Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Phonetics 33 (2005) 311–334

Journal of Phonetics

www.elsevier.com/locate/phonetics

The production and perception of laryngealized vowels in Coatzospan Mixtec

Chip Gerfen[a,*], Kirk Baker[b]

[a]*Department of Spanish, Italian, and Portuguese, The Pennsylvania State University, 211 Burrowes Building, University Park, PA 16802, USA*

[b]*Linguistics Department, The Ohio State University, 222 Oxley Hall, 1712 Neil Avenue, Columbus, OH 43210-1298, USA*

Received 9 November 2000; received in revised form 12 November 2004; accepted 29 November 2004

Abstract

This paper examines the production and perception of contrastively laryngealized vowels in Coatzospan Mixtec (CM), an Otomanguean language of southern Mexico. A production study focuses on four dimensions of the phonetics of contrastively laryngealized vowels for eight speakers: f0 excursions, amplitude drop, amplitude differences between the first and second harmonics, and overall vowel duration. The production study reveals that the implementation of laryngealized vowels is highly variable within and across speakers, and that these vowels cannot be characterized as creaky voiced, but rather must be viewed as produced along a continuum on which creaky voicing constitutes only the most extreme of one of the ends. Strikingly, it is shown that many tokens are realized with highly subtle cues, involving only minimal drops along the f0 and amplitude dimensions. Three perception experiments conducted in a field setting using synthetic speech examined the contributions of f0 drops and amplitude in cueing the percept of laryngealized vowels. These studies agree with the production data in that they reveal speakers to be highly tuned to small changes along the f0 and amplitude dimensions in the experimental stimuli. Taken together, the production and perception data (1) add to our knowledge of the phonetics of an underdescribed and endangered language; (2) add to our knowledge of how phonation types such as laryngealization on vowels can be implemented cross-linguistically; and (3) provide a more finely grained notion of the array of acoustic cues that implement the categorical, contrastive phonological property of vowel laryngealization.

*Corresponding author. Tel.: +1 814 863 9542; fax: +1 814 863 7944.
E-mail addresses: hug2@psu.edu (C. Gerfen), kbaker@ling.ohio-state.edu (K. Baker).

0095-4470/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2004.11.002

1. Introduction

This study examines the production and perception of laryngealized vowels in Coatzospan Mixtec (CM), an Otomanguean language of southern Mexico. As Gerratt and Kreiman (2001) note, discussion of phonation types, both in focus and terminology, can vary widely by intellectual disciple. Following Ladefoged (1983), we employ the term 'laryngealized' as a general term to refer to a particular linguistic use of non-modal phonation, in which the vocal folds are held more stiffly than in modal (i.e., normal) vowel production. This phonation type is also described by Ladefoged as glottalization or creaky voicing. As Blankenship (2002) points out, however, vowel laryngealization does not always yield audible creak. For example, Blankenship notes that in Mazatec (also an Otomanguean language), contrastively laryngealized vowels are not consistently produced with creak. Blankenship thus distinguishes between the general phenomenon of vowel laryngealization and actual creaky phonation, the latter of which can occur in the production of laryngealized vowels, but which is not itself necessarily a requisite component or goal of laryngealization. Anticipating our discussion below, we note that laryngealized vowels in CM are normally produced without audible creak. Bearing this in mind, we follow Blankenship by using the terms laryngealized and glottalized interchangeably in referring to the non-modal manner of phonation employed in CM, while reserving 'creaky voicing' for tokens produced with audible creak.

As is common throughout the Otomanguean language family, CM employs non-modal phonation contrastively for vowels. Like other Mixtec languages, morphemes in CM have two canonical forms, CVV and CVCV, which are traditionally referred to as couplets (Pike, 1948; Longacre, 1957; Josserand, 1982). As first described by Pike and Small (1974), CM laryngealized vowels surface contrastively in both couplet classes. This is seen in minimal pairs such as the CVV pair [ɲḭː] 'scratch' versus [ɲiː] 'buy' or the CVCV pair [lḛndʲu] 'tangled' versus [lendʲu] 'dirty'. Note that the form [lendʲu] syllabifies as [le.ndʲu]. There is no voiced stop series in CM; rather, the voiced/voiceless contrast for obstruent stops is manifest as a contrast between voiceless and prenasalized stops. CVV-type couplets can be comprised of either a long vowel or two different vowels. In both cases, laryngealization is contrastive, as seen in pairs such as [ɲḭː] 'scratch' versus [ɲiː] 'buy' noted above and [tḛu] 'rotten' versus [teu] 'bench'. In couplets such as [tḛu] 'rotten' containing distinct vowels, contrastive laryngealization is always realized on the first vowel. A similar pattern is found in CVCV morphemes in that laryngealization also surfaces only on V1. Thus, we find morphemes such as [lḛndʲu] 'tangled', with a laryngealized V1, while forms such as *[lendʲṵ] are proscribed in the language. Additionally, in CVCV couplets contrastive laryngealization is limited to couplets in which the medial consonant is phonetically voiced, such as [lḛndʲu] 'tangled' or [ða̰βi] 'toss'. In couplets in which the medial consonant is voiceless, the language lacks a laryngealization contrast, except in couplets in which the medial consonant is the voiceless fricative [ʃ]. This exception is exemplified in minimal pairs such as [ʃɨ̰ʃɨ] 'mushroom' and [ʃɨʃɨ] 'coati'.

Phonetically, as we discuss below, there is a notable amount of variability in the production of CM laryngealized vowels. In general terms, however, these vowels are produced with what sounds like a sequencing of modal phonation and laryngealization. This sequencing yields what have been described as echo vowels, because the period of laryngealization is often followed by a brief, rearticulated vowel of identical quality as the vowel which precedes the laryngealized gesture.

From a phonological perspective, the lack of laryngealization throughout the entirety of the vowel raises the question of whether the most appropriate characterization of the system resides in the positing of a contrast between modal and laryngealized vowels or, alternatively, whether the system might better be characterized via the positing of a glottal stop consonant phoneme in the language.

Gerfen (1999) argues on phonological grounds for an analysis in which laryngealization is a vowel feature in CM. Among other arguments, Gerfen notes that positing a glottal stop in the CM inventory raises a number of undesirable phonotactic complications. For example, in CVCV couplets such as [ndʲɨ̰ʃɨ] 'pimple', a consonantal analysis would yield a representation such as [ndʲɨ/ʃɨ], thus rendering |ʔ| as the only licit coda consonant in the language. Alternatively, in order to avoid positing glottal stop codas, one might treat the echo-vowel phenomenon as evidence of trisyllabicity in what we have called CVCV morphemes, e.g., [ndʲɨʔɨʃɨ] 'pimple'. From a distributional perspective, this approach is equally problematic, however, as it offers no explanation for why the vowel following the glottal stop in CVCV-type couplets must be a copy of the initial vowel in the couplet, or for why the only trisyllabic morphemes in the language are forms containing a second syllable in which the onset consonant is a glottal stop. Though the particular analysis of how laryngealized vowels are phonologically licensed differs with Macaulay and Salmon's approach, Gerfen's treatment of these as laryngealized vowels rather than vowels followed by glottal stop accords with Macaulay and Salmons (1995), who argue for treating laryngealization as a vowel feature across the Mixtec languages in general.

A more plausible explanation for the sequencing of modal and laryngealized phonation is found in Silverman's (1995, 1997) observation that sequencing of this type is functionally valuable in what he labels 'laryngeally complex' languages—languages which employ contrastive modal and non-modal phonation for vowels, together with lexically contrastive tone. According to Silverman, the sequencing of phonation types affords speakers the opportunity to implement phonation contrasts (either breathy or laryngealized) as well as tone distinctions in ways that maximize auditory recoverability of both types of contrast. Like the varieties of Mazatec discussed by Silverman (1995, 1997) and Blankenship (2002), CM is laryngeally complex. That is, laryngealized vowels in CM are not the implementation of a particular tone, as is the case, e.g., for the Vietnamese low, constricted tone (Nguyen, 1970). Rather, in CM both contrastive High and Low tones can surface as modal or laryngealized. In fact, this is clearly evidenced in whistle speech—which the first author has heard employed frequently by CM speakers during his field work. When using whistle speech, speakers whistle only the tone patterns of words. Importantly, they whistle through (i.e., factor out) vowel laryngealization completely, so that modal and laryngealized couplets with the same lexical tone pattern are whistled identically. See Gordon and Ladefoged (2001) for a review of non-modal phonation types and their implementation cross-linguistically, both in contrastive and non-contrastive contexts.

Auditorily, laryngealized vowels exhibit a great deal of variation. On one end of the continuum, the most robustly laryngealized tokens are realized with audible creak. Due to the sequencing with modal phonation described above, these tokens sound comparable to an American English utterance with an intervocalic glottal stop, e.g., [ʌʔow] 'uh oh'. Figs. 1 and 2 provide examples of tokens realized with creak for two different speakers. In Fig. 1, we provide a waveform and wide band spectrogram of [ɲṵː] 'ground', produced by BE, while Fig. 2 shows a waveform and spectrogram of the same word produced by TO. Both are males in their twenties. Arrows are

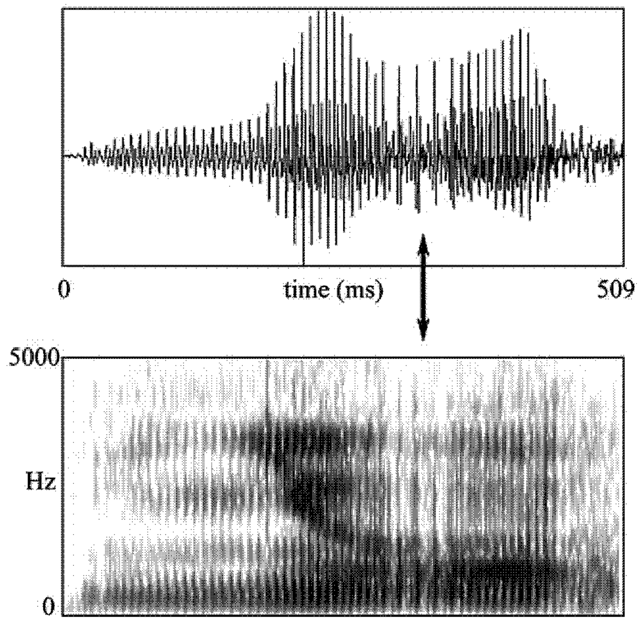

Fig. 1. An example of vowel laryngealization with audible creak in the form [ɲṵː] 'ground', produced by BE, a male in his mid-twenties.

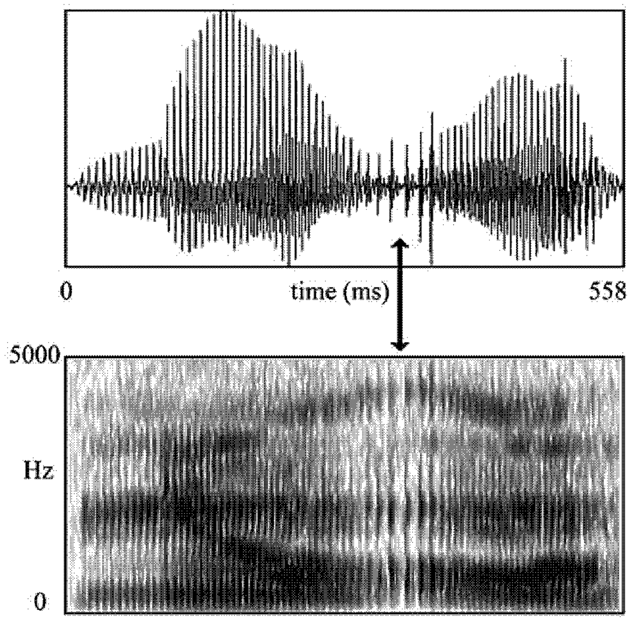

Fig. 2. An example of vowel laryngealization with audible creak in the form [ɲṵː] 'ground', produced by TO, a male in his mid-twenties.

provided for convenience to indicate the visibly creaky voiced portions of the vowel in both tokens.

On the other end of the continuum are tokens produced with a far more subtle laryngealization. Rather than sounding as though they are realized with anything resembling a glottal stop, these instances sound more like they are produced with a weakening of the vowel or a slight hesitation in the vowel's production. These tokens are manifestly not creaky voiced. Fig. 3 exemplifies one of the more subtle cases (as well as a clear case of within speaker variation). The figure provides another token of [ɲṵː] 'ground', also produced by BE, whose creaky voiced token of the same word can be seen in Fig. 1.

Three points are worth noting in Fig. 3. First, we include an f0 track at the bottom of the figure in order to show that the form is realized without a sharp drop in f0, but rather with a relatively flat f0 that is sustained throughout the vowel. Second, by contrast to the token in Fig. 1, although there is a clear amplitude drop in the vowel, once the amplitude dips it remains fairly flat for the duration of the token, with only a slight rise in amplitude towards the end of the vowel. Finally, as seen in the figure, there is far less obvious evidence in either the waveform or the spectrogram signaling the presence of laryngealization. Nevertheless, as indicated by the area marked by

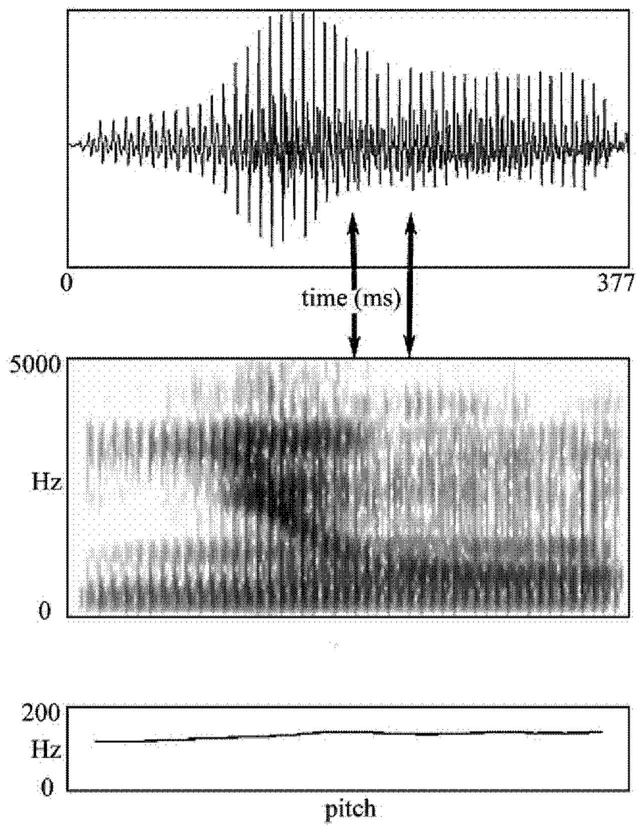

Fig. 3. An example of subtle vowel laryngealization without audible creak in the form [ɲṵː] 'ground', produced by BE, a male in his mid-twenties.

arrows extending from the waveform to the spectrogram, we can observe a loss of energy in the higher frequencies of the spectrogram that is coincident with an alternation of successive higher and lower amplitude glottal pulses for about four cycles in the wave.

While Figs. 1–3 provide examples of the extremes of the continuum, Fig. 4 provides an example of the variability in the data that is representative of neither extreme. In this case, the figure shows the laryngealized vowel (V1) from a token of [ndʲɨ̰ʃɨ] 'pimple', produced by speaker TO, whose creaky vowel in [ɲṵː] 'ground' is shown in Fig. 2. This form is typical of most of the laryngealized vowels in the data in that it is produced without audible creakiness but with clear excursions in amplitude and f0. That is, as can be seen in the figure, the sequencing described above is most obviously signaled upon visual inspection of the waveform, spectrogram, and f0 contour by a drop in both amplitude and f0, followed in this case by a subsequent rise in both, as indicated by the arrows. Auditorily, such tokens yield a clear sense of two beats, but they do not sound creaky voiced.

Finally, for purposes of comparison, Fig. 5 provides an example of a non-laryngealized vowel (V1) for the same speaker, TO, in the minimally distinct [ndʲɨʃɨ] 'dead'. As shown in the figure, the modal vowel does not exhibit either the signature f0 or amplitude dip found in its laryngealized counterpart in Fig. 4.

In short, laryngealized vowels in CM exhibit a wide range of variation. On one end of the continuum, some tokens are robustly laryngealized in that they are produced with audible creak that can sound glottal stop-like to speakers of English. On the other end of the continuum, we find many tokens in which the laryngealization is quite subtle and difficult to hear for speakers not already familiar with the language. Between these two endpoints, we also encounter many tokens with dips and rises of varying magnitudes in amplitude and f0—dips and rises which yield the

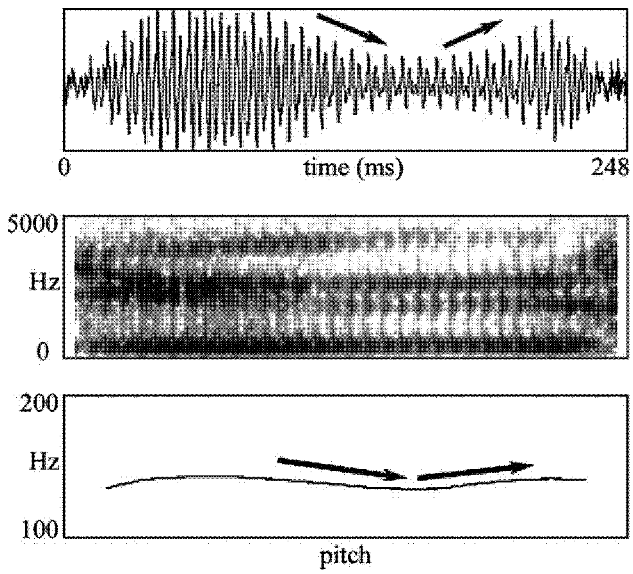

Fig. 4. An example of the laryngealized vowel in the form [ndʲɨ̰ʃɨ] 'pimple', produced by TO, a male in his mid-twenties. The vowel is produced without audible creak, but with clear excursions in both amplitude and f0, as indicated by the arrows.

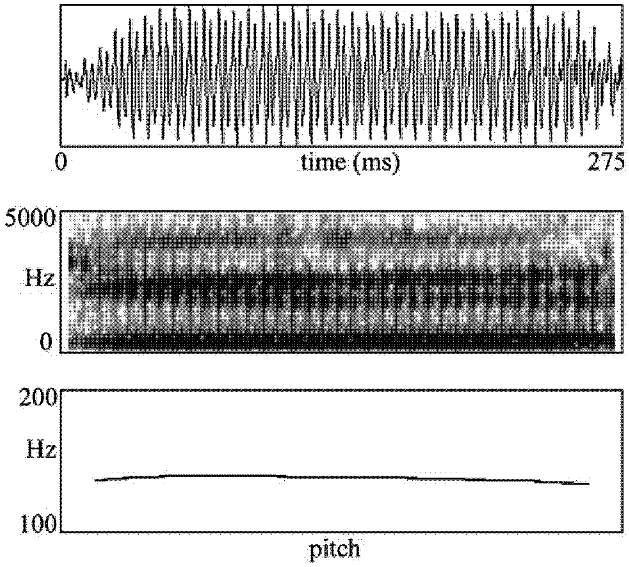

Fig. 5. An example of the modal vowel (V1) in the form [nd^{j}ɨʃɨ] 'dead', produced by TO, a male in his mid-twenties. The vowel lacks excursions in both f0 and amplitude as compared to its larygealized counterpart in Fig. 4.

perception of two clear beats (i.e., the echo vowel effect) but which do not result in audible creakiness. In the following section, we present the results of a controlled production study which examines in closer detail the contributions of four phonetic dimensions to the implementation of CM laryngealized vowels.

2. The production study

In this section, we examine the phonetic implementation of laryngealized vowels, focusing specifically on four aspects of their realization: excursions in amplitude and f0, the duration of laryngealized versus modal vowels, and the spectral differences between laryngealized and modal vowels.

2.1. Data collection

The experimental items under study here are comprised of three minimal pairs containing either a contrastively laryngealized or modal vowel. These are shown in Table 1. The minimal pairs were matched for lexical tone in order to avoid potential confounds introduced by distinct tonal specifications. The morphemes used here were selected to exhibit the range of contrastive possibilities for laryngealization in the language as described above. That is, one minimal pair is of the CVV-type; one is of the CVCV-type with a medial voiced consonant; and one is of the CVCV-type in which the medial consonant is the voiceless fricative /ʃ/.

The experimental words were embedded in a larger list of 72 CM words and phrases, which was randomized six times and presented to speakers in PowerPoint on a laptop computer in the field in

Table 1
Experimental items

Phonetic form	Orthographic form	Gloss	Phonetic form	Orthographic form	Gloss	Tone pattern
[ndʲɨ̰ʃi]	ndɨhxɨ	‘pimple’	[ndʲɨʃɨ]	ndɨxɨ	‘dead’	LH
[ɲṵː]	ñuhu	‘ground’	[ɲuː]	ñuu	‘village’	HH
[kḭni]	kihni	‘tie down’	[kini]	kini	‘push’	HH

Mexico. A single experimental item was presented per slide. The six randomized lists were read aloud by eight native CM speakers (five females and three males) who participated in the experiment, for a total of 8 speakers $\times$ 6 words $\times$ 6 repetitions. Speakers ranged in age from 22 to 45 years old at the time of recording. Though the participants have some Spanish knowledge, all use Mixtec as their primary language in daily life and are strongly Mixtec dominant. In presenting the data, we employed a modified IPA transcription (e.g., using [x] for the palatoalveolar fricative [ʃ] and [h] for the laryngealized vowel marker) designed to take advantage of literacy skills in Spanish and of some familiarity with a transcription system that has been employed by missionaries who have worked in the community. There currently exists no standard written form of CM.

Speakers were first familiarized with the words included in the list by a native speaker trained in the transcription system. Additionally, in the PowerPoint presentation, a Spanish gloss was included in a smaller font below the CM word. Speakers were informed that this was a study of the way that people say words in CM and were asked to speak in a relaxed, normal fashion. They were free to correct themselves if they made a mistake reading a word, and they determined the pace of the experiment by clicking the mouse to proceed to the next slide in the presentation. All instructions were given to the listeners in CM by a native speaker. Recordings were made with a Shure SM10A, close-talking, dynamic, unidirectional, head-worn microphone on a Marantz PMD 222 cassette recorder. Speakers were recorded in a quiet room. The data were digitized on a PowerMacintosh 8600 at 22 kHz with 16 bit sampling and later analyzed in Praat.

2.2. *Dips in amplitude and f0*

Visual examination of waveforms, spectrograms, and f0 contours revealed that the most dependable acoustic signature of laryngealization across tokens and speakers is a clear amplitude drop. Whether or not the drop in amplitude is accompanied by a slight versus a larger pitch drop, or by the onset of visibly creaky phonation is more variable, as illustrated in Figs. 1–4. In this section, we describe in detail the contributions of dips in amplitude and f0 to the implementation of laryngealized vowels in CM.

In order to measure the magnitude of f0 and amplitude drops, we identified landmarks in laryngealized vowels for peak and valley points during the production of the laryngealized part of each vowel. Specifically, using the intensity and f0 tracking functions of Praat, we located cursors at the peak and valley points of the f0 contour and intensity envelope, calculating the magnitude of decline in each dimension as the difference between the peak and valley measures. Fig. 6 provides an illustration of the location of peak and valley markers for f0 and intensity for a token of [ɲṵː] ‘ground’ produced by FE, a female speaker in her early forties.

Focusing first on the f0 dimension, we observe a great deal of variability, ranging from negligible f0 decreases, to deep f0 dips, to tokens with creaky phonation for which the f0 tracking function failed to calculate an f0 contour. Specifically, of the 144 tokens recorded, 24% of the data (35 tokens) exhibited f0 drops of 10 Hz or less, and 53% of the data (77 tokens) exhibited drops of 20 Hz or less. A few tokens were not measured. Five tokens were excluded because they had slight rises in f0 during the implementation of the laryngealized vowel. One token exhibited a 9 Hz rise, while the other four tokens had generally flat f0 contours with rises of between 1 and 3 Hz. On the other end of the continuum, eight tokens were realized with heavy creak, such that Praat's f0 tracker failed to track an f0 contour throughout the laryngealized portion of the vowel. For three tokens with no f0 peak and subsequent valley due to a flat f0 contour, measurements were taken at the amplitude landmarks. Finally, it is interesting to note that all of the tokens recorded were deemed to be acceptable tokens containing laryngealized vowels (i.e., not speech errors) by the native speaker who assisted the first author in running the experiment.

Speakers exhibit a wide range of variation with respect to the amount of f0 drop employed. Although some very large drops in f0 are attested, there is a large degree of variability across the data, with many tokens exhibiting far less dramatic changes in the f0 contour of the vowel. Table 2 provides the mean f0 drop and mean percentage f0 drop by speaker.

Similarly, amplitude drop also exhibits a wide range of variation. Eight of the tokens measured are extremely subtle, exhibiting negligible dips of 1 dB or less, and 34 tokens (24% of the data) are realized with a drop of 3 dB or less. On the other end of the continuum, six tokens are produced with more dramatic declines of over 15 dB. Overall, 74 of the tokens (51% of the data) are realized with a drop of 6 dB or less. Table 3 provides a breakdown of the mean drop in intensity during the production of glottalized vowels for each speaker.

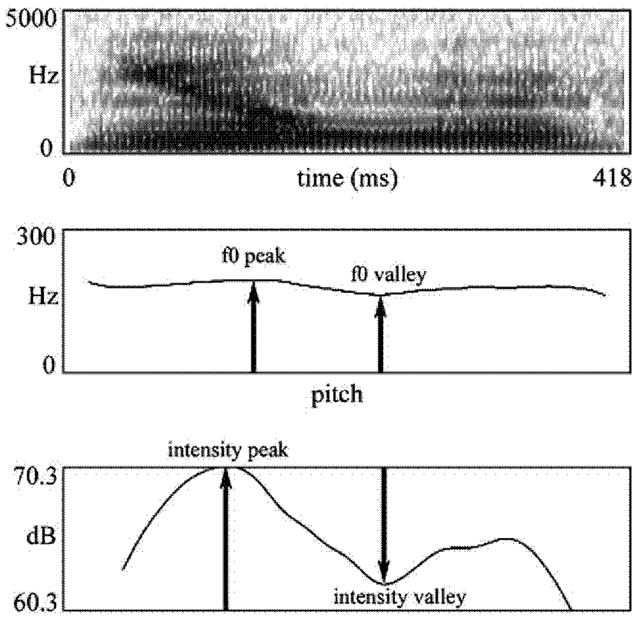

Fig. 6. Landmarks for peak and valley measurements of f0 and intensity for a token of [ɲṵː] 'ground' produced by FE, a female speaker in her early forties.

Table 2
Percentage f0 drop and mean drop in Hz by speaker

Speaker	Sex	Mean %f0 drop	Mean f0 drop (Hz)	SD (Hz)
MP	F	7.56	15.72	9.31
JA	F	10.89	24.35	21.57
FE	F	11.99	23.83	11.32
MR	F	14.05	29.76	35.39
LP	F	20.31	41.94	14.92
SP	M	5.09	7.13	4.35
BE	M	7.21	9.31	7.15
TO	M	24.56	38.93	21.84

Table 3
Mean intensity drop by speaker

Speaker	Sex	Mean dB drop	SD (dB)
MP	F	3.01	1.58
LP	F	5.43	2.46
FE	F	5.61	3.26
JA	F	8.49	2.07
MR	F	10.20	4.68
BE	M	3.52	2.28
SP	M	4.43	3.84
TO	M	9.03	3.25

In summary, the data show a wide range of variability within and across speakers. Declines in f0 and amplitude range from deep drops in both dimensions to tokens with extremely subtle dips. These subtle dips are particularly interesting in that they indicate that vowel laryngealization need not be implemented with robust cues in the f0 and intensity domains. The production evidence suggests that CM speakers might be expected to exhibit a highly tuned sensitivity along the dimensions of both f0 and amplitude as cues to the perception of the contrast between laryngealized and modal vowels in their language, anticipating our perception studies in Section 3.

2.3. Vowel duration

On listening to the forms produced in the experimental context, laryngealized vowels generally sound shorter than their modal counterparts. This impression accords with the first author's general impressions on hearing the contrast in naturalistic contexts in the field for many other laryngealized and modal forms. To test our sense of a length distinction, we measured the duration of the laryngealized vowels and compared them with their modal counterparts. As can be seen in the means in Table 4, modal vowels do exhibit a general pattern of being longer.

To test the durational differences statistically, we pooled the tokens, calculating means for each speaker for the duration of laryngealized versus modal vowels. A paired samples t-test by speaker conducted in SPSS reveals that the duration difference is significant for the laryngealized versus

Table 4
Mean duration in ms for modal versus laryngealized vowel pairs for each speaker

	[ɲuu]	SD	[ɲṵː]	SD	[kini]	SD	[kḭni]	SD	[ndʲɨʃɨ]	SD	[ndʲɨ̰ʃɨ]	SD
BE	288	38	246	20	179	8	164	25	177	11	175	20
FE	435	41	336	14	212	26	175	15	217	16	170	15
JA	409	32	369	41	200	15	200	16	196	12	177	9
LP	383	34	393	42	198	16	158	22	191	26	152	5
MR	463	51	335	24	238	42	189	15	195	35	206	27
MP	400	46	334	18	198	19	165	19	194	29	162	19
SP	355	23	288	24	161	17	136	16	155	4	129	13
TO	483	25	434	33	250	14	246	15	245	15	211	10

modal vowel contrast (df $= 7$, $t = -6.374$, $p = 0.001$). Additionally, we conducted paired samples t-tests on the mean duration values for each minimal pair by speaker to see whether in all three cases the duration was significantly affected by the underlying laryngealized or modal property of the vowel. In each case, the result was significant ([ɲṵː]<[ɲuː]: df $= 7$, $t(7) = -4.112$, $p = 0.005$; [kḭni]< [kini]: df $= 7$, $t(7) = -4.073$, $p = 0.005$; [ndʲɨ̰ʃɨ]<[ndʲɨʃɨ]: df $= 7$, $t(7) = -3.414$, $p = 0.011$). Our results thus confirm our impressionistic sense that modal vowels are longer than their laryngealized counterparts in CM.

2.4. Spectral measures

As noted above, in most cases these vowels do not sound as though they are pronounced with creaky voicing, where we take creak to be characteristic of highly irregular, low-frequency glottal pulsing (cf. Catford, 1964, 1977; Laver, 1980, 1994; Kirk, Ladefoged, & Ladefoged, 1984; Henton & Bladon, 1988; Gobl, 1989; Ladefoged & Maddieson, 1996; Ní Chasaide & Gobl, 1997). As Blankenship (2002) shows, however, laryngealized vowels—though not necessarily realized with audible creak—are often characterized by spectral properties which distinguish them from their modal counterparts. Of particular interest here is the assumption that vowel laryngealization involves a tensing of the vocal folds, yielding a smaller ratio of the open phase of the vocal folds with respect to the complete cycle (see Childers & Lee, 1991), and thus yielding an acoustic effect in which energy is lost in the first harmonic with respect to the energy of the second harmonic.

To see whether this was the case for CM laryngealized vowels, we measured the amplitude of H1 and H2 during the laryngealized portion of each of our laryngealized vowels. For comparison, we measured H1 and H2 at the center of the steady state portion of their modal counterparts. Power spectra were generated in Praat over a 50 ms selection in both the laryngealized and modal vowels and indicate that H1 has relatively more energy with respect to H2 in the modal vowels than in their laryngealized counterparts.

To test the H1–H2 differences statistically, we pooled the tokens, calculating means for each speaker for the H1–H2 difference in dB for laryngealized versus modal vowels. A paired samples t-test by speaker conducted in SPSS reveals that the H1–H2 difference is significant for the laryngealized versus modal vowel contrast (df $= 7$, $t = -4.893$, $p = 0.002$). Additionally, we conducted paired samples t-tests on the mean H1–H2 values for each minimal pair by speaker to

Table 5
Mean H1–H2 intensity difference in dB by speaker for the word pair [ɲṵː] versus [ɲuː]

	[ɲṵː]	SD	[ɲuː]	SD
BE	−12.42	3.55	−17.07	3.04
FE	−8.32	3.35	1.10	0.62
JA	−10.53	7.67	8.28	3.06
LP	−9.50	3.34	−1.82	1.29
MR	−8.57	12.71	8.25	1.63
MP	−8.58	1.42	−2.48	1.65
SP	−10.75	1.28	−4.03	1.13
TO	−13.72	3.04	−3.68	2.05

Table 6
Mean H1–H2 intensity difference in dB by speaker for the word pair [kḭni] versus [kini]

	[kḭni]	SD	[kini]	SD
BE	−16.72	2.78	−15.33	4.32
FE	−7.35	2.89	0.8	0.71
JA	−2.55	4.42	3.85	2.31
LP	−7.98	3.4	−1.28	1.77
MR	3.4	1.95	4.38	1.73
MP	−7.75	3.91	−1.63	2.94
SP	−9.77	5.94	−3.97	0.1
TO	−17.22	2.27	−8.87	2.04

see whether in all three cases the H1–H2 difference was significantly affected by the underlying laryngealized or modal category of the vowel. In each case, the result was significant ([ɲṵː] versus [ɲuː]: df = 7, $t(7) = -3.497$, $p = 0.010$; [kḭni] versus [kini]: df = 7, $t(7) = -5.526$, $p = 0.001$; [ndʲɨ̰ʃɨ] versus [ndʲɨʃɨ]: df = 7, $t(7) = -3.549$, $p = 0.009$). Our results thus confirm the hypothesis that laryngealized vowels will be realized with less energy in H1 relative to H2 than will their modal counterparts. Means tables and standard deviations for each word pair by speaker are provide in Tables 5–7.

2.5. Summary of production study

Our production study has examined four phonetic dimensions: f0 drop, amplitude drop, vowel duration, and H1–H2 energy differences. All appear to contribute to the implementation of the lexical contrast between laryngealized and modal vowels in CM. In the cases of vowel duration and H1–H2 spectral differences, we provide statistical evidence which shows that laryngealized vowels are significantly shorter than their modal counterparts and that laryngealized vowels have relatively less energy in H1 as compared to H2 than do their modal counterparts. The spectral finding is in keeping with Blankenship's (2002) results, e.g., for Mazatec, a language with contrastively laryngealized vowels that is genetically related to CM. Importantly, we also show that laryngealization is not tantamount to creaky phonation and that there are, in fact, many

Table 7
Mean H1–H2 intensity difference by speaker for the word pair [ndʲḭʃi] versus [ndʲɨʃɨ]

	[ndʲḭʃi]	SD	[ndʲɨʃɨ]	SD
BE	−12.85	1.39	−15.82	0.96
FE	−13.47	1.73	−5.30	3.80
JA	−13.53	2.55	−9.53	3.25
LP	−15.70	1.15	−10.32	2.97
MR	−12.35	1.78	−3.97	1.10
MP	−15.25	1.96	−8.52	2.08
SP	−11.38	1.10	−6.32	1.13
TO	−13.70	3.81	−11.40	0.53

tokens in which the implementation of laryngealization is quite subtle, involving minimal dips in the dimensions of f0 and amplitude that cannot simply be viewed as consequences of the implementation of creaky voicing. The second part of this paper begins to explore the issue of how CM speakers perceive the laryngealized/modal contrast. In particular, we focus on the contribution of f0 and amplitude drops as a means of probing CM speakers' sensitivity to subtle manipulations along both of these dimensions.

3. Perception experiments

This section discusses the results of three perception experiments using synthesized CM forms. As we discuss above in the production study, CM laryngealized vowels are characterized by both amplitude and f0 drops. And as we noted, these dips cannot simply be viewed as consequences of the implementation of creaky voicing, since creak is not a necessary component in the realization of laryngealized vowels in the language. Given that we have found expected spectral differences between laryngealized and modal vowels (presumably attributable to increased tensing of the vocal folds), an interesting question arises as to whether such spectral cues are necessary for the perception of these vowels in CM. Specifically, we are led to ask whether amplitude and f0 drops can cue the perception of laryngealized versus modal vowels when we eliminate durational and spectral cues to the laryngeal/modal contrast. Additionally, focusing on f0 and amplitude cues, we investigate whether these two properties can be teased apart. That is, can either cue the perception of larygealized vowels independently, or, e.g., might we best view the amplitude drop as a natural consequence of f0 dips rather than a primary cue in its own right?

The goal of these three experiments is thus to examine whether amplitude and f0 cues can trigger the perception of laryngealized vowels in CM and to subsequently examine more closely the relative contributions of each in so doing.

3.1. Experiment 1

The purpose of Experiment 1 is to test whether or not either amplitude or fundamental frequency alone is a sufficient cue for laryngealization in CM.

3.1.1. Methods

Stimuli were synthesized on the basis of acoustic measurements made from previously digitized tokens of [ɲṵː] 'ground' produced by 4 CM speakers (two females and two males). To make sure that the selected tokens were clear examples of laryngealized forms, four CM speakers (not the subjects whose voices were used) listened to them and identified each as laryngealized [ɲṵː] 'ground'. The four tokens selected for synthesis were all continuously voiced and none showed evidence of irregular glottal pulsing.

Following Hillenbrand and Houde's (1996) study of intervocalic glottal stop perception in English, six versions of each token were synthesized: (a) f0 dip/amplitude dip, (b) f0 dip/flat amplitude, (c) flat f0/amplitude dip, (d) flat f0/flat amplitude, (e) flat f0/inverted amplitude, and (f) inverted f0/flat amplitude. In order to eliminate any potentially confounding cues to laryngealization remaining in the spectrum of the original laryngealized form, we selected an underlyingly non-laryngealized token of the minimally distinct [ɲuː] 'village' produced by each of our four speakers, and mapped the amplitude and f0 values of each speaker's laryngealized [ɲṵː] 'ground' token onto their underlyingly non-glottal [ɲuː] 'village' token. Residual-driven pitch-synchronous LPC analysis and resynthesis (16 kHz sample rate) were performed with Kay Elemetric's Analysis Synthesis Laboratory (ASL) to produce a laryngealized token with the same f0, amplitude, and durational contours as the original underlyingly glottal form for each speaker. The f0, amplitude, and durational parameters of the synthesis conditions containing f0 and amplitude dips are exemplified in Fig. 7.

The flat f0 and amplitude conditions were created by interpolating a line from the first amplitude/f0 peak to the second amplitude/f0 peak in the underlyingly laryngealized tokens, removing the dip. The inverted synthesis conditions were created by rotating the original f0 and amplitude contours about the line interpolated from peak to peak, generating f0 and amplitude contours with maxima in the inverted forms corresponding to minima in the underlyingly laryngealized forms.

3.1.2. Listening test

Listeners for all three experiments consisted of 31 adult CM speakers (approximately 1% of the CM speech community as a whole) ranging in age from 19 to their mid-40 s. The tokens were randomized for each listener and delivered at a comfortable volume through a pair of Sony

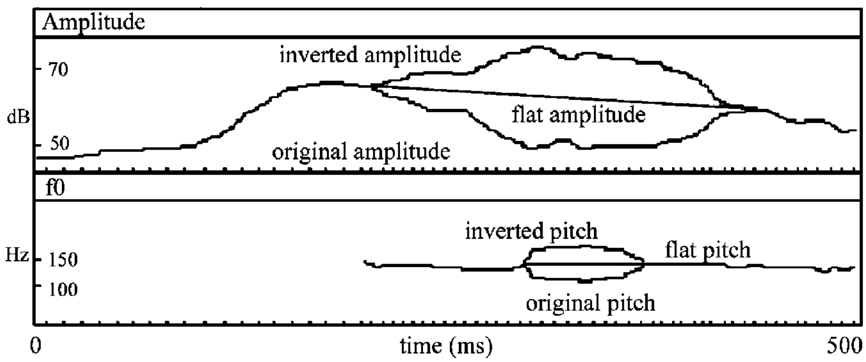

Fig. 7. Sample f0 and amplitude contours used for synthesis.

MDR-7506 headphones, with the volume held constant for all listeners. The listeners were tested individually in a quiet room. Though it was impossible to conduct hearing tests on the listeners, none displayed any evidence of hearing impairment. The subjects were asked to indicate verbally whether the word that they heard was [ɲṵː] 'ground' or [ɲuː] 'village'. The experimenter recorded their responses.

3.1.3. Results and discussion

The results of Experiment 1 are shown in Fig. 8. The amplitude dip/f0 dip condition had the highest laryngealized identification rate at 99%, while 93% and 89% of the listeners identified the flat f0/amplitude dip condition and flat amplitude/f0 dip condition as laryngealized forms, respectively. The flat amplitude/flat f0 condition was identified as laryngealized by 2% of the listeners.

A binary logistic regression analysis of the variance on the labeling data showed a highly significant effect for synthesis condition, $F(5, 744) = 211.06$, $p < 0.0001$. More interestingly, post hoc pairwise comparisons using binary logistic regression on each of the synthesis conditions, with a Bonferroni correction for multiple comparisons of 0.003, showed the following: (a) all of the stimuli involving dips are more likely to be identified as laryngealized than the inverted or flat condition; (b) there are no significant differences between the amplitude dip/f0 dip condition and either of the single-cue dip conditions; (c) the flat amplitude/inverted f0 condition is more likely to be identified as laryngealized than the flat f0/inverted amplitude condition or the flat amplitude/flat f0 condition. It is also interesting to note that 47% of the CM listeners judged the inverted f0 condition to be laryngealized. This might be taken to suggest that along the f0 dimension a two-beat signal can induce the percept of laryngealization, though clearly the f0 dip is the stronger and more reliable condition. (Alternatively, the 47% split might suggest that speakers determined that the stimulus mapped equally badly to both categories and thus responded near chance.)

In short, these results indicate that amplitude and f0 drops, in the absence of other potentially contrast enhancing cues (such as manipulations along continua involving spectral or overall durational differences in the vowels) will cue the perception of laryngealized vowels in CM.

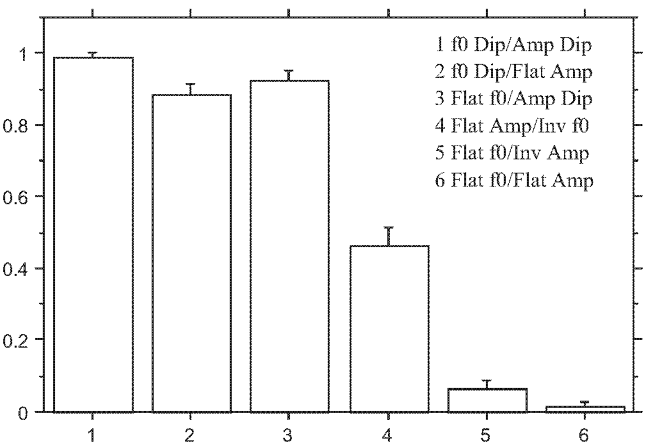

Fig. 8. Bar graph of the laryngealized response rates across the six synthesis conditions (95% confidence error bars).

Importantly, we show that either cue alone can cue the perception of these vowels; i.e., we show that amplitude dips need not be accompanied by f0 drops in order for speakers to perceive laryngealization. In Experiment 2, we examine in a more finely grained fashion the sensitivity of CM speakers to both amplitude and f0 drops in the cueing of laryngealized vowels.

3.2. Experiment 2

The purpose of Experiment 2 is to measure labeling functions using continua that vary systematically in the amount of amplitude and fundamental frequency dips.

3.2.1. Methods

The signals for Experiment 2 comprised two 16-step continua varying in the amount of amplitude dip and f0 dip, respectively. In this experiment, all tokens were synthesized from TO's underlyingly non-laryngealized form [ɲuː] 'village' as in Experiment 1.

Two sets of stimuli with continua varying incrementally along the amplitude and f0 dimensions were synthesized. For the first continuum, an amplitude baseline of 75 dB was established at vowel onset. Beginning 100 ms after vowel onset, amplitude was decreased linearly (in dB) for 50 ms. At the minimum value in the dip, amplitude was held constant for 90 ms and then increased linearly for 50 ms to the original baseline amplitude, where it was held constant for 110 ms until vowel offset. The magnitude of the amplitude dip ranged from 0 to 7.5 dB in 0.5 dB increments. For these stimuli, f0 was held constant at 152 Hz for all signals in the amplitude continuum throughout the vowel duration.

The second set of stimuli varied in the magnitude of an f0 dip with the same durational parameters as the first continuum. The f0 value at the onset and end of the dip was 152 Hz, and the dip's magnitude varied from 0 to 30 Hz in 2 Hz increments. Amplitude was held constant at 75 dB for all of the signals in the f0 continuum for the duration of the vowel. Vowel formants were not altered in any of the signals.

3.2.2. Listening test

The tokens were randomized for each listener for a total of 31 scramblings of the 32 tokens (2 continua × 16 steps). As in Experiment 1, the subjects were asked to indicate verbally whether the word that they heard was [ɲṵː] 'ground' or [ɲuː] 'village'.

3.2.3. Results and discussion

The results of Experiment 2 are shown in Fig. 9. Based on linear interpolation of the 50% point of the group identification function, the phonetic boundary for a laryngealized vowel occurs at 4.1 dB. This result is again consistent with the production data in which amplitude plays a key role. That is, CM listeners are clearly attuned to change along the amplitude dimension as a means of cueing laryngealization. In looking at the graph, we see as well that the laryngealized response pattern rises fairly linearly.

Based on linear interpolation of the 50% point of the group identification function, the phonetic boundary for a laryngealized vowel occurs at 1.8 Hz. Though the frequency of the 50% point in the group identification might seem quite low, as Handel (1993, p.65) points out, at normal speaking levels, people are able to detect changes in frequency from as low as 0.5 to 4 Hz.

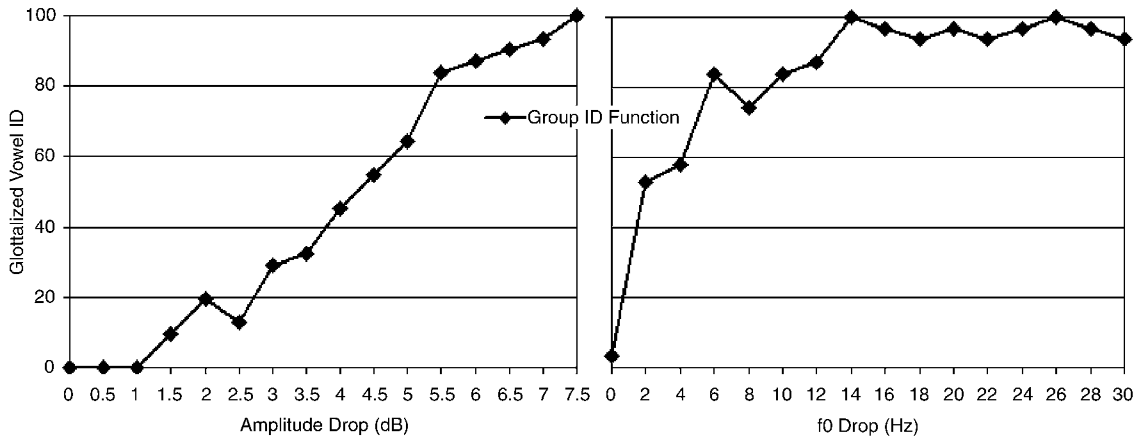

Fig. 9. Group labeling functions for the amplitude and f0 continua.

In looking at the response patterns, we see that while the 50% point does indicate that CM listeners attend to small changes in the frequency dimension, the 50% point does not appear to mark a sharp boundary between the laryngealized and modal vowel categories.

Upon examining these results, then, we note that speakers are quite sensitive to manipulations along both dimensions independently, and that fairly subtle changes along either can trigger the perception of laryngealized vowels. In this sense, the 50% identification point in the f0 continuum is striking in that it reveals speakers to be highly tuned to f0 dips. However, it is important to recognize that some degree of both f0 and amplitude drop are usually present in most laryngealized tokens in naturally occurring speech. With this in mind, in Experiment 3, we turn to the potentially additive effects of these two factors, focusing on the shift in the identification function on the amplitude continuum when f0 is varied incrementally.

3.3. *Experiment 3*

The purpose of Experiment 3 is to examine the additive effects of combining f0 and amplitude cues in the perception of laryngealization.

3.3.1. *Methods*

Four 11-step continua were synthesized (as per the description for Experiment 1) from the non-laryngealized form [ɲuː] 'village' produced by TO, which varied in the magnitude of an amplitude dip in 0.75 dB steps from 0 to 7.5 dB following the contour shapes in Experiment 2. The four continua were synthesized with f0 dips of 0, 1, 2, or 3 Hz.

3.3.2. *Listening test*

The tokens were randomized for each listener for a total of 31 scramblings of the 44 tokens (4 continua × 11 steps). As before, the subjects were asked to indicate verbally whether the word that they heard was [ɲṵː] 'ground' or [ɲuː] 'village'.

3.3.3. Results and discussion

Fig. 10 contains the labeling results for the four amplitude continua in each frequency condition.

The phonetic boundary shifts from the 0 Hz f0 continuum to the 1 Hz f0 continuum are significant, $F(3, 28) = 65.6$, $p < 0.0001$. These statistics were determined by calculating the 50% identification boundary individually for each of the 31 subjects for each of the four continua. We determined the 50% identification boundary for each subject by taking a 3-data point simple moving average. Once this moving average remained above a 50% laryngealized vowel identification rate, we interpolated the identification boundary from within this 3-data point window. We then ran a one-factor analysis of variance on the 31 individual identification boundaries to test for significance against the continua.

The shift in the group identification functions can be seen by comparing the 0 Hz f0 continuum (50% identified laryngealized vowel at 4 dB) to the 1 Hz continuum (50% identified laryngealized vowel at 2.25 dB). This shift toward shallower f0 functions is not maintained for the 2 and 3 Hz continua; instead, by the time a 2 Hz f0 dip is introduced, more than 50% of the subjects reported hearing a laryngealized vowel even with a 0 dB amplitude dip. There is no significant difference between the 2 and 3 Hz f0 continua. Nevertheless, the results for the 0 and 1 Hz continua show that combining an f0 drop with an amplitude drop provides a stronger cue for laryngealization than does the amplitude cue alone. The lack of a difference between the 2 and 3 Hz conditions for CM listeners can be attributed to the sensitivity to smaller change along the f0 dimension that CM listeners show.

4. General discussion and conclusions

In this study, we have examined the phonetics of laryngealized vowels in CM in order to gain a clearer understanding of the dimensions relevant to their production and perception. In our

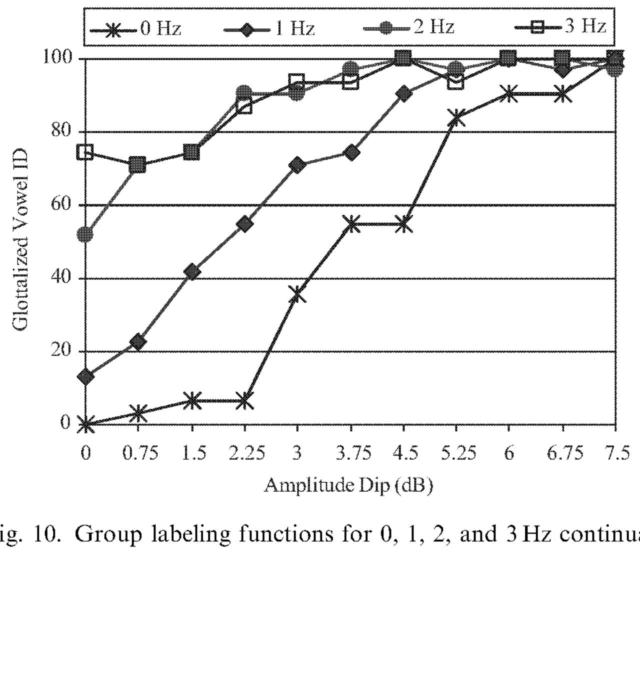

Fig. 10. Group labeling functions for 0, 1, 2, and 3 Hz continua.

production study, we show that CM speakers exhibit a wide range of variability in implementing contrastive laryngealization on vowels. What becomes quickly clear from the data is that laryngealization cannot be equated with creaky voicing. Rather, creaky voicing constitutes one end of the continuum along which laryngealized vowels are produced, a finding in keeping with the results of Blankenship (2002). Despite the variability, our data revealed significant differences between laryngealized vowels and modal vowels both with respect to vowel length and in their spectra. Laryngealized vowels are generally shorter than their modal counterparts, and their spectral properties show a predicted loss of energy in H1 with respect to H2 when compared to the spectra of modal vowels. In our perception experiments, we demonstrate that the other two properties that we initially observed upon inspection of waveforms and spectrograms as being characteristic of laryngealized vowels—dips in f0 and amplitude—can each be shown to trigger the perception of laryngealization in the absence of spectral and durational cues. In fact, we show speakers to be strikingly sensitive to subtle changes along both dimensions, and this finding is in keeping with the large number of subtly laryngealized tokens that we found both within and across speakers in our production data.

One interesting issue raised by our study regards our treatment of these vowels as laryngealized. Recall from Section 1 that we classify these vowels in CM as laryngealized based on two phonological considerations: either glottal stop constitutes the only licit coda in CM or, alternatively, we must allow for trisyllabic morphemes in forms containing glottal stop. As we note, under either approach, one would still have to explain why the vowels surrounding the glottal stop would have to be the same. We argue that the identity of the echo vowel follows trivially from treating CM as a laryngeally complex language, i.e., a language in which both contrastive phonation type and contrastive tones co-occur on vowels. In this sense, the echo vowel phenomenon is not the instantiation of a separate, identical vowel being produced after a glottal stop consonant, but rather, the continued production of a single vowel gesture.

As a reviewer points out, however, one might alternatively claim that glottal stop can simply co-occur with contrastive tone in the same way that we posit that the co occurrence involves the sequencing of laryngealized phonation with contrastive tone. While this approach would fail to explain the identity effects of the vowels surrounding the glottal stop, the reviewer notes that there are other phonological and phonetic considerations which might militate in favor of the glottal stop approach, or, at least, render the choice between the two options less than clear. Phonologically, e.g., it is well-known that codas are often restricted to a subset of segments in languages, so it is perhaps not surprising that the only codas in CM are glottal stops. And phonetically, the duration difference between the modal and laryngealized vowels accords with what we might expect to find in cases of closed syllable vowel shortening (Maddieson, 1984); i.e., the duration difference would appear to follow naturally from a view in which the laryngealized vowels are actually shortened vowels in syllables closed by a glottal stop. Offering indirect support of the closed syllable shortening view, the reviewer also notes that comparative duration data indicate that laryngealized vowels are longer in Mazatec (Kirk et al., 1984) or of non-distinct duration in Mon-Khmer (Thongkum, 1988), rather than of slightly shorter duration as reported here.

The limited nature of the cross-linguistic duration sample, however, is insufficient to warrant the drawing of any firm conclusions; rather, one might equally argue that the laryngealized vowels of CM reveal that cross-linguistically, such vowels can be longer, shorter, or equal in duration to their modal counterparts. The issue of closed syllable shortening, however, merits further consideration. Crucially, in CV̰CV morphemes such as [ʃḭʃɨ] 'mushroom', the vowel shortening account rests on the non-controversial assumption of a [ʃɨʔ.ʃɨ] syllable structure. Such an account becomes problematic, however, when we turn to the treatment of CV̰ː morphemes such as [ɲṵː] 'ground'. Specifically, a shortening story would be predicated on a representation such as [ɲuʔ], where glottal stop surfaces in coda position. Consider, however, the spectrograms in Figs. 1–3. What is clearly visible upon inspection is that the rearticulated vowel following the glottal stop in such forms is not simply a short, echo vowel as is the case in CV̰CV couplets, but rather a longer, full vowel articulation. Simply put, the production of CV̰ː forms such as [ɲṵː] 'ground' is not consequent with a syllable structure such as [ɲuʔ], given that phonetically a full vowel would somehow be assumed to follow the glottal stop that is hypothesized to be in coda position. Rather, to salvage a glottal stop analysis and make the syllabic structure reflect how such words are produced, such forms are better viewed as parsed into two CV syllables, i.e., [ɲu.ʔu]. Paradoxically, however, this parse undermines the argument for closed syllable shortening, since the glottal stop occupies the onset position of the second syllable. To repair the problem, of course, one might stipulate that glottal stop is syllabified as [ɲuʔ.u] in these forms. However, it is important to note that such a stipulation would make recourse of a highly marked syllabification pattern (i.e., a syllabification pattern in clear violation of the Onset Principle) driven entirely by the desire to maintain a uniform closed syllable shortening account of the durational differences between laryngealized and modal vowels in both classes of morphemes in CM. At best, then, the argument for closed syllable shortening becomes circular in the case of laryngealized CV̰ː couplets.

One telling source of external evidence that informs this issue derives from the whistle speech forms of CM words. As we note in Section 1, speakers often whistle the tone patterns of words to communicate with each other. They are also quite adept at employing whistle speech in order to highlight tone patterns of elicited forms for the first author during his fieldwork. Of interest here is how speakers treat intervocalic consonants when whistling the tone patterns of words. In forms such as [ʃɨʃɨ] 'coati', which contain only modal vowels indisputably separated by a consonant, speakers whistle the tone of each vowel separately, pausing between each whistled tone. For forms such as [ʃḭʃɨ] 'mushroom', containing a laryngealized vowel, speakers ignore the laryngealized vowel and whistle a tone for each vowel, also pausing between each whistled tone. In fact, the words [ʃɨʃɨ] 'coati' and [ʃḭʃɨ] 'mushroom' are whistled identically; both are low toned forms, distinguished only by the contrast in vowel laryngealization. By contrast, in CVː forms such as [ɲuː] 'village', speakers whistle the tone pattern with one continuous whistle. Crucially, the whistling is not broken by any pause, since there is no intervocalic consonant as in words such as [ʃɨʃɨ] 'coati'. Of most relevance here is that speakers also whistle laryngealized forms such as [ɲṵː] 'ground' without pausing. The pattern of whistling without a pause holds as well for CVV forms containing non-identical vowels, such as the modal [teu] 'bench' and the laryngealized [tḛu] 'rotten'. Under a glottal stop analysis, this is surprising. Specifically, all other intervocalic consonants in the language yield a sequence of two whistles, one for each vowel on either side of the consonant, while glottal stop does not pattern as a consonant in triggering this effect. By contrast, the pattern is expected under the view that CM contains laryngealized vowels as well as contrastive tone; whistle speech simply consists of whistling the tones. In forms such as [ɲṵː] 'ground', there is no intervening consonant to break up the whistling.

A more general issue at stake in this discussion revolves around the role of phonetic data in determining the phonological status of glottalization in CM. That is, from a phonetician's perspective, it would be of particular interest if the production and perception data were sufficient to dictate our decision regarding the phonological status of the laryngealization phenomenon under scrutiny. In this regard, e.g., the subtle realization of glottalization is quite similar to the results of Hillenbrand and Houde's (1996) English glottal stop perception experiments, a point to which we will return below. In addition, multiple realizations of glottal stop are commonly attested and of the kind we find in the laryngealized vowels in CM. By contrast, the reviewer notes that realizing creak as f0 or amplitude drop may be novel and unmotivated when we can treat laryngealization as glottal stop. We adopt a different view. Specifically, as we note at the outset of this study, Blankenship (2002) shows that laryngealized vowels are not always produced with creaky voicing. That is, we do not assume that laryngealized vowels and creaky phonation are interchangeable terms. Rather, our findings accord with those of Blankenship (2002), who argues for making a distinction between a broader concept of laryngealized vowels and actual creaky phonation. Blankenship notes that creak is not best viewed as the goal of laryngealization, but rather, as a side effect, i.e., as one of a number of cues that can signal the implementation of laryngealized vowels. In this sense, we should not be surprised to find that larnygealized vowels display multiple and sometimes subtle realizations, as do glottal stops, a similarly general category label. In short, if laryngealized vowels are, like glottal stop, realized in multiple ways, determining the status of laryngealized vowels such as those we have examined here requires not simply close phonetic scrutiny but also an understanding of the phonological role they play in a given language.

Much research, e.g., has demonstrated that allophonic and phonemic distinctions involving categorical features (such as [voice]) are signaled by multiple acoustic cues and that the particular expression of these cues varies according to the phonologies of individual languages (Denes, 1955; Abramson & Lisker, 1970; Haggard & Ambler, 1970; Klatt, 1973; Stevens & Klatt, 1974; Lisker, Liberman, Erickson, Dechovitz, & Mandler, 1977; Repp, 1979; Jusczyk, 1993; Bradlow, 1994; Flege, Bohn, & Jang, 1997; Ingram & Park, 1997; Kluender & Lotto, 1998; Lyzenga & Horst, 1998). In this context, our results contribute to a more finely grained understanding of the array of phonetic features contributing to the implementation of a categorical distinction in phonation type. It is also well-known that the way speakers of different languages categorize phonetic continua is conditioned by the number and nature of phonological contrasts in their respective languages (see, e.g., Terbeek, 1977; Jusczyk, 1993; Bradlow, 1994; Flege et al., 1997; Ingram & Park, 1997; Kluender & Lotto, 1998; Lyzenga & Horst, 1998). Since laryngealization marks a lexical contrast in CM, there is good reason to expect speakers to be quite sensitive to both the amplitude and f0 changes that implement the contrast. At the same time, our data add to the growing body of evidence of the ways that languages can differ systematically at the level of fine phonetic detail (see, e.g., Pierrehumbert (1999) and references therein). Given our CM results, it would be interesting to examine other varieties of Mixtec in which laryngealized vowels are contrastive in order to test for whether differences emerge in the roles played by amplitude and f0 (as well as spectral and durational cues) in their production and perception.

While our study adds to our knowledge of how vowel laryngealization is implemented in languages in which it marks a lexical contrast, it is also interesting to note connections between our findings and research in English, which does not have laryngealized vowels but in which glottal stops surface, playing quite different roles than that of vowel laryngealization in CM.

Though a review of the various contexts for glottal stopping in English is beyond the scope of this discussion, as Priestly (1976) observed, glottal stops characterized by complete and sustained vocal fold adduction are only found in the most pronounced realization of glottal stop. Rather, glottal stopping most frequently involves partial rather than complete glottal adduction (Pierrehumbert & Talkin, 1992; see also Fischer-Jørgensen (1989) on Danish), resulting in highly irregular and low frequency glottal pulsing. This scenario yields multiple acoustic correlates, including lowered f0, amplitude drop, spectral tilt and a flattened source spectrum (see, e.g., Gobl, 1989), all of which have also been associated with laryngealized phonation (cf. Blankenship, 2002; Catford, 1964, 1977; Laver, 1980, 1994; Henton & Bladon, 1988; Ladefoged & Maddieson, 1996). As we note above, the overall picture is that there can be significant phonetic overlap between categories that are labeled as glottal stops versus laryngealized vowels cross-linguistically.

In this sense, our perception study can be profitably compared to the perceptual study of English glottal stopping in V–V hiatus contexts by Hillenbrand and Houde (1996) mentioned above. In their focus on the role of amplitude in cueing glottal stops in English, Hillenbrand and Houde show that both amplitude and f0 drops separately can cue the percept of English glottal stop. Their results contrast, e.g., with observations made by Pierrehumbert (1994) and Huffman (1998) that irregular glottal pulsing is the most reliable acoustic cue for identifying glottalization in the wave form and with claims by Pierrehumbert and Frisch (1997) in English synthesis work that while a sharp f0 drop alone can cue glottalization, reduced amplitude and spectral tilt are contrast-enhancing features at best. Similar to what we find for CM vowel laryngealization, Hillenbrand and Houde (1996) show for English that amplitude alone can also trigger the percept of glottal stop. In their recent study of English glottalization in normal speakers, e.g., Redi and Shattuck-Huffnagel (2001), show that English speakers implement glottalization highly variably, both with respect to rate of glottalization and preferred acoustic cues. The picture that emerges as well for glottal stop allophony is one of a continuum of variability along multiple dimensions in a fashion similar to what we have examined in laryngealized vowels in CM.

Finally, we note that our results yield a number of directions for further investigation of vowel laryngealization in CM. First, as we have done for the dimensions of f0 and amplitude, further research will target the possible roles played by overall duration and spectral cues in the perception of these vowels. Additionally, we note that while we synthesized our stimuli to produce acceptable sounding experimental tokens along the [ɲṵː] versus [ɲuː] continuum, we held both the slope of the dips and the time course of the dips, basins, and rises for f0 and amplitude constant. It remains to be seen how the time course of change along these dimensions affects the perception of the phonological contrast in phonation type. For example, we should expect that, all other things being equal, a steeper slope should more robustly cue laryngealization than a more gradual decline in either amplitude or frequency. A more complete picture of how laryngealization is perceived will necessarily involve experimental designs which focus on the potential contributions of such temporal factors.

Acknowledgements

We are grateful to many members of the Coatzospan Mixtec communinty, but wish to express our special thanks to Njau Piña, without whom these studies could not have been conducted. We are also grateful to Wendy Rizzo and audiences at the 2001 meeting of the Linguistic Society of America and at the University of North Carolina at Chapel Hill for their feedback on some of the material discussed here. Finally, we thank Arthur Abramson, two anonymous reviewers, and Doug Whalen for their many helpful comments and suggestions.

References

Abramson, A., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. In: H. Bohuslav, R. Milan, & J. Premysl (Eds.), *Proceedings of the sixth international congress of phonetic sciences, Prague* (pp. 569–573). Prague: Acad. Pub. House of the Czech. Acad. of Sciences.

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

Bradlow, A. R. (1994). *Language-specific and universal aspects of vowel production and perception: A cross-linguistic study of vowel inventories*. Ph. D. dissertation, Cornell University.

Catford, J. (1964). Phonation types. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, & L. M. Trim (Eds.), *In honour of Daniel Jones* (pp. 26–37). London: Longmans.

Catford, J. (1977). *Fundamental problems in phonetics*. Bloomington: Indiana University Press.

Childers, D. G., & Lee, C. K. (1991). Vocal quality factors: Analysis, syntheses, and perception. *Journal of the Acoustical Society of America*, *90*, 2394–2410.

Denes, P. (1955). Effects of duration on the perception of voicing. *Journal of the Acoustical Society of America*, *27*, 761–764.

Fischer-Jørgensen, E. (1989). Phonetic analysis of the stod in standard Danish. *Phonetica*, *46*, 1–59.

Flege, J., Bohn, O., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.

Gerfen, C. (1999). *Phonology and phonetics in Coatzospan Mixtec*. Dordrecht: Kluwer Academic Publishers.

Gerratt, B. R., & Kreiman, J. (2001). Toward a taxonomy of nonmodal phonation. *Journal of Phonetics*, *29*, 365–381.

Gobl, C. (1989). A preliminary study of acoustic voice quality correlates. *Speech Transmission Laboratory Quarterly Status and Progress Report*, *4*, 9–27.

Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics*, *29*, 383–406.

Haggard, M., & Ambler, S. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, *47*, 613–617.

Handel, S. (1993). *Listening*. Cambridge, MA: MIT Press.

Henton, C. G., & Bladon, A. (1988). Creak as a socio-phonetic marker. In L. M. Hyman, & C. N. Li (Eds.), *Language, speech and mind: Studies in honour of Victoria Fromkin* (pp. 3–29). London: Routledge.

Hillenbrand, J., & Houde, R. (1996). Role of f0 and amplitude in the perception of intervocalic glottal stops. *Journal of Speech and Hearing Research*, *39*, 1182–1190.

Huffman, M. K. (1998). Segmental and prosodic effects on coda glottalization. *Journal of the Acoustical Society of America*, *104*, 1818.

Ingram, J., & Park, S. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, *25*, 343–370.

Josserand, J. K. (1982). *Mixtec dialect history*. Ph. D. dissertation, Tulane University.

Jusczyk, P. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, *21*, 3–28.

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

Klatt, D. (1973). Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception. *Journal of the Acoustical Society of America*, *53*, 8–16.

Kluender, K., & Lotto, A. (1998). General contrast effects in speech perception: effect of preceding liquid on stop consonant identification. *Perception and Psychophysics*, *60*(4), 602–619.

Ladefoged, P. (1983). The linguistic use of different phonation types. In D. M. Bless, & J. H. Abbs (Eds.), *Vocal fold physiology. Contemporary research and clinical issues* (pp. 351–360). San Diego: College-Hill Press.

Ladefoged, P., & Maddieson, I. (1996). Recording the phonetic structures of endangered languages. *UCLA Working Papers in Phonetics*, *93*, 1–7.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Lisker, L., Liberman, A., Erickson, D., Dechovitz, D., & Mandler, R. (1977). On pushing the voice-onset-time (VOT) boundary about. *Language and Speech*, *20*, 209–216.

Longacre, R. (1957). Proto-Mixtecan. *International Journal of American Linguistics*, *23*, 1–195.

Lyzenga, J., & Horst, J. W. (1998). Frequency discrimination of stylized synthetic vowels with two formants. *Journal of the Acoustical Society of America*, *104*, 2956–2966.

Macaulay, M., & Salmons, J. (1995). The phonology of glottalization in Mixtec. *International Journal of American Linguistics*, *61*, 38–61.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Nguyen, D. L. (1970). *Vietnamese pronunciation*. Honolulu, Hawaii: University of Hawaii Press.

Ní Chasaide, A., & Gobl, C. (1997). Voice source variation. In W. Hardcastle, & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 427–461). Cambridge: Blackwell Publishers.

Pierrehumbert, J. (1994). Prosodic effects on glottal allophones. In O. Fujimura (Ed.), *Vocal fold physiology* (pp. 39–60). San Diego: Singular Press.

Pierrehumbert, J. (1999). What people know about the sounds of language. *Studies in the Linguistic Sciences*, *29*, 111–120.

Pierrehumbert, J., & Frisch, S. (1997). Synthesizing allophonic glottalization. In J. P. H. van Santen, R. W. Sproat, J. P. Olive, & J. Hirschberg (Eds.), *Progress in speech synthesis* (pp. 9–26). New York: Springer.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Doherty, & R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, degment, prosody* (pp. 90–119). Cambridge: Cambridge University Press.

Pike, K. (1948). *Tone languages*. Ann Arbor: University of Michigan Press.

Pike, K., & Small, P. (1974). Downstepping terrace tone in Coatzospan Mixtec. In R. Brend (Ed.), *Advances in Tagmemics* (pp. 105–134). Amsterdam: North Holland.

Priestly, T. (1976). A note on the glottal stop. *Phonetica*, *33*, 268–274.

Redi, L., & Shattuck-Huffnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, *29*, 407–429.

Repp, B. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, *22*, 173–189.

Silverman, D. (1995). *Phasing and recoverability*. Ph. D. dissertation, University of California at Los Angeles.

Silverman, D. (1997). Laryngeal complexity in Otomanguean vowels. *Phonology*, *14*, 235–261.

Stevens, K., & Klatt, D. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, *55*, 653–659.

Terbeek, D. (1977). A cross-language multidimensional scaling study of vowel perception. *UCLA Working Papers in Phonetics*, *37*.

Thongkum, T. (1988). Phonation types in Mon-Khmer languages. In O. Fujimura (Ed.), *Vocal fold physiology: Voice production, mechanisms and functions* (pp. 319–334). New York: Raven Press.