Variation in Persian Vowel Systems

Corey Miller, University of Maryland
cmiller6@umd.edu

Abstract

The vowel systems employed by different varieties of Persian across time and space exhibit a great deal of variation. This study attempts to describe that variation diachronically, in the spirit of Pisowicz (1985), and synchronically by examining the three major national varieties: Farsi, Dari and Tajik. We interpret the variation encountered through general principles of vowel shifting, as described by Labov (1994), from an Early New Persian baseline. We thus trace the historical development of the vowel systems in the major varieties of Persian spoken in Iran, Afghanistan and Tajikistan, clarifying the relationships among the varieties and the extent of their adherence to such principles.

Keywords: Persian, Farsi, Dari, Tajik, vowels, phonetics, phonology

1. Introduction: Early New Persian Vowel System[1]

The history of Persian, a south-western Iranian language, is generally divided into three major periods: Old Persian, the language of the Achaemenids (558–330 BC), Middle Persian, the language of the Sassanids (224–651 AD), and New Persian, beginning in the seventh century AD, following the Arab Conquest (Windfuhr 2009: 445). The baseline vowel system from which we will be departing has been attributed both to Early New Persian (Windfuhr 2009: 457) and Classical Persian (Pisowicz 1985, Thiesen 1982). Both the dating and use of these terms is subject to divergence in the literature. Windfuhr and Perry (2009: 533) date Early New Persian (ENP) to 1100–1300 AD, while Windfuhr (2009: 447) assigns it to the 10th and 11th centuries. There is similar divergence in the dating of Classical Persian; Windfuhr provides several possibilities, such as between the 13th and 16th centuries (2009: 447), the 13th and 15th centuries (1979: 166), and along with Perry, between 1300 and 1600 (Windfuhr and Perry 2009: 533). However, according to Paul (2002: 21), Classical Persian "is not a well-defined linguistic term. It is a literary term only, but there is no agreement among Iranians or non-Iranians on what it actually denotes." As we will see, the presence of divergences from this system in the periods associated with Classical Persian shows the necessity of ascribing this baseline to ENP. The vowel system of this variety, which we represent in **bold** in order to facilitate comparison with other varieties, is represented in Figure 1. The positions of the vowels are approximate and derived from Pisowicz (1985) and Horn (1901). In particular, Horn (1901: 19) presents evidence from Judeo-Persian that the short **a** had a fronted

[1] I would like to thank Erik Anonby, Youli Ioannesyan, Carina Jahani, and Geoffrey Haig for valuable comments and suggestions.

quality. In this and subsequent charts, arrows emanating from diphthongs are only intended to show the direction of the glides, rather than the location of their final targets. We transcribe diphthongs with offglides rather than vocalic targets because at this stage we have not investigated their phonetic details. Note that several researchers, including Jahani and Korn (2009: 648) in a discussion of similar phenomena in Balochi, prefer to label **aj** and **aw** as VC sequences due to their phonotactics. Okati (2012: 178–179) provides an excellent survey of the treatment of these sounds in various Iranian languages and dialects.

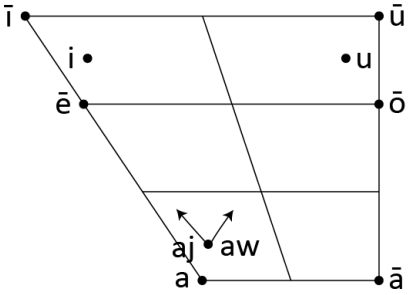

Figure 1. Vowel system of Early New Persian

The establishment of this baseline variety precedes the colonial Persian of India and the Ottoman Empire of the 15th–19th centuries (Windfuhr 1979: 166). Indeed, this vowel system is essentially reflected in the contemporary Persian of Afghanistan, also known as Dari (Farhadi 1955: 7), and the Persian of India in the Mogul period (16th–19th centuries), of which Phillott (1919) offers a comprehensive description; however, it differs markedly from contemporary Iranian Persian (CIP), or Farsi, as we will discuss below.

Since the Persian writing system, based on Arabic, does not provide any insight into changes in the pronunciation of these vowels, it is useful to consult sources which reference foreign languages to assist in establishing when this baseline was in effect. Meier (1981: 71–72) cites useful supporting evidence from Hamza al-Isbahani, who died prior to 970 AD. Writing in Arabic, al-Isbahani notes eight sounds found in Persian, but not Arabic. The only vowels included among these are **ō** as in **bō** (بو 'smell', CIP /bu/) and **ē** as in **sēr** (سیر 'full', CIP /sir/).

Also useful in establishing prior pronunciation norms are early Latin transcriptions of Persian. One of these is the *Codex Cumanicus* (CC), a Latin-Persian-Turkish glossary compiled in the first half of the fourteenth century in Crimea. The Persian section of the manuscript was compiled by an unnamed Franciscan monk of Italian origin (Bodrogligeti 1971: 10). Bodrogligeti normalized the transcriptions found in the CC on the basis of evidence internal and external to the codex; we provide his normalized transcriptions here, with some modifications to represent consonants in the International Phonetic Alphabet (IPA). The CC offers plentiful examples of the existence of a distinction between the *majhul* (Persian/Arabic مجهول 'unknown') vowels, **ē** and **ō**, as distinct from their *ma'ruf* (Persian/Arabic معروف 'known') counterparts, **ī** and **ū**, so named to indicate their presence or absence in Arabic.

While Pisowicz (1985: 73), based on the presence of certain conversational elements of the material in the CC, believes that the informants were native Persians, he notes other research (Monchi-Zadeh 1969: 14) suggesting that the informants were Cumans, a Turkic people of the Eurasian steppe. Be that as it may, a sample of data from the CC indicating the distinctions among these vowels, contrasted with CIP, is shown in Table 1.

Vowel	Word	CC	CIP
ē	دیر 'late'	dēr	dir
	هیچ 'nothing'	hētʃ	hitʃ
ī	کیسه 'bag'	kīsa	kise
	تقویم 'calendar'	taqwīm	taɣvim
ō	کوه 'mountain'	kō	kuh
	پوشش 'garment'	pōʃīʃ	puʃeʃ
ū	موم 'wax'	mūm	mum
	گلو 'throat'	galū	gælu

Table 1. *Majhul* and *ma'ruf* vowels in the Codex Cumanicus and Contemporary Iranian Persian

As mentioned previously, areas where Persian spread early on retain the *majhul* vowels. According to Phillott (1919: 23), "The *majhul* sounds *o* and *e* are still preserved in the Persian spoken by Afghans and Indians, but they are now unknown in Persia." Early Turkish borrowings from Persian indicate that the variety from which they were borrowed had maintained the *majhul* vowels. For example, Stein (2006: 147–148) cites *dost* 'friend' (cf. CIP دوست /dust/), *horoz* 'rooster' (cf. CIP خروس /xorus/), and *meşin* 'sheepskin leather' (cf. CIP میشن /miʃæn/). Kurmanji Kurdish is an example of a north-western Iranian language with a vowel system that maintains the *majhul* distinctions: e.g. *pêş* 'front' vs. *şîr* 'milk' and *roj* 'day' vs. *dûr* 'far' (Haig and Opengin 2012: 12–13).

2. Development of the Contemporary Iranian Vowel System

We will now explore the development of the contemporary vowels of Iranian Persian from the ENP model. Since ENP exhibits distinctions that are merged in different ways in other varieties, it is useful to view the ENP vowels as key vowels in a sense derived from Wells's (1982) description of English. For example, Wells (1982: 120) defines several lexical sets, each identified by a keyword, which behave the same way "in respect of the incidence of vowels in different accents." So words in the BATH set (consisting of words like *bath, path, staff* and *grass*) are generally pronounced with /æ/ in North American English and /ɑ/ in Southern British English, whereas words in the TRAP set (e.g. *cat, back, mass*) are pronounced with /æ/ in both dialects. Returning to Persian, one can consider a lexical set consisting of words containing ENP **ē** and consider the transformations these words have undergone in other dialects. Labov (e.g. 1994: 164–165) has employed a similar notion called "word classes", in order to facilitate comparison of dialects. Thiesen (1982: 9) developed a vowel notation that enables the reading of classical Persian poetry with

either a classical or a contemporary Iranian Persian pronunciation. As we will discuss below, such types of notation will be useful for developing a pandialectal pronunciation dictionary. Following Labov's notation for depicting vowels according to word classes, when comparing the ENP vowels to modern dialects, we place the ENP vowels in **bold**.

The ENP vowel system described above contrasts with that of CIP, as shown in Figure 2 (based on data from Majidi and Ternes 1999, Rees 2008 and Jahangiri 2000). As for the relative chronology of the changes between ENP and CIP, Windfuhr (1979: 144) states, "many questions actually have hardly been asked yet, such as those concerning the time of, and the conditions for, the lowering of the short high vowels to *e, o* ..." Pisowicz's (1985) study is a monumental effort in addressing this question, and we provide relevant examples from work cited in his study below. Note that we do not distinguish between long and short vowels in the modern system, since it appears that in contrast to the ENP system, the vowel system is currently based on quality rather than quantity (Lazard 1957, Toosarvandani 2004). Lazard (1957) characterizes the vowels derived from the ENP long vowels as "stable" and those derived from the ENP short vowels as "unstable". Compared to the stable vowels, the unstable vowels are more subject to fluctuations in quality and quantity.

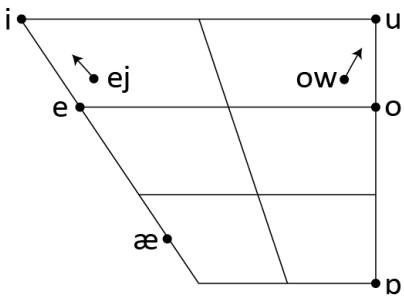

Figure 2. Vowel system of Contemporary Iranian Persian

The direction of changes between ENP long vowels and CIP is shown in Figure 3. As can be seen, all of these changes result in raising: **ē** > i, **ō** > u[2] and pre-nasally under certain linguistic, social, and stylistic conditions, **ā** > u (Modaressi-Tehrani 1978: 74–109, Perry 1996: 274, Miller 2011). In Labov's (1994: 116) vowel shifting framework, these changes accord with Principle I: in chain shifts, long vowels rise. In fact, Labov (1994: 116) notes that this is the most robust of his principles of vowel shifting. According to Labov (1994: 118), a chain shift "is a change in the position of two phonemes in which one moves away from an original position that is then occupied by the other." Below we point out where the vowel shifts that are described result in this kind of situation, although at this stage we have not resolved the relative ordering of the shifts; e.g. whether these are push or pull chains (Labov 1994: 199–200).

[2] Haig and Opengin (2012: 42–43) describe the raising of /o:/ to /u:/ in Shemzinani Kurdish, a chain shift that was set in motion once the original /u:/ had been fronted to /y/.

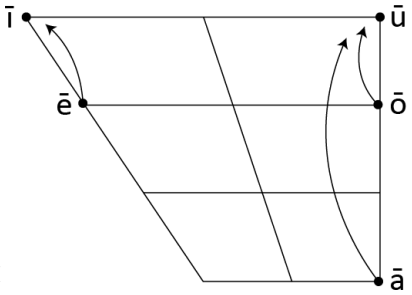

Figure 3. Long vowel raising between ENP and CIP

Note that the raising of **ē** > i and **ō** > u have resulted in mergers with CIP /i/ and /u/ which are descended from ENP **ī** and **ū**, respectively. These mergers have resulted in the following sets of words with distinct vowels in ENP and identical vowels in CIP, as shown in Table 2.

Word	ENP	CIP
سیر 'full'	**sēr**	sir
سیر 'garlic'	**sīr**	sir
شیر 'lion'	**ʃēr**	ʃir
شیر 'milk'	**ʃīr**	ʃir
رود 'river'	**rōd**	rud
بود 'was'	**būd**	bud

Table 2. Merger of ENP **ē**, **ī** and **ō**, **ū** in CIP

We will now explore the relative chronology and conditions for the shift from the ENP long vowel values to the contemporary ones. Meier's (1981: 97) analysis of relevant rhymes in Persian poetry establishes that the merger of **ō**/**ū** preceded that of **ē**/**ī**, based on the persistence of failing to rhyme **ē**/**ī** after rhymes of **ō**/**ū** had become common. On the basis of such evidence, Perry (1996: 271) claims that *majhul* **ō** disappeared by the end of the 12th century and *majhul* **ē** disappeared between the late 15th and early 17th centuries. One example cited of a poet rhyming **ō**/**ū**, but not **ē**/**ī**, is Rumi, who lived in the thirteenth century AD. Here is one of several *abyāt* (sg. *beit*, analogous to a distich or couplet) provided by Nicholson (1926: 424) of rhyming words with **ō** and **ū**:

لطف حق این شیررا و گوررا (**gōrrā**)
الف دادست این دو ضد دوررا (**durrā**)

The grace of God has given amity to this lion and wild-ass
these two far distant contraries[3]

[3] Text: Nicholson 1925–1940, Vol. 1, p. 80, line 1294, translation: Nicholson 1925–1940, Vol. 2, p. 72, line 1294.

In contrast, the following *beit* alludes to the maintenance of the **ē/ī** distinction. While acknowledging that Rumi came from Balkh (in present-day Afghanistan), a region where the *majhul* distinctions most probably held sway, Meier (1981: 97) notes that he composed his poetry in the west (i.e. Anatolia), and took advantage of the full range of variational possibilities in his work:

کار پاکان را قیاس از خود مگیر (**magīr**)
گرچه باشد در نوشتن شیر، شیر (**ʃīr**)

Do not measure the actions of holy men by (the analogy of) yourself,
though *shér* (lion) and *shír* (milk) are similar in writing[4]

As an indication that the status of **ē/ī** was in flux, however, Browne (1895: 239) offers the following *beit* as an example of Rumi's rhyming **ē** and **ī** in the same words:

آن یکی شیریست کآدم میخورد
و این یکی شیریست کآدم میخورد

That one is a *shīr* (lion) which eats man,
while this one is a *shīr* (milk) which man eats

The CC also sheds light on an earlier, non-absolute, stage in the mergers, revealing the linguistic contexts in which they first took hold. For example, while Bodrogligeti (1971: 46–47) notes that **ē** has generally been retained in the CC, he indicates that sometimes it surfaces as /ī/, especially in the verbal prefix **mē** (می) before stem-initial **ā**, as in /mīāmīzam/ 'I mix' (میامیزم). He also notes that **ō** has generally been retained, but sometimes surfaces as ū: e.g. /pūst/ 'skin' (پوست), /frāmūʃ/ 'forget' (فراموش). Finally, while noting that **ā** has generally been retained, he provides a few examples where it has raised to /ō/ before nasals: e.g. /xōm/ 'unripe' (خام), /paʃaxōna/ 'bed curtain' (پشه خانه). Bodrogligeti's (1971) presentation of the CC's snapshot of sound change in progress is notable by its reference to various phonetic and morphological environments which later became irrelevant once the mergers reached completion, such as in the case of the merger of *majhul* and *ma'ruf* vowels in CIP.

A Latin transcription of a Persian Koran (LPK) by an unnamed Spaniard in the early part of the 17th century provides a later snapshot of the state of the changes described here (Bodrogligeti 1961). Table 3 below illustrates some of the data in this work. The column labelled LPK employs the symbols used in the manuscript. The *j* is equivalent to IPA *x*, and there is usually no indication of vowel length. As can be seen, there is variation in the text between forms more similar to the those found in ENP and those found in CIP. For example, Bodrogligeti (1961) provides examples of both the preservation of *majhul* **ē** and its merger with **ī**, even for the same word. He also provides some interesting data regarding the raising of **ā** before nasals.

[4] Text: Nicholson 1925–1940, Vol. 1, p. 18, line 263; translation: Nicholson 1925–1940, Vol. 2, p. 18, line 263. The author uses an acute accent to indicate length.

Some words appear to have the intermediate target *o*, whereas others reach the modern pre-nasal raising target, *u*. The words in which this raising is exemplified show some variation with respect to modern practice. While the morpheme نامه ‘letter, book’ does undergo raising in some modern words, e.g. روزنامه ‘newspaper’ /ruznume/ (Peisikov 1960: 83), the animate plural morpheme ان ān has been claimed to rarely be subject to raising, due t o its replacement by the general plural اه hā in the colloquial language (Kahn and Bernstein 1981: 136).

Phenomenon	Word	LPK	CIP
majhul **ē** preserved	ریختن ‘pour’	rejten	rixtæn
majhul **ē** variation	کنید ‘do’ (2nd person plural)	kuned, kunid	konid
majhul **ē** > i	می (present/durative prefix)	mi	mi
3. **ā** > o,u / __ [+nasal]	نامه ‘book’	nome, nume	nɒme, ruznume ‘newspaper’ (Peisikov 1960)
ā > o / __ [+nasal]	دانا ‘knowing’	dona	dɒnɒ
ā > u / __ [+nasal]	نیکان ‘good men’	nikun	nikɒn

Table 3. Variation in LPK

The direction of changes between ENP short vowels and CIP is shown in Figure 4. Two of these changes, **i** >e and **u** > o, result in lowering.[5] This is in accordance with Labov’s (1994: 116) Principle II: in chain shifts, short vowels fall. The chain aspect here is that the /e/ and /o/ positions are vacant due to the CIP long vowel shifts described above where **ē** > i and **ō** > u. However, two changes involving ENP **a** do not adhere to Principle II. In general, **a** has raised to /æ/ in CIP, while word-finally it has raised even further to /e/ (Perry 1996: 272-273). Labov (1994: 116) notes that while this principle of short vowels falling applies to most available examples, there are exceptions.

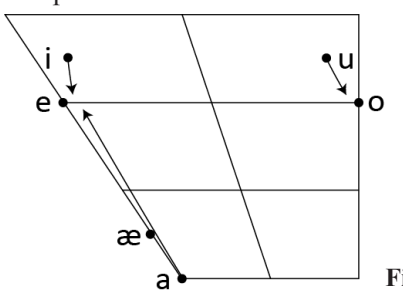

Figure 4. Short vowel changes between ENP and CIP

These changes are also foreshadowed in earlier documents. Horn (1901: 19) cites a Judeo-Persian text from Ahvaz, Khuzestan, from 1021 AD where **a** is written with the Hebrew letter *yod* to indicate a raised initial vowel in /kärd/ (כירד ‘did’), /häst/ (היסת ‘is’) and /färmān/ (פירמאן ‘order’).[6] While noting that ENP **a** was retained in

[5] Thanks to Bruce Hayes for suggesting an investigation of this phenomenon.

the majority of cases, Bodrogligeti (1971: 43) cites several examples from CC of **a** having raised to what he also transcribes with ä, which is presumably /æ/ (Peisikov 1960 uses the same symbol to transcribe the CIP short a): e.g. /kärdäm/ (کردم 'I did'), /zän/ (زن 'woman'). Regarding word-final **a**, Bodrogligeti (1971: 43) notes that /a/ is predominant, but there was raising in /ʃambä/ (شنبه 'Saturday') in CC. In the later LPK (Bodrogligeti 1961), /e/ is very common for ENP **a** in all positions, except in the environment of uvulars, as noted by Pisowicz (1985: 80); e.g. /kerdenha/ (گردن ها 'necks') vs. /katere/ (قطره 'drop').[7]

While observing that the majority of ENP **i** vowels in CC are retained, e.g. /gil/ (گل 'earth'), /dil/ (دل 'heart'), Bodrogligeti (1971: 44) notes some examples of lowering to /e/: e.g. /xedmat/ (خدمت 'service'), /āʃeq/ (عاشق 'in love'). Similarly, while reporting that generally ENP **u** is retained, e.g. /gul/ (گل 'flower'), /gurg/ (گرگ 'wolf'), he provides examples of lowering to /o/: e.g. /honarmand/ (هنرمند 'skillful'), /moʃt/ (مشت 'fist'). Similarly in LPK, **i** and **u** are generally retained, but here are lowered examples of both: /ke/ (که 'that'), /jodauenda/ (خداوندا 'O God').[8]

The direction of changes between ENP diphthongs (or VC sequences) and CIP is shown in Figure 5. The chain aspect here is that the ENP diphthongs are now occupying the space previously occupied by ENP **ē** and **ō**, which have shifted to /i/ and /u/ respectively. These changes do not adhere to Labov's (1994: 116) Principle IIA: in chain shifts, the nuclei of upgliding diphthongs fall, as exemplified in the English Great Vowel Shift where Middle English ī > Shakespearean /ej/ > modern /aj/ as in *bite* and Middle English ū > Shakespearean /ow/ > modern /aw/ as in *foul* (Jespersen 1949: 232). Labov (1994: 116–117) finds that this principle applies to a larger number of cases than Principle II (in chain shifts, short vowels fall), and ultimately restates the input to this principle as "short nuclei of upgliding diphthongs". We have no reason to believe the nuclei of these diphthongs in ENP were not short, so these facts remain in opposition to Labov's Principle IIA. However, it does not seem to be a coincidence that ENP **a** and **aj** are raised in tandem in CIP.

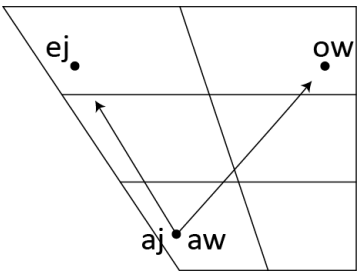

Figure 5. Diphthongs in CIP

[6] Horn's transcriptions are retained. Since Horn does not provide Judeo-Persian script, similar examples were taken from Margoliouth (1899). Thanks to Rachel Strong and Melissa Fox for help finding relevant examples in this document.

[7] Original LPK transcription retained; vowel length generally not distinguished. As Bodrogligeti (1961: 265–266) notes, the transcriber exhibits variation in the transcription of uvulars and velars, in this case we see that the transcriber has used the same symbol, k, for the velar represented by گ /g/, and the uvular represented by ق /q,ɢ/.

[8] Original LPK transcription retained; LPK j = IPA x; vowel length generally not distinguished.

Jahangiri (2000: 69–79) discusses the monophthongization of both /ej/ -> [e] and /ow/ -> [o] in contemporary Tehrani Persian. He finds that monophthongization is more common with /ow/ than /ej/, and that the more formal the style, and the higher the social group, the less the monophthongization encountered. Jahangiri mentions a further process among working class Tehranis whereby /ow/ is monophthongized to [o] and then raised to [u], e.g. [ʃuhar] for standard CIP /ʃowhar/ (شوهر 'husband'). He stresses that this process only occurs to [o]'s resulting from monophthongization and not to [o]'s in general. Given that it would be reasonable to assume that [o]'s deriving from the diphthong /ow/ would be fairly long to start with (while Jahangiri discusses the lengthening resulting from monophthongization of /ej/ he does not mention it with respect to /ow/), this raising to [u] fits with Principle I, whereby long vowels rise. Its failure to apply to "short" [o], or those that have not resulted from monophthongization, fits with Principle II, which expects short vowels to fall rather than rise.

3. Afghan Persian

The standard Kaboli Afghan Persian, or Dari, vowel system is generally the same as that of ENP, though it is not clear whether the short vowels have retained their original quality. In Table 4, we compare the qualities assigned by several studies, adding phonemic equivalents using IPA symbols when the authors provide foreign example words. We see that the majority of transcription variation is in the mid vowels.

ENP	Farhadi (1955) French example	Neghat (1993) English example	Rees (2008) acoustic study	Kiseleva and Mikolaichik (1978)	Henderson (1972)
a	*salle* a	*up* ʌ, ə[9]	a	a	æ, a
i	*geste* ɛ	*sit* ɪ	ɛ, ɪ	e	ɛ
u	*poste* ɔ	*put* ʊ	ʊ	o	ʊ
ā	Swedish *dag* ɒ[10]	*fall* ɒ[11]	ɑ, ɔ	â	ɔ, ɑ
ē	*maire* ɛː[12]	*pay* e	e	ê	e
ī	*dit* i	*see* i	i	i	i
ō	*taux* o	*note* o	o	ô	o
ū	u	*moon* u	u	u	u

Table 4. Transcription of Dari vowels

Farhadi (1955: 8) observes that short /a/ retains its quality word-finally, and does not raise to /e/ as in CIP. However, he notes several words where ENP **a** does cor-

[9] Cf. Ladefoged (1982: 28–30).
[10] Farhadi's choice of a Swedish example is interesting. French traditionally has a distinction between *pâte* /pɑt/ and *patte* /pat/, however this distinction is becoming less reliable in metropolitan France. In addition to providing a surer exemplar of a back /ɑ/, Farhadi probably sought to indicate rounding. According to Haugen (2009), "Swedish *a* has moved closer to *å*, being backed and rounded."
[11] American English *fall* has a range of pronunciations, including /ɑ/, /ɒ/ and /ɔ/.
[12] Following /r/ serves as a *consonne allongeante*, lengthening the preceding vowel (Walker 1984: 26).

respond to Dari /ɛ/: آتش ‘fire’ /ātɛʃ/, چشم ‘eye’ /tʃɛʃm/, دست ‘hand’ /dɛst/. Also, in Dari, **ā** does not raise to u before nasals. Finally, in contrast to CIP, majhul **ē** and **ō** are preserved, as shown in Table 5. However, Farhadi (1975: 15) notes the word /ʃirdān/ ‘faucet’, literally ‘lion mouth’, exhibiting the CIP merger, which he ascribes to Iranian influence.

Word	Afghan pronunciation	CIP pronunciation
نیست ‘is not’	/nest/	/nist/
گیتی ‘universe’	/geti/	/giti/
ابرو ‘eyebrow’	/abro/	/æbru/
روز ‘day’	/roz/	/ruz/

Table 5. Preservation of **ē** and **ō** in Dari

The diphthongs **aj** and **aw** are preserved in Dari, e.g. پیدا ‘find’ /pajdā/, چلو ‘rice’ /tʃalaw/, though Farhadi (1955: 8–9) notes a set of words where the spoken language uses /ē/ and the “traditional educated pronunciation” uses both /ē/ and /aj/: e.g. امید ‘hope’ /omed, omajd/ and پریشان ‘distressed’ /pɛreʃɒn, pɛrajʃɒn/. Similarly, he notes variation between /o/ and /aw/: e.g. روشن ‘clear’ /roʃan, rawʃan/, روقن ‘oil’ /roqan, rawqan/. This variability, combined with examples like /ʃirdān/, indicates that the Dari vowel system may be in motion along the lines established earlier for CIP.

4. Tajik Persian

The Tajik vowel system, depicted in Figure 6, represents divergences from the ENP system in ways that are different from CIP (Perry 2005). The majhul **ē** and **ō** are preserved as /e/ and /u̇/ (for this symbol, see below) respectively, but ENP short **i** and long **ī** have merged as /i/, and ENP short **u** and long **ū** have merged as /u/. Tajik /o/ represents a raised and rounded ENP **ā** in all positions, in contrast to the socially and stylistically gradient ā and u of CIP (Modaressi-Tehrani 1978: 74–109, Perry 1996: 274, Miller 2011). Just as a qualitative system replaced a quantitative one in CIP, a different set of mergers has resulted in a qualitative system in Tajik.

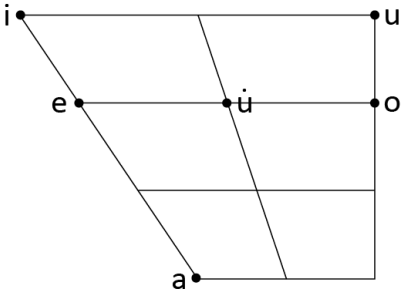

Figure 6. Tajik vowel system

Perry notes that Tajik /u̇/ is also used for borrowings from Uzbek (in turn derived from common Turkic ö and ü). The vowel chart placement and symbol for this vowel are taken from Perry, who reprises a symbol used by Lazard (1956). Accord-

ing to Perry (2005: 18), /u̇/ "lies phonetically between [u] and [y] ... higher than French *peu* (/ø/) ... less rounded and more lax than /u/." Lazard (1956: 126) simply says it is lower and fronter than /u/. In any case, it is clear that it is fronted with respect to **ō**, which is our first example of Labov's (1994: 116) Principle III: in chain shifts, back vowels move to the front.

In other Persian dialects we see fronting as well. For example Okati et al. (2009) report on **ū** fronting to [ʉ] and [ʏ] in Iranian Sistani. Miller and Moats (2011) report on data from a Herati (Afghanistan) speaker with a merger between **ū** and **ō** (cf. Farhadi 1955: 10 and Ioannesyan 2007: 268) realized as [y] as shown in Table 6. LeCoq (1989: 250) mentions other dialects where **ū** fronting is attested, including unrounding to /i/ in Semnani, e.g. /pil/ 'money', and Mazanderani, e.g. /dir/ 'far'. Haig and Opengin (2012: 13, 42) cite Kurdish dialects where the cognate sound of **ū** is fronted to [øː] or [yː] (Shemzinani) and others where it additionally unrounds to [iː] (Badinani in northern Iraq and south-eastern dialects of Kurmanji in Turkey). In apparent contrast to a view whereby fronted variants derived from an original /u/, Bodrogligeti (1961: 267) ascribes to Németh the supposition that ENP **u** had the quality [y], based on the presence of ü in Persian words borrowed into Turkish.

ENP vowel	Herati
ū	بودیم /bydim/ 'we were'
ō	دوست /dyst/ 'friend'
ō	فراموش /farāmyʃ/ 'forget'

Table 6. Herati fronting

5. Synthesis

The vowel transformations between ENP and the various contemporary dialects we have discussed so far could be encoded in a pandialectal pronunciation key as shown in Table 7. Due to the preliminary nature of this aspect of the study, we have employed a mixture of IPA and traditional symbols, and in some cases, we have included a set of possible symbols encountered in the literature. Each word's pronunciation could be provided underlyingly in ENP, and the various contemporary dialects' pronunciations could be generated automatically by a rule-based system effecting the transformations indicated in the key, and then presented to users in one or more of the contemporary dialects in an electronic or online dictionary. We have not discussed consonantal variation, but a similar approach could be taken to account for the presence for example of /v/ in CIP and /w/ in Dari.

Of course, such a presentation ignores the fact that there are more complicated divergences among the dialects. For example, there are idiosyncratic differences between Dari and CIP, such as خنک 'cool', pronounced /xonak/ in CIP and /xʊnʊk/ in Dari. While the first vowel could be represented with ENP **u**, the variation in the second vowel cannot properly be derived using the method described above. Another example is the word عوض 'exchange', which is pronounced /avaz/ in CIP and

ENP	CIP	Kaboli	Herati[13]	Tajik
ū	u	ū	y, u	u
ō	u	ō	y, ō, u	u̓
ā	ɒ, u (before nasals under certain conditions)	ɒ	ɒ, u (before and after nasals), a (unstressed)	o
ē	i	ē	ē, i	e
ī	i	ī	i	i
a	æ, e (word-finally)	a	æ, ɛ (word-finally)	a
i	e, ɛ	ɛ, ɪ	ɪ, e, ɛ	i
u	o	ʊ, o	u	u

Table 7. Rudiments of a pandialectal pronunciation key for Persian

/ēwaz/ in Dari. In this case, Farhadi (1955: 20) indicates that there is compensatory lengthening in Dari resulting from the non-pronunciation of ع.

6. Conclusion

We have attempted to show the utility of the ENP vowel system in understanding the variation among the vowel systems of several diachronically and synchronically separated varieties of Persian which are derived from it. The vowel changes have been considered with respect to Labov's (1994) principles of vowel shifting, and they appear both to corroborate his findings and to provide some alternative directions of change in particular cases. Foreign-language descriptions of Persian in different periods have been adduced to capture sound change in progress in cases where the contemporary situation exhibits sound changes that have reached completion. Finally, a practical way of using this knowledge to provide learners and scholars with a dynamic picture of Persian vowels in a number of distinct varieties has been suggested in the form of a pandialectal pronunciation dictionary. In the future, we hope by means of a pandialectal acoustic study to shed light on the exact quality of vowels, as well as the appropriate symbols to use to describe them when viewed from a holistic Persian perspective.

Abbreviations

CC: *Codex Cumanicus*
CIP: Contemporary Iranian Persian
ENP: Early New Persian
IPA: International Phonetic Alphabet
LPK: Latin transcription of a Persian Koran

References

Bodrogligeti, A. (1961). "The Persian Translation of the Koran in Latin Letters." *Acta Orientalia Academiae Scientiarum Hungaricae* 13, pp. 261–276.

Bodrogligeti, A. (1971). *The Persian Vocabulary of the Codex Cumanicus*. Budapest: Akadémiai Kiadó.

Browne, Edward G. (1895). "Review of *A Grammar of the Persian Language*, by John T. Platts." *Journal of the Royal Asiatic Society* 27(1), pp. 235–245.

Farhadi, Abd-ul-Ghafur (1955). *Le Persan parlé en Afghanistan*. Paris: Klincksieck.

Farhadi, A. G. Rawan (1975). *The Spoken Dari of Afghanistan*. Kabul: Peace Corps.

Haig, Geoffrey, and Öpengin, Ergin (2012). "Kurmanji Kurdish in Turkey: Structure, Varieties and Status" (unpublished manuscript).

Haugen, Einar (2009). "Danish, Norwegian and Swedish." In: Comrie, Bernard (ed.), *The World's Major Languages* (2nd edition). London: Routledge, pp. 124–144.

Henderson, Michael M. T. (1972). *Dari (Kabul Persian) Phonology*. PhD dissertation, University of Wisconsin.

Horn, Paul (1901). "Neupersische Schriftsprache." In: Geiger, Wilhelm and Kuhn, Ernst (eds), *Grundriss der iranischen Philologie* 1:2. Strassburg: Trübner, pp. 1–200.

Ioannesyan, Youli (2007). "Situating the Khorasani Dialects within the Persian-Dari-Tajiki Dialect Continuum." In: Lewis, Franklin, Sharma, Sunil, and Moayyad, Heshmat (eds), *The Necklace of the Pleiades*. Amsterdam: Rozenberg, pp. 267–278.

Ioannesyan, Youli (2009). *Afghan Folktales from Herat*. Amherst, New York: Cambria Press.

Jahani, Carina, and Korn, Agnes (2009). "Balochi." In: Windfuhr, Gernot L. (ed.), *The Iranian Languages*. London: Routledge, pp. 634–692.

Jahangiri, Nader (2000). *A Sociolinguistic Study of Persian in Tehran*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Jespersen, Otto (1949). *A Modern English Grammar on Historical Principles*. London: George Allen and Unwin.

Kahn, Margaret, and Bernstein, Jared (1981). "Progress on the uN Change in Persian: Dæbestan vs. Tabestun." *Acta Orientalia Academiae Scientiarum Hungaricae,* 35(1), pp. 133–140.

Kiseleva, L., and Mikolaichik, V. (1978). *Dari-Russkii Slovar*. Moscow.

Labov, William (1994). *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.

Ladefoged, Peter (1982). *A Course in Phonetics*, Second Edition. New York: Harcourt Brace Jovanovich.

Lazard, Gilbert (1956). "Caractères distinctifs de la langue Tadjik."′ *Bulletin de la Société de Linguistique de Paris* 52(1), pp. 117–186.

Lazard, Gilbert (1957). *Grammaire du Persan Contemporain*. Paris: Klincksieck.

LeCoq, Pierre (1989). "Le classement des langues irano-aryennes occidentales." In: de Fouchécour, Charles-Henri and Gignoux, Philippe (eds), *Études irano-aryennes offertes à Gilbert Lazard* [Studia Iranica Cahier 7]. Paris: Association pour l'avancement des études iraniennes, pp. 247–264.

Majidi, Mohammad-Reza, and Ternes, Elmar (1999). "Persian." In: *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press, pp. 124–125.

Margoliouth, D. S. (1899). "A Jewish-Persian Law Report." *The Jewish Quarterly Review* 11(4), pp. 671–675.

Meier, Fritz (1981). "Aussprachefragen des älteren Neupersisch." *Oriens,* 27–28, pp. 70–176.

Miller, Corey (2011). "A Holistic Treatment of /ān/ to [un] in Persian." *Proceedings of ICPhS XVII*, Hong Kong, pp. 1386–1389.

Miller, Corey, and Moats, Najiba (2011). "Convergence on [u] across Persian Dialects." Poster presented at International Conference on Iranian Linguistics 4, Uppsala, Sweden.

Modaressi-Tehrani, Yahya (1978). *A Sociolinguistic Analysis of Modern Persian*. Ph. D. dissertation, University of Kansas.

Monchi-Zadeh, Davoud (1969). *Das Persische im Codex Cumanicus* [Studia Indoeuropaea Upsaliensia, 1]. Uppsala: Acta Universitatis Upsaliensis.

Neghat, Muhammad Nassim (1993). *Dari-English Dictionary*. Omaha: University of Nebraska.

Nicholson, Reynold A. (1926). "On Editing the *Mathnawi*." *Bulletin of the School of Oriental and African Studies*, 4(2), pp. 421–430.

Nicholson, Reynold A. (1925–1940). *The Mathnawi of Jalálu'ddín Rúmí.* London: Brill.

Okati, Farideh, Ahangar, Ali Abbas, and Jahani, Carina (2009). "Fronting of /u/ in Iranian Sistani." *Orientalia Suecana*, 58, pp. 120–131.

Okati, Farideh (2012). *The Vowel Systems of Five Iranian Balochi Dialects* [Studia Linguistica Upsaliensia, 13, Studia Iranica Upsaliensia, 20]. Uppsala: Acta Universitatis Upsaliensis.

Paul, Ludwig (2002). "A Linguist's Fresh View on Classical Persian." *Cahiers de Studia Iranica,* 26, pp. 21–34.

Peisikov, L. S. (1960). *Tegeranskii Dialekt*. Moscow.

Perry, John R. (1996). "Persian in the Safavid Period: Sketch for an *Etat de Langue*." In: Melville, Charles (ed.), *Safavid Persia: The History and Politics of an Islamic Society*. London: I.B. Tauris & Company, pp. 269–283.

Perry, John R. (2005). *A Tajik Persian Reference Grammar*. Leiden: Brill.

Phillott, D. C. (1919). *Higher Persian Grammar*. Calcutta: Baptist Mission Press.

Pisowicz, Andrzej (1985). *Origins of the New and Middle Persian Phonological Systems*. Krakow: Jagiellonian University.

Rees, Daniel A. (2008). *Towards Proto-Persian: An Optimality Theoretic Historical Reconstruction*. PhD dissertation, Georgetown University.

Stein, Heidi (2006). "Palatal-Velar Vocalism of Arabic-Persian Loanwords in 16th-Century Ottoman Turkish." In: Johanson, Lars and Bulut, Christiane (eds), *Turkic-Iranian Contact Areas: Historical and Linguistic Aspects*. Wiesbaden: Harrassowitz, pp. 143–157.

Thiesen, Finn (1982). *A Manual of Classical Persian Prosody*. Wiesbaden: Harrassowitz.

Toosarvandani, Maziar Doustdar (2004). "Vowel Length in Modern Farsi." *Journal of the Royal Asiatic Society*, Third Series 14(3), pp. 241–251.

Walker, Douglas C. (1984). *The Pronunciation of Canadian French*. Ottawa: University of Ottawa Press.

Wells, J. C. (1982). *Accents of English 1*. Cambridge: Cambridge University Press.

Windfuhr, Gernot (1979). *Persian Grammar: History and State of its Study*. The Hague, Paris and New York: Mouton.

Windfuhr, Gernot (2009), "Persian." In: Comrie, Bernard (ed.), *The World's Major Languages* (2nd edition). London: Routledge, pp. 445–459.

Windfuhr, Gernot, and Perry, John (2009). "Persian and Tajik." In: Windfuhr, Gernot (ed.), *The Iranian Languages*. London: Routledge, pp. 416–544.