. Participant and word-pair were included as random intercepts. The interaction between glottalization proportion and duration was not significant ($p = 0.456$) and was removed from the model. Tukey-adjusted pairwise comparisons revealed significant differences among all levels of glottalization proportion (Table VIII). Vowels with first-half glottalization showed a significantly higher probability of eliciting rearticulated responses than those with first-third glottalization ($b = 1.04$, $p = 0.001$), and first-third glottalization in turn led to more rearticulated responses than first-fifth glottalization ($b = 1.26$, $p < 0.001$).

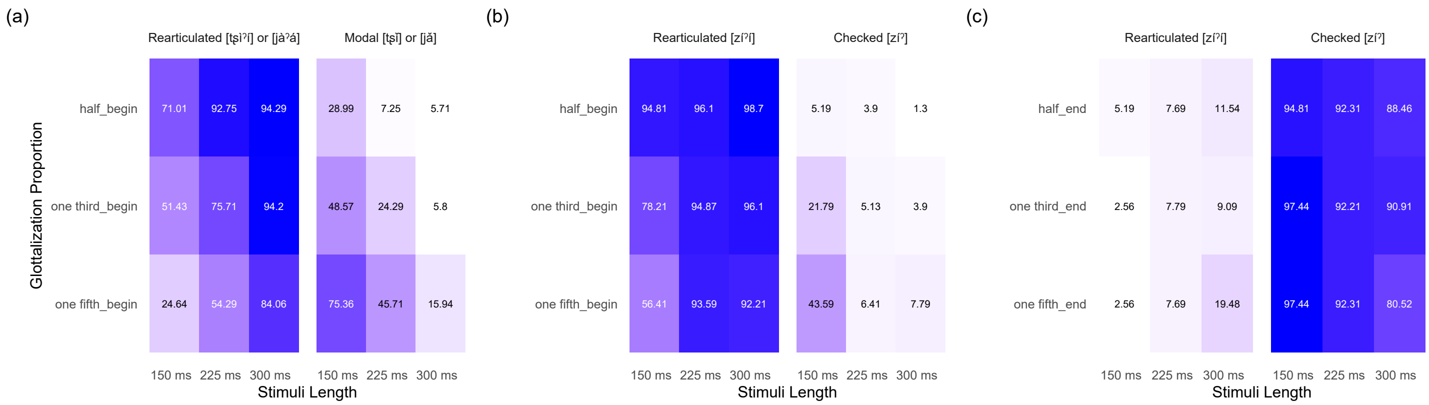

FIG. 16. Percentage (%) of rearticulated vs. modal (a) and rearticulated vs. checked (b) for increasing the proportion of glottalization from one fifth, one third, to half at the beginning of the vowel. Percentage (%) of rearticulated vs. checked (c) for increasing the proportion of glottalization from one fifth, to one third, then to half at the end of the vowel.

TABLE VIII. Pairwise comparisons of glottalization proportion from the mixed-effects logistic regression model for the effect of glottalization proportion at the beginning of the vowel. Reference level: Modal phonation = 0; Rearticulated phonation = 1.

Contrast	Estimate	SE	df	z ratio	p value	
half – one third	1.04	0.30	∞	3.53	0.0012	***
half – one fifth	2.30	0.30	∞	7.60	<0.0001	***
one third – one fifth	1.26	0.26	∞	4.86	<0.0001	***

The results for the distribution of rearticulated and checked responses across durations and vowel-initial glottalization proportions are shown in Figure 16b. We fit a generalized linear mixed-effects model with phonation response as the dependent variable and the proportion of vowel-initial glottalization and duration, along with their interaction, as fixed effects. Participant was included as a random intercept. Statistical results are reported in Table IX. One-third and one-half glottalization at the beginning of the vowel on average elicited more rearticulated responses than one-fifth glottalization (*third_halfvsfifth*, $b = 1.374$, $p < 0.001$), and one-half glottalization elicited more rearticulated responses than one-third glottalization (*halfvsthird*, $b = 1.114$, $p = 0.034$). On average, duration of 225 ms and 300 ms elicited more rearticulated responses than 150 ms (*length225_300vs150*, $b = 1.982$, $p < 0.0010$), same as the duration effect found in Experiment II. The interaction between glottalization proportion and duration was significant. The difference between one-fifth, and the average of one-third and one-half glottalization was larger at 150 ms than at the average of 225 ms and 300 ms, indicating that increases in glottalization proportion have a stronger effect on promoting rearticulated perception when the vowel is short.

The results for the distribution of rearticulated and checked phonations across duration and glottalization-proportion conditions at the end of the vowel are shown in Figure 16b. The data were analyzed using a generalized linear mixed-effects model with phonation response as the dependent variable and the proportion of vowel-final glottalization and duration, along with their interaction, as fixed effects. A random intercept was included for each participant. An ANOVA comparison with a simpler model excluding the interaction term indicated no improvement in model fit ($p = 0.508$), so the interaction was removed. As summarized in Table X, the proportion of glottalization at the end of the vowel did not significantly influence listeners' phonation judgments. Across all proportions and duration conditions, the responses favored the checked category, as reflected by the positive intercept

TABLE IX. Fixed-effect estimates from the mixed-effects logistic regression model predicting the effect if vowel-initial glottalization proportion on rearticulated and checked phonation responses. Reference level: Checked phonation = 0; Rearticulated phonation = 1.

Predictor	Estimate	Std. Error	z value	Pr(> $\|z\|$)	
(Intercept)	3.198	0.369	8.672	< 0.001	***
third_halfvsfifth	1.374	0.361	3.802	< 0.001	***
halfvsthird	1.114	0.527	2.114	0.034	*
length225_300vs150	1.982	0.362	5.475	< 0.001	***
length300_225	0.382	0.532	0.718	0.473	
third_halfvsfifth:length225_300vs150	-1.407	0.652	-2.159	0.031	*
third_halfvsfifth:length300_225	1.021	0.988	1.034	0.301	
halfvsthird:length225_300vs150	-1.175	0.959	-1.225	0.221	
halfvsthird:length300_225	0.932	1.448	0.644	0.520	

estimate (*Intercept*, $b = 3.094$, $p < 0.001$), corresponding to an estimated probability of 96% eliciting the checked phonation response.

D. Discussion

Vowel-initial glottalization is not typically associated with either rearticulated or checked phonation in Yate e Zapotec. However, when the proportion of glottalization at the beginning of the vowel increases, listeners show a stronger bias toward rearticulated phonation. This holds for both the modal vs. rearticulated contrast and the checked vs. rearticulated contrast, indicating that rearticulated phonation can be perceived at vowel onset when the glottalized portion is long enough.

TABLE X. Fixed-effect estimates from the mixed-effects logistic regression model predicting the effect of vowel-final glottalization on the perception of phonation. Reference level: Rearticulated response = 0, Checked response = 1

Predictor	Estimate	Std. Error	z value	Pr(> $\|z\|$)	
(Intercept)	3.094	0.335	9.226	< 0.001	***
third_halfvsfifth	0.385	0.304	1.268	0.205	
third_half	-0.283	0.381	-0.743	0.457	
length225_300vs150	-1.259	0.405	-3.107	0.002	**
length300vs225	-0.684	0.331	-2.066	0.039	*

In terms of the effect of glottalization proportion at the end of the vowel, two hypothese were posited. The first hypothesis was increasing the duration of vowel-final glottalization would shift perception toward rearticulated phonation. The results does not support that hypothesis, because extending the glottalized portion at the end of the vowel did not increase rearticulated responses. The second hypothesis was that longer glottalization at vowel-final position reinforces a checked percept. This hypothesis remains viable with more evidence required. Since checked responses were already near ceiling in this context, any further strengthening effect may have been masked. Additional evidence, using stimuli that avoid ceiling performance, will be needed to determine whether extended glottalization at the end of the vowel indeed enhances checked phonation perception.

VI. EXPERIMENT IV

Experiment I showed that, when glottalization occurs at the end of the vowel, a full glottal stop elicits a higher proportion of checked responses than a shorter glottalization gesture, suggesting that a stronger degree of glottalization increases the probability of eliciting checked phonation. We next asked whether a stronger degree of glottalization at the middle of vowel would also facilitate the perception of rearticulated phonation. To do so, we created two degrees of vowel-medial glottalization. The weak glottalization condition included a small, gradual f0 dip and a slight reduction in intensity. The strong glottalization condition included a larger, irregular f0 dip combined with a stronger intensity reduction. Participants completed a two-choice identification task judging whether each token sounded modal or rearticulated by choosing between [ʂǐ] vs. [tʂì͡ʔí], and between [jǎ] vs. [jà͡ʔá].

A. Stimuli

Figure 17 presents the stimuli used to examine whether differences in the degree of vowel-medial glottalization affect listeners' perception of rearticulated phonation. Panels 17(a)-(c) and (d)-(f) are the stimuli with weak and strong glottalization for the [ʂǐ] vs. [tʂì͡ʔí] contrast; (g)-(i) and (j)-(l) are the stimuli with weak and strong glottalization for the [jǎ] vs. [jà͡ʔá] contrast.

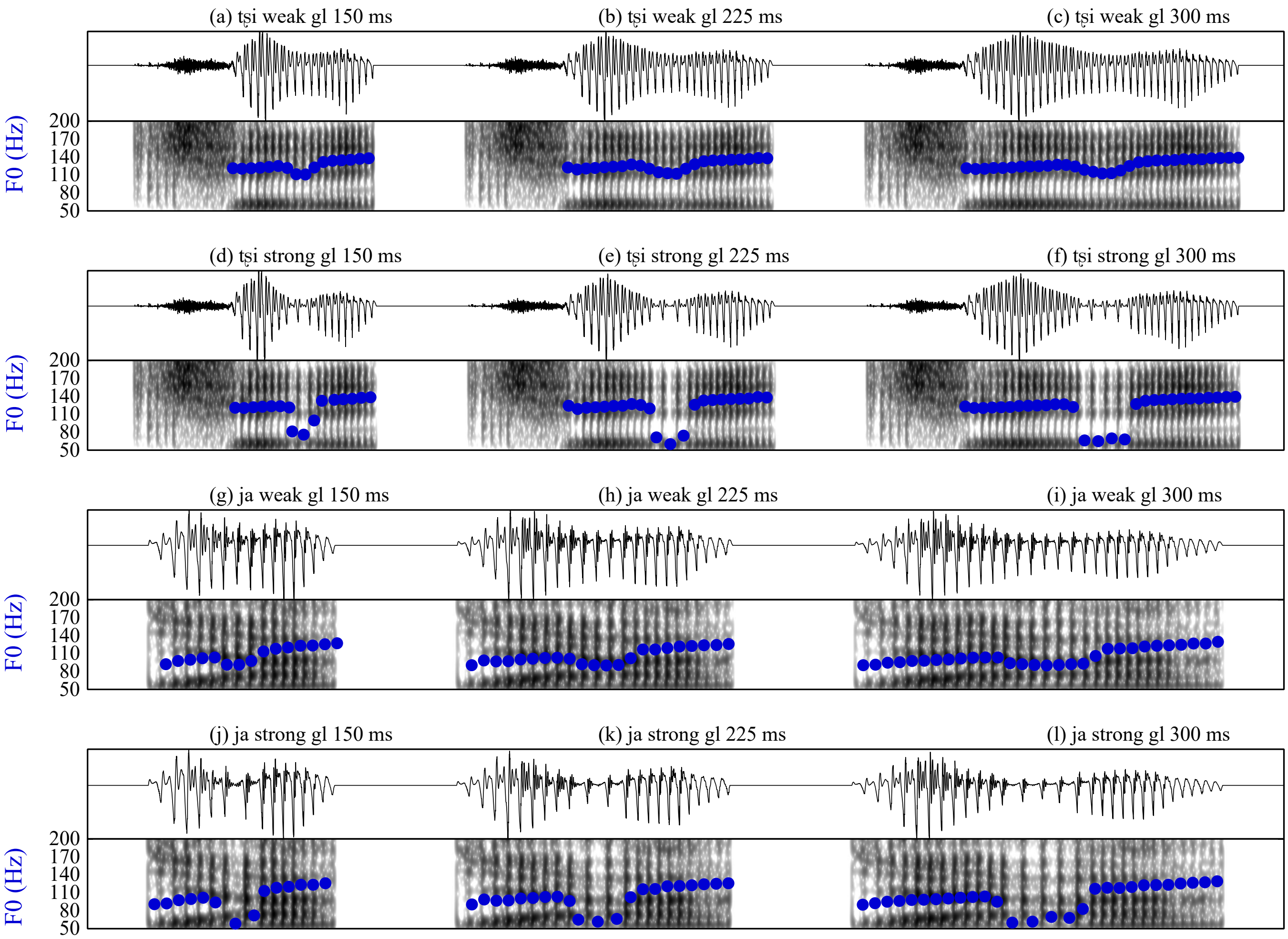

FIG. 17. Waveforms and spectrograms for the resynthesized stimuli testing the degree of glottalization on the perception of rearticulated phonation.

B. Participants and Procedure

The participants and procedure for Experiment IV were the same as in Experiment II. For the [ʂ̩ǐ] vs. [tʂ̩ʻi͡ʔʼi] contrast, we obtained 321 responses in total (3 durations * 2 glottalization degrees * 3 repetitions * 16 participants − 3 missing responses + 3 durations * 2 glottalization degrees * 1 repetition * 6 participants). For the [jǎ] vs. [jʻa͡ʔʼa] contrast, we obtained 96 responses (3 durations * 2 glottalization degrees * 2 repetitions * 8 participants).

Responses from the two contrast pairs were combined for analysis, yielding 417 observations in total.

C. Results and Discussions

The proportions of modal and rearticulated responses across duration and glottalization-intensity conditions are shown in Figure 18. The data were analyzed using a generalized linear mixed-effects model with phonation response as the dependent variable and glottalization intensity, duration, and their interaction as fixed effects. Random intercepts were included for subjects and word pairs. The random intercept for word pair was dropped because its variance is near zero and cause singularity in the model. The interactions were dropped because they do not significantly improve the model. As shown in Table XI, stronger glottalization elicited significantly more rearticulated responses than weaker glottalization (*strongglvsweakgl*, $b = 1.206$, $p < 0.001$). This pattern mirrors the results of Experiment I and demonstrates that listeners are sensitive not only to the presence and timing of glottalization, but also to its magnitude. In both experiments, increasing the strength of glottalization shifted perception toward non-modal phonation: greater glottalization in vowel-offset position favored checked phonation, while stronger glottalization in medial position favored rearticulated phonation.

VII. GENERAL DISCUSSION

This study is the first to use strictly controlled stimuli to systematically investigate the perceptual cues underlying contrastive phonation categories in a Zapotec language. Across

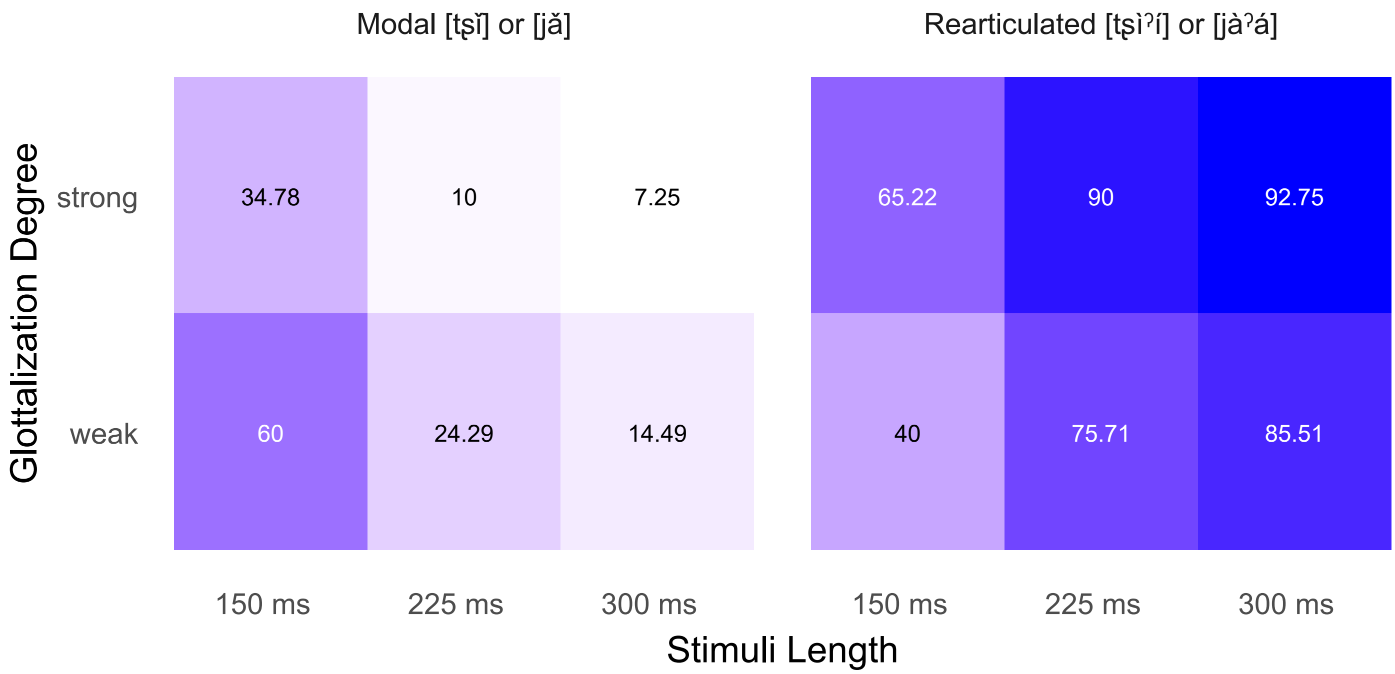

FIG. 18. Percentage (%) of rearticulated vs modal responses for weak and strong glottalization in three duration conditions.

TABLE XI. Fixed-effect estimates from the mixed-effects logistic regression model predicting the effect of glottalization magnitude on phonation response.

Predictor	Estimate	Std. Error	z value	$\Pr(>\|z\|)$	
(Intercept)	1.603	0.281	5.696	< 0.001	***
strongglvsweakgl	1.206	0.286	4.220	< 0.001	***
length225_300vs150	2.211	0.297	7.436	< 0.001	***
length300vs225	0.653	0.388	1.684	0.092	

four experiments, we provide the first controlled evidence that glottalization position, glottalization degree, and duration jointly shape the perception of multiple glottalized phonation contrasts. Experiment I employed near-minimal pairs that distinguished modal, rearticulated, and checked phonation but differed in tone, allowing us to probe the effects of glottalization position, glottalization degree, and vowel duration. The results revealed consistent patterns: vowel-final glottalization promotes checked phonation, while medial (non-initial and non-final) glottalization promotes rearticulated phonation. Longer vowels favor rearticulated responses, whereas shorter vowels favor checked responses. Importantly, glottalization position exerted a stronger influence than duration on phonation categorization. Listeners were also sensitive to the magnitude of glottalization at vowel edges: at vowel ends, a strong glottal cue (full glottal stop plus glottalization) elicited the highest checked response rates, followed by a full glottal stop alone, and then by vowel-final glottalization without a full stop.

Experiments II, III, and IV further consolidated the findings of Experiment I. Experiment II employed strict minimal pairs contrasting modal–checked and modal–rearticulated phonation, confirming that both duration and the presence versus absence of glottalization play key roles in phonation perception, with the glottalization cue exerting a stronger influence than duration. The minimal-pair contrast between checked and rearticulated phonation in Experiment II additionally showed that checked phonation is predominantly elicited when glottalization occurs at the very end of the vowel, whereas rearticulated phonation is predominantly elicited when glottalization occurs in non-final positions.

Experiment III addresses the ambiguity observed for 1/5 glottalization in Experiment I and in Experiment II at 150 ms. We hypothesized that 1/5 glottalization may be ambiguous because the glottalized interval is too short to elicit a rearticulated percept—particularly when the entire vowel is only 150 ms. Experiment III systematically manipulated the proportion of vowel-initial and vowel-final glottalization. For both the rearticulated vs. modal contrast and the rearticulated vs. checked contrast, increasing the proportion of glottalization from one fifth to one third to one half increased the likelihood of rearticulated responses. In contrast, when glottalization occurred at the end of the vowel, increasing its proportion did not significantly change the response pattern: checked phonation remained the predominant category across all durations of glottalization. This pattern likely reflects a ceiling effect: vowel-final glottalization already strongly favors checked perception, leaving little room for further increase. Although the inferential statistics did not reveal a significant effect of glottalization proportion at the vowel end, the descriptive results suggest the same directional tendency. The condition yielding the lowest proportion of checked responses in this set was the 300 ms, final one-fifth glottalization condition (80.52%). When the glottalization proportion increased to the final third and final half of the vowel, the percentage of checked responses increased to 90.91% and 88%, respectively. This pattern suggests that lengthening glottalization at the vowel end may further strengthen checked perception, but the effect is difficult to detect due to the ceiling already created by final-phase glottalization. Taken together, these results support the hypothesis that increasing glottalization proportion strengthens the percept associated with the vowel phase in which the cue occurs: expanding glottalization at non-final positions increases rearticulated perception, whereas expanding glottalization at vowel-final position likely increases checked perception. Future work can reduce ceiling and floor effects by creating more ambiguous cue combinations. For example, when testing vowel-final glottalization, future studies could pair long glottalization with tonal cues favoring the rearticulated category; conversely, when testing vowel-medial glottalization, extremely short vowels combined with tonal cues favoring checked phonation may provide a more sensitive test of duration effects. Such designs would allow a clearer evaluation of whether glottalization proportion continues to modulate perceptual categorization under less asymmetrical conditions.

Experiment IV complements the findings from Experiment I by further testing the role of glottalization degree. In Experiment I, increasing the degree of glottalization at the end of the vowel resulted in more checked responses. Experiment IV extended this question to vowel-medial glottalization by creating two levels of glottalization strength and testing their perceptual consequences. The results show that increasing the degree of glottalization in the middle of the vowel increases the probability of rearticulated responses. These findings indicate that listeners are sensitive not only to the presence and position of glottalization, but also to its magnitude. Strengthening the glottalization cue reinforces the perception of the glottalized category—checked at vowel-final position and rearticulated at vowel-medial position.

The results also contribute to the typology of perceptual cues for glottalized tones and phonations. When comparing Yate'e Zapotec to other languages reviewed in Section I, we find parallels with Vietnamese (Brunelle, 2009), where the position of glottalization influences the perception of rearticulated vs. checked phonation, and with Mandarin (Huang, 2018), Sgaw Karen (Brunelle and Finkeldey, 2011), and Taiwanese Min (Zhang and Lu, 2023), where duration also contributes to listeners' categorization. In contrast, Yate'e Zapotec differs from White Hmong (Garellek *et al.*, 2013) and Xiapu Min (Chai *et al.*, 2023), where listeners rely more heavily on duration than on glottalization when perceiving low creaky tones. These perceptual patterns align with production tendencies: in White Hmong, creaky tone is not consistently produced with creaky voice, and in Xiapu Min, f0 and duration pattern more strongly with tone categories than glottalization. By comparison, in Yate'e Zapotec, glottalization timing—particularly at vowel-medial and vowel-final positions—more reliably differentiates rearticulated and checked phonation than duration, suggesting a close match between the acoustic realization of glottalization and its perceptual weighting in this language.

The finding that the relative position of glottalization within the vowel has a stronger effect than duration on phonation perception carries important implications for phonological representation. Specifically, it supports the need for phonological frameworks that encode sub-segmental timing, allowing different phases of glottalization to be represented for checked vs. rearticulated phonations (Browman and Goldstein, 1989; Shih and Inkelas, 2019). Future work can probe even more fine-grained manipulations of glottalization timing and duration, and develop abstract phonological representations that incorporate these sub-segmental distinctions to capture the three-way phonation contrast characteristic of Zapotec languages.

ACKNOWLEDGMENTS

Thanks to the family of Lilia Fernandez and Fidel Bernal for hosting me to complete this work. Thanks to all the participants from San Francisco Yateé for participating in my study. Thanks to Professor Marc Garellek and Professor Richard Wright for the guidance in the study design and manuscript. Thanks to the audience at Sound System of Latin America 4 for the suggestions. This research was supported by Endangered Languages Documentation Programme (GLB0038).

AUTHOR DECLARATIONS

Conflict of Interest

The author has no conflicts to disclose.

Ethical Approval

Consent forms have been obtained from all participants in this study. The study was approved by the Institutional Review Board of the University of Washington (protocol code: STUDY00020307; date of approval: July 11, 2024).

DATA AVAILABILITY

The data used for analysis and the programming scripts are are available in Open Science Foundation at https://doi.org/10.17605/OSF.IO/SA2TD.

[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.

[4]In the training data, the condition with the most observations has 34 observations. To capture the majority phonation response out of the three phonation types in a condition with 34 observations, the minimum observed number of the majority response need to be larger than 11.33 (34/3).

[5]Among the 27 participants, 17 completed the [tʂ̥ǐ] vs. [tʂ̥ʻi͡ˀʼi] contrast only, with three repetitions per stimulus. The remaining 10 participants completed both the [tʂ̥ǐ] vs. [tʂ̥ʻi͡ˀʼi] and [jǎ] vs. [jʻa͡ˀʼa] contrasts. For these participants, the [tʂ̥ǐ] contrast included three repetitions and the [jǎ] contrast included two repetitions.

[6]All the missing trials in Experiment II and the following experiments were due to one participant requested to end the experiment early.

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.

Boersma, P., and Weenink, D. (**2023**). "Praat: doing phonetics by computer [Computer program]" http://www.praat.org/.

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

Browman, C. P., and Goldstein, L. (**1989**). "Articulatory gestures as phonological units," Phonology **6**(2), 201–251, https://www.cambridge.org/core/product/identifier/S0952675700001019/type/journal_article, doi: 10.1017/S0952675700001019.

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.

Shih, S. S., and Inkelas, S. (**2019**). "Autosegmental Aims in Surface-Optimizing Phonology," Linguistic Inquiry **50**, 137–196, doi: 10.1162/ling_a_00304.

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.