



Laryngealization and Breathiness in Persian

Vahid Sadeghi

Department of English, Imam Khomeini International University

vsadeghi5603@gmail.com

Abstract

Persian has sequences of two vowels separated by an intervening glottal consonant (/h/ or /ʔ/). The VG(lottal)V sequence becomes reduced in certain occurrences, with the perceptual effect of the loss of the glottal consonant. The purpose of this study is to provide an acoustic description of VGV sequences in reduced forms. A production study examined three acoustic measurements of phonation types: H1-H2, H1-F1, and F0. The measurements were made at 15 ms time intervals throughout the second vowel to determine the time course of phonation effect. The issue of interest is what properties of VGV remain where G is lost. It is shown that /ʔ/ will be preserved as vowel laryngealization and /h/ as breathiness.

Index Terms: spectral tilt, laryngealization, breathiness

1. Introduction

It is commonly suggested that Persian intervocalic glottal consonants are deleted in informal speech [11]. This deletion, however, is not accompanied by lengthening of the preceding vowel, contrary to coda glottal deletion that triggers compensatory lengthening¹. This view asserts that glottal consonants /ʔ/ and /h/ are completely lost in vowel-consonant-vowel context. It further suggests that glottal consonants in this context are not realized as creaky or breathy phonations on their neighboring vowels, rather they show normal voicing that is continuous throughout the glottal constriction gesture. In other words, it is assumed that VGV sequences exhibit modal values of phonation differences along the entire length of the sequence, and that there is no interval of non-modal phonation types corresponding to the entire length of the glottal consonants. Thus, under this view, the lexically contrastive minimal pairs such as /saheb/ "owner" and /saʔeb/ "competent" should have the same acoustic manifestation. Previous research has mainly addressed the phonetic interpretation of glottal consonants in coda position. It is commonly held that Persian glottal consonants in syllable coda undergo vowel lengthening [1], [9], [10], [11]. However, the phonetic interpretation of the phenomenon has been subject to much debate. Earlier studies found compensatory lengthening (CL) as involving the deletion or weakening of glottal consonants, followed by the lengthening of the preceding vowel. More recent studies, however, interpret CL as a gradient process in which different magnitude of glottal gesture is realized in speech from a weak through complete deletion of glottal consonants. Evidence for this view comes from experimental studies in which CL data were compared

1- The asymmetry may be explained by moraic phonology. Coda glottals are considered as moraic, hence when deleted, the remaining moras may be reassigned to the preceding vowel to maintain syllable weight. Intervocalic glottals, on the other hand, are not moraic, so their deletion does not trigger compensatory lengthening due to changes in syllable weight [1], [10].

against non-CL data using various F0 and spectral tilt measurements [1], [9], [10].

In the present paper a production experiment was conducted to examine measurements of phonation differences in the acoustic manifestation of VGV sequences. Values of F0 and spectral slope differences were computed at several time intervals across V2 for instances of VGV and VCV sequences. It was assumed that the vowels in VCV sequences exhibit modal voice quality, and thus their values could be taken as reference values of modal phonation. The key questions examined in the present production experiment are: (1) Are glottal consonants completely lost or simply reduced in VGV sequences? (2) If reduced, what properties of G are preserved for VʔV and VhV sequences separately?

2. Background of Persian phonology

Persian is a six-vowel system with three lax vowels /æ, e, o/ and three tense vowels /a, i, u/. Persian, also contains twenty three consonants of which two are glottal: /h/ and /ʔ/. Syllables are made according to CV(C)C template, although some phonologists suggest the optionality of syllable onset due to the predictability of glottal stop in onset position [1]. Main stress falls on the final syllable except for verbs with specific inflectional affixes.

3. Experiment

3.1. Subjects and tokens

The experimental stimuli consisted of three sets of disyllabic CV.CVC words as exemplars of each phonation type. There were ten words in each set. The words in the first and the second set contained an intervocalic glottal stop [ʔ] and glottal fricative [h] respectively. The words in the third set had an intervocalic stop [d] as examples of modal phonation. All words carried stress on the second syllable. The words were embedded in ten sentence frames with the main stress falling on the embedded words. 10 native speakers, aged between 20-40, with no known knowledge of linguistics, participated in the experiment were asked to utter the randomized sentences colloquially at a normal rate.

The first 90 ms portion of vowels in each related VCV / VGV tokens was designated for comparison. The voiced portion of the target vowels was tagged at 15 ms time intervals. The criteria for vowel onset was that vowels must be voiced and show F1 and F2 on the spectrogram. Voiced portions without a clear F1 and F2 were excluded.

3.2. Recordings

The recordings were made in a sound proof chamber using a Sony cardioid microphone. The microphone was placed at a height equal to and 15 cm in front of the subject's mouth.

Digitization was performed at 20 KHz. Separate recordings were made for V?V/ VCV and VhV/VCV. The data recorded for each included 200 tokens (10 speakers \times 2 syllable templates \times 10 words).

Table 1: *The Persian sample words: Target vowels are underlined>*

After /?/	After /h/	After /d/
saʔ <u>eb</u> : "competent"	sa <u>heb</u> : "owner"	sade <u>g</u> : "honest"
zaʔ <u>er</u> : "pilgrim"	za <u>her</u> : appearance	sade <u>r</u> : "export"
jaʔ <u>el</u> : "forger"	ja <u>hel</u> : "fool"	ʔade <u>l</u> : "righteous"
saʔ <u>æt</u> : "watch"	sa <u>hat</u> : "presence"	ʔade <u>t</u> : "habit"
ʃæʔ <u>æf</u> : "joy"	kæ <u>hæf</u> : "cave"	sæ <u>dæf</u> : "shell"
zæʔ <u>if</u> : "weak"	næ <u>hif</u> : "thin"	ræ <u>dif</u> : "ordered"
mæʔ <u>ɑf</u> : "living"	mæ <u>hal</u> : "impossible"	mæ <u>dɑr</u> : "orbit"
ʃuʔ <u>ur</u> : "wisdom"	zu <u>hur</u> : "realization"	sud <u>ur</u> : "issuing"
sæʔ <u>id</u> : "prosperous"	ʃæ <u>hid</u> : "martyr"	mæ <u>did</u> : "long"
ʃaʔ <u>er</u> : "poet"	ta <u>her</u> : "clean"	ga <u>der</u> : "powerful"

3.3. Analysis

Phonation types are quantified through a number of phonetic parameter [8], [12]. These include fundamental frequency and spectral tilt, among others. Phonation types are commonly associated with lowering of fundamental frequency. For example, laryngealized voice is associated with lowered f_0 values in many languages including Capuga [3] and Coatzospan Mixtec [5]. Additionally, f_0 tends to be lower in breathy vowels, presumably due to the need to slacken the folds to promote continued voicing in spite of the static posterior separation [7].

Spectral tilt, i.e. the degree to which intensity drops off as frequency increases, is also known to differentiate phonation types in a number of languages including Jalapa Mazatec [2], Tagalog [2], Gujarati [4] and !Xoo [6], [8], irrespective of whether nonmodal phonation is a contrastive property of sounds. Spectral tilt can be quantified by comparing the amplitude of the first harmonic to that of the second and higher frequency harmonics (H1-H2 and H1-F1). It is assumed that breathy vowels which are characterized by an increased open quotient and a gradual closing gesture have the greatest drop off in energy as frequency increases, while laryngealized vowels, potentially associated with a decreased open quotient and a more precipitous closing gesture, display the largest boost in energy as frequency is increased [12].

It follows that if VGV involves the complete deletion of glottal consonants, then the measurements of the F_0 changes as well as the changes in the amplitude of the first two harmonics or the first harmonic and the first formant over the course of the second vowel should have similar values for VGV and VCV tokens. On the other hand, if we assume that VGV involves glottal weakening, then F_0 and the differences H1-H2 and H1-F1 should change substantially throughout the duration of the second vowel and that F_0 and spectral tilt values at the initial time intervals corresponding to the duration of glottal consonants in VGV should be significantly different from those in VCV.

Phonetically, as we discuss below, there is some amount of variation in the production of VGV sequences. However, in general terms, they sound like a sequencing of modal-nonmodal-modal phonation. Figure 1 shows the spectrograms, FFT spectra and F_0 contours for the word [saʔeb] "competent" produced by two different speakers: RE and FA. The spectra,

taken over a 25 ms window, were centered at 20th of the vowel [e].

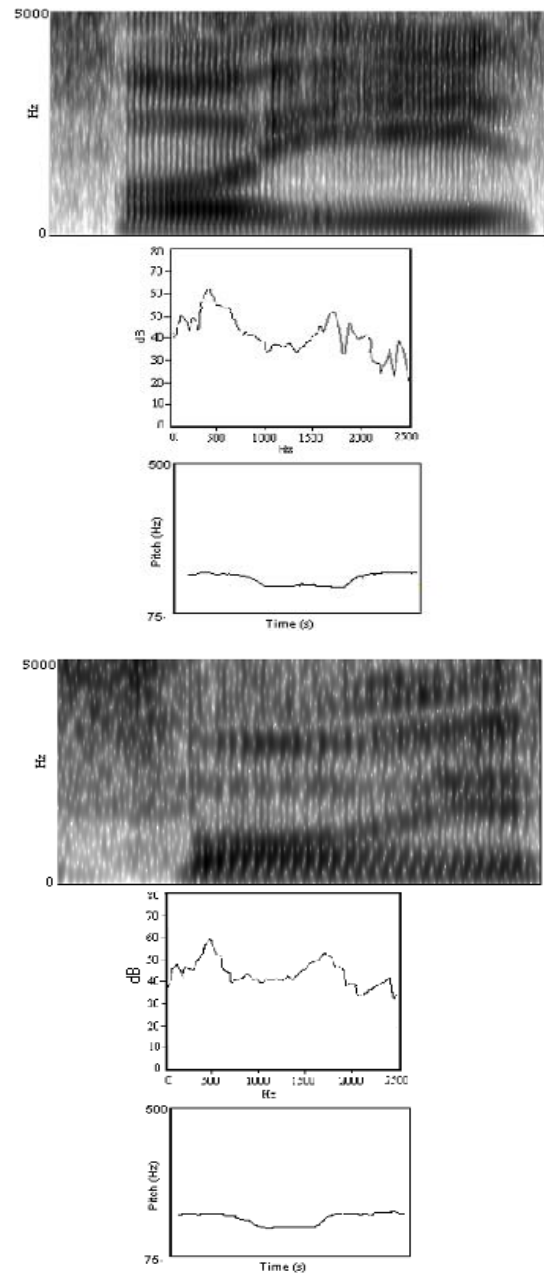


Fig. 1: *An example of a vowel laryngealization (V2) with audible creak (upper) and without audible creak (lower) in the form [saʔeb], produced by RE and FA respectively.*

The spectrogram in Fig.1 (upper) shows a creaky laryngealized vowel visually reflected in increased distance between the vertical striations reflecting pitch pulses before modal voicing commences on the latter portion of the vowel [e]. By contrast to the token in Fig. 1(upper), the spectrogram in Fig. 1(lower) shows no visibly creaky voiced portion intervening the two vowels, yet the laryngealized portion exhibits decreased overall acoustic intensity relative to the adjacent modal portions similar to 1 (upper). Although the spectrograms of the creaky and non-creaky laryngealized tokens are different, the F_0 contours and FFT spectra are similar. Both F_0 contours signal the sequencing of phonation

differences by a drop in F0, followed by a subsequent rise. In addition, both spectra display a steeply positive spectral slope due to a fall off in energy at H1; that is, the difference between the amplitude of the fundamental frequency and higher frequency harmonics in both tokens is negative.

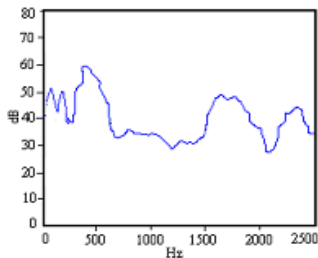


Fig. 2: FFT spectrum for the vowel [e] sampled at 45 ms to the vowel in the form [saʔeb], produced by RE.

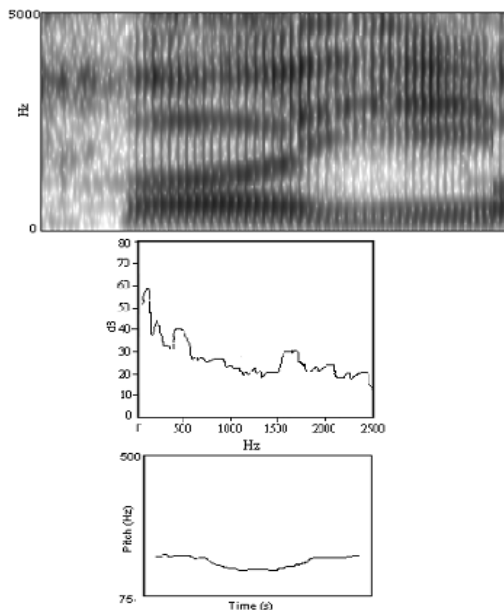


Fig. 3: An example of vowel breathiness (V2) in the form [saheb], produced by RE.

It is, also apparent from the spectrograms in Fig. 1 that the non-modal phase of laryngealization persists beyond the expected portion the glottal stop. In fact, the laryngealized phonation is not confined to a short interval at the beginning of the vowel; rather, glottal stop has extended its non-modal voice quality over half of the duration of the vowel [e]. Displayed in Fig. 2 are FFT spectrum of the vowel [e] sampled at about 45 ms to the vowel, produced by RE. Like the spectra in Fig. 1, this spectrum exhibits laryngealized values of spectral differences as reflected in the negative amplitude differences in H1-H2 and H1-F1.

Figure 3 shows the spectrograms, FFT spectra and F0 contours for the word [saheb] "owner" produced by the same speakers. The spectrogram shows a clear case of breathiness, realized at the first half portion of the vowel [e] with decreased intensity, as well as aperiodic energy at higher frequencies. The same portion is, also characterized by the lowering of the fundamental frequency in the F0 contour. Furthermore, by contrast to the spectra of the laryngealized tokens in Fig. 1, the [e] spectrum in [saheb] shows a deeply negative spectral tilt due to a considerable loss of energy at H2 and higher

frequency components, meaning that the difference between the amplitude of the fundamental frequency and higher frequency harmonics is positive in the breathy token.

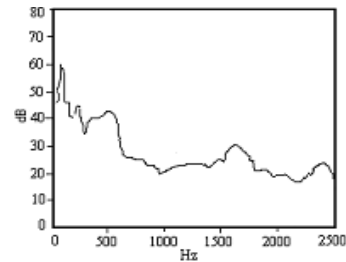


Fig. 4: FFT spectrum for the vowel [e] sampled at 45 ms to the vowel in the form [saheb], produced by RE.

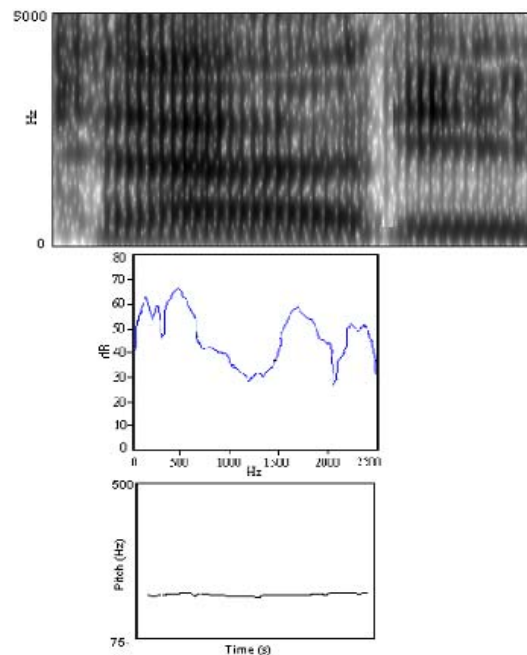


Fig. 5: An example of the modal vowel (V2) in the form [sadeG], produced by RE.

In addition, like the laryngealized tokens in Fig. 1, the non-modal phonation of the breathy vowel [e] in [saheb] is not just localized to the glottal consonant position; rather, it spreads onto the half portion of the following vowel. Fig. 4 shows FFT spectrum of the vowel [e] sampled at about 45 ms into the vowel by the same speaker. The substantial positive differences in H1-H2 and H1-F1 indicate that the glottal gesture has not yet adjusted toward a more modal setting at 45 ms.

For purposes of comparison, Fig. 5 provides an example of a non-laryngealized non-breathy vowel (V2) for the speaker RE in the minimally distinct word [sadeG] "honest". As shown in the figure, the token does not exhibit either F0 and intensity drop or steeped spectral slope found in the laryngealized and breathy counterparts in Fig. 1 and Fig. 2.

Examination of data of the kind given above suggests that that VGV sequences involve a sequencing of modal-nonmodal-modal phonation. The first portion of the sequence is characterized by modal quality corresponding to the entire length of the vowel preceding glottal consonants (V1). This portion is, then, followed by a phase of non-modal phonation of identical quality as the vowel that follows glottal

consonants (V2). The non-modal portion is characterized by laryngealization in V?V and breathiness in VhV. The sequence, then, resumes modal quality up to the end of V2. VGV sequences can, then, be characterized as involving two vowels of modal voice quality separated by an intervening period of non-modal phonation in the first half portion of the second vowel. Following Blankenship [2], I employ the term laryngealization as a general term and distinguish it from creaky phonation that may occur in the production of laryngealized vowels, but is not itself a requisite component or goal of laryngealization.

Table 2: means of overall vowel length as well as means of F0 and spectral tilt values at 15 and 45 time intervals for /V/ in VCV and VGV tokens.

		Mean		t		p	
		15	45	15	45	15	45
H1-H2	VCV	2.7 (0.6)	2.4 (0.7)	21	9.7	0.00	0.00
	V?V	1.0 (0.6)	1.3 (0.7)				
H1-F1	VCV	-4.4 (0.8)	-4.2 (1.0)	24	16.5	0.00	0.00
	V?V	-7.9 (1.0)	-7.4 (1.1)				
F0	VCV	132(17.2)	133(19.7)	3.8	4.4	0.00	0.00
	V?V	126(21.6)	132(23.9)				
L	VCV	110(16.9)		-1.0		0.28	
	V?V	113(19.4)					

		Mean		t		p	
		15	45	15	45	15	45
H1-H2	VCV	2.5 (0.8)	2.1 (0.7)	-31	-28	0.00	0.00
	VhV	6.3 (1.2)	6.2 (1.4)				
H1-F1	VCV	-3.9 (1.1)	-4.1 (1.0)	-86	-75	0.00	0.00
	VhV	9.7 (1.4)	9.8 (1.2)				
F0	VCV	129(18.8)	130(20.4)	5.7	6.1	0.00	0.00
	VhV	121(19.6)	120(17.3)				
L	VCV	112(19.6)		-1.4		0.28	
	VhV	114(22.2)					

A Fast Fourier transform (FFT) was calculated over a 25.6 ms window centered at each tag. Thus, the span of the first window commenced about 3 ms after the onset of the vowel formants. The FFT used a 512 point Hamming window. Values of F0 and spectral parameters were computed for vowels in each related CV/GV tokens: (1) H1-H2 amplitude difference at each time interval; (2) H1-F2 amplitude difference at each time interval; (3) the average F0 values calculated over each time interval, and (4) vowel duration.

4-3- Result

For each of the parameters H1-H2, H1-F1 and F0, a paired sample t-test was conducted on the mean values calculated at each time interval across the 80 ms vowel portion. Another paired-samples T-test was conducted to test the significance of difference in mean duration values for /V/ in VCV / VGV tokens.

Results indicate that the means of H1-H2, H1-F1 for the vowels in VGV are significantly different from the vowels in VCV except for the last two time intervals. This shows that while the vowels in VCV maintain a modal value of phonation measurements through most of their duration, the vowels in VGV display a non-modal laryngeal setting for the first 60 ms of their duration, but adjust into a modal modal setting after 60 ms. For VhV data, the spectral differences H1-H2 and H1-F1 increase over this time interval, leading to a steeply negative spectral slope, while for V?V, the differences decrease over the same interval, causing a steeply positive spectral slope.

Results, also show significant F0 drop for both VhV and V?V data over the first 60 ms interval. In addition, t-test shows no significant duration differences for modal versus breathy or laryngealized vowel contrast.

Table 3: Probability values from T-tests at each time interval for H1-H2, H1-F1, and F0

		15	30	45	60	75	90
H1-H2	?V vs CV	0.00	0.00	0.00	0.00	0.11	0.13
	hV vs CV	0.00	0.00	0.00	0.00	0.20	0.11
H1-F1	?V vs CV	0.00	0.00	0.00	0.00	0.08	0.35
	hV vs CV	0.00	0.00	0.00	0.00	0.18	0.30
F0	?V vs CV	0.00	0.00	0.00	0.00	0.82	0.15
	hV vs CV	0.00	0.00	0.00	0.00	0.35	0.11

5- Discussion

A significant finding for VGV data acoustic structure was substantial changes in spectral tilt and F0 values during the first half portion of the vowel, as opposed to non-VGV data. This suggests that VGV data cannot simply be viewed as involving complete deletion of glottal consonants; rather, it seems that they involve allophonic variation of modal phonation through magnitude reduction of glottal gesture. As glottal gesture is not satisfied to reach the intended target, i.e. stop or fricative, a laxer laryngeal setting is adopted and glottal consonants arise as allophonic variants of the following vowel. In addition, minimal pairs such as [sa?eb]/[saheb], which were previously assumed to have similar acoustic manifestation, are shown to exhibit acoustic contrast due to substantial differences in spectral measurements. Meanwhile spectral tilt values and F0 are phonetically significant acoustic cues that can be used to recognize words involving VGV sequences. Hence, the result of this study can be used to improve ASR.

6. References

- [1] Bijankhan, M. 2000. Farsi Vowel Compensatory lengthening: An Experimental Approach. Proc. 5th ICSLP, Beijing, China.
- [2] Blankenship, B. 2002. The timing of nonmodal phonation in vowels. *Journal of Phonetics* 30, 163-191.
- [3] Docherty, G. 1992. The timing of voicing in British English. Berlin: New York.
- [4] Fischer-jørgensen, E. 1967. Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguistics* 28, 71-139.
- [5] Gerfen, C., Baker, K. 2005. The production and perception of laryngealized vowels in Coatzospan Mixtec. *Journal of Phonetics* 33, 311-334.
- [6] Gordon, M., Ladefoged, P. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29, 383-406.
- [7] Klatt, D., Klatt, L. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87, 820-57.
- [8] Ladefoged, P. 1983. The linguistic use of different phonation types. In D. Bless & J. Abbs (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (PP. 351-360). San Diego: College Hill Press.
- [9] Sadeghi, V. 2007. Compensatory Lengthening in Persian. *The 16th International Congress on Phonetic Sciences*, Saarbrücken, Germany.
- [10] Sadeghi, V. 2008. Compensatory Lengthening in Persian: The timing of non-modal phonation. ISCA conference on Experimental Linguistics, Athens.
- [11] Shademan, Sh. 2003. Glottal-Deletion and Compensatory Lengthening in Farsi. *UCLA Working Papers in Phonetics*, No. 104, 61-81.
- [12] Stevens, K. N. 1977. Physics of laryngeal behavior and larynx modes, *Phonetica* 34, 267-79.