Cue weight in the perception of Trique glottal consonants

Christian DiCanio[a)]
Haskins Laboratories, 300 George Street, New Haven, Connecticut 06511

(Received 16 November 2012; revised 4 December 2013; accepted 30 December 2013)

This paper examines the perceptual weight of cues to the coda glottal consonant contrast in Trique (Oto-Manguean) with native listeners. The language contrasts words with no coda (/V:/) from words with a coda glottal stop (/V?/) or breathy coda (/Vɦ/). The results from a speeded AX (same–different) lexical discrimination task show high accuracy in lexical identification for the /V:/-/Vɦ/ contrast, but lower accuracy for the other contrasts. The second experiment consists of a labeling task where the three acoustic dimensions that distinguished the glottal consonant codas in production [duration, the amplitude difference between the first two harmonics (H1-H2), and F_0] were modified orthogonally using step-wise resynthesis. This task determines the relative weight of each dimension in phonological categorization. The results show that duration was the strongest cue. Listeners were only sensitive to changes in H1-H2 for the /V:/-/Vɦ/ and /V:/-/V?/ contrasts when duration was ambiguous. Listeners were only sensitive to changes in F_0 for the /V:/-/Vɦ/ contrast when both duration and H1-H2 were ambiguous. The perceptual cue weighting for each contrast closely matches existing production data [DiCanio (2012 a). J. Phon. **40**, 162–176] Cue weight differences in speech perception are explained by differences in step-interval size and the notion of adaptive plasticity [Francis *et al.* (2008). J. Acoust. Soc. Am. **124**, 1234–1251; Holt and Lotto (2006). J. Acoust. Soc. Am. **119**, 3059–3071].
 [http://dx.doi.org/10.1121/1.4861921]

PACS number(s): 43.71.Es, 43.71.Hw, 43.71.An [BRM] Pages: 884–895

I. INTRODUCTION

In human language, a number of different phonetic cues may be used to distinguish between phonological categories. One of the primary questions in phonetic and psycholinguistic research concerns the relative importance of different phonetic cues in the identification of these categories. For instance, there are multiple cues to the voicing contrast in English. The vowel preceding the [d] in the word "*had*" is longer than the vowel preceding the [t] in the word "*hat*." However, [d] also typically contains phonetic voicing during closure, while [t] does not. So, to what extent does a listener pay attention to the duration of the vowel compared to closure voicing? Both phonetic cues are useful to distinguish [d] from [t] (and consequently "*had*" from "*hat*"), but one of the cues may be more useful to listeners than the other. Cues which are more perceptually relevant for listeners are said to be *weighted* more heavily.

Generally, researchers have assumed that those aspects of the speech signal which are most significant in speech production will be strong cues in speech perception (Stevens and Blumstein, 1981). In addition, the psychoacoustic salience of the acoustic dimension and its degree of distributional variance strongly determine the perceptual ranking of cues (Mirman *et al.*, 2004; Holt and Lotto, 2006; Clayards *et al.*, 2008; Francis *et al.*, 2008). The psychoacoustic salience of certain acoustic dimensions, like VOT (voice onset time) and F_0, is well-established, especially in relation to common phonological contrasts. For instance, in the perception of word-initial voicing, most English listeners use VOT as a cue to a greater degree than they use F_0 as a cue (Francis *et al.*, 2008). However, the degree to which these acoustic cues are recruited in the perception of this contrast varies somewhat by speaker (Massaro and Cohen, 1977).

The current study examines the relative weight of different acoustic cues in the perception of final glottal consonants in Itunyoso Trique, an Oto-Manguean language spoken in Mexico. The language contrasts three rime types: long vowels (V:), vowels with coda glottalization (V?), and vowels with coda breathiness (Vɦ). In speech production, these contrasts are distinguished by the duration of the modal vowel, H1-H2 (the difference in amplitude between the first and second harmonics), and F_0 (DiCanio, 2012a). Such data predicts that duration will be a cue for all contrasts, but pitch will only be used to distinguish long vowel rimes from vowels with coda breathiness, as this latter contrast is produced with greater pitch perturbation effects. This paper reports on an experiment designed to determine whether these differences also affect listeners' perception of these contrasts. The first experiment consists of a speeded AX discrimination task, where listeners were asked to distinguish between words with unmodified rime types. This task served as a baseline to determine how well native listeners discriminated among the contrasts. The second experiment consisted of a labeling task (also 2AFC), where three phonetic cues (vowel duration, pitch, and spectral tilt) were independently manipulated. This experiment serves to test the relative weight of different acoustic cues in the perception of glottal consonants.

Despite a plethora of recent studies focusing on the production of glottal consonants and non-modal phonation type, relatively little work has focused on the perceptual relevance of the different acoustic characteristics that researchers have identified (Hillenbrand and Houde, 1996; Gerfen and Baker, 2005;

a)Author to whom correspondence should be addressed. Electronic mail: dicanio@haskins.yale.edu

Frazier, 2009; Esposito, 2010; Kreiman *et al.*, 2010; Kuang, 2011; Kreiman and Gerratt, 2012; Garellek *et al.*, 2013). One of the main findings in this work is that, like segmental contrasts, the perception of phonation type is an acoustically complex phonological contrast involving multiple, contextually dependent cues. In addition to spectral differences related to voice quality, listeners are sensitive to F_0, signal-to-noise ratio, and intensity.[1] While researchers have investigated how well non-native listeners perceive certain cues to non-modal phonation, relatively few studies have examined native listener sensitivity (though, see Esposito, 2010; Kreiman *et al.*, 2010). The current study is unique in this regard. Exploring native listener perception is not only relevant for the empirical study of voice quality, but it pertains to larger issues in speech perception. Notably, work on phonetic cue integration has argued that listeners are sensitive to the temporal dynamics of different cues in word recognition (McMurray *et al.*, 2008). Acoustic cues, mostly coarticulatory, which occur earlier in a word, are used by listeners to make earlier lexical decisions. The relative weight of such cues may determine how available they are for listeners in spoken word recognition.

A. Predictions from Trique speech production

Itunyoso Trique is an Oto-Manguean language spoken in Oaxaca, Mexico (DiCanio, 2008). The language has an inventory of nine lexical tones which are phonologically and morphologically contrastive. The phonology of the tonal system is discussed at length in DiCanio (2008). All syllables in Itunyoso Trique are open with the exception of two possible glottal consonant codas. Word-final syllables may be open, contain a coda glottal stop, or coda breathiness which varies with [h] in careful speech. Examples of this contrast are given in Table I. The contrast between these rime types spans the lexicon and is also involved in morphological alternations marking person.

There is evidence from speech production data that F_0, H1-H2, and duration play a role in distinguishing these rimes. H1-H2 is the difference in amplitude between the first and second harmonics. When there is greater constriction between the vocal folds, as in tense or creaky voice, their closure is faster and results in the stronger excitation (higher amplitude) of the higher harmonics in the spectrum. When there is less constriction between the vocal folds, as in breathy or lax voice, their closure is slower and results in weaker excitation (lower amplitude) of the higher harmonics in the spectrum. Therefore, one can use the difference between the amplitude of a lower harmonic (H1) and a higher one (H2) to evaluate voice quality (Ladefoged *et al.*, 1988).

TABLE I. Glottal consonant contrast on words of different types. Tone is marked here as a superscript after the syllable on which it occurs. "5" reflects the highest tone level and "1" reflects the lowest tone level.

Long vowel rime (/V:/)	Glottalized rime (/V?/)	Breathy rime (/Vfi/)
nːe^{3} *plough*	nːeʔ^{3} *straw rope*	nːefi^{3} *toothless*
kːã^{3} *squash*	kːãʔ^{3} *spicy hominy*	kːãfi^{3} *sandal*
ɾu^{3}mi^{3} *ball, coil*	ɾu^{3}miʔ^{3} *eclipse*	ɾu^{3}mifi^{3} *lumpy*
si^{4}ki^{4} *scab*	si^{3}kiʔ^{3} *to move oneself*	si^{3}kifi^{3} *thing (abstract)*
stĩ^{4} *fingernail*	stĩʔ^{4} *our (du.) fingernail*	stĩfi^{35} *my fingernail*
le^{4}tu^{43} *to bother*	le^{4}tuʔ^{4} *we (du.) bother*	le^{4}tufi^{4} *I bother*

In Itunyoso Trique, vowels are shortest preceding a glottal stop, slightly longer when before a breathy coda (though the transition between the two is gradient), and longest in open syllables. In DiCanio (2012a), the degree to which the glottal codas (/ʔ/ and /fi/) were coarticulated with the preceding vowel on the rime was investigated. The results from this work show that both the glottal stop and breathy coda cause changes on the preceding vowel's voice quality and on syllable's tone, though to different degrees. On /Vʔ/ rimes, the glottal stop is abrupt and lowers $F_0 \sim 20$ Hz on the preceding vowel's tone, but these were limited in duration, occurring in the 20–30 ms immediately preceding glottal constriction. In terms of spectral tilt, H1-A3 values (the amplitude difference between the first harmonic and the strongest harmonic in the third formant) were approximately 5 dB lower on the vowel preceding glottal stop than on vowels in an open syllable, but no significant differences were found for the H1-H2 measure. On /Vfi/ rimes, the breathy coda [fi] occurs with an F_0 lowering of ~20 Hz as well, but this lowering spans a much longer duration of the rime (~83 ms). Significant changes in spectral tilt were also observed. H1-H2 values were ~2–3 dB higher on the breathy rime than on an open syllable rime, while H1-A3 values were ~5 dB higher. The production data from Trique predicts that there will be a greater influence of spectral tilt and F_0 on the perception of breathy rimes in Trique than on rimes with a glottal stop. It also predicts that duration may play a slightly larger role in the perception of glottal stop rimes in Trique than on breathy rimes.

B. Perceptual cue weight

Previous work has found the relative weight of perceptual cues to be primarily determined by their auditory distinctiveness and distributional variance (Mirman *et al.*, 2004; Holt and Lotto, 2006; Clayards *et al.*, 2008; Francis *et al.*, 2008). While cue variance relates to its distribution, auditory distinctiveness is a more ambiguous concept. It could reflect (i) psychoacoustically salient contrasts which are perceived well by the mammalian ear, e.g., rise-time differences (Macmillan, 1987), (ii) a large step size difference along a particular acoustic dimension, e.g., sine wave tones at 100 vs 300 Hz, or (iii) the dynamic nature of the cue itself, e.g., an upward tone sweep vs a level tone.

Each of these manifestations of auditory distinctiveness are relevant to the weight of a specific acoustic cue. In a series of experiments involving perceptual learning, researchers investigated the degree to which VOT and F_0 were used in the perception of English voicing contrasts (Francis *et al.*, 2008). They observed that the difference in step size within the acoustic dimension was the most significant predictor of its perceptual weight. They argued that, since VOT was more psychoacoustically salient to listeners than F_0, it obtains more attentional resources. Furthermore, the temporal structure of an acoustic cue may also determine its perceptual salience. In a study examining the categorization and discrimination of nonspeech sounds, listeners were

more sensitive to dynamically changing spectral cues than steady-state spectral cues (Mirman *et al*., 2004). This finding remained robust even after listeners received extensive training which biased them toward using the steady-state cue. Such findings are relevant to the current paper since the acoustic differences between the different Trique rime types vary in magnitude and will, as a result, also vary in step size.

Implicit in studies on cue weight is the notion that the perceptual task may play a role. In experiments where the experimenter has reduced the degree to which a particular cue may be informative, listeners are forced to attend to other acoustic dimensions during categorization (Holt and Lotto, 2006). Tasks like these artificially parallel the natural variability that listeners perceive in non-experimental settings, but the task condition is seldom explicitly controlled. In a recent study investigating the acoustic cues of stop voicing among English and Korean listeners, Kong and Edwards (2011) found that English listeners paid little attention to F_0 as a cue compared to VOT. However, their attention to F_0 increased in a context where VOT information was ambiguous. In a related study examining the production and perception of English voicing contrasts, Shultz *et al*. (2012) found that speakers who produced more robust VOT differences across voicing categories were less likely to use F_0 as a cue. These studies share a general finding: Listeners rely on more subordinate cues when dominant cues are ambiguous. In the current study, the role of the categorization task was more explicitly controlled to examine this effect. Attention to certain cues was naturally modified by excluding different acoustic dimensions across different blocks.

C. Acoustic cues and glottal contrasts

The production of glottal consonants varies greatly across different languages and speech contexts within the same language. Glottal plosives are produced with complete glottal closure only in careful speech (Pierrehumbert and Talkin, 1992; Ladefoged and Maddieson, 1996; Gordon and Ladefoged, 2001) or in certain prosodic contexts, like word-final position in Trique (DiCanio, 2008). A more typical variant of a glottal stop is creaky phonation which will, at times, approximate complete glottal closure. The same variability is found in the production of a glottal fricative, which may be produced as breathy phonation, e.g., the English word "*ahead*" [əɦɛd] (Gordon and Ladefoged, 2001; Blankenship, 2002). In a study investigating lenition of /h/ and glottalization in English, researchers found that /h/ was most often produced as vocalic breathiness in prosodically weak contexts (Pierrehumbert and Talkin, 1992), though no studies to date have directly investigated its perceptual cues. Given that glottal consonants are often produced as localized spans of non-modal phonation type, studies on phonation perception are relevant here.

Both F_0 and intensity are used in the perception of glottalization in Coatzospan Mixtec, a language related to Trique (Gerfen and Baker, 2005). Researchers found that intervocalic glottalization was often minimally distinguished by small F_0 and intensity perturbations. Glottalized tokens were realized with an amplitude dip between 3–10 dB and an F_0 dip between 7–40 Hz. Similar F_0 and amplitude perturbations to those found in the production task were resynthesized and presented to listeners in a labeling task. Results showed that even small differences in F_0 and amplitude caused abrupt shifts in categorization responses. When F_0 and amplitude were independently manipulated, a 5.5 dB or 6 Hz difference resulted in a glottalized identification response of 80%. When these cues were dependent, differences of 1–2 Hz and 2.25 dB were sufficient to cause comparable shifts in identification. These results are quite similar to those found previously in English (Hillenbrand and Houde, 1996). Here, English listeners were able to perceive the presence of glottalization in synthetic stimuli using only small F_0 and amplitude perturbations.[2]

Mazatec, an Oto-Manguean language like Trique, has a three-way phonation type contrast among breathy, modal, and creaky vowels. In a study using Mazatec data with neutralized F0 patterns and duration, Esposito (2010) found that Gujarati listeners used spectral tilt[3] cues more consistently than listeners of English and Spanish. While the listeners from all these languages used H1-H2 as a cue for categorizing Mazatec tokens, the Gujarati listeners were more sensitive to changes in this cue. As Gujarati has a two-way phonation type contrast between breathy and modal vowels, Esposito argues that native language experience with phonation type contrasts confers an advantage in the discrimination of phonation type. In a study of Yi, a language with tense and lax phonation types, Kuang (2011) investigated the production and perception of tone and phonation categories among native listeners using EGG (electroglottography) and acoustic measures. Kuang found that native listeners relied most heavily on closed quotient (CQ) derived measures for categorizing Yi words into the phonation categories. CQ is most heavily correlated with the amplitude of the first harmonic (H1) and the H1-H2 measure (DiCanio, 2009; Kuang, 2011; Esposito, 2012). Each of these studies argue that H1-H2 is a significant cue in the perception of phonation type contrasts.

A greater number of speech production studies on phonation type have examined the role of F_0 and duration in non-modal phonation. Breathy phonation is associated with F_0 lowering in various languages (Gordon and Ladefoged, 2001; DiCanio, 2009; Garellek and Keating, 2011; Esposito, 2012) as is creaky phonation (Hillenbrand and Houde, 1996; Gordon and Ladefoged, 2001). Tense phonation is typically associated with F_0 raising (Gordon and Ladefoged, 2001; DiCanio, 2009). Phonation types may also differ in terms of duration (Gordon and Ladefoged, 2001; DiCanio, 2009), though relatively few studies have examined it as a perceptual cue. Tense and breathy-tense vowels are shorter in duration than both breathy and modal vowels in Takhian Thong Chong (DiCanio, 2009), breathy vowels are longer in duration than modal vowels in Gujarati (Fischer-Jørgensen, 1967), and breathy vowels are longer than modal and creaky vowels in Jalapa Mazatec (also Oto-Manguean; Kirk *et al*., 1984).

The present experiments test the weight of three acoustic cues (duration, pitch, and H1-H2) in the perception of the glottal consonant contrast. Unlike most previous studies investigating glottal contrasts, these experiments were conducted with native listeners and investigated the role of

duration in perception. Moreover, unlike previous work relying on multi-dimensional scaling (Esposito, 2010; Kuang, 2011), experiment 2 (Sec. III) independently tests the relevance of acoustic cues using an orthogonal three-dimensional design with resynthesized natural tokens embedded in natural sentences. As contrast with previous work, this design allows us to investigate the perceptual relevance of specific acoustic cues when others are ambiguous and explicitly examine listener adaptation in the perception of subordinate cues.

II. EXPERIMENT 1: LEXICAL DISCRIMINATION TASK WITH UNMODIFED STIMULI

A. Methodology

A speeded AX discrimination task was used. Three pairs of unmodified stimuli were presented to listeners in two different orders. The stimuli consisted of two minimal triplets contrasting a long open vowel with a breathy coda rime and a glottal stop coda rime, shown in Table II. These stimuli were recordings of words in isolation from a 28 year old male who had typical realizations of each of the words tested.

For each triplet, three possible pairings of two tokens were presented to listeners, e.g., Vː-Vʔ, Vː-Vɦ, Vɦ-Vʔ. Given that stimulus presentation order plays a role in tonal perception, each pair was presented in two possible orders. Each triplet was presented in a separate block, with six different trials (3 comparisons × 2 orders) presented five times each (30 different tokens). An additional 30 same tokens were presented within each block as well for a total of 60 trials/block. The two blocks were presented pseudo-randomly to participants. Half the participants completed the /nneː^{3}/ triplet block first and the /kkãː^{3}/ triplet block last, while the other half did the reverse. A practice session consisting of 24 trials preceded the 2 blocks.

Following the procedure for a speeded AX discrimination task, stimuli pairs were presented with a short ISI of 100 ms, as in recent work on tone perception (Huang and Johnson, 2010). The experiment was presented using a speeded design in order to control for higher level linguistic perception (Pisoni, 1973). As the stimuli were real words with different frequencies of occurrence, a speeded design is predicted to prevent lexical information from being accessed in the discrimination task. The stimuli should be evaluated by listeners solely on their acoustic differences. Listeners were instructed to respond as quickly as possible by pressing either a button labeled "I" for "*igual*" (same) or a button labeled "D" "*diferente*" (different) on the keyboard. Response time was also recorded starting at the onset of the second stimulus. The computer keyboard was used for collecting response times since a response box was unavailable. Note that we are interested here in the relative differences in response times among conditions and not in specific response times. Moreover, errors in response time (RT) estimation introduced by computer keyboards are largely offset by the much greater between-subject variability in RT (Damian, 2010).

TABLE II. Stimuli used in laryngeal contrast discrimination task doubled consonants mark geminates.

nneː^{3} "*plough*"	nnɦi^{3} "*toothless*"	nneʔ^{3} "*straw rope*"
kkãː^{3} "*squash*"	kkãɦi^{3} "*sandal*"	kkãʔ^{3} "*ground corn*"

Fifteen participants (nine female, six male) were recruited from the town of San Martín Itunyoso, where the experiment was conducted. All participants were bilingual in Trique and Spanish without any reported history of speech or hearing disorders. It should be noted that Trique is the language used in most day-to-day communication in San Martín Itunyoso. It is also the first language learned by all members of the community. Spanish is learned once children attend school and it is relegated only to the academic sphere and interactions with outsiders. Most Trique speakers have no literacy in their native language, but are able to read and write in Spanish. While instructions were given in Spanish, the experiment involved no reading. While there was a mismatch between the instruction language and the language for the stimuli, note that the Trique contrast does not exist in Spanish. All participants were between 18–40 years of age (mean age 24.7 years). The experiment was presented using Psyscope (Cohen *et al.*, 1993) on the a MacBook Pro laptop computer (Apple Computer, Cupertino, CA) over Sennheiser HD448 headphones (Sennheiser USA, Old Lyme, CT) in a quiet room in San Martín Itunyoso.

B. Results

In general, participants perform with great accuracy in lexical discrimination tasks, with correct discrimination performance usually between 80%–100%. Errors in discrimination usually reflect acoustic confusability between stimuli. Yet, of the 15 participants, 4 performed near chance level, with discrimination performance under 60%. The remaining 11 participants had normal discrimination performance above 80%. It was clear from this data that the sub-optimal performance by the four participants reflected their misunderstanding of the task. With exception of the Vː-Vʔ comparison, where two of the four participants performed at 65% accuracy, each of these participants performed equally poorly (<60% accuracy) on all block comparisons. Their performance during practice trials was equally poor. Thus, data from these four participants was not considered in the analysis of the experimental data. Removing these participants' data had the benefit of eliminating all RT values below 400 ms, most of which were spurious.

The average discrimination performance across "different" trials for the remaining 11 Trique listeners was 93.9%. However, there were differences between the stimuli comparisons (Vː-Vʔ, Vː-Vɦ, Vɦ-Vʔ). The Vː-Vɦ contrast was discriminated most accurately (97.7%), followed by the Vː-Vʔ contrast (94.5%), followed by the Vɦ-Vʔ contrast (89.5%). These data along with listener sensitivity values are shown in Fig. 1.

The data were statistically analyzed using a logistic mixed effects model with condition, order of stimulus presentation, and stimulus triplet as factors, and subject treated as a random effect. P levels were calculated using a Markov

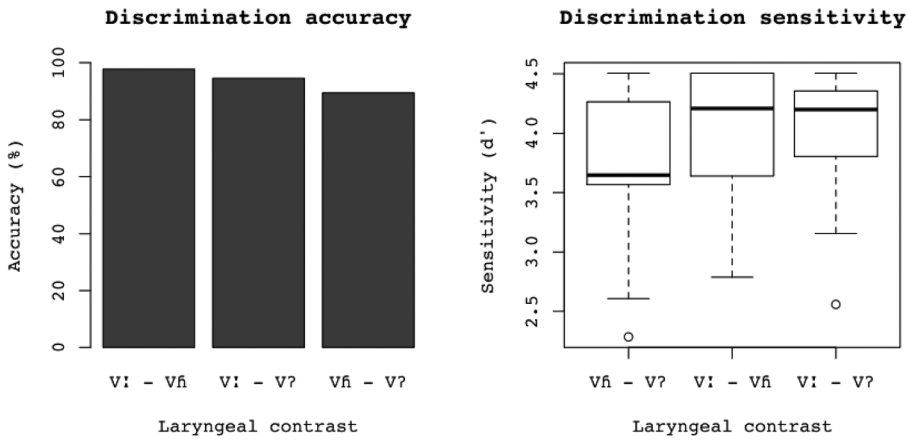

FIG. 1. Performance in Lexical Discrimination Task.

chain Monte Carlo simulation (Baayen, 2008). Mixed effects logistic modeling (MELM) was used because it allows one to perform a *by-participant* repeated measures design with discrete distributions and, unlike arcsine transformations with analysis of variance, MELM properly estimates probabilities when the proportion of responses approach ceiling (Jaeger, 2008). A significant main effect of condition was found for the Vː-Vfi comparison ($z[792] = 2.1$, $p < 0.05$). This reflected greater accuracy in discriminating this pair than the other pairs. No effects of presentation order or the stimulus triplet were observed. Discrimination sensitivity (d') was also computed for each condition (Macmillan and Creelman, 2005). Differences in sensitivity were statistically evaluated using a linear mixed effects model with condition as the fixed effect and subject as a random intercept. No significant differences in sensitivity were found, though sensitivity was lowest for the Vʔ-Vfi comparison. Note, however, that participant responses were close to ceiling for the Vː-Vfi and Vː-Vʔ comparisons.

RT data was recorded along with the discrimination performance. Reaction times which fell greater than two standard deviations away from the mean were excluded, following methods discussed in Baayen (2008). This involved removing just three observations (0.5% of the data). The logarithm of the RT [log(RT)] was statistically analyzed using a linear mixed effects model with condition and order of stimulus presentation as factors, and subject treated as a random effect. Recall that RT was recorded from the onset of the second stimulus and the stimuli differ in duration. The duration of stimuli introduces a bias in recording RTs, as shorter RTs would occur in contexts where a shorter duration stimulus occurred as the second stimulus. To control for the difference in the duration of the stimuli, only stimulus pairs where the shorter duration token was presented first, e.g., /Vfi/ followed by /Vʔ/, were analyzed. In contrast to the analysis of discrimination accuracy, no effect of condition on reaction time was found.

C. Discussion

These findings demonstrate that the Trique glottal consonant contrast was well discriminated by native listeners. Furthermore, the higher discrimination accuracy observed for the Vː-Vfi comparison is noteworthy. Work on Trique glottal consonant production showed that a greater number of cues distinguished this particular contrast than the other contrasts (DiCanio, 2012a). The accuracy of lexical discrimination corresponds with the number of acoustic correlates used for the contrast in production.

It is worth noting that each of the excluded participants who performed at chance level were female. The eldest three participants belonged to this group as well (ages 39, 34, 29). This observation may reflect a clear educational advantage for accurate performance on lexical decision tasks in indigenous communities. Older Trique women are less likely to have attended secondary school and, as a result, typically have had less interaction with computers than men of the same age. Moreover, tasks where listeners are asked to evaluate token similarity on psychoacoustic grounds instead of making lexical decisions may be less natural for speakers of indigenous languages. Three of 18 Trique subjects were removed from a tonal discrimination task in a previous study for similar reasons (DiCanio, 2012b).

Like many phonological contrasts, the perception of the glottal coda contrast in Itunyoso Trique involves a complex relationship between different acoustic cues. Gordon and Ladefoged (2001) catalog a number of different acoustic correlates for phonation type differences, but little is known regarding their perceptual importance for listeners. For instance, spectral tilt differences may significantly correlate with tonal contrasts, but unless speakers systematically vary phonation type independently from tone, they do not attune to this cue (Kreiman *et al.*, 2010). Given that native Trique speakers use spectral tilt and pitch differently in the production of each of the glottal coda contrasts, these cues should have different perceptual weight. The following experiment examines this particular question.

III. EXPERIMENT 2: PERCEPTUAL WEIGHT OF ACOUSTIC CUES FOR GLOTTAL CONSONANT CODAS

A. Methodology

1. Stimuli

The stimuli for the experiment were composed from a minimal triplet of words in Trique contrasting only in rime type and matched for tone: [nne^3] "*plough*," [nneʔ^3]

"*straw rope*," and [nnefi^{3}] "*toothless*." All participants were familiar with these words, a fact which was verified during the practice session. A recording of each word was made in an identical carrier sentence: [*TARGET* ka^{3}tah^{3} ɾi^{3}ə̃^{32}-ɾeʔ^{1}], "*I said TARGET to you*." The word order here (subject-verb-object) reflects a focus construction where emphasis is placed on the target word. The unfocused word order in Trique is verb-subject-object. Each sentence was spoken by a 27 year old male native speaker. The target words were excised and acoustically analyzed in order to compare them to typical examples found in the language; this is found in DiCanio (2012a). The duration, F_0 contour, and H1-H2 values were examined on the vowel in each target word. H1-H2 closely corresponds to the closed quotient portion of the glottal cycle and is widely used in the analysis of voice quality (Blankenship, 2002; DiCanio, 2009, 2012a; Esposito, 2010). Moreover, recent work has demonstrated that this measure is one of the most robust cues listeners use in their perception of non-modal phonation type (Esposito, 2010; Kreiman *et al.*, 2010). Breathy voice has higher H1-H2 values than modal (or creaky) voice, while creaky voice has lower H1-H2 values than modal voice (or breathy voice). The rime in the word [nːefi^{3}] was produced with steadily increasing H1-H2 values and a slightly falling F_0 contour compared to the long vowel rime in the word [nːe^{3}], which was produced with level H1-H2 values and a flat F_0. The vowel preceding the glottal stop coda, in the word [nːeʔ^{3}], was produced with slightly lower H1-H2 values and overall higher F_0 compared to the long vowel rime. This higher F_0 value was somewhat different from the production data shown in DiCanio (2012a), where tone /3/ preceding a glottal stop coda was produced with slightly lower pitch target. However, this previous study did not find any significant pitch differences for tones preceding a glottal stop coda. Overall, the stimuli closely matched the observed average values discussed in DiCanio (2012a). Figure 2 shows the process of spectral tilt resynthesis at four equal continuum steps.

Note that the magnitude of spectral tilt differences between the naturally produced /Vː/ and /Vʔ/ stimuli (−8 dB) is smaller than the magnitude between the /Vː/ and /Vfi/ stimuli (+16 dB). There were notable differences in duration and F_0 as well. The vowel in the /Vː/ stimulus had an average F_0 of 192 Hz with a 10 Hz roll-off in the last 10% of the vowel duration. The vowel in the /Vfi/ stimulus had an average F_0 of 188 Hz with a 25 Hz fall across the second half of the rime (195 → 170 Hz). The vowel in the /Vʔ/ stimulus had a higher average F_0 of 208 Hz with a 6 Hz roll-off in the last 10% of the vowel duration. The duration of the voiced portion of the /Vː/ rime was 297 ms, while the duration of the voiced portion of the /Vfi/ and /Vʔ/ rimes was 137 ms and 89 ms, respectively. Note that because the /Vfi/ rime is produced as a vowel with increasing breathy phonation, this rime duration here includes breathy phonation. Since the vowel in the /Vʔ/ rime is produced with relatively modal phonation, this duration is shorter.

Each of these stimuli were resynthesized along three dimensions: F_0, H1-H2, and duration. A linearly interpolated continuum was created between the vowel of the target word [nːe^{3}] "*plough*" and the vowel duration target, F_0 contour target, and H1-H2 contour target in the other two words. Three sets of acoustic comparisons were made for two pairs of words ([nːe^{3}] "*plough*" vs [nːeʔ^{3}] "*straw rope*," and [nːe^{3}] "*plough*" vs [nːefi^{3}] "*toothless*"). In the first comparison, F_0 and duration were manipulated, in four and six steps, respectively, while H1-H2 was neutralized. In the second comparison, H1-H2 and duration were manipulated, in four and six steps, respectively, while F_0 was neutralized. In the third comparison, H1-H2 and F_0 were manipulated, in six and four steps, respectively, while duration was neutralized. In each of these comparisons, the third cue was neutralized to a value corresponding to the average between the two stimuli endpoints, i.e., average duration of the vowel in [nːe^{3}] and [nːefi^{3}]. The grouping of pairwise comparisons allows us to vary each of the acoustic dimensions independently, but avoids creating an excessively large number of stimuli to present to participants (cf. Zhang and Francis, 2010).

Three dimensions (duration, H1-H2, and F_0) were resynthesized for the stimuli. Duration resynthesis was done by hand. Portions of the vowel were excised at equally spaced time intervals to reduce its duration. Care was taken to not remove portions of the vowel during the consonant-vowel

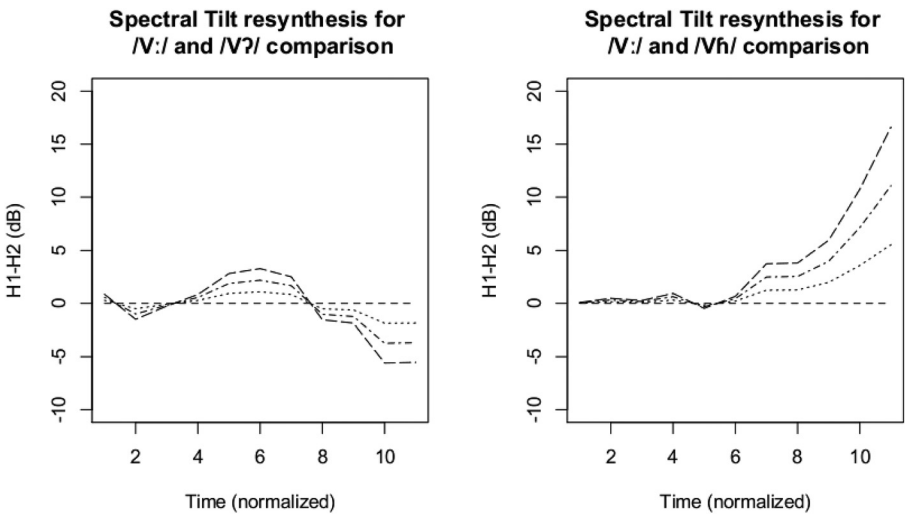

FIG. 2. Resynthesis of spectral tilt values from baseline for each condition. Each line type reflects a different set of values for a step in the spectral tilt continua.

transition. This method maintains the overall shape of the vowel. Separately, F_0 contours were created by linearly interpolating F_0 contours from the naturally produced rime types. Since the F_0 contour on a shorter vowel, e.g., in the word [nːeʔ^3], is produced over a shorter duration than the F_0 contour in the word [nːe^3], F_0 was adjusted independently from duration in the current experiment. In order to do this, a third-order polynomial was fit to each different F_0 contour at each step along the duration continuum, using MATLAB (MATLAB, 2009). To control for potential F_0 discontinuities across the word, separate polynomial equations were fit for onsets and vowels. These polynomials were then used to calculate F_0 values for words of different durations. These F_0 values were then imported into Praat (Boersma and Weenink, 2013) and used as the inputs for F_0 resynthesis. Words which varied in duration along the established continua were resynthesized with these F_0 values using a Praat script written by the author. This process required that duration manipulation precede both F_0 and H1-H2 resynthesis.

Following this step, H1-H2 resynthesis was done for each token using a MATLABscript. This script modifies spectral tilt by adjusting only the amplitude of the first harmonic (H1) dynamically across the duration of the vowel using 11 separate windows for each vowel. The adjustments of the first harmonic made for the resynthesized token corresponded to the set of linearly interpolated H1-H2 values along a six- or four-step continuum. While this results in a general increase of the first harmonic relative to the entire spectrum, it should be noted that the slope of the entire spectrum is also a strong cue for phonation type (Kreiman *et al.*, 2007; Kreiman and Gerratt, 2012). For all stimuli, H1-H2 was the last dimension resynthesized.

2. Stimulus presentation and procedure

Resynthesized stimuli were placed back into the original carrier sentence ([*TARGET* ka^{3}tah^3 ɾi^{3}ə̃^{32}-ɾeʔ^1], "*I said TARGET to you*."). A total of 24 (6 × 4) stimuli were created for each comparison of the 2 stimuli pairs. Each stimulus was presented twice within the experiment for a total of 48 trials. These comparisons were presented in six blocks: F_0 × duration for /Vː-Vɦ/, H1-H2 × duration for /Vː-Vɦ/, F_0 × H1-H2 for /Vː-Vɦ/, F_0 × duration for /Vː-Vʔ/, H1-H2 × duration for /Vː-Vʔ/, and F_0 × H1-H2 for /Vː-Vʔ/. A total of 288 trials were presented to each participant (48 × 6). The experiment was preceded by a 32 trial practice session with mixed stimuli from each of the blocks. This practice session was identical for all participants.

Each trial consisted of the target word embedded in the carrier sentence. Subjects responded via a labeling task where they were asked to identify which word they heard by pressing a key corresponding to a picture on the screen. The picture corresponded to one of two targets of the stimulus set: a plough and a toothless woman for the /Vː/-/Vɦ/ continuum or a plough and a straw rope for the /Vː/-/Vʔ/ continuum. The practice session was particularly useful for the purpose of training participants to use the picture stimuli. In each trial, the pictures appeared on the screen for 500 ms prior to the stimulus sentence so that the listeners could acquaint themselves with the response options. Subjects were instructed to respond as quickly as possible. Identification responses and RTs were recorded for all trials starting from onset of the trial, 500 ms before the stimulus presentation. Subjects were compensated for their participation.

3. Subjects

A total of 14 native listeners (6 female, 8 male) were recruited for the experiment in San Martín Itunyoso, Mexico. The participants for this experiment were a different cohort from those who participated in experiment 1. Subject age ranged from 18–39 years old with no history of speech or hearing disorders. All participants were bilingual in Trique and Spanish, the latter of which was used to provide instructions. Two of the participants in the current experiment were illiterate in both Trique and Spanish. For this reason, all instruction was given verbally to Trique listeners in Spanish. For Trique words, listeners pressed a key corresponding to a picture on the screen. Unlike the first experiment, all subjects understood the task.

B. Results

1. Identification

Identification responses were statistically analyzed in a two-factor logistic mixed effects model where each manipulated phonetic cue was treated as a fixed effect and participant was treated as a random variable. The interaction between the cues was included in the model as well. The alpha level was set to 0.05. In the first set of results, H1-H2 was ambiguous while F_0 and duration were manipulated. This data is shown in Fig. 3.

For the analysis of the data, the effect of dimension on identification was evaluated across steps, e.g., d1…d6 for duration, s1…s4/s6 for H1-H2, and p1…p4 for F_0. The results reflect a comparison between the initial level (d1,s1,p1) and the other levels in the continua. For the /Vː/-/Vɦ/ comparison in Fig. 3 (on the left), there was a significant effect of duration on lexical identification (at d4, $z = -3.6$, $p < 0.001$***; at d5, $z = -4.7$, $p < 0.001$***; at d6, $z = -4.4$, $p < 0.001$***). Rimes with shorter vowel duration were perceived as breathy. There was no significant main effect of F_0 on identification, demonstrating that listeners did not use F_0 cues to identify the target stimuli. For the /Vː/-/Vʔ/ comparison in Fig. 3 (on the right), there was a significant effect of duration on lexical identification (at d3, $z = -2.0$, $p < 0.05$*; at d4, $z = -2.4$, $p < 0.05$*; at d5, $z = -4.3$, $p < 0.001$***; at d6, $z = -4.6$, $p < 0.001$***). Rimes with shorter vowel duration were perceived as glottalized. Similar to the breathy rime condition, there was no significant effect of F_0 on identification. When H1-H2 is neutralized, F_0 is not used as a cue for listeners to distinguish between long vowel and glottalized rimes. Duration is a very strong cue for listeners. No interactions between duration and F_0 were found.

Figure 4 shows the results from blocks where F_0 was ambiguous while duration and H1-H2 were manipulated. For the

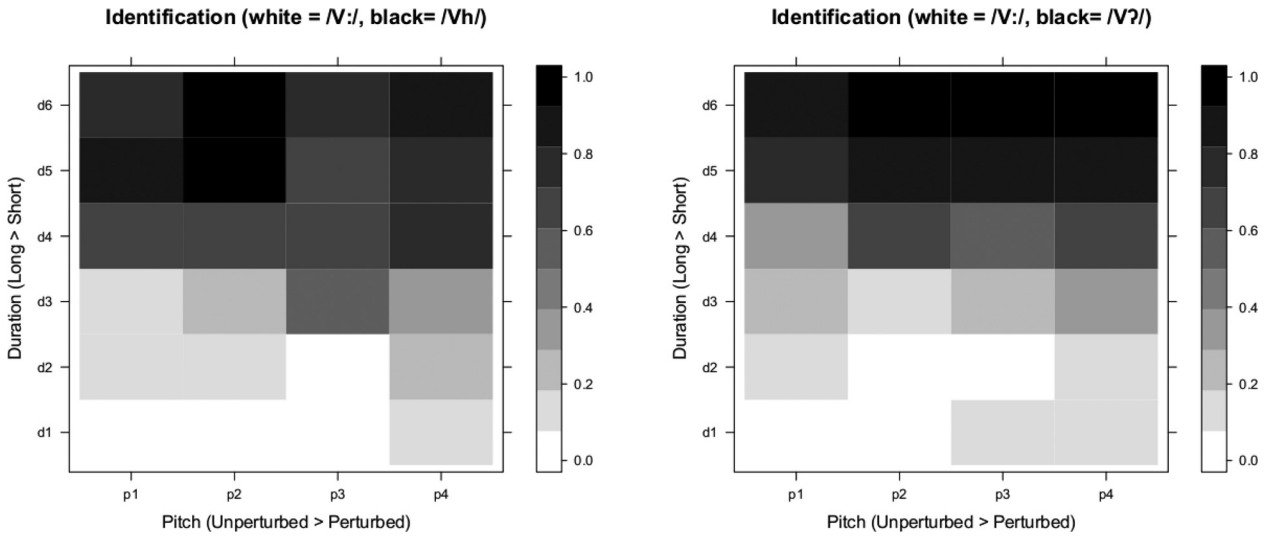

FIG. 3. Effect of F_0 and duration manipulation on identification. Legend shading represents proportion identification for all participants.

/V:/-/Vɦ/ comparison, there was a significant main effect of duration on identification (at d4, $z = -3.6$, $p < 0.001$***; at d5, $z = -5.2$, $p < 0.001$***; at d6, $z = -5.2$, $p < 0.001$***). Words with shorter rimes were perceived as breathy. While there was no significant main effect of H1-H2 on identification, there was a significant interaction between duration and H1-H2. Stimuli resynthesized to sound as breathy as the target /Vɦ/ stimulus (largest H1-H2 increase, s4) were more often perceived with breathy rimes (at d3 × s4, $z = -2.3$, $p < 0.05$*; at d4 × s4, $z = -2.3$, $p < 0.05$*). Thus, when duration was ambiguous along the stimuli continuum, there was an effect of H1-H2 on identification. This effect is most noticeable at step d3, where duration is more ambiguous. H1-H2 was used as a cue by listeners only when duration *and* F_0 were ambiguous. I return to this observation in Sec. IV.

With respect to the /V:/-/V?/ comparison here, there was a significant effect of vowel duration on identification (at d4, $z = -3.0$, $p < 0.01$**; at d5, $z = -4.9$, $p < 0.001$***; at d6, $z = -5.1$, $p < 0.001$***). Stimulus words with shorter rimes were more often perceived as glottalized than those with longer rimes. No effect of H1-H2 was observed for this comparison, nor were interactions between H1-H2 and duration significant.

Figure 5 shows the results from blocks where duration was ambiguous while F_0 and H1-H2 were manipulated. For the /V:/-/Vɦ/ comparison, there were no significant main effects of either F_0 or H1-H2. There were two interactions though. Where H1-H2 was ambiguous (at s3), there was a small effect of F_0 on identification (at p2, $z = -2.5$, $p < 0.05$*; at p4, $z = -1.9$, $p = 0.06$). For the /V:/-/V?/ comparison, there was a significant main effect of H1-H2 on identification, but only when H1-H2 was manipulated to the end of the continuum (at s6, $z = -2.8$, $p < 0.01$**), i.e., when the largest differences in H1-H2 were observed. Otherwise, listeners were unable to categorize these stimuli using these acoustic dimensions.

For both comparisons where duration was ambiguous, participants did not categorize stimuli as discretely in terms

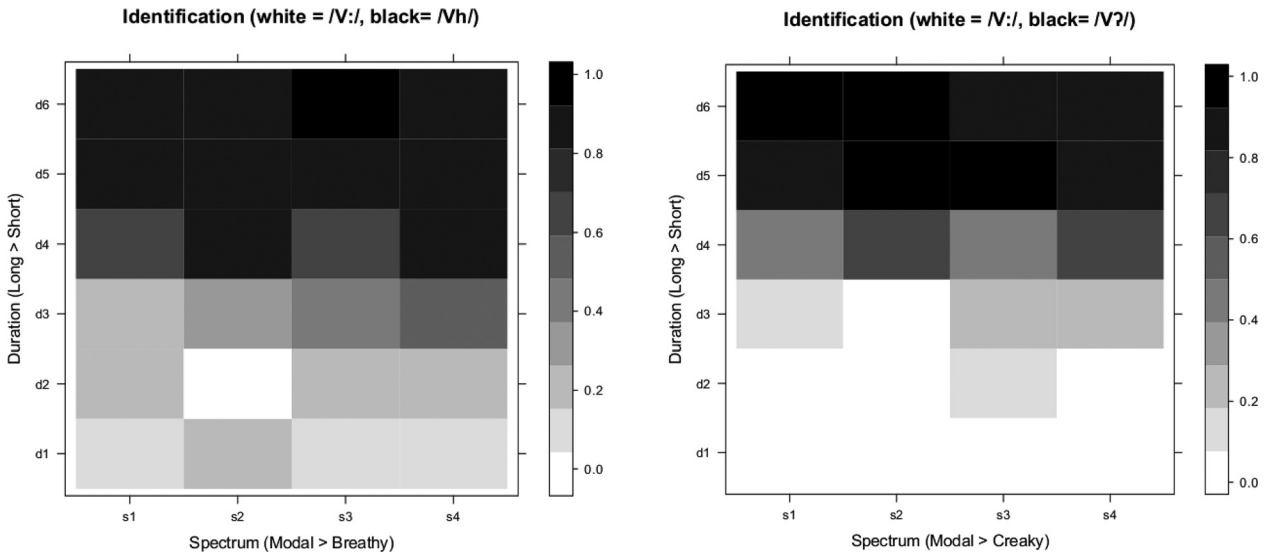

FIG. 4. Effect of H1-H2 and duration manipulation on identification. Legend shading represents percent identification for all participants.

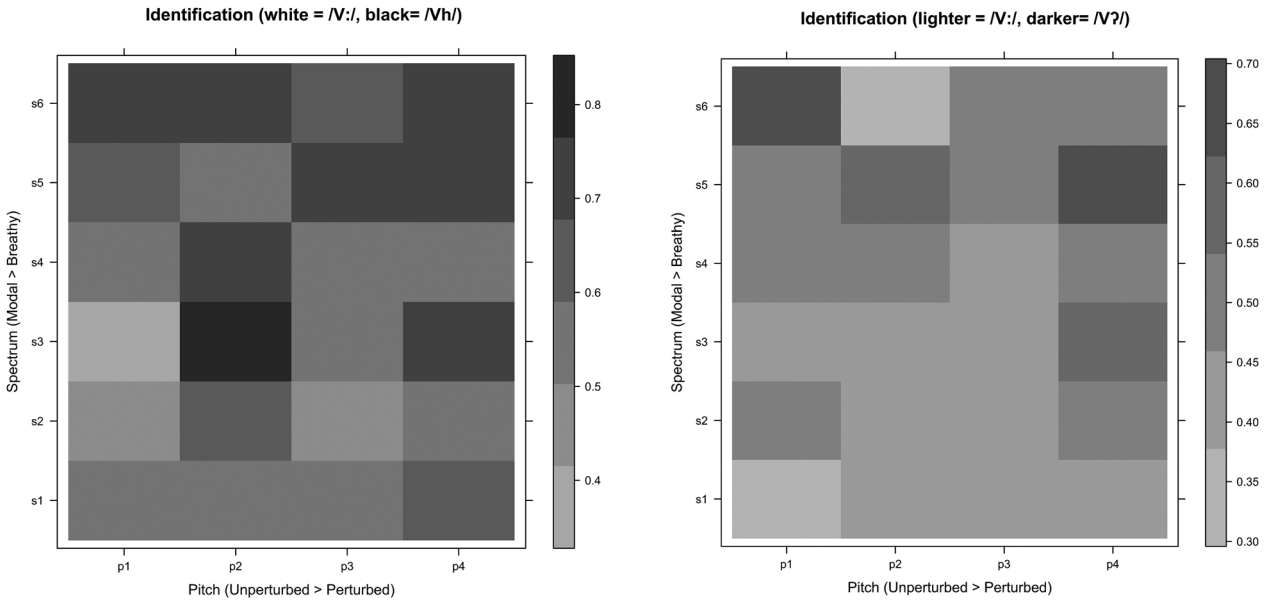

FIG. 5. Effect of H1-H2 and F_0 manipulation on identification. Legend shading represents percent identification for all participants.

of different phonetic cues as they had in blocks where duration had been manipulated. This is apparent from the sharpness of the identification responses. For stimuli which varied in duration and another cue, identification fell between 0%–97%. For stimuli which did not vary in duration, identification fell between 32%–84%. These stimuli were more ambiguous to listeners in these trials. Without duration as a cue to identifying minimal rime pairs, participants performed more poorly at identification.

2. RT

Log(RT) was analyzed with a linear mixed effects model with two independent variables (condition nested under laryngeal contrast) and subject treated as a random effect. Reaction times which fell greater than two standard deviations away from the mean were excluded, following methods discussed in Baayen (2008). This reduced the data by 4.2% ($4032 > 3862$ data points). The data is shown in Fig. 6.

The main effect of glottal consonant contrast was not significant, but there was a significant interaction between contrast type and condition ($t[3862] = 3.1$). For the pitch × H1-H2 condition in the /V:/-/V?/ contrast, RTs were significantly slower. As we observed in Sec. III B 1 where the identification results are shown, listeners did not use the dimensions in this condition for the categorization of this contrast.

3. Summary

The results show a clear preference for duration as a cue for listeners in distinguishing between rime types in Itunyoso Trique. In blocks where duration was manipulated (Figs. 3 and 4), participants treated short duration stimuli as containing a glottal consonant (either /V?/ or /Vʔ/) and long duration stimuli as non-glottal (/V:/). In blocks where duration was neutralized across all stimuli (Fig. 5), participants' identification decisions were less discrete. In these particular blocks, where listeners more heavily relied on H1-H2 as a cue in identification, listeners were also slower at identifying stimuli as a non-glottal or rime with a glottal consonant. Finally, when H1-H2 and duration were neutralized (Fig. 4), participants relied on pitch as a cue in identification.

IV. GENERAL DISCUSSION

The results from experiment 2 suggest that Trique listeners treat the identification of glottalized rimes (/V?/)

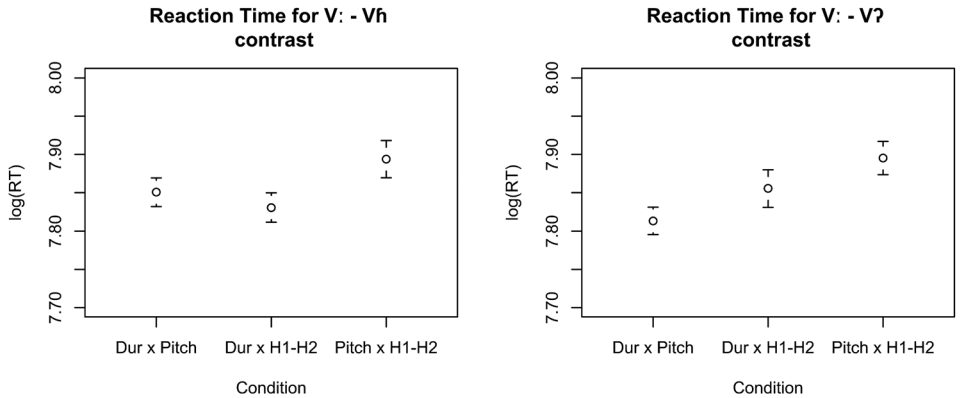

FIG. 6. Effect of condition on reaction time across glottal consonant contrasts.

differently than the identification of breathy rimes (/Vɦ/). In the former case, pitch is never used as a cue in rime identification and H1-H2 is only used when it is substantially lowered. In the latter case, both H1-H2 and pitch may be used in rime identification. These results are schematized in the cue-ranking shown in Table III.

The fact that duration was a strong cue for both stimuli comparisons helps explain the results from the discrimination task. Listeners were worst at discriminating the /VʔV/-/VɦV/ contrast. Duration was less helpful as a cue to this contrast because the vowels preceding the glottal codas are of similar duration. Vowel duration is rarely examined as a perceptual cue to glottal consonants, though there is a reason why this cue would be used in Itunyoso Trique: The final glottal consonants in the language are the only possible codas. Closed syllables often have shorter vowel duration than open syllables (Maddieson, 1985). This phonetic cue is useful for categorization of glottal and non-glottal rimes by Trique listeners.

Moreover, while the magnitude of H1-H2 changes differed across stimuli comparisons, listeners were able to use spectral tilt as a secondary cue. This finding jibes well with the data from speech production, where spectral tilt differences were more robust than F_0 differences. The lower status of pitch as a perceptual cue for glottal consonants also fits with the other phonetic contrasts in the language. Itunyoso Trique has a robust lexical tone system, so pitch might only be recruited as a secondary cue in glottal consonant perception when other cues are ambiguous, which is what we observed here. In particular, coda breathiness co-occurs with F_0 declination similar to the declination patterns found with falling tones /43/ and /32/ (DiCanio, 2012a). Nevertheless, the fact that pitch can be used as a secondary cue in glottal coda perception fits with predictions made in the literature on tonogenesis, where the loss of glottal codas conditions tonal changes on the preceding syllable (Dürr, 1987; Hombert *et al.*, 1979; Kingston, 2011).

Even though fewer cues were used by listeners to identify /VʔV/ stimuli, participants' reaction times while identifying these stimuli were more significantly affected by changes in duration. In the block where duration was neutralized, participants were slower to categorize these stimuli. These results suggest that duration is a *stronger* cue for the former contrast than for the latter. Such a finding would reflect an inverse relationship between cue weight and the number of cues used to signal a phonological contrast. As the number of possible cues increases, the strength of each individual cue slightly decreases. As a result, individual cues may be stronger for certain contrasts when fewer subordinate cues are available.

These experiments support the hypotheses that adaptive plasticity and step size determine relative cue weight in the perception of phonological contrasts (Repp, 1982; Holt and Lotto, 2006; Francis *et al.*, 2008). The secondary cues were not used by listeners when the main cue of duration was available. When this main dimension was ambiguous either across the experimental block or at a particular step within the block, listeners are forced to attend to other, less informative acoustic dimensions. As the data for experiment 2 was modeled on the natural variability across different stimuli, there were inherent differences in step size. First, the vowel duration differences were slightly greater in the /Vː/-/VʔV/ comparison than in the /Vː/-/VɦV/ comparison. Second, spectral tilt and pitch differences were greater in the /Vː/-/VɦV/ comparison than in the /Vː/-/VʔV/ comparison. The result of using this natural variability is a reduction in step size for those cues which differ less in production. Across the stimuli, those cues with smaller step sizes were used less by listeners in categorization.

TABLE III. Weight of phonetic cues by contrast.

/Vː/ vs /Vɦ/: Duration > H1-H2 > Pitch
/Vː/ vs /VʔV/: Duration > H1-H2

V. CONCLUSIONS

In general, the experiments demonstrate a close link between the relative importance of an acoustic cue during speech production and its weight in speech perception. The perceptual results here correspond well with production data discussed in DiCanio (2012a). More coarticulatory overlap was found on /Vɦ/ rimes than on /VʔV/ rimes. In the former case, H1-H2 gradually increased across the rime duration along with slight pitch perturbations. In the latter case, glottalization was abruptly timed so that the pitch on the preceding vowel was not significantly affected by the presence of the glottal coda. There were some small, but significant effects of the coda on spectral tilt (H1-A3). These effects were not as strong as the effects on /Vɦ/ rimes though. Spectral tilt and F_0 were used in the perception of the /Vː/-/Vɦ/ contrast to a greater degree than in the /Vː/-/VʔV/ contrast. Listeners are sensitive to the coarticulatory cues between rime types and use these cues in perception.

The cues examined here are not comprehensive. Glottal consonant codas, like most phonological contrasts, involve a much larger set of phonetic cues that can be tested within one or two experiments. Cues such as intensity, jitter, and following closure duration were excluded from the present study. For instance, in the stimuli context, a voiceless stop followed the target word, [*TARGET* ka^{3}tah^{3} ɾi^{3}ə̃^{32}-ɾeʔ^{1}], "*I said TARGET to you*." The duration of silence following the glottal codas was longer than the duration of silence following the coda-less target word /nneː^{3}/ "*plough*." Just as duration on the vowel preceding the glottal stop coda was a strong cue for perception, silence in the transition from the glottal consonant to the following stop may also be perceptually relevant. Furthermore, the production of coda glottalization involves a short duration increase in jitter (30–40 ms) immediately preceding glottal closure. This jitter was not tested in the current experiment, though it has been found to be a significant cue to glottalization in Yucatec Maya (Frazier, 2009). Last, the production of non-modal phonation often involves decreases in global amplitude in the speech signal (Gordon and Ladefoged, 2001). The data here were normalized for intensity. Each of these cues may be relevant

to the perception of glottal coda contrasts, but they were not examined here.

The notion of adaptive plasticity predicts that the availability of the examined acoustic cues for categorization of glottal consonant contrasts will heavily depend on the speech context. For instance, at a faster speech rate, durational differences may be more ambiguous while vowel-glottal coarticulation may increase, causing greater differences in voice quality. In such a context, we would predict that secondary, non-durational cues increase in perceptual weight. In order to adapt to this type of context, the phonetic categories encoding the contrast must be flexible (Repp, 1982; Holt and Lotto, 2010). Further explorations into such context-dependency of this sort will shed light on the degree to which phonological contrasts involve dynamic representations.

ACKNOWLEDGMENTS

Funding for this work was provided by a grant from the Fyssen Foundation, *The Cross-Linguistic Perception of Tone and Phonation Type*, while the author was a postdoctoral researcher at Laboratoire Dynamique du Langage at Université Lyon 2. Later support for this work was provided by National Science Foundation Grant No. 0966411 to Haskins Laboratories (Whalen, Principal Investigator). Special thanks are given to François Pellegrino for assistance with experimental design and for his script for spectral tilt resynthesis. Thanks are also given to David Braze, Grant McGuire, and John Kingston, for comments on earlier versions of this work.

[1]A plethora of other acoustic correlates to phonation type and glottal consonants have been observed as well, notably vowel quality and duration, but at the time of writing, their perceptual relevance has not been examined.

[2]Though, it should be noted that English listeners required larger perturbations than Mixtec listeners did for a similar change in their categorization response. This suggests an effect of language experience on sensitivity to these cues; while glottalization is phonologically contrastive in Mixtec, it is not phonemically contrastive in English.

[3]Such as H1-H2, H1-A1, H1-A2, etc.

Baayen, R. H. (**2008**). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R* (Cambridge University Press, Cambridge, UK), pp. 1–353.

Blankenship, B. (**2002**). "The timing of nonmodal phonation in vowels," J Phonetics **30**, 163–191.

Boersma, P., and Weenink, D. (2013). "Praat: Doing phonetics by computer (Version 5.1) [computer program]", http://www.praat.org (Last accessed 10/19/2013).

Clayards, M., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (**2008**). "Perception of speech reflects optimal use of probabilistic speech cues," Cognition **108**, 804–809.

Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (**1993**). "Psyscope: A new graphic interactive environment for designing psychology experiments," Behav. Res. Methods, Instrum., Comput. **25**, 257–271 (Version B57 for OSX, 2010), available at http://psy.cns.sissa.it/ (Last viewed 6/1/2010).

Damian, M. F. (**2010**). "Does variability in human performance outweigh imprecision in response devices such as computer keyboards?," Behav. Res. Methods **42**, 205–211.

DiCanio, C. T. (**2008**). "The phonetics and phonology of San Martín Itunyoso Trique," Ph.D. thesis, University of California, Berkeley.

DiCanio, C. T. (**2009**). "The phonetics of register in Takhian Thong Chong," J. Int. Phonetic Assoc. **39**, 162–188.

DiCanio, C. T. (**2012a**). "Coarticulation between tone and glottal consonants in Itunyoso Trique," J. Phonetics **40**, 162–176.

DiCanio, C. T. (**2012b**). "Cross-linguistic perception of Itunyoso Trique tone," J. Phonetics **40**, 672–688.

Dürr, M. (**1987**). "A preliminary reconstruction of the Proto-Mixtec tonal system," Indiana: Contributions to the Ethnology and Archaeology, Linguistics, Social Anthropology, and History of Indigenous Latin America **11**, 19–60.

Esposito, C. (**2010**). "The effects of linguistic experience on the perception of phonation," J. Phonetics **38**, 306–316.

Esposito, C. (**2012**). "An acoustic and electroglottographic study of White Hmong tone and phonation," J Phonetics **40**, 466–476.

Fischer-Jørgensen, E. (**1967**). "Phonetic analysis of breathy (murmured) vowels in Gujarati," Indian Linguist. **28**, 71–139.

Francis, A. L., Kaganovich, N., and Driscoll-Huber, C. (**2008**). "Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English," J. Acoust. Soc. Am. **124**, 1234–1251.

Frazier, M. (**2009**). "The production and perception of pitch and glottalization in Yucatec Maya," Ph.D. thesis, University of North Carolina, Chapel Hill.

Garellek, M., and Keating, P. (**2011**). "The acoustic consequences of phonation and tone interactions in Jalapa Mazatec," J. Int. Phonetic Assoc. **41**, 185–205.

Garellek, M., Keating, P., Esposito, C. M., and Kreiman, J. (**2013**). "Voice quality and tone identification in White Hmong," J. Acoust. Soc. Am. **133**, 1078–1089.

Gerfen, C., and Baker, K. (**2005**). "The production and perception of laryngealized vowels in Coatzospan Mixtec." J. Phonetics **33**, 311–334.

Gordon, M., and Ladefoged, P. (**2001**). "Phonation types: A cross-linguistic overview," J. Phonetics **29**, 383–406.

Hillenbrand, J. M., and Houde, R. A. (**1996**). "Role of F_0 and amplitude in the perception of intervocalic glottal stops." J. Speech Hear. Res. **39**, 1182–1190.

Holt, L. L., and Lotto, A. J. (**2006**). "Cue weighting in auditory categorization: Implications for first and second language acquisition," J. Acoust. Soc. Am. **119**, 3059–3071.

Holt, L. L., and Lotto, A. J. (**2010**). "Speech perception as categorization," Atten., Percept., Psychophys. **72**, 1218–1227.

Hombert, J. M., Ohala, J. J., and Ewan, W. (**1979**). "Phonetic explanations for the development of tones," Language **55**, 37–58.

Huang, T., and Johnson, K. (**2010**). "Language-specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners," Phonetica **67**, 243–267.

Jaeger, T. F. (**2008**). "Categorical data analysis: Away from ANOVAS (transformation or not) and towards logit mixed models," J. Mem. Lang. **59**, 434–446.

Kingston, J. (**2011**). "Tonogenesis," in *Blackwell Companion in Phonology*, edited by M. van Oostendorp, C. J. Ewen, E. Hume, and K. Rice (Blackwell, Oxford, UK), Vol. 4, Chap. 97, pp. 2304–2333.

Kirk, P. L., Ladefoged, P., and Ladefoged, J. (**1984**). "Using a spectrograph for measures of phonation types in a natural language," UCLA Working Papers in Phonetics **59**, 102–113.

Kong, E. J., and Edwards, J. (**2011**). "Gradient perception of laryngeal contrast in stops in English and Korean: Eye-tracking evidence," J. Acoust. Soc. Am. **129**, 2418.

Kreiman, J., and Gerratt, B. R. (**2012**). "Perceptual interaction of the harmonic source and noise in voice," J. Acoust. Soc. Am. **131**, 492–500.

Kreiman, J., Gerratt, B., and Antoñanzas Barroso, N. (**2007**). "Measures of the glottal source spectrum," J. Speech, Lang., Hear. Res. **50**, 595–610.

Kreiman, J., Gerratt, B. R., and ud Dowla Khan, S. (**2010**). "Effects of native language on perception of voice quality," J. Phonetics **38**, 588–593.

Kuang, J. (**2011**). "Production and perception of the phonation contrast in Yi," Master's thesis, UCLA.

Ladefoged, P., and Maddieson, I. (**1996**). *Sounds of the World's Languages* (Blackwell, Oxford, UK), pp. 1–425.

Ladefoged, P., Maddieson, I., and Jackson, M. (**1988**). "Investigating phonation types in different languages," in *Vocal Physiology: Voice Production, Mechanisms and Functions*, edited by O. Fujimura (Raven, New York), pp. 297–317.

Macmillan, N. A. (**1987**). "Beyond the categorical/continuous distinction: A psychophysical approach to processing modes," in *Categorical*

Perception: The Groundwork of Cognition, edited by S. R. Harnad (Cambridge University Press, Cambridge, UK), Chap. 2, pp. 53–85.
Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide*, 2nd ed. (Erlbaum, Hillsdale, NJ), pp. 1–492.
Maddieson, I. (**1985**). "Phonetic cues to syllabification," in *Phonetic linguistics: Essays in Honor of Peter Ladefoged*, edited by V. A. Fromkin (Academic, New York), pp. 203–221.
Massaro, D. W., and Cohen, M. M. (**1977**). "Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction," Percept. Psychophys. **22**, 373–382.
MATLAB (**2009**). *Version 7.9.0.529 (Release 2009b)* (The Mathworks Inc., Natick, MA).
McMurray, B., Clayards, M. A., Tanenhaus, M. K., and Aslin, R. N. (**2008**). "Tracking the time course of phonetic cue integration during spoken word recognition," Psychonomic Bull. Rev. **15**, 1064–1071.
Mirman, D., Holt, L. L., and McClelland, J. L. (**2004**). "Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues," J. Acoust. Soc. Am. **116**, 1198–1207.
Pierrehumbert, J., and Talkin, D. (**1992**). "Lenition of /h/ and glottal stop," in *Papers in Laboratory Phonology 2: Gesture, Segment, Prosody* (Cambridge University Press, Cambridge, UK), pp. 90–117.
Pisoni, D. B. (**1973**). "Auditory and phonetic memory codes in the discrimination of consonants and vowels." Percept. Psychophys. **13**, 253–260.
Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**, 81–110.
Shultz, A. A., Francis, A. L., and Llanos, F. (**2012**). "Differential cue weighting in perception and production of consonant voicing," J. Acoust. Soc. Am. **132**, EL95–EL101.
Stevens, K. N., and Blumstein, S. E. (**1981**). "The search for invariant acoustic correlates of phonetic features," in *Perspectives on the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ), pp. 1–38.
Zhang, Y., and Francis, A. L. (**2010**). "The weighting of vowel quality in native and nonnative listeners' perception of English lexical stress." J. Phonetics **38**, 260–271.