

Shinji Ido

Nagoya University
ido@gsid.nagoya-u.ac.jp

Standard Tajik, or Modern Literary Tajik as it was called during the Soviet era, was established in the nineteen twenties and thirties based largely on the dialects of the Bukhara-Samarkand area, which was at the time the undisputed cultural centre of the Tajik-speaking population. Dushanbe, the current capital of Tajikistan, was then a small village with a population of only a few hundred and had no cultural heritage comparable to that of Bukhara or Samarkand.¹ Bukharan Tajik, whose phonology is described in this paper, is a variety of Tajik that played a particularly influential role in the phonological standardization of Tajik, which took place for the most part in 1930. For instance, the Scientific Conference of Uzbekistan Tajiks of 1930 resolved that the dialect of Bukhara must be the designated basis of the sound and orthography of literary Tajik (*baroji tajjori va kanfiransijaji ilmiji istalinobod 1930: 2*). In August the same year, the Linguistic Conference held in the then newly established Tajik Soviet Socialist Republic also adopted a similar resolution that establishes the ‘language of the Tajiks of Samarkand and Bukhara’ as the reference point in establishing the literary (i.e. standard) pronunciation (Halimov 1974: 126). According to Bergne (2007: 82), ‘the same Linguistic Conference of 22 August 1930 in Stalinabad decided that the phonetic base for the language had better be the dialect of Bukhara’.^{2,3} Thus, the Bukharan Tajik of today is the direct descendant of the variety of Tajik which served as a primary basis of standard Tajik phonological norms; and hence differs little from standard Tajik phonologically and phonetically.

The resolutions mentioned above were passed at the respective conferences despite the fact that both Bukhara and Samarkand were within the territory of Uzbekistan, in which they would remain to the present day. This has resulted in an interesting phonological discordance between standard Tajik and many dialects spoken in Tajikistan. Bukharan Tajik and the dialect of Samarkand belong to the Northern dialects, which share basically the same phoneme inventory. Since standard Tajik is based on the dialects of the Bukhara-Samarkand area, its phoneme inventory is also basically that of the Northern dialects. On the other hand, Dushanbe is just outside the area where the Northern dialects are spoken, and as a result, the phoneme inventory of the dialects spoken in the area where the capital city is situated does not coincide with that of standard Tajik. For example, the Northern dialects, and hence also standard Tajik, have the phoneme /ø/, but the close-mid central vowel is either an unstable phoneme or not a phoneme at all in the dialects spoken in the area where Dushanbe is situated. (A dialect in which the merger of /ø/ and /u/ has been recorded is spoken in a village located nine kilometres north of Dushanbe, suggesting the southern limit of the Northern dialects is further north.)⁴ The phoneme inventory of the Northern dialects does not coincide with that of the dialects

¹ A 1925 census records the population of Dushanbe at two hundred and forty two (Spisok 1925: 4, cited in Komatsu 2004: 18).

² Dushanbe had been renamed Stalinabad, after Joseph Stalin, in late 1929.

³ Bergne does not cite the source of this information.

⁴ See Sokolova (1949: 19–20) and Rastorgueva (1964: 39–40) for detail.

spoken by the people who have dominated Tajik political life after the civil war (1992–1997) either, because they are speakers of the Southern dialects.⁵

Despite its status as a basis of standard Tajik, Bukharan Tajik survives today primarily as a spoken variety with no standardized writing system. This is due, at least partly, to the political isolation of Bukhara from Tajikistan and to the fact that only a small fraction of Tajik speakers in Bukhara receive education in or on the standardized Tajik of Tajikistan. Bukhara is situated in the western-most corner of the Tajik-speaking area with another sizeable concentration of the Tajik-speaking population more than two hundred kilometres away and Tajikistan even further. The standardized Tajik of Tajikistan has a very limited presence in Bukhara, in which the language of administration, officialdom, education, and publication is now firmly Uzbek, with Russian making up a decreasing share of it. Many Bukharan Tajik speakers appear to have little desire to acquire the standardized Tajik of Tajikistan though they typically acknowledge its prestige by attaching to it such positive attributes as ‘pure’ (i.e. not heavily ‘Uzbekified’) and ‘proper’.

Today the use of Bukharan Tajik is limited to everyday interpersonal communication. In formal registers, Bukharan Tajik speakers usually resort to using a language in which they have written proficiency, namely Uzbek. Since Bukharan Tajik is a spoken variety without a written norm, written Bukharan Tajik is confined to private spheres, such as personal correspondence, shopping lists, and diaries, where Bukharans write in their respective idiolects (as there is neither an orthography nor a standardized grammar shared among Bukharan Tajik speakers).

Bukharan Tajik has diverged considerably from standard Tajik in its grammar and lexicon. The Tajik language of the media in Tajikistan is decreasingly intelligible to average Bukharan Tajik speakers, who have replaced a number of basic lexical items of Tajik (e.g. the words for ‘cloud’, ‘eyebrow’, and ‘to wait’) with their Uzbek and Russian counterparts and who have no knowledge of some of the grammatical constructions that are in frequent use in the Tajik media. For example, deontic modality marking with *бояд* ‘must’, future tense marking with *хостан* ‘want; will’, and superlative degree marking with *-тарин*, all of which have much currency in standard Tajik, are unknown to the average speaker of Bukharan Tajik. Not surprisingly, such lack of grammatical knowledge has a significantly adverse effect on the intelligibility of standard Tajik by Bukharan Tajik speakers.

Virtually every Bukharan Tajik speaker is bilingual in Bukharan Tajik and Uzbek, the Turkic language with which Tajik has been in intensive contact for centuries.⁶ Language mixing, i.e. code switching and code mixing (Ritchie & Bhatia 2004: 336–337), takes place even in households where every member is a native speaker of Bukharan Tajik. However, Bukharan Tajik–Uzbek bilingualism is not limited to those who have Bukharan Tajik as their first language – native Uzbek speakers who grow up in the city of Bukhara usually acquire some command of Bukharan Tajik, which they use either passively or actively. It may also be worth noting that Bukharan Tajik enjoys some prestige in Bukhara province as the language of city dwellers.

The formant data of Bukharan Tajik vowels that appear in this paper come mostly from two informants. One of them, to whom most of the recorded voices accompanying this paper also belong, is a female native Bukharan Tajik speaker in her early twenties. The other informant

⁵ Given the strong regionalism in Tajik politics (see Jonson 2006: 41–43, 112), the standard Tajik pronunciation may change, perhaps in favour of speakers of the Southern dialects. From this viewpoint, it is interesting that the latest orthography (Ministry of Culture of the Republic of Tajikistan 2011) gives the reading of ⟨*ӯ*⟩ as *у-и дароз* ‘long u’, because ⟨*ӯ*⟩ was originally introduced into the Tajik alphabet as the orthographical representation of /ø/, which the Southern dialects lack.

⁶ Russian is also understood to varying degrees by most Bukharan Tajik speakers. However, Bukharans speak Russian primarily to communicate with Russian monolinguals, whose number in Bukhara has dwindled over the past few decades. The active use of Russian is therefore limited particularly among young Bukharan Tajik speakers. This is despite the large number of Russian loan words that are retained in the lexicon of Bukharan Tajik.

is a male native Bukharan Tajik speaker in his mid-twenties. The informants, both of whom are bilingual in Bukharan Tajik and Uzbek, will be referred to in this paper simply as ‘the female informant’ and ‘the male informant’, respectively.

Bukharan Tajik and Persian

Tajik is closely related to Dari and to Persian within the western subgroup of the Iranian languages. Tajik is officially called *забони тоҷикӣ* ‘the Tajik language’ in the Republic of Tajikistan, but it has been frequently referred to as the Central Asian dialect of Persian (Halimov 1974: 30–31). Today, whether one regards Tajik as a language distinct from Persian or as a variety of it depends more on sentiment and politics than it does on science.⁷ The stance of the government of Tajikistan on this issue too has swung around in the past. In the language law of the Tajik SSR adopted on 22 July 1989, the word *форсӣ* ‘Persian’ in brackets was added to the term *забони тоҷикӣ* ‘the Tajik language’ only to be removed later when the law was revised in 1992 (Tojnews 2009).

However, regardless of the linguistic unity of Tajik and Persian or lack thereof, Bukharan Tajik arguably merits a separate name from the Persian of Iran, because the mutual intelligibility between Bukharan Tajik and Persian appears to be confined to grammatically and lexically simple utterances. None of the following morphemes and words, all of which are present in the ‘North Wind and the Sun’ passage at the end of this paper, exists in Persian: the ablative case suffix /ban/, dative suffix /ba/, postposition /kati/ ‘with’, noun-forming suffix /tʃi/ ‘-er, -ist’ (< Uzbek suffix *-chi*), Russian loanwords /spor/ ‘argue’ (< Russian noun *спор*), /kansakanso/ ‘eventually’ (< Russian phrase *в конце концов*), /atkazat/ ‘give up’ (< Russian verb *отказать(ся)*), /srazu/ ‘straightaway’ (< Russian adverb *сразу*), Uzbek loanwords /kutʃ/ ‘strength’ (< Uzbek noun *kuch*), /keličmič/ ‘agree on’ (< Uzbek participle *kelishmish*), and /içqilib/ ‘in short, to sum up’ (< Uzbek adverb *ishqilib*). A few dozen Bukharan Tajik auxiliary verbs, e.g. /mondan/ ‘to stay’ in /dida mondan/ ‘suddenly saw’, are also foreign to Persian. The prenominal relative clause, two instances of which are found in the following sentence taken from the ‘North Wind and the Sun’ passage, is no less foreign to Persian:

[çimol taraf-ban meomadagi] çamol [budagi] kutʃ-aç kati çamol kunont
north side-ABL come.PRPT wind was.PSPT power-3SG with blew.3SG
‘The North Wind blew with all its strength.’

Consonants

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Uvular	Glottal
Plosive	p b		t d				k g	q	
Nasal	m			n					
Tap				r					
Fricative		f v		s z	ç (z)			χ ʁ	h
Affricate					tʃ dʒ				
Approximant						j			
Lateral approximant				l					

⁷ For example, the general tendency among Iranists seems to be to postulate the linguistic unity of Dari, Tajik, and Persian, typically emphasizing the role New Persian once played as a common literary language in the regions that are now Afghanistan, Tajikistan, and Iran.

p	/pur/	‘full’	r	/røz/	‘day’
b	/bur/	‘Cut!’	s	/søz/	‘Burn!’
m	/mur/	‘Die!’	z	/zøz/	‘force’
			ç	/çøz/	‘salty’
t	/tar/	‘wet’			
d	/dar/	‘door’	ʈ	/ʈoj/	‘tea’
n	/nar/	‘male’ (used usually in reference to male dogs)	ɖ	/ɖoj/	‘place’
k	/køz/	‘blind’	χ	/χam/	‘bent’
g	/gøz/	‘tomb’	ɣ	/ɣam/	‘sorrow’
q	/qøz/	‘supernatant fluid of yoghurt’	h	/ham/	‘also’
f	/raft/	‘S/he went’	l	/laχ/	‘piece of meat’
v	/rav/	‘Go!’	j	/jaχ/	‘ice’

/t/ and /d/ are articulated at both the upper teeth and alveolar and hence are denti-alveolar.

The consonants in the column for postalveolars are in fact alveolo-palatals. The number of native words in which /z/ occurs is small in Tajik in general, and is even smaller in Bukharan Tajik, hence the brackets. The female informant recalls only the following six words as native Bukharan Tajik words in which /z/ occurs: /laʁzidan/ ‘to slip’, /azdaχo/ ‘dragon’, /azdar/ ‘dragon’, /ʁizʁak/ ‘a kind of stringed musical instrument’, /ʁazʁuz/ ‘the sound of wood sawing’, /ʁizʁiz/ ‘the sound of wood sawing’.⁸ She does not recall any native Bukharan Tajik words with word-initial /z/, which is nevertheless abundant in her loanword vocabulary. Commonly used loanwords with word-initial /z/ include /zdat/ < Russian *ждать* ‘to wait’, which is used in the light verb construction /zdat kardan/ ‘to wait’ and /zarit/ < Russian *жарить* ‘to fry’, which is also used in the light verb construction /zarit kardan/ ‘to fry’.⁹ A number of words with /z/ in Tajik have /ɖz/ in their Bukharan Tajik counterparts¹⁰ (for example, Tajik *мижа* /mizʁa/ ‘eyelash’ corresponds to Bukharan Tajik /midɖʁa/¹¹), which partly explains why the occurrence of /z/ in Bukharan Tajik is largely confined to the loanword vocabulary. /ʁizʁiz/ ‘the sound of sawing wood’ and /ʁitʁitʁ/ ‘a sound symbolic word denoting crowdedness’ are one of few minimal pairs in the native Bukharan Tajik lexicon that involve /z/.

Buzurgzoda (1940: 44) reports of Tajik dialects in which /f/ and /v/ are phonetically [ɸ] and [β], respectively, with the implication that the bilabial fricatives in the dialects are Uzbek and/or Kirghiz in origin. In contrast, Bukharan Tajik, which is also in intensive contact with Uzbek, does not utilize [ɸ] or [β]. However, the consonant phoneme inventory of Bukharan Tajik is broadly similar to that of Uzbek, with the former lacking the /ɸ/, /β/, and /ŋ/ of Uzbek and the latter lacking the /f/ and /v/ of Bukharan Tajik.

The pharyngeal fricatives that have been reported to exist in many Tajik dialects (Rastorgueva 1964: 51–52, 166), including the dialect of Samarkand Jews (Ešniëzov 1977: 65), are absent in Bukharan Tajik.

The word-initial vowel is regularly preceded by the glottal plosive, which, however, is not a phoneme in Bukharan Tajik.

The difference in voice onset time (VOT) between the voiced word-initial plosives /b d g/ and their voiceless counterparts /p t k/ is large in the recordings of the above-listed words,

⁸ I excluded Russian loanwords such as /zempir/ ‘cardigan’ and /ziket/ ‘vest’ from the words the female informant provided.

⁹ The light verb /kardan/ ‘to do’ has a shorter form /kadan/, with which it often alternates.

¹⁰ The /ɖz/–/z/ alternation is widespread among the Northern dialects (see T. N. Xaskašev 1983: 83; Fajzov 1985: 4).

¹¹ The /d/ in /midɖʁa/ is assimilated to the alveolo-palatal place of articulation.

with their VOTs ranging from approximately –80 to 20 milliseconds for the voiced plosives and from approximately 40 to 80 milliseconds for the voiceless plosives (which actually carry aspiration).

Vowels

The Bukharan Tajik vowel inventory consists of six phonemes, namely /i e a ə o u/. Figure 1 shows the formant frequencies of the vowels which the informants produced in the test words of /saχ/ ‘hard’, /se/ ‘three’, /si/ ‘thirty’, /soχt/ ‘s/he made’, /səχt/ ‘s/he burnt’, and /suq/ ‘evil eye’. Unfortunately, the small size of the native lexicon that is shared among all Bukharan Tajik speakers meant that no native context was found in which every one of the six vowels could be placed, hence the different contexts (/sV/, /sVχ/, /sVχt/) of the test words. Two of these words, namely /se/ and /si/ were randomized along with thirteen other monosyllabic numerals to form four lists of words, which the informants read aloud. The other four test words were similarly randomized with twenty-three other native Bukharan Tajik words to form another four lists of words. Consequently, a total of four repetitions were recorded for each of the test words.¹²

The most noticeable allophonic variation observed in the vowel phonology of Bukharan Tajik is that of /i/, whose allophone [i̠] occurs following and/or preceding a uvular consonant, e.g. in /βiʒβiʒ/ (see the preceding section).

The reduction of /i/, /u/, and /a/ that has been repeatedly documented in the Tajik linguistic literature (e.g. in Sokolova 1949: 32–33; Bobomurodov 1978: 6–7; T. N. Xaskašev 1983: 89–91; T. Xaskašev 1985: 42) is also observed in Bukharan Tajik. In unstressed open syllables, /i/ and /u/, and to a lesser degree /a/, are more susceptible to vowel reduction than the other three vowels, namely /e/, /o/, and /ə/. Compare, for example, /be'raχm/ ‘merciless’, in which the unstressed vowel /e/ is not reduced, with /bi'rində/ ‘rice’, where the unstressed /i/ is reduced/centralized to [ə] (Figure 2). The immunity of /e/, /o/, and /ə/ to radical reduction may possibly have some relation to the fact that they all descend from Early New Persian long vowels, while the other vowels, namely /a/, /i/, and /u/, descend either wholly or partly from Early New Persian short vowels (see Windfuhr 1987: 457–458 for detail).

One peculiarity of the male informant’s vowel system that differentiates it from the female informant’s vowel system is the relative backness of the male informant’s /ə/ (see Figure 1).¹³ One may be tempted to identify the back realization of /ə/ in his vowel system as a characteristic of the vowel system of male Bukharan Tajik speakers in general. However, it is not clear whether this is the case, as the formant data obtained from a group of ten male Bukharan Tajik speakers (Figure 3) do not exhibit the same degree of /ə/-backness as those obtained from the male informant – in fact, formant data shown in Figure 3 appear to point to a vowel system of male Bukharan Tajik speakers where the six vowels are more evenly distributed in the vowel space.¹⁴

It is also worth mentioning that the two plots in Figure 3 exemplify the ‘anatomically unexplained’ (Diehl et al. 1996: 188) formant similarities of the male and female close

¹² Formant frequencies were measured with Praat (Boersma & Weenink 2011). The calculation of mean formant values was made on the NORM website (Thomas & Kendall 2007).

¹³ The vowel /ə/ is historically Early New Persian ‘long o’ that shifted forward in a chain shift, which also raised Early New Persian ‘long a’ to the position of /o/.

¹⁴ The test words were /du/ ‘two’, /se/ ‘three’, /tʃor/ ‘four’, /da/ ‘ten’, /bis/ ‘twenty’, and /imrəz/ ‘today’ (for one male informant, /imrəz/ was replaced by /hamin rəz/ ‘this very day’). Totals of 65 and 124 tokens were obtained from nine females and ten males, respectively. Factors such as the number of repetitions per test word and number of tokens per speaker were not controlled for. (Thus, while it seems positive that the male informant’s /ə/ in Figure 1 is more back than the general Bukharan average, it should be borne in mind that the methods employed for obtaining formant data shown in Figure 3 do not warrant an ideally controlled comparison with data shown in Figure 1.) The female informants’ years of birth ranged from 1944 to 1989, while the male informants’ years of birth ranged from 1947 to 1997.

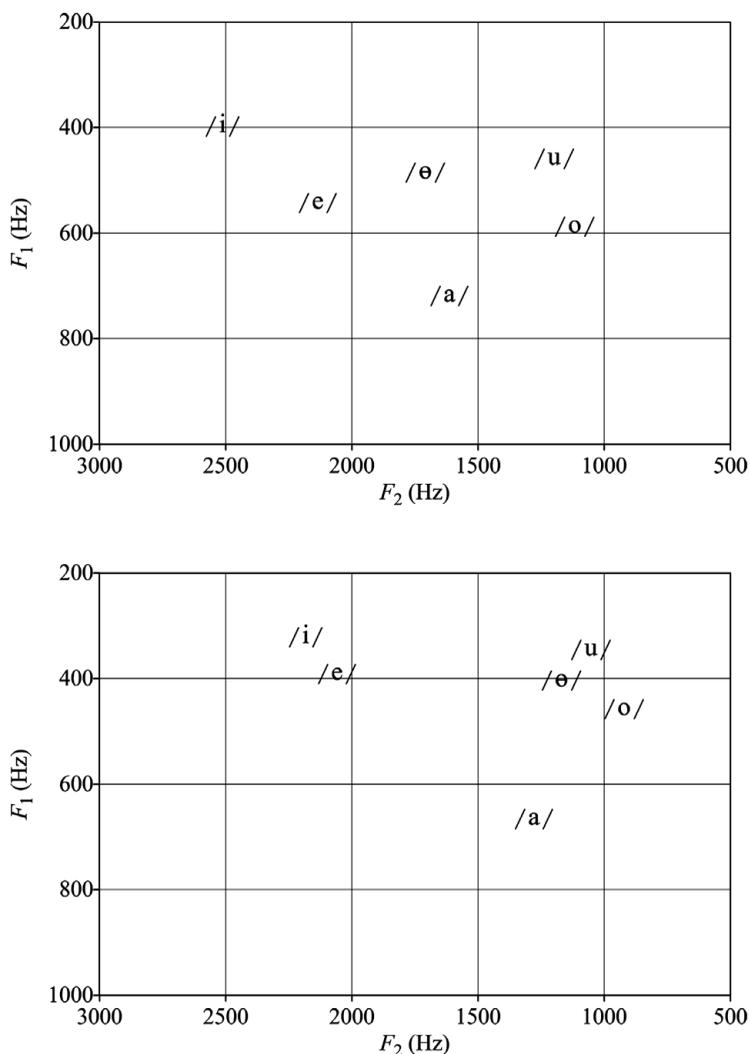


Figure 1 Mean F1 and F2 values of the Bukharan Tajik vowels in /sax/ 'hard', /se/ 'three', /si/ 'thirty', /soxt/ 's/he made', /sext/ 's/he burnt', and /suq/ 'evil eye' produced by the female informant (top) and the male informant (bottom).

rounded back vowel which have attracted repeated attention among phoneticians (e.g. Fant 1966, Simpson & Ericsson 2003).

Bukharan Tajik and Uzbek vowel systems

Since Bukharan Tajik speakers are bilingual in their native language and Uzbek, in the next few paragraphs, I attempt a preliminary comparison of the vowel systems of Bukharan Tajik and Uzbek as they are represented in the speech of the two Bukharan informants. Uzbek, like Bukharan Tajik, has six vowel phonemes.

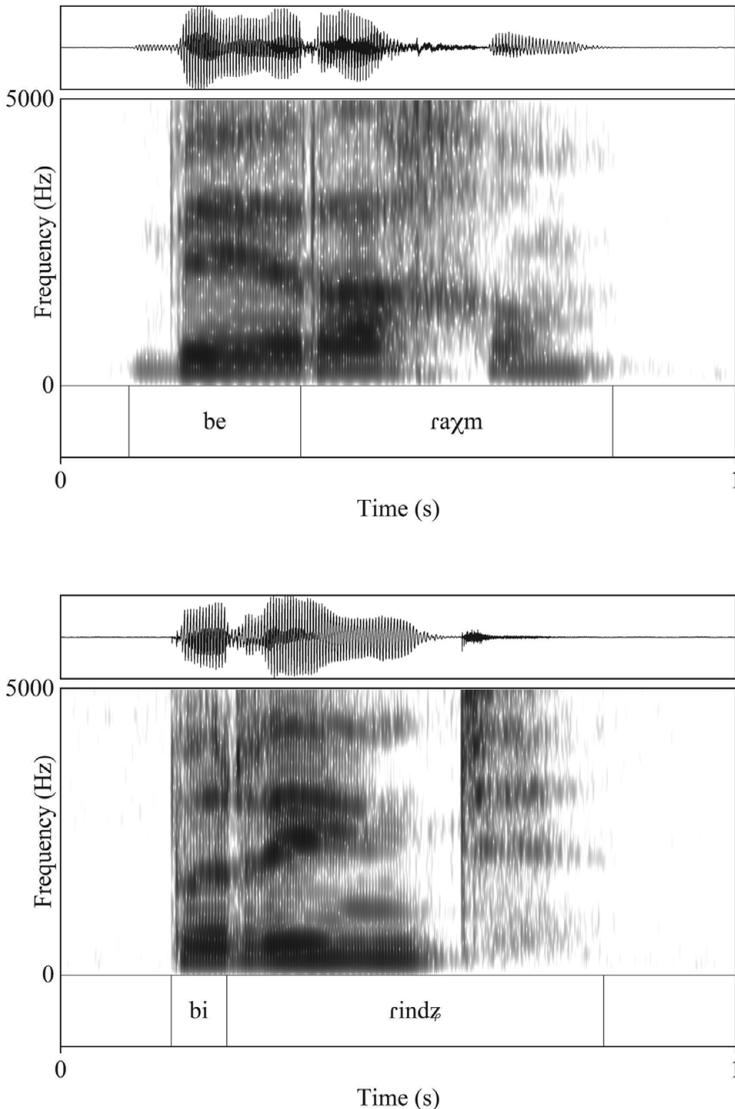


Figure 2 /be'raxm/ 'merciless' (top) and /bi'rindz/ 'rice' (bottom) produced by the female informant.

Observe Figure 4, which shows the formant frequencies of the six Uzbek vowels as they were produced by the informants in the Uzbek test words *uz* 'Tear off!', *ez* 'Crush!', *iz* 'trace', *oz* 'few' *o'z* 'self', and *az*, the last of which is a meaningless word in Uzbek.¹⁵

A comparison of Figure 1 with Figure 4 reveals an interlingual consistency between the informants' Bukharan Tajik and Uzbek vowel systems – their Bukharan Tajik vowels exhibit a

¹⁵ These words were randomized along with eleven other monosyllabic native Uzbek words to form four lists of words, which the informants read aloud. Consequently, a total of four repetitions was recorded for each of the test words. Incidentally, /az/ appears in a few seldom-used phrases borrowed into Uzbek from Tajik, where /az/ is the preposition meaning 'from'. The Bukharan Tajik counterpart of the preposition is /a/ 'from'.

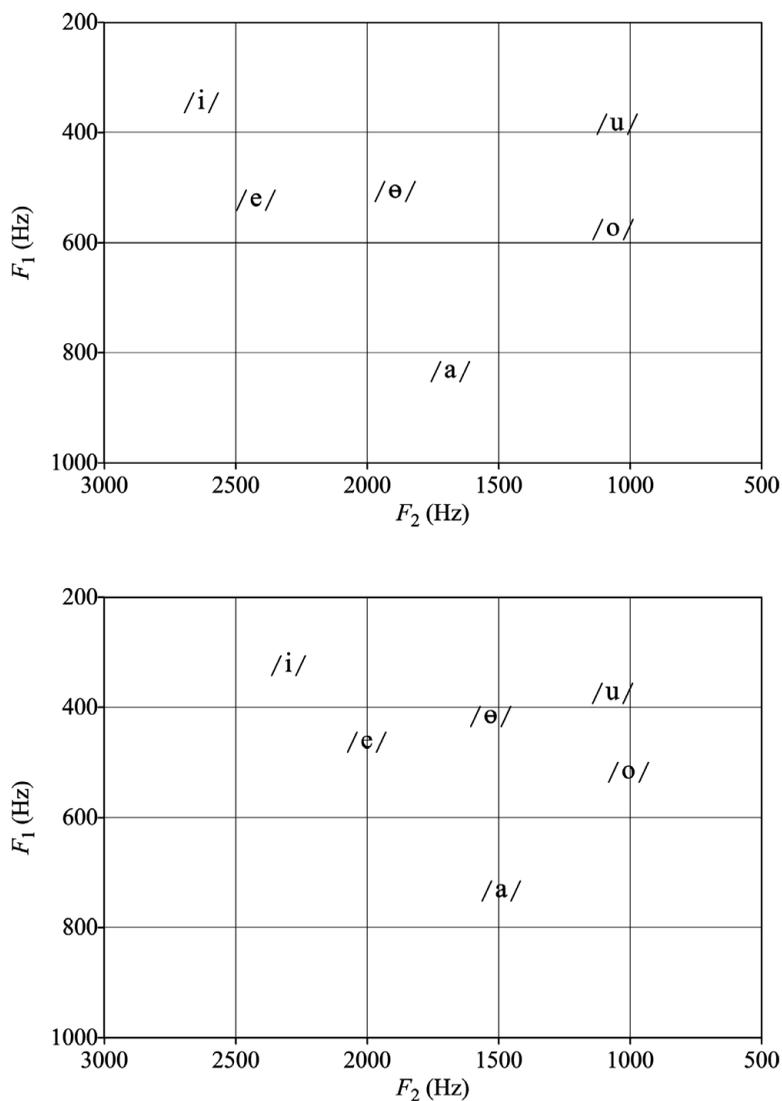


Figure 3 Mean F1 and F2 values of the Bukharan Tajik vowels produced by nine female speakers (top) and ten male speakers (bottom).

marked resemblance to their Uzbek vowels in terms of their positions relative to one another in the F1-F2 space. The Bukharan Tajik vowels of the male informant in particular are virtually identical with his Uzbek vowels. Does this, then, mean that Uzbek happens to have a vowel system that is identical with that of Bukharan Tajik? The formant values of standard Uzbek vowels plotted in Figure 5 indicate that this is not the case – the vowel system of the Tashkent dialect of Uzbek, which is the phonetic basis of standard Uzbek, clearly differs from the Uzbek vowel systems of the Bukharan informants shown in Figure 4.¹⁶

¹⁶ The formant frequencies of the six Tashkent Uzbek vowels in the /Vz/ context were obtained using the same procedure as for Figure 4. For obtaining the formant values in the plot on the bottom of Figure 5,

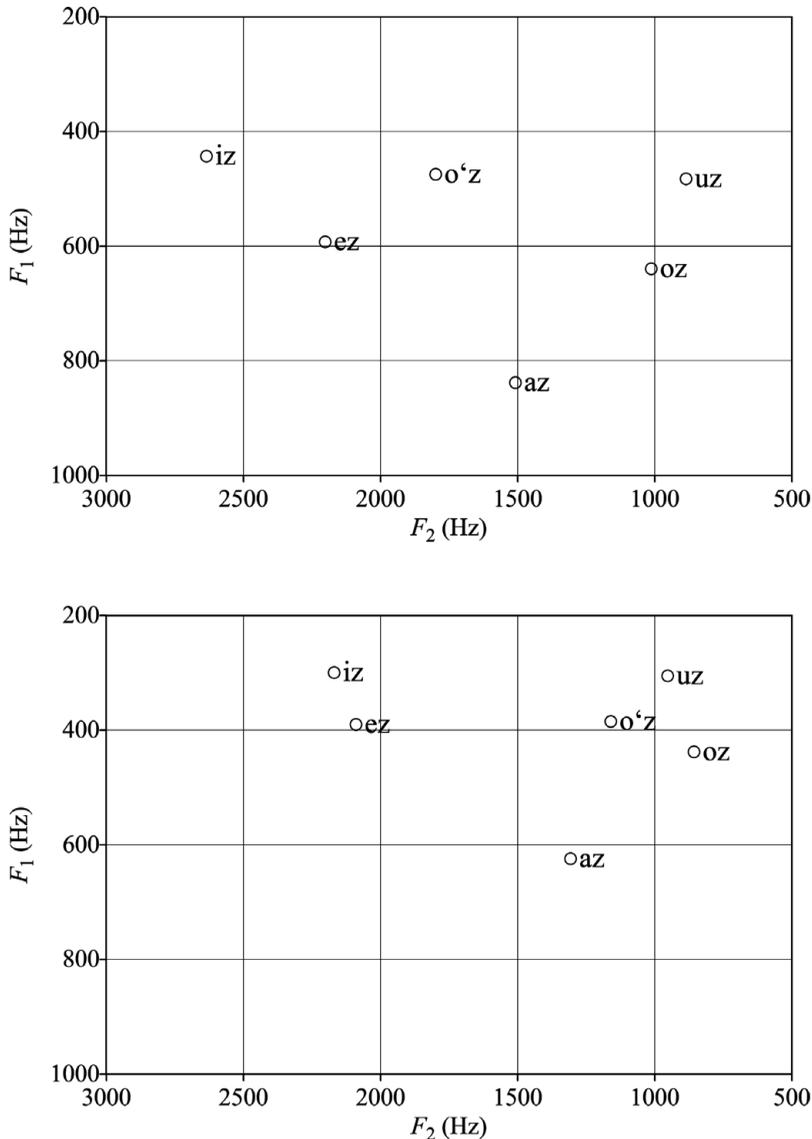


Figure 4 Mean F1 and F2 values of the Uzbek vowels in the /Vz/ context produced by the female informant (top) and the male informant (bottom).

In other words, the Bukharan informants' Uzbek vowel systems have non-standard features that render them practically identical with the Bukharan Tajik vowel system. Two such non-standard features can be identified in their Uzbek vowel systems. One of them is the closeness of the vowel in *oz* 'few'. The vowel in *oz*, which is open-mid in standard Uzbek, is close-mid in the Uzbek vowel systems of the Bukharan informants and thus coincides with Bukharan Tajik /o/. This observation is in agreement with Mirzaev's (1969: 19, 28, 30)

the Uzbek letters for the six vowels were randomized to form four lists of the vowels, with which four repetitions of each of the vowels were recorded.

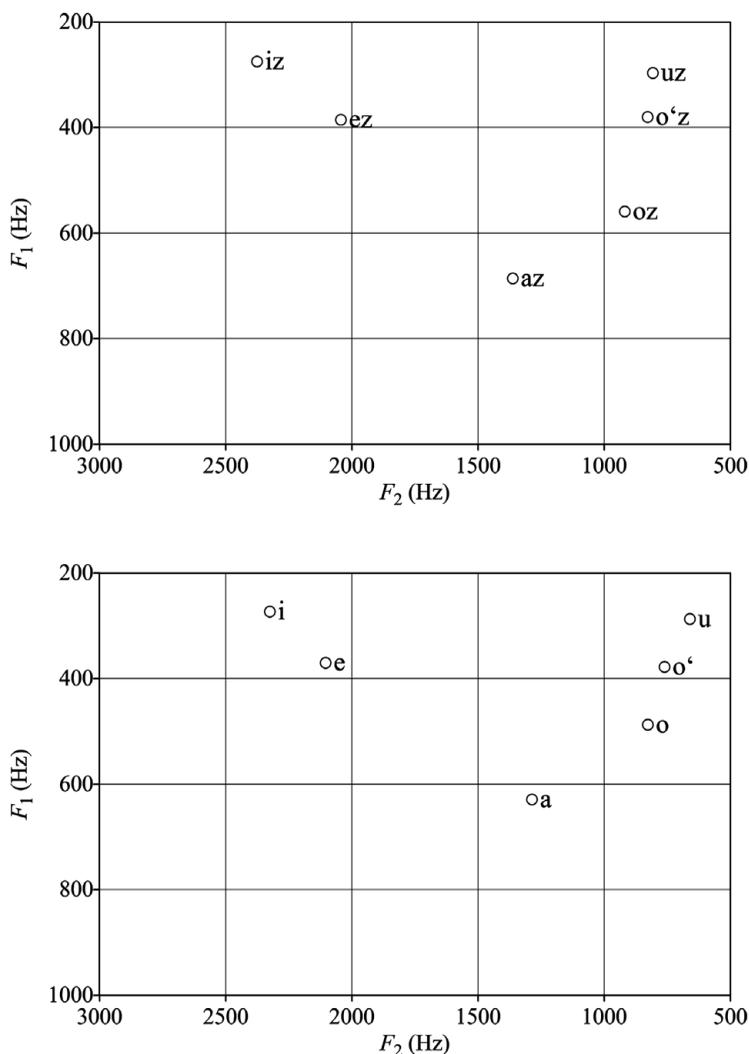


Figure 5 Mean F1 and F2 values of the Uzbek vowels produced in the /Vz/ context (top) and in isolation (bottom) by a male native Uzbek speaker (born 1985) from Tashkent.

remark that the vowel is very close to Tajik /o/ in the variety of Uzbek spoken by Bukharan Tajik-Uzbek bilinguals. The other feature is the frontness of the vowel in *o'z* 'self' relative to the back vowels in *oz* 'few' and *uz* 'Tear off!'. While the vowel in *o'z* is a fully back vowel in standard Uzbek, it is more front in the Bukharan bilinguals' Uzbek vowel systems, and coincides with Bukharan Tajik /ø/. This observation is endorsed by Bobomurodov's (1978: 13) remark that the Uzbek close-mid rounded vowel differs from Tajik close-mid rounded /ø/ in being a back vowel.

Thus, the closeness of the vowel in *oz* 'few' and the relative frontness of the vowel in *o'z* 'self' distinguish the Bukharan informants' Uzbek vowel systems from the standard Uzbek vowel system. These are at the same time the features that render their Uzbek vowel systems practically identical with the Bukharan Tajik vowel system of /i e a ø o u/.

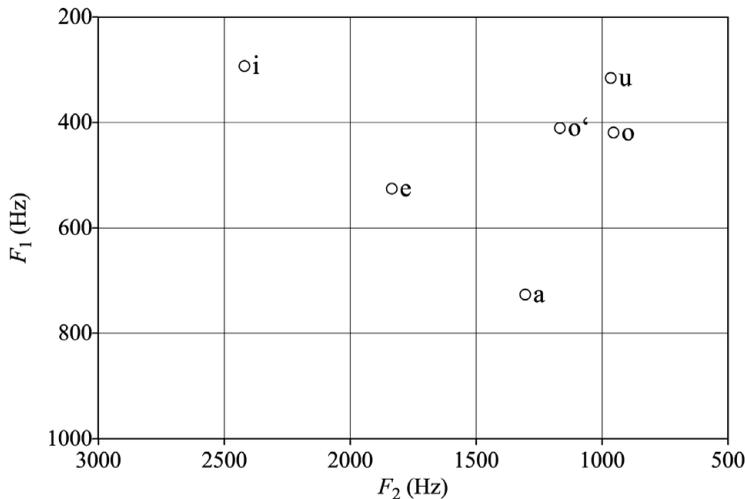


Figure 6 Mean F1 and F2 values of the Uzbek vowels produced in isolation by a male native Uzbek speaker (born 1988) from Bukhara.

Interestingly, these two non-standard features are also found in the Uzbek vowel system of a native Uzbek speaker from Bukhara district (Figure 6). The male native Uzbek speaker has Uzbek as his first language and is bilingual in Bukharan Tajik.¹⁷

Thus, the three bilinguals from Bukhara, regardless of their first languages, utilize similar Uzbek vowel systems which are characterized by their resemblance to the Bukharan Tajik vowel system. This allows the speculation that Central Asian Iranian–Turkic language contact has reorganized the vowel systems of Tajik and Uzbek in the Bukhara area in such a way that they resemble each other. There are a few lines of evidence that seem to suggest that this is the case. I present some of this evidence below, though this is admittedly fragmentary, not least because no wide-scale study has been carried out in Bukhara that involves measurement of vowel formant frequencies.

One piece of evidence comes from Central Asian Arabic. According to Tsereteli (1970), in the dialect of Arabic spoken in Bukhara province, a couple of vowel shifts occurred where ‘C[ommon] A[rabic] /ā/’ and ‘/aw/’ shifted to a position close to that of Tajik /o/ and the position of Tajik /ø/, respectively. If, as Tsereteli claims, the shifts took place due to Tajik influence, it shows that reorganization of a vowel system can and did occur in the Bukhara area. Perhaps more importantly, it shows that all of the Iranian, Turkic, and Semitic varieties in the Bukhara area have /ø/ as a phoneme (Bukharan Tajik /ø/, the non-back Uzbek vowel in *o‘z*, and Central Asian Arabic /aw/ > /ø/). This is significant because /ø/ is not used in all (six-vowel) Uzbek dialects that are classified as belonging to the same dialect group as the Bukhara dialect, nor is it widespread among Tajik dialects.^{18,19} The fact that /ø/ exists in

¹⁷ The Uzbek speaker is from Galaosiyo, a town about 10 kilometres from the city centre of Bukhara. The quality of the recording from which the formant data of his vowels were obtained is less than ideal, but the formant data, which are plotted in Figure 6, allow a rough comparison with those obtained from other informants. The vowels were recorded with no randomization of tokens, two repetitions of each vowel, and some background noise. The mean formant value of the vowel in *az* in particular should only be taken as a very rough approximation because the speaker’s phonation of two tokens for *az* are highly volatile in the recorded audio data.

¹⁸ See Rešetov & Šoabdurahmonov (1978: 29–42) for various classifications of Uzbek dialects.

¹⁹ The Central and Southern dialects generally lack the phoneme, and even in the area that is halfway between Bukhara and Dushanbe, a dialect exists (or existed in the mid-twentieth century) that lacks the

genetically different languages in Bukhara appears to lend further support to the speculation that language contact shifted vowels in different languages in Bukhara, resulting in their utilization of vowel systems that resemble one another.

Another piece of evidence is the number of vowels in Tajik and Uzbek dialects. The number of vowel phonemes vary in Tajik and Uzbek dialects. Uzbek dialects with six vowel phonemes are most conspicuous in the Tajik-Uzbek contact area, with other dialects typically utilizing larger inventories of vowels (see Šoabdurahmonov 1971: 397–398; Rešetov & Šoabdurahmonov 1978: 44–46; Rajabov 1996: 99–100). The picture is pretty much the same for Tajik dialects, with six-vowel dialects distributed prominently in areas where Tajik-Uzbek bilingualism has been the norm (Rastorgueva 1964: 31–41). This, together with the lack of vowel harmony in Uzbek six-vowel dialects, which has often been ascribed to Tajik/Iranian influence (e.g. Johanson 2006), may be seen as circumstantial evidence that language contact influenced the vowel systems of Tajik and Uzbek in bilingual areas (of which Bukhara is one) to be more like each other.

In sum, the two informants and an Uzbek speaker from Bukhara use similar Uzbek vowel systems, which are characterized by their resemblance to the Bukharan Tajik vowel system. This resemblance may be ascribed to the language contact in Bukhara, as evidence appears to suggest that the language contact induced Bukharan Tajik and Uzbek vowel systems to resemble one another.

Vowel length

Vowel length was a point of contention during the period of standardization of Tajik, when Tajik linguists were in disagreement as to whether the long vowels /i:/ and /u:/, which originate in Early New Persian, were to be credited the status of phonemes in standard Tajik. Indeed, the dialects on which standard Tajik is based were not homogeneous in their degree of retention of the historical vowel length distinction.²⁰ The Samarkand dialect has been explicitly classified as a dialect without /i:/ by Buzurgzoda (1940: 33). Rastorgueva (1964: 33), on the other hand, classifies the same dialect as a dialect without /u:/. As for Bukharan Tajik, Kerimova (1959: 6) lists twelve Bukharan Tajik words as words in which the vowel length distinction of Early New Persian close vowels is retained. However, only a few of the words in the list, such as /çiča/ ‘glass’, /surat/ ‘appearance’, and /sina/ ‘breast’ are currently in common use among Bukharan Tajik speakers. The vowel length distinction in the close vowels is therefore only marginally, if at all, significant in terms of phonology in the Bukharan Tajik of today. Nevertheless, the vowel length distinction is, albeit only phonetically, still present in the pronunciation of Bukharan Tajik. As can be observed in Table 1, the average vowel durations of the unstressed /i/ and /u/ in /çiča/ and /surat/ are greater than those in /çičtan/ ‘sit’ and /surχak/ ‘reddish’ which do not contain any vowels that originate in Early New Persian long vowels.²¹ (For comparison, the average durations of stressed /i/ and /u/ are also shown in the table, using /çič/ ‘six’ and /surχ/ ‘red’ as examples.)

The greater lengths of the unstressed vowels in /çiča/ and /surat/ relative to those in /çičtan/ and /surχak/ cannot be ascribed to the different syllabic contexts (i.e. open and

phoneme (see Sa’dulloev 1970: 25). /ə/ does occur as a phoneme in the South-Eastern dialects of Tajik, which are, however, not directly relevant to the discussion here because they are geographically remote from the Iranian–Turkic contact area (and hence also from Bukhara) and are in contact with the Pamir languages whose vowel systems are diverse (see Payne 1989: 426).

²⁰ To my knowledge, there are no dialectological studies on the Bukhara-Samarkand area that present minimal pairs involving vowel length distinction.

²¹ These words were among the twenty-seven words in the list of Bukharan Tajik words that the informants read aloud. The order of words in the list was randomized to make four lists, all of which the informants read aloud, giving four tokens for each word. Each vowel duration shown in the table is the average of the vowel durations measured in the four tokens (except the average for /surχak/ of the female informant, for which the number of tokens was five).

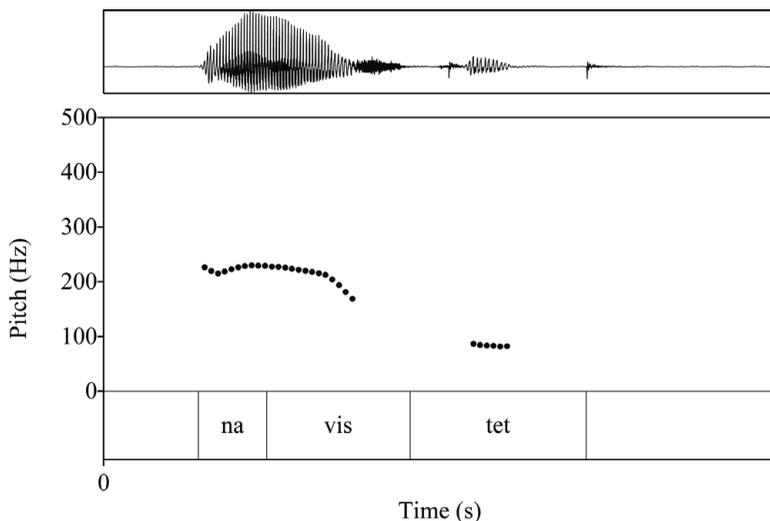
Table 1 Average duration (s) of /i/ and /u/ in different words.

	The female informant	The male informant
/çiča/	0.0691	0.0958
/çičtan/	0.0437	0.0372
/çič/	0.0948	0.1207
/surat/	0.1392	0.1201
/surçak/	0.0402	0.0466
/surç/	0.0832	0.0884

close) in which the vowels occur, because the unstressed /i/ and /u/ in the open syllables in the words /çifo/ ‘recovery of health’ and /sukut/ ‘silence’ produced by the female informant are not long.²² In fact, in the female informant’s pronunciation, the unstressed /i/ and /u/ in /çifo/ and /sukut/ are invariably devoiced or elided outright – in contrast, the vowel length distinction of Early New Persian close vowels retained in /çiča/ and /surat/ apparently prevents the unstressed /i/ and /u/ from being reduced in the context in which they are prone to reduction (see the third paragraph of this section).

Intonation

An intonation pattern in Bukharan Tajik that is of particular interest is the one for the polar interrogative. Bukharan Tajik has the sentence-final yes–no question particle /mi/, which it borrowed from Uzbek, but the borrowed particle apparently did not put the rising intonation for the yes–no question into complete disuse.²³ The particle /mi/ is used frequently in Bukharan Tajik, but the rising intonation for the yes–no question, which is exemplified in Figure 8, is used just as frequently. Note that the pitch contour of the declarative (Figure 7) lacks the pitch rise observable in the pitch contour of the polar interrogative (Figure 8).

**Figure 7** /navistet/ ‘You wrote’ produced by the female informant.

²² Three repetitions for each word were obtained from the female informant.

²³ The particle /mi/ was introduced into standard Tajik through the literary works of Sadridin Ayni (1878–1954) (Halilov 1977: 28).

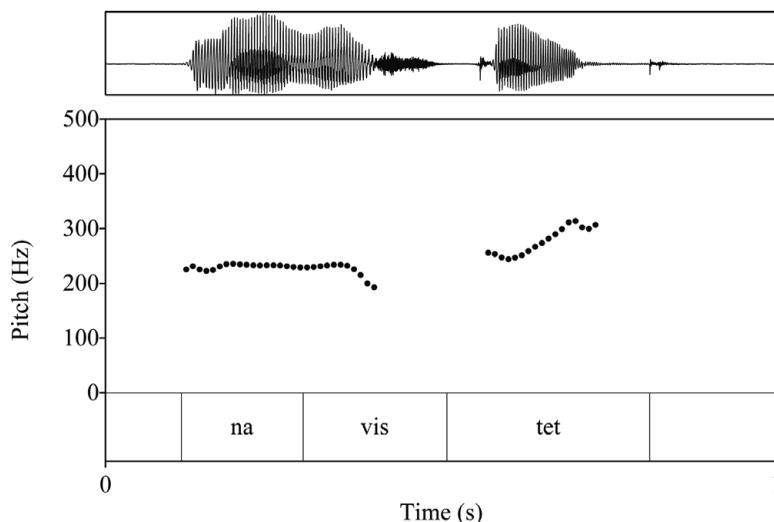


Figure 8 /navistet/ 'Did you write?' produced by the female informant.

Transcription of recorded passage 'The North Wind and the Sun'

Phonetic transcription

çimol tarafban meomadagi çamol oftob kati hamdigaraçba man kutçnok gufta spor kardaçtagi budan || hamun pajtba kaçvs palto pçidagi sajohtaçija dida mondan | ki hamin sajohtaçija paltoça kaçonda tonat hamun kutçnok gufta qaror qabul kardanba keliçmiç kardan || çimol tarafban meomadagi çamol budagi kutçaç kati çamol kunont lekin sajohtaçi dastaç kati paltoça zitç dasgirift || kansakansof çimol tarafban meomadagi çamol kutç sarf kardanban atkazat kad || badi oftob buromadan hamma dçoj garm çudanban bad sajohtaçi paltoça srazu kaçidanba qaror kad || ieçilib çimol tarafban meomadagi çamol oftob a çudaç dida kutçnok budageçban tan gitanban ilodçaç namond ||

Phonetic transcription with interlinear gloss

çimol taraf-ban meomadagi çamol oftob kati hamdigar-aç-ba man kutçnok
north side-ABL come.PRPT wind sun with each.other-3SG-DAT I powerful
gufta spor kardaçtagi bud-an || hamun pajt-ba kaçvs palto pçidagi
say.PFGER argue.PRPT were-3PL that.very time-DAT thick coat wear.PSPT
sajohtaçi-ja dida mond-an | ki hamin sajohtaçi-ja palto-ç-a
traveller-ACC see.PFGER stayed-3PL that this.very traveller-GEN coat-3SG-ACC
kaçonda ton-at hamun kutçnok gufta qaror qabul kardan-ba
take.off.PFGER can-3SG that.very powerful say.PFGER decision to.accept-DAT
keliçmiç kard-an || çimol taraf-ban meomadagi çamol budagi kutç-aç kati
agreed-3PL north side-ABL come.PRPT wind was.PSPT power-3SG with
çamol kunont lekin sajohtaçi dast-aç kati palto-ç-a zitç dasgirift ||
blew.3SG but traveller hand-3SG with coat-3SG-ACC tight grasped.3SG

kansakansof ċimol taraf-ban meomadagi ċamol kutċ sarf kardan-ban
 eventually north side-ABL come.PRPT wind power to.spend-ABL

atkazat kad || bad-i oftob buromadan hamma dċoj garm ċudan-ban bad
 gave.up.3SG after-IZ sun to.come.out all place hot to.become-ABL after

sajohattċi palto-ċ-a srazu kaċidan-ba qaror kad || iċqilib ċimol
 traveller coat-3SG-ACC straightaway to.take.off-DAT decided.3SG in.sum north

taraf-ban meomadagi ċamol oftob a ħud-aċ dida kutċnok
 side-ABL come.PRPT wind sun from self-3SG see.PFGER powerful

budageċ-ban tan gitan-ban ilodċ-aċ na-mond ||
 was.PSPT.3SG-ABL to.acknowledge-ABL means-3SG NEG-stayed.3SG

Abbreviations

3SG	third person singular	IZ	izafet
3PL	third person plural	NEG	negative
ABL	ablative	PFGER	perfective gerund
ACC	accusative	PRGPT	progressive participle
DAT	dative	PRPT	present participle
GEN	genitive	PSPT	past participle

Acknowledgements

This research was partially supported by a Grant-in-Aid for Young Scientists (B) (22720162) from the Ministry of Education, Culture, Sports, Science and Technology in Japan. I gratefully acknowledge the helpful comments of Adrian P. Simpson and two anonymous reviewers.

References

- varoċi tajjorċi va kanfransijaji ilmiji istalinobod (برای نیاری به کنفرنسیه علمی استالین آباد). 1930, June. *rahşari doniř*, 6(33), 1–4.
- Bergne, Paul. 2007. *The birth of Tajikistan: National identity and the origins of the republic*. London: I. B. Tauris.
- Bobomurodov, řakar. 1978. *Садоноки "Ў" ва мавċен он дар системаи вокализми забони адабии тоċик (Дастури методċи)*. Dushanbe: Universiteti Davlatii Tojikiston ba nomi V. I. Lenin.
- Boersma, Paul & David Weenink. 2011. Praat: Doing phonetics by computer (Version 5.2.23). <http://www.praat.org/>.
- Buzurgzoda, Lutfullo. 1940. *Fonetikaji zavoni adaviji toċik*. Stalinobod & Leningrad: Nařrijoti davlatiji Toċikiston.
- Diehl, Randy L., Björn Lindblom, Kathryn A. Hoemeke & Richard P. Fahey. 1996. On explaining certain male–female differences in the phonetic realization of vowel categories. *Journal of Phonetics* 24, 187–208.
- Eřniċzov, M. 1977. *Диалектологияи тоċик (Қисми якум)*. Dushanbe: Universiteti Davlatii Tojikiston ba nomi V. I. Lenin.
- Fajzov, Mahram. 1985. *Таджикское литературное произношение*. Dushanbe: Doniř.
- Fant, Gunnar. 1966. A note on vocal tract size factors and non-uniform F-pattern scalings. *STL-QPSR* 7(4), 22–30.
- Halilov, Abdusalom. 1977. *Ҳиссаċаċо дар забони адабии ҳозираи тоċик*. Dushanbe: Doniř.
- Halimov, Soleh. 1974. *Садриддин Айни ва баъзе масъалаҳои инкишофи забони адабии тоċик*. Dushanbe: Irfon.

- Johanson, Lars. 2006. Uzbek. In Keith Brown (ed.), *Encyclopedia of language and linguistics*, 2nd edn., 287–290. Amsterdam: Elsevier Science.
- Jonson, Lena. 2006. *Tajikistan in the New Central Asia: Geopolitics, great power rivalry and radical Islam*. London: I. B. Tauris.
- Kerimova, Aza Alimovna. 1959. *Говор таджиков Бухары*. Moscow: Izdatel'stvo Vostočnoj Literatury.
- Komatsu, Hisao. 2004. Chūō ajia no chōbō. In Ichirō Iwasaki, Tomohiko Uyama & Hisao Komatsu (eds.), *Gendai chūō ajiaron*, 1–26. Tokyo: Nihon Hyōronsha.
- Ministry of Culture of the Republic of Tajikistan. 2011. Қоидаҳои имлои забони тоҷикӣ. http://www.vfarhang.tj/index.php?option=com_content&view=category&layout=blog&id=37&Itemid=54&lang=tj (accessed 25 December 2011).
- Mirzaev, Mustaqim M. 1969. *Ўзбек тилининг Бухоро гурппа шевалари*. Tashkent: Fan.
- Payne, John. 1989. Pāmīr languages. In Rüdiger Schmitt (ed.), *Compendium linguarum Iranicarum*, 417–444. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Rajabov, Nazar. 1996. *Ўзбек шевашунослиги*. Tashkent: O'qituvchi.
- Rastorgueva, Vera Sergeevna. 1964. *Опыт сравнительного изучения таджикских говоров*. Moscow: Nauka.
- Rešetov, Viktor Vasil'evič & Šonazar Šoabdurahmonov. 1978. *Ўзбек диалектологияси*. Tashkent: O'qituvchi.
- Ritchie, William C. & Tej K. Bhatia. 2004. Social and psychological factors in language mixing. In Tej K. Bhatia & William C. Ritchie (eds.), *The handbook of bilingualism*, 336–352. Malden, MA & Oxford: Blackwell.
- Sa'dulloev, B. 1970. Доир ба садонокҳои шеваҳои Фелон. In Razzoq Ğafforov & Ğaffor Jūraev (eds.), *Масъалаҳои шевашуносии тоҷик (Ҷилди I)*, 23–34. Dushanbe: Doniш.
- Simpson, Adrian P. & Christine Ericsdotter. 2003. Sex-specific differences in f0 and vowel space. *16th International Congress of the Phonetic Sciences (ICPhS XVI)*, Saarbrücken, 933–936.
- Šoabdurahmonov, Šonazar. 1971. Ўзбек шевалари ва уларни ўрганиш ҳақида умумий маълумот. In Šonazar Šoabdurahmonov (ed.), *Ўзбек халқ шевалари луғати*, 386–407. Tashkent: Fan.
- Sokolova, Valentina Stepanovna. 1949. *Фонетика таджикского языка*. Moscow: Izdatel'stvo Akademii Nauk SSSR.
- Spisok. 1925. *Список населенных мест Узбекской ССР и Таджикской АССР (Выпуск I. Таджикистан)*. Samarkand: Central'noe Statističeskoe Upravlenie Uzbekistana.
- Thomas, Erik R. & Tyler Kendall. 2007. NORM: The vowel normalization and plotting suite. <http://ncslaap.lib.ncsu.edu/tools/norm/> (accessed 22 December 2011).
- Tojnews. 2009. /09/07. ҲМИТ хостори дар Қонуни забон чорӣ шудани вожаи «форсӣ» аст, 7 September 2009. http://tojnews.org/taj/index.php?option=com_content&task=view&id=264 (accessed 24 December 2011).
- Tsereteli, Giorgi V. 1970. The influence of the Tajik language on the vocalism of Central Asian Arabic dialects. *Bulletin of the School of Oriental and African Studies* 33(1), 167–170.
- Windfuhr, Gernot L. 1987. Persian. In Bernard Comrie (ed.), *The world's major languages*, 445–459. London: Routledge.
- Haskašev, Talbak Nabotovič. 1983. *Фонетикаи забони адабии ҳозираи тоҷик (Қисми I. Садонокҳо ва ҳамсодаҳо)*. Dushanbe: Universiteti Davlatii Tojikiston ba nomi V. I. Lenin.
- Haskašev, Tolib. 1985. Фонетика. In Šarofiddin Rustamov & Razzoq Ğafforov (eds.), *Грамматикаи забони адабии ҳозираи тоҷик (Ҷилди I. Фонетика ва морфология)*, 13–78. Dushanbe: Doniш.