VOWEL ONSET CHARACTERISTICS AS A FUNCTION OF VOICE AND MANNER CONTRASTS IN PERSIAN CORONAL STOPS

Barry Heselwood and Zahra Mahmoodzade

Abstract

Acoustic and electrolaryngographic analysis was carried out on vowel onsets after coronal stops produced by seven male native speakers of Persian (Farsi) in syllable initial position. In addition to voice onset time (VOT), we present measures of pitch (Fx), closed quotient (Qx) and spectral tilt (ST).

Results show that VOT distinguishes between voiced and voiceless stops, and between plosives and affricates. Pitch at vowel onset, as measured electrolaryngographically, distinguishes voiced from voiceless plosives but not voiced from voiceless affricates; it does, however, distinguish the plosives from the affricates. Closed quotient also distinguishes voiced from voiceless plosives but not voiced from voiceless affricates; it distinguishes /d/ from /dʒ/ but not /t/ from /tʃ/. The spectral tilt measure distinguishes voiced from voiceless stops but not plosives from affricates.

Closed quotient and spectral tilt were found to be closely positively correlated but with some evidence that they are at least partially independent. This independence may enable variation between whispery voice and breathy voice at the onset of vowels after phonologically voiceless stops, and variation between whispery voice and modal voice at vowel onset after phonologically voiced stops, particularly /dʒ/.

1. Introduction

While different phonatory qualities having a contrastive phonological function have been quite extensively studied in many languages (Gordon & Ladefoged, 2001), the coarticulatory effect of the phonological voice category of a consonant on the texture of phonation in adjacent vowel off-sets and onsets has been the subject of a relatively small number of studies looking at only a few languages. Ní Chasaide & Gobl (1999) report finding clear cross-linguistic differences in the five western European languages they investigated – Swedish, German, English, French and Italian, with four informants for each language. Most relevant to this present study, they found that vowel onsets following a phonologically voiceless stop were much breathier in their German data than in their English and Swedish data, despite all three languages having aspirated realisations of voiceless stops (Ní Chasaide & Gobl, 1999: 136-9). The effect was evident up to about 30ms into the vowel. Aerodynamic influences from a preceding consonant have in fact been detected as far as 50-75ms into a following vowel in American English (Löfqvist, Koenig & McGowan, 1995: 57), and fundamental frequency differences as much as 100ms (Ohde, 1984: 224).

Ní Chasaide & Gobl's results show that phonatory quality at vowel onset after a plosive cannot be predicted language-universally from the presence of aspiration. This is consistent with Kingston & Diehl's (1994: 432) contention that pitch at vowel onset is independent of glottal aperture during the closure and release phases of a stop. This paper looks at how voice characteristics differ in vowel onsets after voiced and voiceless coronal plosives and affricates in Persian, a language in which these phenomena have not yet been studied. Voiced stops are typically realised in Persian as voiceless unaspirated ([b̥, d̥, d̥ʒ̊, g̊]), and voiceless stops as voiceless aspirated ([p^h, t^h, tʃ^h, k^h]) (Mahootian, 1997: 287-8). The picture from the literature is that, compared to phonologically voiced stops, phonologically voiceless stops cross-linguistically are realised with a longer positive VOT (Lisker & Abramson, 1964; Keating, 1984), the vowel following it begins with a higher pitch (Lea, 1973; Ohde, 1984; Kingston & Diehl, 1994), the phonatory quality is breathier with low closed quotient values (Ní Chasaide & Gobl, 1999) and the spectral tilt is more steeply negative (Scobbie, Gibbon, Hardcastle & Fletcher, 2000). The purpose of the current investigation is to see if Persian conforms to these general trends, but perhaps more importantly for phonetic theory to see to what extent these parameters might be at least partially independent of one another. For example, in addition to breathiness varying cross-linguistically after aspirated stops, lower rather than higher pitch at vowel onset has been reported after aspiration (Downing & Gick, 2001, cited in Jansen, 2004: 53), and the pitch at vowel onset after a voiceless unaspirated stop seems to depend on whether it realises a phonologically voiced or a phonologically voiceless stop (Kingston & Diehl, 1994: 435). Furthermore, relatively low closed quotients have been found alongside positive spectral tilt (Heselwood, 2007: 20-21).

It should be made clear at the outset, however, that it is not our purpose to make any specific claims about the relative perceptual importance of the parameters investigated but to try to establish correspondences between parameter values and the distinctive phonological categories of voiced, voiceless, plosive and affricate which, together with the coronal place feature, define the coronal stops /t, d, tʃ, dʒ/ in Persian. Our study could of course be followed up in the future with an investigation of their possible perceptual roles as cues to consonant recognition. The vowel onset in a stop+vowel sequence is the first high amplitude event after the much lower amplitudes of the closure and release phases of the stop. This fact may give it an importance in the perception of the sequence such that any cues located there, and relating to the type of stop preceding it, might be expected to contribute to a particular stop percept. Amerman & Parnell (1984) found that when listeners had to rely on the vowel onset to identify a preceding stop they only performed at chance levels. But they did not report on listeners' ability, using only vowel onset information, to distinguish between voiced and voiceless stops, or between plosives and affricates.

1.1 Plosive-affricate contrasts

Plosives and affricates can be classed as oral stops because they both involve a complete maintained closure between an active and a passive articulator accompanied also by a velic closure. Intra-oral air pressure builds up during the closure and is released when the active articulator is moved away. Differences in the relative durations of the closure and release phases are an important factor in distinguishing plosives from affricates. Plosives typically have a longer closure and shorter release; affricates reverse this relationship and show a shorter closure with a longer release (Kochetov & Lobanova, 2007: 54). The release phase of an affricate is typically longer than that of a long-lag plosive but there is also a difference in the source of the fricative noise. In long-lag plosives, the fricative noise is aspiration which is generated in the glottis, possibly enhanced also by cavity friction as the air passes through the oral chamber, whereas in affricates it is created by the same articulators that formed the closure. However, the division between affricate and plosive tokens is not always as sharp as the foregoing description might imply. Voiceless affricates sometimes have a short period of aspiration between the decline of the homorganic friction and the onset of voicing (Jansen, 2004: 59), indicating that the vocal folds have not come together until after the articulators have moved far enough apart not to induce local turbulence. Conversely, plosives often exhibit some affrication at the start of the release phase (Abercrombie, 1967: 147-8) which may then be followed by aspiration.

Against this background of differences in the closure and release phases of plosives and affricates, vowel onset parameters are investigated to see whether they vary systematically and significantly across the plosive-affricate category distinction following the release of a stop.

1.2 Voicing contrasts

Voicing contrasts in stops are common in the languages of the world (Ladefoged & Maddieson, 1996: 48). Persian has in common with many languages, including English, a two-way voice contrast, normally classified phonologically as voiceless and voiced. Languages with two-way contrasts vary between basing the contrast on a prevoiced-short lag voice onset time (VOT) distinction, as in French and other Romance languages, or on a short lag-long lag VOT distinction as in English and other Germanic languages, and also Persian (Mahootian, 1997: 287-8). The phonological category ‘voiced’ can therefore mean, in phonetic terms, either prevoiced or short lag, and the category ‘voiceless’ can mean either short or long lag (Keating, 1984). Where there is long lag, the lag time is filled with aspiration noise but the relationship between aspiration and long lag VOT is not altogether clear. Whether aspiration is there merely as a consequence of the glottis remaining open to effect a long lag, or whether the glottis remains open to enable aspiration to occur, is not an easy question to answer. That is to say, should we regard the long lag as the ‘target’ phenomenon with aspiration as a side effect that merely fills in the time, or should we regard aspiration as the ‘target’ which is facilitated by a long VOT lag? Results of perception tests reported in Repp (1979) suggest that presence of aspiration, and its relative amplitude, does in fact contribute to voicing perception in stops but the problem for experimental investigation is that one cannot manipulate the VOT value while maintaining the aspiration duration value, at least not without radically altering the spectral character of the aspiration. Whichever of the two phenomena is primary, it has nonetheless been well established that VOT is a robust indicator of phonological voicing categories in natural speech (Lisker & Abramson, 1964; Hayward, 2000: 108). It is not, however, the only one.

Other phonetic parameters that have been shown to vary with voicing contrasts in plosives include closure phase duration (Fry, 1979: 122), F1 cutback (Gimson, 1980: 156; Kent & Read, 1992: 120-21), pitch (Haggard, Ambler & Callow, 1970; Lea, 1973; Ohde, 1984; Kingston & Diehl, 1994), spectral tilt (Scobbie, Gibbon, Hardcastle & Fletcher, 2000: 198), and amplitude of noise burst (Gimson, 1980: 156). This study, after establishing VOT as an indicator of voicing category, looks at vocal fold vibration frequency (Fx) at vowel onset, the closed quotient (Qx) of the glottal cycle at vowel onset, and spectral tilt (ST) at vowel onset. Qx has in fact been identified as correlating with spectral tilt (Gordon & Ladefoged, 2001: 399; Hanson, Stevens, Kuo, Chen & Slifka, 2001: 459). The purpose of this study is to investigate whether these parameters vary systematically and significantly with the voiced-voiceless category distinction at the onset of a vowel following the release of a stop.

1.3 Persian coronal stops

Among the 23 consonants of Standard Persian (Mahootian, 1997: 286-7) are ten stops, structured phonematically as shown in Table 1.

Table 1. Persian stop consonants

Labial	Coronal		Dorsal		Guttural
	Dental	Post-alveolar	Antero-dorsal	Postero-Dorsal	Glottal
p b	t d	tʃ dʒ	k g	ɢ	ʔ

As can be seen, the correlation of voice applies to all except the postero-dorsal and glottal stops. The voiceless member of each pair is realised initially with aspiration and a long-lag VOT, the voiced member with either short-lag VOT or pre-voicing (Mahootian, 1997: 287-8).

The stops under investigation in this study are /t, d, tʃ, dʒ/, i.e. the set of coronal stops. According to Samareh (2006: 40, 64), the same active articulator, the tip and blade of the tongue, is used for the dentals and for the post-alveolars. Mahootian (1997: 287) considers that /t, d/ vary between apico-dental and apico-alveolar although Majidi & Ternes (1999: 124) and Samareh (1992: 40) only specify dental. The affricates /tʃ, dʒ/ are postalveolar (Mahootian, 1997: 287; Majidi & Ternes, 1999: 124).

2. Methodology

2.1 Informants

Seven male native speakers of Standard Persian (Farsi Me'yar), the variety spoken in Tehran and by most educated people in Iran, formed the experimental sample. All were born in Iran and currently live in Leeds, UK. Some relevant biographical information is given for each informant in Table 2.

Table 2. The subjects in this study.

Informant	Age in years	Languages spoken	Level of education
HA	36	Persian, Kurdish (bilingual) English (L2)	PhD student
JK	24	Persian, English (bilingual)	BSc
MK (brother of JK)	27	Persian, English (bilingual)	MSc
MN	40	Persian, Gilaki (bilingual) English (L2)	PhD student
PP	35	Persian, English (L2)	PhD student
BS	27	Persian, English (L2)	MSc
YP	34	Persian, English (L2)	PhD student

2.2 Elicited data

The data are a subset of a larger set of utterances collected for the second author's PhD thesis. The four target words for the current study were therefore embedded in a longer list of words meaning that the speakers' attention was not drawn to them as an explicit minimal quadruplet. Each speaker pronounced the words on the list four times yielding 16 tokens of the target words per speaker, 28 tokens of each word and 112 tokens altogether. The words are listed in (1):

(1) /tang/ تنگ *tang* ‘narrow’
/dang/ دنگ *dang (onomat.)* ‘clang’
/tʃang/ چنگ *chang* ‘claw’
/dʒang/ جنگ *jang* ‘war’

The words were controlled for context by having the same syllabic rhyme [-aŋg] and being produced as isolated citation forms separated by pause. Each word therefore formed a separate intonation group, and each was pronounced with a falling tone.

The productions were recorded onto a computer using an electret headset microphone to capture the acoustic signal, and gold-plated electrodes to capture the laryngographic signal. The recording took place in the Language Research Laboratory in the Department of Linguistics and Phonetics at the University of Leeds. The signals were digitised using Speech Studio software at a sampling rate of 16kHz for obtaining the VOT, Fx and Qx values; Speech Station 2 software was used for the spectral tilt values.

2.3 Measures

2.3.1 Voice onset time (VOT)

Voice onset time is one of the most used measures in speech acoustics because it has been found to distinguish reliably between categories of voicing in many different languages (Lisker & Abramson, 1964). It identifies the timing of the onset of vocal fold vibration in relation to the release burst of a stop consonant.

To obtain a VOT value, the stop burst was identified on the acoustic waveform with reference to an accompanying spectrogram, and the onset of voicing was determined from the synchronised larynx (Lx) waveform by locating the first peak. The portion of waveform defined by these two events was highlighted manually and the duration value taken. By using the Lx waveform to locate voice onset, the risk of mistaking multiple bursts, or other non-laryngeal events, for voicing on the acoustic waveform was avoided.

2.3.2 Rate of vocal fold vibration (Fx)

Fx was chosen as a measure of pitch because it is calculated directly from the Lx waveform and is less prone to error than acoustic measures using autocorrelation where formant transition frequency shifts can interfere (Kent & Read, 1992: 82) and pitch-halving and pitch-doubling may occur (Johnson, 1997: 35-6).

Fx at the onset of the following vowel was determined from the periods of the first four full glottal cycles after stop release. Measurement was done automatically by the Speech Studio programme’s multi-dimensional voice profile (MDVP) which gives the mean from the four cycles, and the minimum and maximum. Averaging over four cycles instead of merely measuring the first cycle ensures there is potentially a meaningful relationship to pitch perception for which three or four consecutive cycles are required (Howard & Angus, 2001: 135-6). Pitch at vowel onset is thought to be determined by degree of vocal fold tension, with lax folds resulting in lower pitch and tense folds resulting in higher pitch (Halle & Stevens, 1971; Ohde, 1984).

2.3.3 Closed quotient (Qx)

The closed quotient (Qx) for the first four full glottal cycles after stop release was calculated by the Speech Studio programme as the upper 70% of the peak amplitude, encompassing all those stages of the cycle where there is some vocal fold contact. There is a close relationship between closed quotient value and phonatory quality, with lower values arising when phonation is more breathy due to less vocal fold contact (Laver, 1980: 134). Qx is expressed as a percentage. A value of 50% indicates that the vocal folds are in contact for half the time period of the cycle.

2.3.4 Spectral tilt (ST)

By also looking at spectral tilt, we can see how differences in phonatory quality shape the spectrum in the region of the lower harmonics where the ear's frequency resolution is most powerful (Hayward, 2000: 140-41). Spectral tilt was calculated from FFT spectra taken at the onset of the vowel by placing the cursor between the first two voicing pulses. Spectral tilt for the purposes of this paper is expressed as the difference between the amplitudes of the first and second harmonics (Hayward, 2000; Gordon & Ladefoged, 2001). Because the same open vowel was used in all elicited words, the problem of harmonic amplitudes being differently influenced by different F1 frequencies (Hanson et al., 2001: 459) was avoided; Esposito (2006: 11) states that measurement of spectral differences in the lower harmonics is most accurate for open vowels. Negative tilt (first harmonic having a higher amplitude than the second harmonic) is a feature of breathy phonation (Ní Chasaide & Gobl, 1997: 442-3) and is a direct consequence of the cartilaginous part of the glottis being open throughout voicing (Hanson et al., 2001: 453). Positive tilt indicates a lack of breathiness.

2.3.5 Statistical tests

Because there was a normal distribution of values, paired sample two-tailed t-tests were run on the results from all the above measures. The variables for the tests are the mean values of the phonemes being compared taken from the same speaker.

3. Results

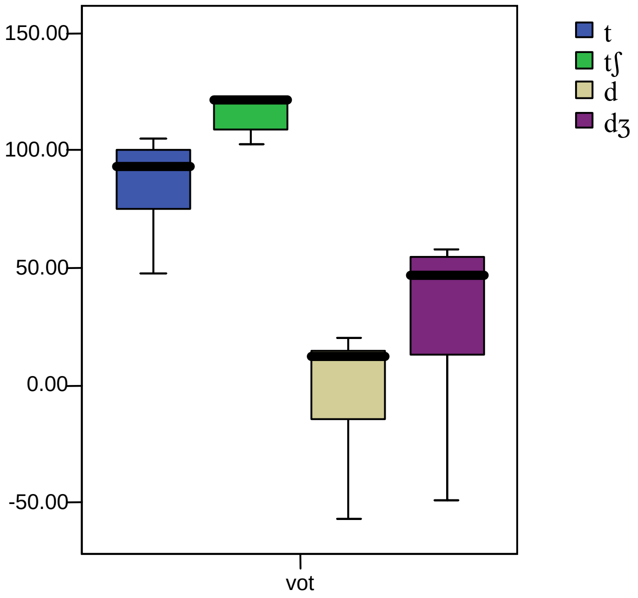
3.1 VOT

Table 3 presents the mean and standard deviation for each speaker's VOT values for /t, tʃ, d, dʒ/, and also the mean and standard deviation for the group. Figure 1 presents the data in box-plot form. A paired-sample two-tailed *t*-test shows that the VOT difference between /t/ and /d/ is significant (t_6=12.7, p<0.001), and also between /tʃ/ and /dʒ/ (t_6=-8.5, p<0.001). VOT is therefore shown to correspond to the voicing contrast in plosives and in affricates. The same statistical test shows that the VOT differences between /t/ and /tʃ/, and also between /d/ and /dʒ/, are also significant (t_6=8.0, p<0.001, and t_6=6.3, p<0.005 respectively). VOT is therefore also shown to correspond to the plosive-affricate manner contrast.

Table 3. Mean VOT values in ms for 4 tokens per phoneme per speaker.

	t		tʃ		d		dʒ	
	mean	SD	mean	SD	mean	SD	mean	SD
HA	105.10	22.33	141.83	17.28	12.05	2.79	55.38	5.41
JK	101.28	11.58	121.75	13.56	19.88	2.63	54.58	10.82
MK	98.68	6.88	121.90	9.00	11.98	0.85	57.98	5.76
MN	75.58	11.51	102.85	11.45	-41.28	85.56	-11.73	31.75
PP	47.70	7.65	89.00	4.07	-57.18	43.06	-48.83	65.9
BS	75.35	8.39	121.70	9.97	15.75	2.56	37.33	19.90
YP	93.30	10.25	115.38	14.05	13.70	3.80	46.88	8.61
group mean	**85.28**	**20.40**	**116.34**	**16.68**	**-3.59 (14.67)**	**31.63**	**27.37 (50.43)**	**41.39**

Figure 1. VOT results in box-plot form. Black lines indicate median values. Scale in milliseconds.



Notably, only two speakers (MN and PP) have negative VOT values for phonologically voiced stops. MN prevoiced both /d/ and /dʒ/ three times out of four, while PP voiced both of them twice out of four. This pattern of a minority of speakers prevoicing their voiced stops, but showing some consistency in doing so, is also reported for English by Docherty (1992). The group means for realisations of these phonemes have been given twice: firstly with prevoiced tokens included, and secondly (in brackets) with them excluded. When they are included it can be seen that they have a big effect on the standard deviation.

3.2 Fx

Table 4 presents three Fx values per speaker: mean Fx across the first four glottal cycles, and also the minimum and maximum values from those four cycles (shortest and longest cycle). Figure 2 presents these data in box-plot form.

Table 4. Fx values in Hz for 4 tokens per phoneme per speaker

	t			tʃ			d			dʒ		
	min	mean	max	min	mean	max	min	mean	max	min	mean	max
HA	122.4	126.2	134.0	121.0	126.8	134.1	118.1	123.5	135.0	123.6	130.2	142.2
JK	135.4	142.5	159.3	140.9	151.8	182.3	131.1	140.9	165.5	135.5	143.6	159.5
MK	107.3	120.8	152.5	112.3	122.2	143.5	102.7	110.7	130.2	112.3	122.7	141.5
MN	152.3	161.9	186.6	154.8	161.7	186.6	143.7	153.0	167.3	148.1	151.5	157.0
PP	135.0	141.1	153.6	137.3	146.5	163.6	131.3	136.9	143.6	131.4	132.2	137.1
BS	126.0	130.6	139.7	131.5	136.3	145.1	122.0	129.1	144.8	130.6	140.7	166.0
YP	137.4	143.4	151.3	143.0	149.4	156.8	136.4	142.4	156.3	149.3	151.6	155.2
group mean	**130.8**	**138.1**	**153.9**	**134.4**	**142.1**	**157.6**	**126.5**	**133.8**	**148.9**	**133.0**	**139.5**	**151.2**

A paired-sample two-tailed *t*-test shows that, in the first four glottal cycles following stop release, the minimum and mean Fx differences between /t/ and /d/, and also between /t/ and /tʃ/, and /d/ and /dʒ/, are significant. The minimum Fx values show a highly significant difference between /t/ and /d/ with $p<0.005$, and a significant difference between the plosives and the affricates with $p<0.05$. None of the Fx measures revealed significant differences between /tʃ/ and /dʒ/. The statistical results for the three measures, minimum Fx, mean Fx and maximum Fx, are given in Table 5 for the /t/-/d/, /t/-/tʃ/ and /d/-/dʒ/ oppositions. For the Fx maximum values, differences do not reach significance for any pair of these phonemes.

Figure 2. Fx results in box-plot form. Black lines indicate median values. Scale in Hz.

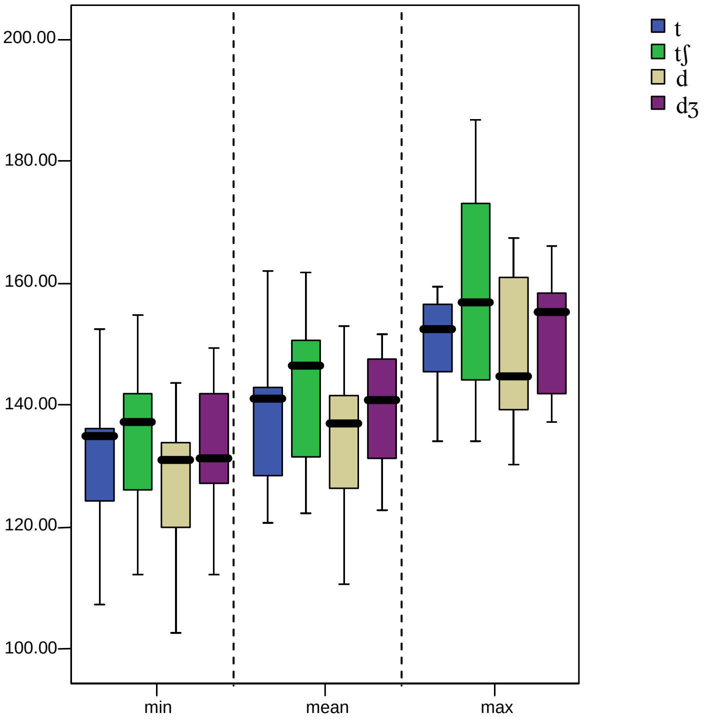


Table 5. t-test results for the Fx measures.

/t/-/d/			/t/-/tʃ/			/d/-/dʒ/		
Min Fx	Mean Fx	Max Fx	Min Fx	Mean Fx	Max Fx	Min Fx	Mean Fx	Max Fx
$t_6=5.2$ $p<0.005$	$t_6=3.0$ $p<0.05$	$t_6=1.1$ $p>0.05$	$t_6=3.6$ $p<0.05$	$t_6=3.0$ $p<0.05$	$t_6=0.9$ $p>0.05$	$t_6=4.1$ $p<0.01$	$t_6=2.0$ $p>0.05$	$t_6=0.5$ $p>0.5$

3.3. Qx

Table 6 presents three Qx values per speaker: mean Qx across the first four glottal cycles, and also the minimum and maximum values from those four cycles (shortest and longest closed phases respectively). Figure 3 presents the data in box-plot form. A paired-sample two-tailed *t*-test shows that the minimum, mean and maximum Qx differences between /t/ and /d/, and between /d/ and /dʒ/, are significant. Only the minimum Qx difference reaches significance for distinguishing /tʃ/ and /dʒ/ with ($p=0.035$) while none reach significance for /t/ and /tʃ/. Table 7 gives the statistical results where significance has been found.

Table 6. Qx values in % for 4 tokens per phoneme per speaker

	t			tʃ			d			dʒ		
	min	mean	max	min	mean	max	min	mean	max	min	mean	max
HA	21.5	31.1	36.7	27.2	32.6	35.8	40.9	44.8	46.6	29.3	37.6	41.4
JK	28.0	38.5	44.4	24.8	34.1	40.1	48.2	43.6	47.1	32.1	39.0	42.9
MK	28.0	43.4	51.0	33.2	44.1	49.8	44.7	51.5	55.3	40.8	49.1	52.5
MN	27.8	41.9	49.0	29.9	40.1	45.7	40.4	45.9	50.0	33.6	37.5	41.1
PP	30.4	39.2	46.3	30.9	41.6	48.7	42.5	47.7	51.3	40.5	43.3	45.8
BS	35.4	41.2	44.6	32.6	40.7	44.8	39.6	48.1	51.9	28.9	40.2	45.2
YP	31.8	37.3	42.2	32.9	38.4	42.3	48.7	55.2	58.7	46.5	50.6	53.0
group mean	**29.0**	**38.9**	**44.9**	**30.2**	**38.8**	**44.0**	**43.6**	**48.1**	**51.5**	**36.0**	**42.5**	**46.0**

Figure 3. Qx results in box-plot form. Black bars show median values. Scale in %.

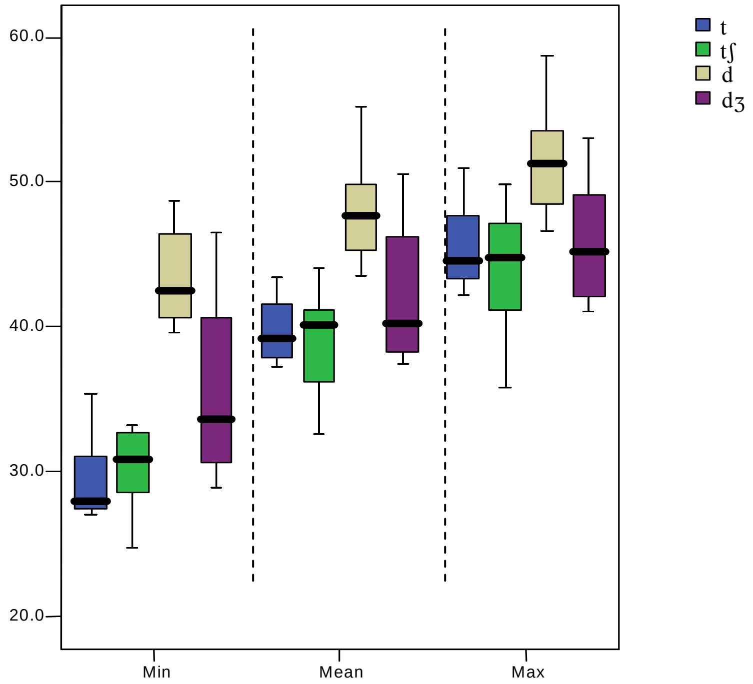


Table 7. t-test results for the Qx measures reaching significance

/t/-/d/			/d/-/dʒ/			/tʃ/-/dʒ/
Min Qx	Mean Qx	Max Qx	Min Qx	Mean Qx	Max Qx	Min Qx
$t_6 = -7.0$ $p<0.005$	$t_6 = -5.0$ $p<0.005$	$t_6 = -3.4$ $p<0.05$	$t_6 = -3.0$ $p<0.05$	$t_6 = -4.6$ $p<0.005$	$t_6 = -4.2$ $p=0.005$	$t_6 = -2.7$ $p<0.05$

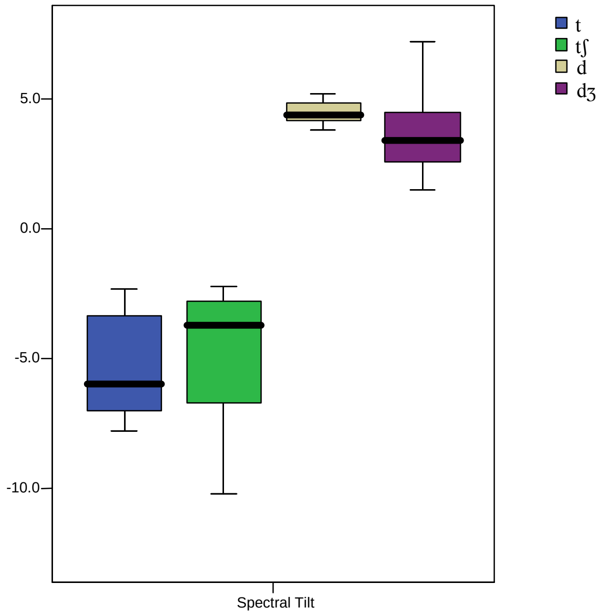
3.4 Spectral tilt

Table 8 presents mean spectral tilt values for each speaker as well as group means. The data are given in Figure 4 in box-plot form. A paired-sample two-tailed *t*-test shows that, at the onset of voicing after release of the stop closure, spectral tilt is significantly different between the voiced and voiceless plosives, and also between the voiced and voiceless affricates ($t_6 = -13.4$, $p<0.001$, and $t_6 = 5.7$, $p=0.001$ respectively). Differences between /t/ and /tʃ/, and between /d/ and /dʒ/, are not shown to be significant. Voiceless stops are characterised by negative tilt and voiced stops by positive tilt.

Table 8. Mean spectral tilt values in dB for 4 tokens per phoneme per speaker

	/t/		/tʃ/		/d/		/dʒ/	
	mean	SD	mean	SD	mean	SD	mean	SD
HA	-3.7	1.8	-2.2	4.9	+4.8	1.2	+1.5	3.2
JK	-6.6	0.9	-6.7	3.0	+4.1	0.9	+4.1	0.7
MK	-2.3	5.3	-3.7	3.0	+4.4	2.0	+7.2	1.8
MN	-6.0	1.4	-2.9	0.9	+4.9	2.3	+2.4	0.4
PP	-7.8	1.3	-6.7	2.3	+3.8	2.2	+3.4	0.5
BS	-3.0	1.2	-2.7	4.5	+5.2	1.2	+2.8	2.1
YP	-7.4	3.4	-10.2	2.7	+4.2	1.6	+4.9	4.0
group mean	**-5.3**	**2.2**	**-5.0**	**3.0**	**+4.5**	**0.5**	**+3.7**	**1.9**

Figure 4. Spectral tilt in relation to voicing and manner. Black bars show median values. Scale in dB.



It can be seen in Table 8 and Figure 4 that /d/ has the lowest standard deviation and in fact had no negative values in any of the 28 tokens measured. There was a single positive token for /t/ (speaker MK), two for /tʃ/ (speakers HA and BS), and three negative tokens for /dʒ/ (speakers HA, BS, YP).

4. Discussion of results

Our results show that when the vowel after a coronal stop begins in Persian, there are acoustic and physiological differences present that correspond to the type of stop, i.e. voiceless versus voiced, and plosive versus affricate. The results for each measure are discussed and then summarised for each correlated pair of stops.

4.1 VOT

The results for VOT are as expected. Their significance for this paper is that they establish that conditions in the glottis are different immediately prior to voice onset for realisations of voiced and voiceless stops with a wider glottal aperture in long-lag tokens compared to short lag tokens (Catford, 1977: 111-16), and that short-lag tokens are realisations of phonologically voiced stops.

Phonologically voiced /d, dʒ/ have either negative or smaller positive VOT values compared to larger positive values for phonologically voiceless /t, tʃ/. This is consistent with a larger glottal opening in realisations of /t, tʃ/ during the stop release meaning that the vocal folds must travel further to meet in the midline so that modal voice can take place. Löfqvist, Koenig & McGowan (1995: 62) explain that vibrations may in fact begin before reaching the midline, in which case the quality of phonation will be different from that found after voiced stops when the vocal folds do not have so far to travel.

However, we do need to give a careful interpretation to these results for VOT in relation to affricates. While we can take the results for the plosives as being important for the voicing contrast, we should take note of Jansen's (2004, 58-61) view that VOT is less important for the voicing contrast in affricates. The high VOT values for /tʃ/ are doubtless a consequence of the duration of the voiceless homorganic friction in the release phase, and the lower values for /dʒ/ (but still in the range of possible plosive long lag values) can be interpreted as a consequence of devoiced affricates having shorter fricative release phases than voiceless affricates (Mahmoodzade & Bijankhan, 2007). It also needs to be noted that higher VOT values are to be expected in postalveolars compared to dentals/alveolars because of the longer section of vocal tract in front of the closure. A larger volume of air has to be shunted out of the mouth chamber before the air behind the closure can escape (Cho & Ladefoged, 1999: 213-14). Nevertheless, the magnitude of the differences in VOT between the dental plosives and the post-alveolar affricates cannot be accounted for simply by this difference.

The presence of both prevoiced and short lag tokens of /d/ and /dʒ/ suggests that, in Persian, voiced coronal stops are variable with respect to phonetic voicing in initial position, but in our data devoiced allophones predominate. In the case of /d/, devoicing means that vocal fold vibration is absent during the closure phase and begins between about 10-20ms after closure release. Where /dʒ/ is concerned, vocal fold vibration begins between about 35-60ms after closure release, meaning that not only is the closure phase devoiced, but also the fricative phase.

Cho & Ladefoged (1999) have shown that within the three contrastive VOT categories languages occupy different parts of the continuum implying that speakers have to acquire knowledge of, for example, the range of long-lag values typical of their language. Our results indicate that long-lag values in Persian are generally at the higher end. Six out of the seven speakers have means of 75ms or more for /t/. Speaker PP exceptionally has a mean of 47ms which can only be noted here as an idiosyncracy since he is judged to speak with a standard Tehran accent.

4.2 Fx

In the plosive pair, vocal fold vibration at vowel onset following /d/ is slower than at vowel onset following /t/ as would be expected from previous research findings that pitch is typically lower after voiced plosives compared to voiceless plosives (Ohde, 1984; Kingston & Diehl, 1994). The pitch-depressing effect of voiced plosives is therefore evident in the Persian data collected for this study. It is found when they are realised with prevoicing and also when realised with a short-lag VOT. In the plosive-affricate pairs, voicing following the affricates /tʃ, dʒ/ has a higher Fx value than when following the plosives /t, d/, a result that appears not to have been reported before. In seeking an explanation for this, it must be remembered that voiced plosives and affricates in Persian are generally devoiced in realisation. The glottis is therefore somewhat open in the release phase, but glottal aperture may need to be wider for affricates compared to their plosive correlates in order to supply sufficient air to generate turbulence at the point of constriction further forward in the vocal tract, and the opening needs to be held for longer because of the longer fricative release phase. To prevent voicing commencing too early, and thereby interfering with the glottal flow required for the fricative release, the vocal folds may be tenser for realisations of /dʒ/ than for realisations of /d/, and tenser folds cause a higher rate of vibration once voicing does begin (Halle & Stevens, 1971). This could explain the lack of a significant difference in minimum Fx at vowel onset after the affricates /tʃ/ and /dʒ/, in stark contrast to the highly significant difference after the plosives /t/ and /d/.

4.3 Qx

The Qx results for the plosives strongly indicate that the glottal tension settings are different after /t/ compared to after /d/, with less adductive tension and medial compression at the onset of voicing after /t/, causing reduced vocal fold contact during the course of the glottal cycle, and a breathy phonatory quality. These results are consistent with the explanation given by Löfqvist, Koenig & McGowan (1995: 62):

> At the release of the oral closure, the glottis is open and in the process of being adducted for the following vowel. The glottal vibrations begin while the glottal area is decreasing, and pulses at vowel onset have large values of peak flow, minimum flow, and the open quotient [the converse of Qx]. During the first part of the vowel, the values of these parameters decrease.

The largest difference in Qx is evident when comparing the minimum value from the four initial cycles. That the minimum Qx value is from the very first cycle can be seen from the Lx waveforms in Figures 5 and 7 and the Qx trace in Figure 6. The smallest difference is found in the maximum Qx values, the last of the four cycles. That is to say, as one gets further into the vowel after /t/, phonation becomes progressively less breathy. Figure 6 shows synchronised Fx and Qx traces for /taŋg/ and /daŋg/ in which it can be seen that as the pitch falls after /t/, the closed quotient rises (left hand dotted ellipsis). After /d/, we find the opposite tendency (right hand dotted ellipsis). This pattern agrees with previous observations in other languages (Lea, 1973; Kingston & Diehl, 1994).

All three Qx measures – minimum, mean and maximum – show significant differences between the voiced plosive /d/ and the voiced affricate /dʒ/, with the affricate having the lower values indicative of breathier phonation. However, no such significant differences were found to distinguish the voiceless plosive and voiceless affricate. A possible explanation for this finding may be sought, as with the Fx results, in the devoicing of the voiced affricate. If the glottal opening during the devoiced fricative phase of /dʒ/ is wider than at the release of the plosive /d/ then we might expect a breathier onset to the following vowel's phonation. A wider glottal opening facilitates greater volume-velocity airflow into the buccal chamber, supplying more air for conversion into a fricative at the place of articulation of the affricate. The Qx measures tend to group /t, tʃ, dʒ/ together with lower values than /d/, as shown above in Figure 3. In acoustic terms, the division corresponds with VOT and the presence of aperiodic noise during the VOT interval, whether in the form of aspiration (/t/), affrication (/dʒ/), or both affrication and aspiration (/tʃ/). The absence of aspiration in realisations of /dʒ/ may be responsible for its slightly higher Qx values. Where there is neither affrication nor aspiration, as with /d/, Qx values are yet higher.

Figure 5. Length of open phase (arrowed) decreases through the first four glottal cycles of a realisation of /t/. Example from speaker HA.

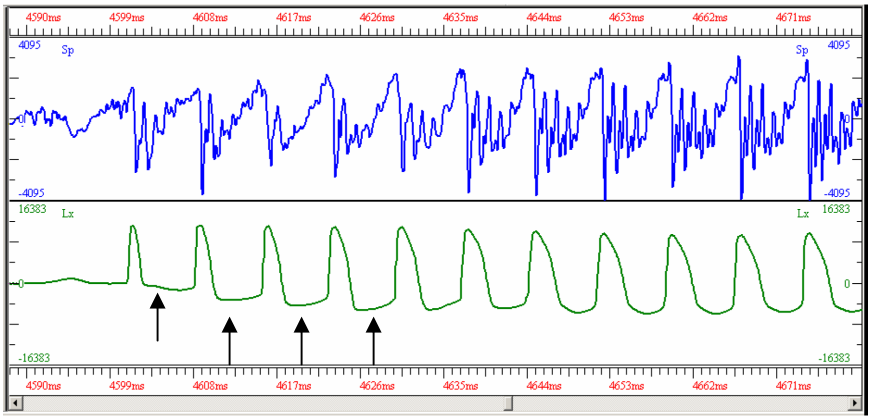
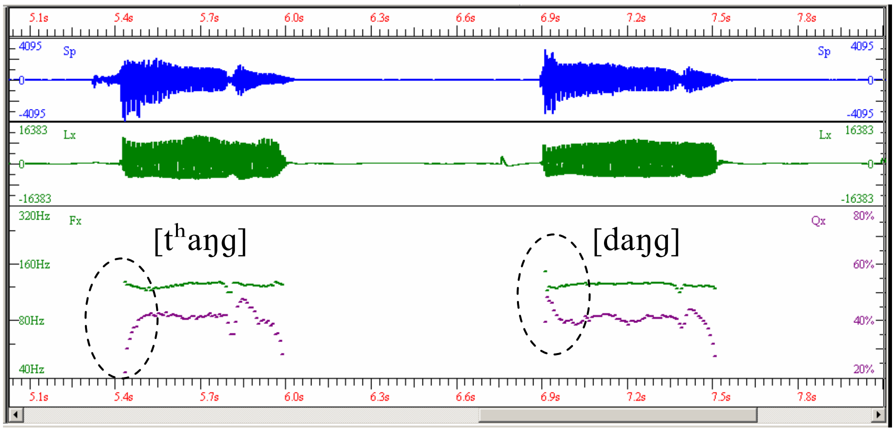


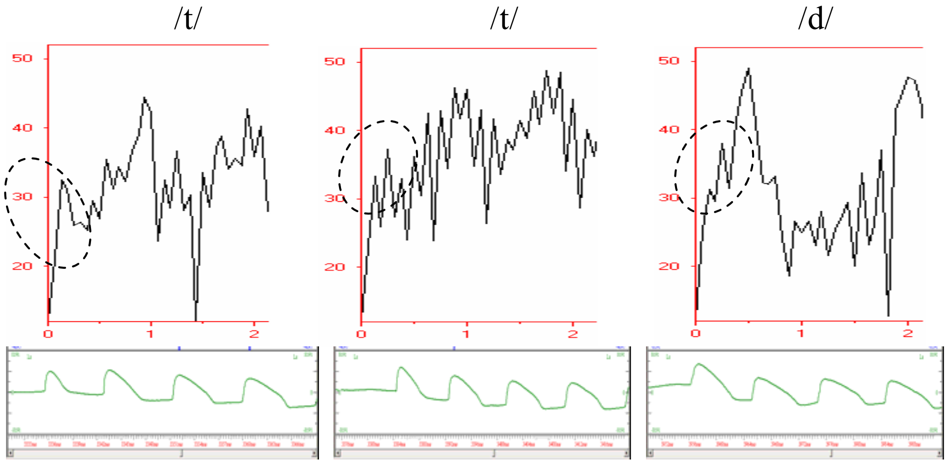
Figure 6. Synchronised Fx (green trace) and Qx (purple trace) for /t/ and /d/. Example from speaker HA.



4.4 Spectral tilt

In talking of the relationship between spectral tilt and glottal cycles, Gordon & Ladefoged (2001: 399) state that ‘the less the amplitude of the second harmonic relative to that of the fundamental, the greater is the open quotient’ and therefore also the lower is the Qx value. We should therefore expect a high degree of correlation between lower Qx values and negative spectral tilt on the one hand, and higher Qx values and positive tilt on the other. Comparison of Tables 6 and 8 do show general correspondences of this kind but we can see, for example, that the speaker with the lowest Qx values for /t/ (speaker HA) does not have the steepest negative spectral tilt – in fact it is speaker PP who exhibits the steepest negative tilt. Similarly, speaker YP has the highest Qx value for /d/ but speaker BS shows the steepest positive tilt. These mismatches may be largely an artefact of the statistical use of mean values. When we look at individual tokens we can observe a more consistent pattern of co-variation of Qx and spectral tilt indicating that both are plausibly effects of the same cause, i.e. a breathy phonatory quality. A bivariate correlation test shows a significant positive correlation (Pearson correlation coefficient = 0.5, $p<0.01$). However, the picture given of phonatory quality by the two measures, Qx and spectral tilt, are not exactly the same. While the Qx measure distinguishes significantly between /d/ and /dʒ/, spectral tilt does not ($p=0.391$). This can only be explained if the two parameters are at least partially independent. In physiological terms, the parameters are likely to be vocal fold tension and arytenoidal abduction. Whispery voice, for example, has a similar arytenoidal gap to breathy voice but greater vocal fold tension (Laver, 1980: 134). According to Esling & Harris (2005: 367), there is also aryepiglottic constriction. Tokens of post-stop vowel onsets exhibiting positive tilt but lower Qx values might therefore be produced with whispery voice. It is possible that this phonatory vowel onset quality is responsible for the small minority of realisations of voiceless /t/ and /tʃ/ that have positive tilt. Figure 7 shows spectral tilt at vowel onset and Lx waveforms of the first four glottal cycles for realisations of /t/ (left and centre) and /d/ (right) from speaker MK. The /t/ in the centre has positive tilt at vowel onset but not as steep as for the /d/ on the right, and a relatively low Qx value but not as low as the breathy voiced /t/ on the left. These characteristics suggest whispery voice. The voiced affricate /dʒ/ typically has a shallower tilt and lower Qx than the voiced plosive /d/ which may also indicate whispery voice in contrast to modal voice. Our results are consistent with there being two continua here – a continuum of vocal fold tension, and a continuum of arytenoidal abduction, but further research needs to be done into the precise relationship between them and the extent to which, and

Figure 7. Spectral tilt (ST) and Qx in breathy voice (left), whispery voice (centre), modal voice (right). Dotted ellipsis identifies first two harmonics. Speaker MK.



ST = -8.5dB
Min Qx = 21.78%

ST = +3.9dB
Min Qx = 35.04%

ST = +6.7dB
Min Qx = 46.03%

conditions under which, they might vary independently. Some evidence for their independence is provided in Heselwood (2007: 18-21) where the Arabic pharyngeal *'ayn* was found sometimes to have lower Qx values occurring with a positive spectral tilt.

The distribution of negative values for voiced /d/ and /dʒ/, and of positive values for voiceless /t/ and /tʃ/, and the low standard deviation of spectral tilt values in the case of /d/, suggest a possible hierarchy of spectral tilt consistency of /d/>/t/>/tʃ/>/dʒ/ although the sample is too small to be confident of it.

4.5 /t/ - /d/

The VOT measure, unsurprisingly, was found to distinguish robustly between voiced and voiceless plosives thus confirming its importance in Persian. When voicing starts at the end of the VOT interval, it is consistently and significantly lower in fundamental frequency for /d/ than for /t/ by around 4-5Hz, about half a semitone, which is well above the twelfth of a semitone difference limen for pitch (Howard & Angus, 2001: 125) but it may or may not be effective as a perceptual cue. Revoile, Pickett, Holden-Pitt & Talkin (1987), for example, found Fo at vowel onset in American English had no influence on perception of stops as voiced or voiceless. In addition to glottal cycles being more rapid after /t/, the ratio of the open and closed phases of the cycles is different. Cycles at vowel onset after /t/ have a relatively shorter closed phase and longer open phase which is reflected in spectral tilt. The acoustic energy created by these glottal vibrations has a fairly steep negative tilt after /t/ but a positive tilt after /d/. These differences can be summed up briefly by saying that vowel onset after /t/, compared to vowel onset after /d/, has:

- a later start relative to stop release by c.50-60ms
- a higher pitch by c.4-5Hz (c.0.5 semitones)
- a lower closed quotient by c.10-13%
- a negative spectral tilt with the second harmonic about 10dB lower in relation to the amplitude of the first harmonic

4.6 /tʃ/ - /dʒ/

The VOT measure also distinguished clearly between voiced and voiceless affricates. Jansen (2004: 58-61) has pointed out that voiced affricates, if phonetically devoiced, can have positive VOT values similar to those for aspirated plosives. VOT therefore has to be interpreted in the light of other phonetic properties distinguishing plosives from affricates, such as the duration of homorganic friction relative to glottal and general cavity friction which we have not looked at in this paper. Rate of vocal fold vibration at vowel onset was not significantly different, nor were closed quotients. However, the tilt in the spectrum at vowel onset after /tʃ/ was significantly negative compared to after /dʒ/. The apparent independence of the closed quotient and spectral tilt results here suggest a whispery voice onset but with a larger arytenoidal opening after /tʃ/. The results for these phonemes can be summarised by saying that vowel onset after /tʃ/, compared to after /dʒ/, is characterised by:

- a later start relative to stop release by c.70-80ms
- a negative spectral tilt with the second harmonic about 8-9dB lower in relation to the amplitude of the first harmonic

4.7 /t/ - /tʃ/

The VOT difference between realisations of the voiceless plosive and the voiceless affricate is significant, with the affricate having the longer time lag between stop release and voice onset. Pitch at voice onset is significantly higher after the affricate than after the plosive. The average difference of 4Hz represents about half a semitone. Closed quotient values are not significantly different, in fact they are virtually identical; the same is true for spectral tilt. In summary, vowel onset after /tʃ/, compared to after /t/, is characterised by:

- a later start relative to stop release by c.30ms
- a higher Fo by about 4Hz, equivalent to about 0.5 semitones

4.8 /d/ - /dʒ/

VOT distinguishes significantly between realisations of the voiced plosive and the voiced affricate, with the affricate having the longer time lag. A mean difference of 5-6Hz is found at vowel onset which is equivalent to a difference in semitones of around 0.7 – the affricate has the higher values. The closed quotient in the glottal cycles after the affricate is significantly lower by around 6%, but this is not reflected in spectral tilt where no significant difference was found. Vowel onset after /dʒ/, compared to after /d/, is characterised by:

- a later start relative to stop release by c.35ms
- a higher Fo by about 5-6Hz, or 0.7 semitones
- a lower closed quotient by about 6%

5. Conclusion

Unsurprisingly, the parameter that consistently distinguishes voiced from voiceless, and also plosive from affricate, is voice onset time. In Persian as in many languages, results indicate that glottal aperture is greater at the release of phonologically voiceless stops, realised with long-lag VOTs and aspiration, than at the release of phonologically voiced stops which are mostly realised with short-lag VOTs and no aspiration; in the case of the voiced affricate /dʒ/, VOT is in the long-lag range because of the devoiced fricative phase. Rate of vocal fold vibration at vowel onset was found to distinguish all pairs except /tʃ/-/dʒ/, while closed quotient distinguished /t/-/d/ and, to a lesser extent, /d/-/dʒ/. Differences in spectral tilt distinguished voiced from voiceless stops but not plosives from affricates. Phonatory quality seems to vary between breathy voice occurring in the first cycles after a phonologically voiceless stop, modal voice in the first cycles after phonologically voiced stops, and whispery voice which can occur after either category but is less common and was not observed after /d/ in our data.

References

Abercrombie, D. (1967) *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Amerman, J.D. & Parnell, M.M. (1984) Variable perceptual potency of the initial 20ms time window. *Journal of Phonetics* 12, 1-7.

Bijankhan, M. (2000) Farsi vowel compensatory lengthening: an experimental approach, In *ICSLP-2000*, vol.2, 507-510.

Bhatia, T. (1993) *The Punjabi language*. London: Routledge.

Catford, J.C. (1977) *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Cho, T. & Ladefoged, P. (1999) Variations and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27, 207-229.

Docherty, G.J. (1992) *The timing of obstruents in British English*. Berlin: Foris Publications.

Downing, L. & Gick, B. (2001) Voiceless tone depressors in Nambya and Botswana Kalang'a. In *Proceedings of the 27th meeting of the Berkeley Linguistics Society*.

Esling, J.H. & Harris, J.G. (2005) States of the glottis: an articulatory phonetic model based on laryngoscopic observations. In Hardcastle, W.J. & Beck, J.M. (eds.) *A figure of speech: a festschrift for John Laver*. Mahwah: Lawrence Erlbaum.

Esposito, C.M. (2006) *The effects of linguistic experience on the perception of phonation*. Unpublished PhD dissertation, University of California, Los Angeles. http://www.linguistics.ucla.edu/faciliti/research/research.html#Dissertations Accessed 03/09/07.

Fry, D.B. (1979) *The physics of speech*. Cambridge: Cambridge University Press.

Gimson, A.C. (1980) *An introduction to the pronunciation of English*. London: Arnold.

Gobl, C. & Ní Chasaide, A. (1999) Voice source variation in the vowel as a function of consonantal context. In Hardcastle, W.J. & Hewlett, N. (eds.) *Coarticulation*. Cambridge: Cambridge University Press. Pp.122-143.

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

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

Halle, M. & Stevens, K.N. (1971) A note on laryngeal features. *Quaterly Progress Report, MIT Research Laboratory Electronics* 101, 198-213.

Hanson, H.M., Stevens, K.N., Kuo, H-K, Chen, M.Y., & Slifka, J. (2001) Towards models of phonation. *Journal of Phonetics* 29, 451-480.

Hayward, K. (2000) *Experimental phonetics*. London: Longman.

Heselwood, B. (2007) The 'tight approximant' variant of the Arabic *'ayn*. *Journal of the International Phonetic Association* 37, 1-32.

Howard, D.M. & Angus, J. (2001) *Acoustics and psychoacoustics*. Oxford: Focal Press, 2nd edition.

Jansen, W. (2004) *Laryngeal contrast and phonetic voicing*. Groningen Dissertations in Linguistics 47. University of Groningen.

Johnson, K. (1997) *Acoustic and auditory phonetics*. Cambridge, Mass.: Blackwell.

Keating, P. (1984) Phonetic and phonological representation of stop consonant voicing. *Language* 60, 286-319.

Kent, R.D. & Read, C. (1992) *The acoustic analysis of speech*. San Diego: Singular Publishing.

Kingston, J. & Diehl, R. (1994) Phonetic knowledge. *Language* 70, 419-454.

Kochetov, A. & Lobanova, A. (2007) Komi-Permyak coronal obstruents: acoustic contrasts and positional variation. *Journal of the International Phonetic Association* 37, 52-82.

Ladefoged, P & Maddieson, I. (1996) *The sounds of the world's languages*. Oxford: Blackwell.

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

Lea, W.A. (1973) Segmental and suprasegmental influences on fundamental frequency contours. In Hyman, L.M. (ed.) *Consonant types and tones*. Los Angeles: University of California Press. 17-70.

Löfqvist, A., Koenig, L.L. & McGowan, R.S. (1995) Vocal tract aerodynamics in /aCa/ utterances: Measurements. *Speech Communication* 16, 49-66.

Lisker, L. & Abramson, A. (1964) A cross-language study of voicing in initial stops. *Word* 20, 384-422.

Mahmoodzade, Z. & Bijankhan, M. (2007) Acoustic analysis of the Persian fricative-affricate contrast. Poster presented at the *International Congress of Phonetic Sciences*, University of Saarbrücken, 6-10 August 2007.

Mahootian, S. (1997) *Persian*. London: Routledge.

Majidi, M-R. & Ternes, E. (1999) Persian. *Handbook of the IPA*. Cambridge: Cambridge University Press. Pp.124-5.

Ní Chasaide, A. & Gobl, C. (1997) Voice source variation. In Hardcastle, W.J. & Laver, J. (eds.) *The handbook of phonetic sciences*. Oxford: Blackwell. Pp. 427-461.

Ní Chasaide, A. & Gobl, C. (1999) Voice source variation in the vowel as a function of consonantal context. In Hardcastle, W.J. & Hewlett, N. (eds.) *Coarticulation*. Cambridge: Cambridge University Press. Pp.122-143.

Ohde, R. (1984) Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America* 75, 224-230.

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

Revoile, S., Pickett, J.M., Holden-Pitt, L.D. & Talkin, D. (1987) Burst and transition cues to voicing perception for spoken initial stops by impaired- and normal-hearing listeners. *Journal of Speech and Hearing Research* 30, 3-12.

Samareh, Y. (2006) *Avashesani-e Zaban-e Farsi (Persian language phonetics).* Tehran: University Publications Centre.

Scobbie, J.M., Gibbon, F., Hardcastle, W.J. & Fletcher, P. (2000) Covert contrast as a stage in the acquisition of phonetics and phonology. In Broe, M. & Pierrehumbert, J. (eds.) *Papers in laboratory phonology V: acquisition and the lexicon.* Cambridge: Cambridge University Press. Pp. 194-207.

Barry Heselwood
Department of Linguistics & Phonetics
University of Leeds
LEEDS LS2 9JT
UK

b.c.heselwood@leeds.ac.uk

Zahra Mahmoodzade
Department of Linguistics
University of Tehran
TEHRAN
Iran

zmahmodzdh@ut.ac.ir