ORIGINAL ARTICLE

Open Access

From non-uniqueness to the best solution in phonemic analysis: evidence from Chengdu Chinese

San Duanmu

Correspondence:
duanmu@umich.edu
University of Michigan, Ann Arbor,
USA

Abstract

The "non-uniqueness" theory assumes that there is no best solution in phonemic analysis; rather, competing solutions can co-exist, each having its own advantages (Chao, Bulletin of the Institute of History and Philology 4: 363–398, 1934). The theory is based on the assumption that there is no common set of criteria to evaluate alternative solutions. I argue instead that such a set of criteria can be established and it is possible to find the best solution. The criteria include riming properties, rime structure, constraints on syllable gaps, phonemic economy, phonetics, syllable sizes, and feature theory. I illustrate the proposal with Chengdu. Four analyses are compared, the "CGV" segmentation, the "CV" segmentation, the "finest" segmentation, and the "CVX" segmentation, and CVX is shown to be the best.

Keywords: Phonemic analysis, Non-uniqueness, Phone segmentation, Syllable gaps, Syllable structure, CVX analysis, Phonological neighborhood density, Chengdu Chinese

1 Introduction: phonemic analysis and the non-uniqueness theory

Phonemic analysis is the foundation of phonology. According to Goldsmith (2011: 193), phonemic analysis has been the "greatest achievement" and "the beginning of all work" in phonology.

Hockett (1960: 90) proposes that a defining property of human language is the use of two coding systems (the duality of patterning): (i) sentences are made of words (or morphemes) and (ii) words are made of phonemes (consonants and vowels). On this view, the first job in phonemic analysis is to segment words into phones, which are then grouped into phonemes. However, there is no agreement on the granularity of segmentation, and a common view is that it can vary from language to language. For example, Chao (1934: 371) suggests that aspirated stops and affricates, such as [tʰ ts tsʰ], need not be segmented further, so that they are each a single sound, or we can segment them into two or three parts each, so that they contain seven sounds [t+h t+s t+s+h], where + indicates a segmentation boundary. Pike (1947a) offers a similar view. Given such options, it is hard to decide what the proper granularity of segmentation ought to be, even for well-known languages. For example, for Swadesh (1935: 149), English diphthongs are single phonemes, but for Trager and Bloch (1941: 234) and Pike (1947b: 151), each English diphthong is made of two phonemes. Similarly, for Wiese (1996), the German affricates [pf ts tʃ dʒ] are four phonemes and

no further segmentation is needed, but for Kohler (1999), they are two phonemes each and should be segmented as [p+f t+s t+ʃ d+ʒ].

Aware of such ambiguity, Chao (1934) argues that there is no best solution in phonemic analysis. Instead, phonemic analysis serves multiple functions, and each function may favor a different solution. In other words, there is no unique solution that serves all functions, hence the "non-uniqueness" theory. Chao's proposal is illustrated by the two approaches in 1, each assuming a different granularity of segmentation. (Here and below, IPA symbols are placed in square brackets throughout).

(1)   Illustration of the non-uniqueness theory of phonemic analysis: competing solutions may co-exist
      a. "Fine" segmentation of [pa ta ka ha sa pha tha kha tsa tsha]
         Segmentation: [p+a t+a k+a h+a s+a p+h+a t+h+a k+h+a t+s+a t+s+h+a]
         Phonemes: [p t k h s a]
         Syllable: CV, CCV, CCCV
         Property: fewer phonemes; more syllable types/sizes
      b. "Coarse" segmentation of [pa ta ka ha sa pha tha kha tsa tsha]
         Segmentation: [p+a t+a k+a h+a s+a ph+a th+a kh+a ts+a tsh+a]
         Phonemes: [p t k h s ph th kh ts tsh]
         Syllable: CV
         Property: more phonemes; simpler syllable structure/size

In general, finer segmentation yields fewer phonemes, but more complex syllable structure, whereas coarser segmentation yields more phonemes and simpler syllable structure. On the non-uniqueness view, each analysis has its own advantage, each has its own shortcoming, and there is no best solution.

Ao (1992) proposes that the ambiguity in segmentation can be solved by reference to morphophonemic alternation. For example, the plural suffix in English creates the alternation between *cat-cats*; therefore, [ts] ought to be split into two sounds. In contrast, the onset [ts] in Chinese is not based on morphology, and so it should not be split. There are questions for Ao's proposal though. First, in German, the onset [ts] is not based on morphology, and word-final [ts] is sometimes based on morphology and sometimes not. Should all [ts] be split in German? Should initial [ts] be kept intact and final [ts] be split? Or should final [ts] be split only if there is a morphological boundary in the middle? Similarly, in Beijing Chinese, the "diminutive" suffix creates 袋-袋儿 [tai]-[taɚ] 'bag' and so [ai] ought to be split to [a+i]. However, in Chengdu Chinese, the diminutive alternation creates 袋-袋儿 [tai]-[tɚ] 'bag' instead. Should [ai] be split in Chengdu? Clearly, morphophonemic alternation alone is insufficient to solve the non-uniqueness problem.

Given the non-uniqueness theory, all sorts of phonemic solutions become legitimate, and people rarely attempt to find out whether some are demonstrably better than others. An unwanted result is that phonemic analysis seems rather arbitrary and not very useful. Thus, the dominant tradition in phonological descriptions of Chinese dialects still focuses on the syllable, with an inventory of onsets, rimes, and tones, but not an inventory of phonemes. Similarly, some Western scholars begin to doubt the methodology of phonemic analysis, or the concept of the phoneme itself. For example, Ladefoged (2001: 170) says that "consonants

and vowels are largely figments of our good scientific imaginations". Similarly, Fowler (2015: 40) says that the basic units of phonology are not phonemes but articulatory gestures.

I shall argue that there is a fundamental flaw in the non-uniqueness theory, which is the assumption that there is no common set of criteria that all phonemic solutions can be measured against. I shall show that such a set of criteria can be established and it is possible to determine the best solution. To illustrate the proposal, I offer an in-depth analysis of Chengdu Chinese, whose phonemic analysis has not received much attention. In Section 2, I outline four possible phonemic analyses of Chengdu. In Section 3, I propose a common set of criteria. In Section 4, I evaluate the four analyses and show that the best solution is identifiable. In Section 5, I discuss two additional issues of interest: variation in syllable duration and reaction time in lexical access. In Section 6, I offer concluding remarks.

2 Four possible phonemic analyses of Chengdu

The data on Chengdu are based on He and Rao 何婉, 饶冬梅 (2014). Differences between this source and other ones are small and have little impact on our discussion. Chengdu has an inventory of 326 syllables (excluding tones and some 40 derived syllables with the diminutive suffix [ɚ]). Based on how the syllables are segmented, there are four possible analyses, which I shall refer to as the "CGV" segmentation, the "CV" segmentation, the "finest" segmentation, and the "CVX" segmentation. For ease of discussion, I shall exclude tones, the "zero initial" (the lack of initial consonant), and the syllabic consonant [z̩] (traditionally transcribed as [ɿ]), which need not concern us here.

2.1 The "CGV" segmentation

You et al. 游汝杰等 (1980) proposes that the phonemic analysis of Chinese need not follow the Western tradition, which segments words or syllables into consonants and vowels. Instead, Chinese syllables should be segmented into the traditional components, which are initials, medials, rimes, and tones, where the initial is the consonant (C) before the medial, the "medial" is a pre-nuclear glide (G), and the rime is the part after the medial (denoted here as V), hence the CGV analysis. In addition, You et al. 游汝杰等 consider such components to be phonemes, even though further segmentation seems possible; for example, the rime [an] is a "rime phoneme", even though it could be split into [a]-[n]. You et al. 游汝杰等 do not discuss Chengdu, but their proposal can be extended to it, shown in 2. The three medial glides [i u y] are the same as those in the rime inventory and are not separately listed. The transcription of He and Rao 何婉, 饶冬梅 (2014) already includes some phonological adjustment. For example, the vowels in [an] and [aŋ] are phonetically [æ] and [ɑ]. I have followed their transcription. Finally, since [ə] does not occur as a monophthong, I have interpreted [ən] as [en] and [əu] as [ou].

(2) The CGV analysis of "phonemes" in Chengdu (omitting tones)

19 "initial phonemes"	[p pʰ m f t tʰ n ȵ ts tsʰ s z k kʰ ŋ x tɕ tɕʰ ɕ]
16 "rime phonemes"	[i u y a o e ai ei au ou an en aŋ oŋ in ɚ]
Maximal syllable	CGV (initial-medial-rime)
Example	穿 [tsʰ][u][an] 'wear', 吹 [tsʰ][u][ei] 'blow'

In the Western approach, the rimes [an en aŋ oŋ in] would perhaps always be segmented into two phonemes each. However, under the non-uniqueness theory, they do not need to.

2.2 The "CV" segmentation (Lee and Zee 2003)

Lee and Zee (2003) follow the Western tradition and segment syllables along consonant-vowel (or vowel-consonant) boundaries. I have extended their analysis of Putonghua to Chengdu and the result is shown in 3.

(3)　The CV analysis of Chengdu phoneme (tones omitted)

19 consonants	[p pʰ m f t tʰ n ɳ ts tsʰ s z k kʰ ŋ x tɕ tɕʰ ɕ]
23 vowels	[i u y a o e ɚ] [ai ei au ou *ia io* ie ua ue ye ya [iai iau iou uai uei]
Maximal syllable	CVC
Example	穿 [tsʰ][ua][n] 'wear', 吹 [tsʰ][uei] 'blow'

Although Lee and Zee always segmented at C-V and V-C boundaries, they kept diphthongs and triphthongs intact.

2.3 The "finest" segmentation

No one has proposed a finest segmentation of Chinese (or any other language). However, since the non-uniqueness theory allows us to split aspirated consonants, affricates, diphthongs, and triphthongs (Chao 1934; Pike 1947a), the finest segmentation is theoretically possible. The analysis is shown in 4, where [h x] are treated as allophones.

(4)　Chengdu phoneme based on the finest segmentation (tones omitted)

12 consonants	[p m f t ɳ k s z ɕ ŋ h]
7 vowels	[i u y a e o ɚ]
Maximal syllable	CCCVVC, CCCVVV
Examples	穿 [t][s][h][u][a][n] 'wear' 吹 [t][s][h][u][e][i] 'blow'

A major feature of the finest segmentation is that it has the smallest number of phonemes, while it has the largest syllable size. For example, its consonant inventory is a little over half of that in the CV analysis and its vowel inventory is less than a third of the latter. On the other hand, it maximal syllable is twice the size as that in the CV analysis.

2.4 The "CVX" segmentation

The CVX segmentation is based on Duanmu (1990, 2007) and Ao (1992), according to which the maximal Chinese syllable has three positions, where C is the onset (including the medial glide) and VX is either VV (a diphthong or long vowel) or VC (a short monophthong and a consonant). The analysis of Chengdu is shown in 5. For reasons given

in Ao (1992) and Duanmu (2007), I have transcribed [tɕᶣ tɕʰᶣ ɕᶣ] as [tɕʷ tɕʰʷ ɕʷ], which does not change the total number of consonants. In addition, since there is no contrast between [tɕʲ tɕʰʲ ɕʲ ŋʲ] and [tɕ tɕʰ ɕ ŋ], I have used the latter. The only case of [nᶣ] is 略 [nᶣe] 'slightly' (He and Rao 何婉, 饶冬梅 2014: 16), which can be pronounced as [nʲo] as well (He and Rao 何婉, 饶冬梅 2014: 114).

(5)   The CVX analysis of phonemes in Chengdu (tones omitted)

39 consonants	[p pʰ m f t tʰ n ŋ̩ ts tsʰ s z k kʰ ŋ x tɕ tɕʰ ɕ] [tʷ tʰʷ nʷ tsʷ tsʰʷ sʷ zʷ kʷ kʰʷ xʷ tɕʷ tɕʰʷ ɕʷ] [pʲ pʰʲ mʲ tʲ tʰʲ nʲ] [nᶣ]
7 vowels	[i u y a e o ɚ]
Maximal syllable	CW, CVC
Examples	穿 [tsʰʷ][a][n] 'wear', 吹 [tsʰʷ][e][i] 'blow'

According to Ao (1992), a consonant-glide unit, such as [tʷ], is a single phoneme both underlyingly and at the surface. According to Duanmu (1990, 2007), a consonant-glide unit is made of two phonemes underlyingly (e.g., [tʷ] is [t]+[u]), which merge into one sound at the surface. Therefore, the analysis above is more similar to Ao's than to Duanmu's.

2.5 Summary

It can be seen that the four analyses just outlined differ a lot. Nevertheless, the non-uniqueness theory considers all of them to be legitimate, because each seems to have its own advantage. For example, the "finest" analysis seems to have the best phonemic economy (although it assumes an oversized syllable structure), and the CVX analysis seems to have the most consistent syllable structure (although it has the largest consonant inventory). There seems to be no consistent set of criteria that all approaches can be measured against.

The treatment of [tɕʲ tɕʰʲ ɕʲ ŋʲ] as [tɕ tɕʰ ɕ ŋ] is worth an additional comment. Since there is no contrast between [tɕʲan tɕʰʲan ɕʲan ŋʲan] and [tɕan tɕʰan ɕan ŋan], there is no reason to use the former transcription in any of the analyses (Ao 1992; Duanmu 1990). As a result, the palatal series are in contrast with the dental series and the velar series, exemplified in 6.

(6)   Minimal contrast among palatal, dental, and velar series in Chengdu

Onset	Palatal	Dental	Velar
Unaspirated	[tɕan] 尖 *jiān* 'sharp'	[tsan] 赞 *zàn* 'praise'	[kan] 干 *gān* 'dry'
Aspirated	[tɕʰan] 千 *qiān* 'thousand'	[tsʰan] 参 *cān* 'join'	[kʰan] 刊 *kān* 'publish'
Fricative	[ɕan] 仙 *xiān* 'angel'	[san] 三 *sān* 'three'	[xan] 寒 *hán* 'cold'
Nasal	[ŋan] 年 *nián* 'year'	[nan] 男 *nán* 'male'	[ŋan] 安 *ān* 'peace'

In Chao (1934), the palatal series is thought to be in complementary distribution with both the dentals series and the velar series, and it is thought to be impossible to determine whether we should (i) group the palatal series with the dental series, (ii) group the palatal series the velar series, or (iii) not group the palatal series with either.

Our representation removes this apparent uncertainty in the phonemic analysis of Chinese. In addition, there is no need to derive [tɕ tɕʰ ɕ] from [tsj tsʰj sj], as proposed by Duanmu (2007).

3 A set of criteria for evaluating phonemic analysis

Phonemic analysis is only part of phonology. Therefore, it is reasonable to expect phonemic analysis to facilitate the analysis of other parts of phonology, or to help explain properties in other parts of phonology. In addition, other things being equal, a phonemic analysis that makes better phonetic predictions is better than one that does not. Based on such considerations, I propose a set of criteria that every phonemic analysis should be measured against.

3.1 Riming property

Chao 赵元任 (1923: 6–7) offers a detailed discussion of whether two syllables rime or not. They are rephrased in 7, where the onset is made of CG (the initial consonant and the medial glide, if any) and the rime is the rest of the syllable (VX, including tone).

(7)   Conditions for two Chinese syllables A and B to rime in Chinese
      a. A and B have the same rime (i.e., same VX).
      b. A and B have different onsets (i.e., different CG).

According to 7, [man3]-[jan3]-[lwan3] (as in 满 *mǎn* 'full', 演 *yǎn* 'perform', and 卵 *luǎn* 'egg', where 3 indicates the third tone in Putonghua) rime with each other. On the other hand, identical syllables such as [wa1]-[wa1] (as in 蛙 *wā* 'frog' and 挖 *wā* 'dig', where 1 indicates the first tone in Putonghua) are not the best riming pairs, at least not in formal poetry.

It can be seen that the riming conditions are hard to state in the CV analysis of phonemes (Lee and Zee 2003). Some examples are shown in 8.

(8)   Difficulty in defining riming pairs by the CV analysis of phonemes

Riming pair	CV representation	CV description
[l][an]-[lu][an]	[l][a][n]-[l][ua][n]	Same initial C; different VC
[m][an]-[l][an]	[m][a][n]-[l][a][n]	Different initial C; same VC
[m][an]-[lu][an]	[m][a][n]-[l][ua][n]	Different initial C; different VC
[m][ai]-[ku][ai]	[m][ai]-[k][uai]	Different initial C; different V

In the CV analysis, there is no consistent way to state what makes two syllables rime, unless reference is made to part of a diphthong or triphthongs, in which case we are in effect saying that diphthongs and triphthongs ought to be segmented.

3.2 Rime structure

Rime properties in Chinese (including Chengdu) are summarized in 9.

(9) Rime structure in Chinese (excluding medial glide)
  a. The rime can be VV (diphthong or long vowel) or VC (short vowel plus C).
  b. The rime cannot be VVC.
  c. The duration of an unstressed syllable is about half that of a stressed syllable.
  d. Unstressed syllables lack diphthongs or a consonant coda.

Examples of 9a are [ai au an in aː] and so on. 9b accounts for the lack of such rimes as [ain aun]. 9c is supported by experimental evidence, such as Lin and Yan 林茂灿, 颜景助 (1980) and Zhu (1995). Examples of 9d are shown in 10, from Gao and Shi 高名凯, 石安石 (1963).

(10) Rime reduction in unstressed syllables in Putonghua (relevant change in boldface)

Reduction	Example
[ai] ➔ [e]	[naut**ai**] ➔ [naud**e**] 脑袋 'head'
[ou] ➔ [o]	[muːtʰ**ou**] ➔ [muːtʰ**o**] 木头 'wood'
[ən] ➔ [ə̃]	[ɕanʂ**ən**] ➔ [ɕanʐ**ə̃**] 先生 'Mr.'

The simplest explanation of 9 is that the Chinese rime is measured by the number of positions it has: a light rime has one position and other rimes have two, shown in 11.

(11) Analysis of Chinese rimes
  a. An unstressed (light) rime has one position.
  b. Other rimes have two positions.
  c. Each rime position can hold one phoneme.
  d. A consonant or a short vowel is one phoneme.
  e. A diphthong (or a long vowel) takes up two rime positions.

A diphthong plus a consonant will take up three rime positions, yet a regular rime only has two positions. Therefore, Chinese lacks such rimes as [ain aun] (except in final position in some dialects, where the rime is lengthened). In addition, a regular CV syllable in fact has a long vowel, such as 妈 [maː] 'mother'.

It can be seen that the generalizations in 11 are easy to state if regular rimes are segmented into two units. In contrast, in the CGV or the CV segmentations, properties of rime structure are hard to generalize.

3.3 Syllable gaps

Syllable gaps refer to those syllables that seem to have the expected phoneme combinations and fall within the expected syllable size yet are not found to occur in the language. For example, 猫 [mau] 'cat', 烧 [sau] 'burn', and 好 [xau] 'good' occur in Chengdu, but [fau] does not, even though all of them are CVV.

Syllable gaps in Chinese, regardless of the dialect, are surprisingly numerous, especially in view of the fact that the syllable inventory in Chinese is quite small. Syllable gaps in Chengdu are shown in 12. To simplify the discussion, I have omitted tones and the retroflex vowel [ɚ].

(12) Syllable gaps in Chengdu

Maximal size	CGVX
C = 20	[p pʰ m f t tʰ n ɳ ts tsʰ s z k kʰ ŋ x tɕ tɕʰ ɕ 0]
G = 4	[i u y 0]
V = 6	[i u y a e o]
X = 5	[i u n ŋ 0]
All combinations	20 × 4 × 6 × 5 = 2400
Occurring	326
Gaps	2074 (86.4%)

If we consider the onset to be made of a consonant C and a glide G, the maximal syllable size is CGVX, where X can be C or V. There are 20 choices for C (19 consonants plus 0, which is lack of C), 4 choices for G, 6 choices for V, and 5 choices for X. This gives a total of 2400 possible syllables, of which just 326 occur. Thus, the percentage of syllable gaps is 86.4%.

Such a high percentage of gaps calls for an explanation. Some discussion has been offered in previous literature, such as Duanmu (1990, 2007), Ma (2003), and Yi and Duanmu (2015), which account for the gaps in terms of constraints that disallow certain combinations of phonemes or features. Let us consider some specific cases in Chengdu.

We begin with affricates. Chengdu has six stops [p pʰ t tʰ k kʰ] and four fricatives [f s z ɕ] ([x] can be seen as an allophone of [h] and is not listed separately). If stops, fricatives, and aspiration are combined freely, there are 24 possible affricates, of which just four occur. The data are shown in 13 and the constraints are shown in 14.

(13) Missing affricates in Chengdu (missing forms in parentheses; occurring forms in boldface)

Unaspirated				Aspirated			
(pf	ps	pz	pɕ)	(pfʰ	psʰ	pzʰ	pɕʰ)
(tf	**ts**	(tz)	**tɕ**	(tfʰ)	**tsʰ**	(tzʰ)	**tɕʰ**
(kf	ks	kz	kɕ)	(kfʰ	ksʰ	kzʰ	kɕʰ)

(14) Constraints on affricates in Chengdu
  a. The stop and the fricative gestures must both involve the tongue tip.
  b. All affricates are voiceless.

The occurring affricates are [ts tsʰ tɕ tɕʰ], as in 杂 [tsa] 'miscellaneous', 擦 [tsʰa] 'wipe', 夹 [tɕa] 'folder', and 掐 [tɕʰa] 'pinch'. Palatal affricates [tɕ tɕʰ] are allowed because a palatal involves both the tongue tip and the tongue body (Browman and Goldstein 1989).

Next, we consider C+[j] combinations. The data are shown in 15 and the constraints are shown in 16. As discussed earlier, we do not count [tɕj tɕʰj ɕj ŋj] as occurring forms, because they do not contrast with [tɕ tɕʰ ɕ ŋ].

(15) Missing C+[j] combinations in Chengdu (missing forms in parentheses; occurring forms in boldface)

Labial	**pj pʰj mj** (fj)
Dental	**tj tʰj nj** (tsj tsʰj sj zj)
Velar	(kj kʰj ŋj)
Palatal	(tɕj tɕʰj ɕj ŋj)
Glottal	(hj)

(16) Constraints on C+[j] combinations in Chengdu
    a. C must be labial or dental.
    b. C cannot contain [+fricative].

The constraints allow just six of the 21 consonants to combine with [j], as in 变 [pjan] 'change', 骗 [pʰjan] 'cheat', 面 [mjan] 'face', 电 [tjan] 'electricity', 天 [tʰjan] 'sky', and 链 [njan] 'chain', respectively.

Next, we consider aspiration. The data are shown in 17 and the constraints are shown in 18. Compared with unaspirated affricates, voiceless fricatives are all aspirated, and we have so indicated them accordingly. Of the 26 forms, 18 satisfy the constraints and do occur.

(17) Gaps in consonant aspiration in Chengdu (missing forms in parentheses; occurring forms in boldface)

	Unaspirated				Aspirated			
Fricative	(f	s)	**z**	(ɕ)	**fh**	**sh**	(zh)	**ɕh**
Nasal	**m**	**n**	**ŋ**	**ɳ**	(mh	nh	ŋh	ɳh)
Stop	**p**	**t**	**k**		**ph**	**th**	**kh**	
Affricate	**ts**	**tɕ**			**tsh**	**tɕh**		

(18) Constraints on consonant aspiration in Chengdu
    a. Voiceless fricatives are aspirated; voiced fricatives are unaspirated.
    b. Stops and affricated can be aspirated.
    c. Nasals are unaspirated.

Finally, we consider the combination of VX in the rime. The data are shown in 19 and the constraints are shown in 20. [ii uu] are treated as gaps because they do not contrast with [iː uː]. To simplify the discussion, I have excluded the retroflex vowel [ɚ], whose distribution is highly restricted (but can be accounted for by an additional constraint).

(19) Gaps in VX combinations in Chengdu (missing forms in parentheses; occurring forms in boldface)

(ii	iu)	**in**	(iŋ)
(ui	uu	un	uŋ)
(yi	yu)	**yn**	(yŋ)
ai	**au**	**an**	**aŋ**
ei	(eu)	**en**	(eŋ)
(oi)	**ou**	(on)	**oŋ**

(20) Constraints on VX combinations in Chengdu
   a. VX cannot both be [+hi].
   b. [n ŋ] do not contrast except after [a].
   c. Otherwise, [n] is used after front vowels and [ŋ] after back vowels.

Assuming six vowels [i u y a e o] and four options for X [i u n ŋ], there are 24 VX combinations, and ten satisfy the constraints and do occur. The six vowels [i u y a e o] occur by themselves as long vowels, too, making the total number of rimes to be 16.

The examples above do not exhaust all constraints on syllable gaps. For example, we have not discussed constraints on triphthongs (see Ma 2003 for a discussion on this issue). Nevertheless, the examples suffice to show that most syllable gaps are accountable, often by a few constraints that can be explicitly stated.

3.4 Phonemic economy vs. onset inventory economy

Phonemic economy is a well-known criterion according to which, other things being equal, an analysis that assumes fewer phonemes is thought to be better than one that assumes more. At first sight, the four analyses under consideration differ a lot in phonemic economy, as the summary in 21 shows, where C is a consonant phoneme (or an "initial" phoneme) and V is a vowel phoneme (or a "rime" phoneme).

(21) Phonemes in four analyses of Chengdu

Analysis	C	V	Total
CGV segmentation	19	16	35
CV segmentation	19	23	42
Finest segmentation	12	7	19
CVX segmentation	39	7	46

In the summary, the "finest" segmentation yields the best phonemic economy by far and the CVX segmentation yields the worst. However, the comparison is misleading, because the finest analysis yields a large number of consonant clusters, whereas the CVX analysis does not. Let us take a close look at the comparison by considering all onsets, shown in 22–24.

(22) Simple and complex onsets in Chengdu in the "finest" analysis

Simple onsets (consonants)	12	[p m f t n ȵ k s z ɕ ŋ h]
Complex onsets (consonant clusters)	27	[ph th ts tsh kh tɕ tɕh] [tw thw nw tsw tshw sw zw kw khw xw tɕw tɕhw ɕw] [pj phj mj tj thj nj] [nɥ]
Total	39	

(23) Simple and complex onsets in Chengdu in the CVX analysis

Simple onsets (basic consonants)	12	[p m f t n ȵ k s z ɕ ŋ h]
Complex onsets (non-basic consonants)	27	[pʰ tʰ ts tsʰ kʰ tɕ tɕʰ] [tʷ tʰʷ nʷ tsʷ tsʰʷ sʷ zʷ kʷ kʰʷ xʷ tɕʷ tɕʰʷ ɕʷ] [pʲ pʰʲ mʲ tʲ tʰʲ nʲ] [nᶣ]
Total	39	

(24) A comparison between finest and CVX analyses

	Simple onsets	Other onsets	Total
Finest	12	27	39
CVX	12	27	39

The comparison shows that the two analyses yield identical results, once all onsets are taken into consideration, which is necessary if we want to account for syllable gaps. The only difference is purely terminological, where "consonants" and "consonant clusters" in the finest analysis are called "basic consonants" and "non-basic consonants", respectively, in the CVX analysis. Similarly, the difference in the IPA transcription, such as [twh] vs. [tʰʷ], has little substantive value, because there is no contrast between them.

3.5 Phonetic facts

It is reasonable to assume that, other things being equal, a phonemic solution is better if it makes better phonetic predictions than other solutions. In addition, it is reasonable to assume the predictions in 25.

(25) Three assumptions with regard to phonetic predictions
    a. Each consonant has a unit of phonetic duration.
    b. Consonants in a cluster are ordered in a temporal sequence.
    c. Articulatory gestures within the same sound are more or less simultaneous.

According to the assumptions, the predictions made by the finest and the CVX analyses are shown in 26 and 27.

(26) Predictions on syllable duration

	Cs in the onset	Prediction	Facts
Finest	1–4	Syllables differ a lot in duration	Incompatible
CVX	1	Syllables differ little in durations	Compatible

(27) Predictions on the ordering of articulatory gestures

	Cs in the onset	Prediction	Facts
Finest	1–4	Gestures in the onset are ordered	Incompatible
CVX	1	Gestures in the onset are simultaneous	Compatible

In the "finest" analysis, a Chinese onset can have from one to four consonants. Therefore, syllables are expected to differ a lot in duration. In contrast, in the CVX analysis, each onset has just one consonant and syllables are expected to have similar durations. Phonetic evidence in general support the CVX prediction (Feng 冯隆 1985). Consider the data in 28, from Ren (1983: 38). The column "Onset" shows the phoneme count in the onset as proposed by the "finest" analysis. The column "Rime" shows the phoneme count in the rime as proposed by the "finest" analysis, where a long vowel counts as two sounds. The target syllable occurs in the first position of a disyllabic word X spoken in a carrier sentence, shown in 29.

(28) Syllable duration in Putonghua (six speakers; phoneme count under "Onset" and "Rime" are based on the "finest" analysis)

Syllable	Onset	Rime	Duration (ms)	Context
[kʰɤː]	2 [kh]	2	353	[kʰɤːɹən] 客人 'guest'
[tʂʷoː]	3 [tʂw]	2	323	[tʂʷoːtsz] 桌子 'table'
[tʰaː]	2 [th]	2	310	[tʰaːmən] 他们 'they'
[tʂʰuː]	3 [tʂh]	2	317	[tʂʰuːtʰou] 锄头 'hoe'
[tʂaŋ]	2 [tʂ]	2	320	[tʂaŋlʷo] 张罗 'busy to host'
[xan]	1 [x]	2	335	[xanɕyː] 含蓄 'implicit'
[kʰʷai]	3 [khw]	2	339	[kʰʷaixʷo] 快活 'happy'

(29) Carrier sentence used by Ren (1983)
朗读__X__给__你__听
lǎngdú__X__gěi__nǐ__tīng
Read__X__for__you__to-listen-to

The data show that, regardless of the onset type, syllable durations differ not much, and there is no discernable correlation between syllable duration and the complexity of the onset.

With regard to articulatory gestures, the relevant facts are that all gestures in the onset are basically simultaneous (Browman and Goldstein 1995; Öhman 1966; Xu 2017; Xu and Liu 2006). For example, in [pʰʲ], the gesture for [h] is simultaneous with that for [j], evidenced by the fact that the formant pattern for [j] is in the aspiration of [h]. In addition, before the release of [p], the tongue position for [j] is already in place. Thus, articulatory facts support the CVX analysis and not the "finest" analysis.

3.6 Consistency in syllable sizes

This criterion involves two parts, shown in 30. In Chinese, heavy syllables (also called regular syllables) are long and carry a lexical tone, whereas light syllables are short and do not carry or retain a lexical tone.

(30) Consistency in syllable sizes
      a. An analysis should distinguish two types of syllables, heavy and light.
      b. An analysis with simpler and fewer syllables types is better.

A comparison among the four analyses under discussion is shown in 31, with regard to their analyses of two heavy syllables, 穿 [tsʰʷan] 'wear' and 快 [kʷai] 'fast' and one light syllable, 过 [kʷo] 'ASP' (an aspect marker), in Chengdu speech.

(31) Consistency in syllable sizes

	Heavy: 'wear'	Heavy: 'fast'	Light: 'ASP'	Evaluation
CGV	CGV [tsʰʷ][u][an]	CGV [k][u][ai]	CGV [k][u][o]	X
CV	CVC [tsʰʷ][ua][n]	CV [k][uai]	CV [k][uo]	X
Finest	CCCGVX, [t][s][h][u][a][n]	CGVX [k][u][a][i]	CGV [k][u][o]	X
CVX	CVX [tsʰʷ][a][n]	CVX [kʷ][a][i]	CV [kʷ][o]	√

The CGV analysis fails to meet the criterion, because it offers the same representation for heavy and light syllables (both 'fast' and 'ASP' are represented as CV). The CV analysis also fails to meet the criterion, because it offers different rime representations for the heavy syllables and the same representation for the heavy syllable 'fast' and the light syllable 'ASP'. The "finest" analysis can distinguish the rime difference between heavy and light syllables (VX vs. V), but it fails 30b, in comparison with the CVX analysis. Therefore, the CVX analysis is the best solution.

3.7 Feature theory (contour features)

In the most common cases of feature representations, within a consonant or vowel, each feature can take just one value. For example, a vowel is either [+ high] or [− high], but not [+ high, − high] or [− high, + high]. To account for the generalization, Hoard (1971: 237) proposes a "principle of simultaneity," according to which all feature values within a consonant or vowel must be simultaneously

implementable. Feature values such as [+ high, – high] and [– high, + high], which are called "contour features," are ruled out because they have to be sequentially ordered, instead of being simultaneously implementable. Duanmu (1994) proposes a similar constraint called the No Contour Principle, which disallows contour feature values.

In a sequence of two consonants, sequences of opposite feature values can occur. For example, in the English word *smoke*, the cluster [sm] has multiple sequences of opposite feature values: [+ fricative][– fricative], [– nasal][+ nasal], and [– voice][+ voice].

If the Chinese onset is made of one consonant, there should be no contour feature in it. If the Chinese onset is made of more than one consonant, we should expect contour features to occur in it. It can be shown that Chinese onsets do not contain contour features (Duanmu 2008, 2016). Let us consider a specific case, which is [tsʰʷ].

There seem to be quite a few contour features in [tsʰʷ]. For example, [t] is not labial but [w] is, and [t] is [– voice] but [w] is [+ voice]. However, a key component in feature theory is underspecification, according to which only contrastive feature values are represented (e.g., Archangeli 1988; Dresher 2009; Steriade 1987). Let us start with the affricate [ts], which I shall assume to be representable without contour features (Duanmu 2008, 2016). Next, we add [h] to [ts]. The most relevant part of [h] is just the glottal feature [+ aspirated]. [ts] is either unspecified for aspiration, in which case we simply add [+ aspirated] to it. Or [ts] could be specified as [– aspirated], in which case we simply change it to [+ aspirated]. Finally, we add [w], whose most relevant part is the labial feature [+ round]. Once again, if [ts] is unspecified for rounding, we simply add [+ round] to it, and if [ts] is specified as [– round], we simply change it to [+ round]. It should be pointed out that [w] (or [u]) need not be specified for voicing, because there is no contrast between [w] and [w̥] (or between [u] and [u̥]). Thus, [tsʰʷ] is voiceless throughout, not [+ voice, – voice].

It can be seen that the CGV segmentation is incompatible with feature theory, because VV rimes, such as [ai], and VC rimes, such as [an], contain contour features. Similarly, the CV segmentation is incompatible with feature theory, because it produces diphthongs, such as [ai], and triphthongs, such as [iai], which contain contour features. In contrast, the finest segmentation and the CVX segmentation are both compatible with feature theory.

3.8 Summary

We have proposed a common set of seven criteria for evaluating any phonemic analysis: riming property, rime structure, syllable gaps, phonemic economy, phonetic facts, syllable complexity, and feature theory. We are now ready to compare competing solutions in the phonemic analysis of Chengdu (or of any language).

4 Evaluating different phonemic proposals and choosing the best solution

Let us now compare the four analyses of Chengdu against the common set of criteria that we have discussed in Section 3. The result is shown in 32

(32) Comparing phonemic solutions with a common set of criteria

	CGV	CV	Finest	CVX
Riming in poetry	√	X	√	√
Rime structure	X	X	√	√
Syllable gaps	?	?	√	√
Phonemic economy	?	?	√	√
Phonetic facts	X	X	X	√
Consistency in syllable sizes	X	X	X	√
Feature theory	X	X	√	√

The definition of riming pairs in poetry depends on the notion of the syllable rime, which does not include the pre-nuclear glide G. Of the four proposals, the CV analysis is the only one that has no segmentation boundary between G and the nuclear vowel, and so it is the only one that cannot properly define the syllable rime.

To account for rime structure, we need to state whether a syllable has one or two positions in the rime. The CGV and CV analyses cannot properly do so (because they do not segment the rime further), while the finest and the CVX analyses can.

It is unclear how the CGV and CV analyses would account for syllable gaps. On the other hand, both the finest and the CVX analyses can.

We have seen that phonemic economy should consider not just the inventory of phonemes but also the inventory of onsets (and the inventory of rimes). Since the CGV and CV analyses do not offer proper segmentation at the onset-rime boundary, it is unclear how they would satisfy phonemic economy. In contrast, we have seen that the finest and the CVX analyses are identical with regard to phonemic economy.

Phonetic facts support the view that the syllable onset is a single unit. The CVX analysis is the only one that is compatible with such facts.

The consistency in syllable sizes is discussed in Section 3.6, where we have seen that the CGV, CV, and "finest" analyses all fail the criterion, while the CVX analysis is the only one that satisfies the criterion.

With regard to feature theory, the CGV and CV analyses produce "phonemes" that contain contour features, such as [ai] and [au]. In contrast, the finest and the CVX analyses do not produce contour features and are both compatible with feature theory.

In summary, we have seen that it is clearly possible to compare different proposals of the phonemic analysis of a language, and the best solution can be determined unambiguously. In the case of Chengdu, there is little doubt that the CVX analysis is the only choice and best solution.

5 Additional issues

In this section, I discuss two additional issues that bear on the present discussion: variation in syllable duration and reaction time in lexical access

5.1 Variation in syllable duration

As discussed earlier, the CVX analysis predicts that all full syllables with an onset are similar in duration, and the prediction is supported by the phonetic data in Ren (1983). On the other hand, the CVX analysis also predicts that syllable duration can differ with

regard to whether the syllable is heavy or light and whether a syllable has an onset, shown in 33.

(33) Two factors that affect syllable duration ("A > B" means A is longer than B)
    a. A full rime VX is longer than a light rime V (i.e., VX > V; CVX > CV).
    b. A syllable with an onset is longer than one without (i.e., CVX > VX; CV > V).

Let us now consider some phonetic data. Shih and Ao (1997: 395) report that, in Mandarin Chinese, "the more phonemes there are in a syllable, the longer the syllable duration is". Their data are shown in 34, based on 19, 150 syllables, and the mean values are estimated from a histogram. Since the only VC rime is VN, I have changed VC to VN in order to be more specific.

(34) Variation in syllable duration in Mandarin (Shih and Ao 1997, Fig. 31.3)

Type	Mean (ms)
CGVN	270
CGVV	245
CVN	240
CGV	230
CVV	225
VN	220
CV	220
VV	215
V	180

Shih and Ao provide no statistics on the mean values. Nevertheless, given 33, most of the variations are expected, shown in 35, where "A ~ B" means A and B have similar durations.

(35) Expected variations in syllable duration

Fact	Comment
CVV > CV	Expected by 33a
CV > V	Expected by 33b
CV ~ VN	Expected by 33a, b
CVV > VV CVN > VN	Expected by 33b
VV ~ VN	Expected by CVX
CGVV ~ CVN ~ CGV ~ CVV	Expected by CVX, if CGV is [CGV:]
CGVN > CGVV	

Shih and Ao use V to represent a monophthong, either long or short, and VV to represent a diphthong. Therefore, syllable types CV and V are likely to have included unstressed syllables, which would explain their shorter mean duration than those with full rimes. On the other hand, CGV is similar to other full syllables, likely because it mostly includes full syllables, more accurately

represented as [CGV:]; this is confirmed by a search through 现代汉语词典 *Xiàndài Hànyǔ Cídiǎn Modern Chinese Dictionary* (Chinese Academy of Social Sciences 中国社会科学院 2005), where unstressed syllables are dominantly CV. The CVX theory also predicts that (i) onsetless full syllables have similar durations and (ii) full syllables with an onset have similar durations, and both (i) and (ii) seem to be true. Overall, therefore, the data are compatible with the CVX analysis. The only thing still unaccounted for is the fact that CGVN is longer than CGVV (and other syllable types), but this fact does not support the view that syllable duration is related to onset complexity.

Wu (2017) also reports that, in Mandarin, "syllables with more phonemes tend to have longer syllable duration". Let us consider her data from trisyllabic words, shown in 36, based on 34 trisyllabic expressions (102 syllables in all). I have made three adjustments. First, I have changed V rimes (a long monophthong) to VV, because all the syllables are heavy. Second, I have corrected an error with the type count of CVV (from 24 to 23) and an error with the type count of CGVG (from 6 to 5). Finally, I have added the percentage of syllables whose main vowel is [a].

(36) Variation in syllable duration in Mandarin (Wu 2017, Table 4.4)

Structure	Type	Token	Mean (ms)	% [a]
VV	1	90	164	0
GVN	5	450	170	60
CVV	23	2160	198	17
CGVV	8	720	200	25
CVG	5	450	202	20
CVN	35	3150	204	23
CGVG	6	450	215	100
CGVN	19	1710	226	100

Each expression (type) was read 90 times (tokens), three times each by 30 speakers. The variation in mean syllable duration can be explained as follows. First, the shortest form VV lacks an onset, which may explain why its duration is a lot shorter than those of others. Second, it is well known that low vowels are longer than non-low vowels (e.g., Feng 冯隆 1985; House 1961; Peterson and Lehiste 1960). In CGVG and CGVN, the main vowel is 100% [a], the only low vowel in Mandarin, whereas in other syllable forms, the main vowel is mostly not [a]; this could explain why CGVG and CGVN have the longest mean duration. The only form to be accounted for is GVN, whose mean is shorter than expected. This form includes two characters, 文 *wén* (in two expressions, 天文学 *tiānwénxué* 'astronomy' and 古文明 *gǔ wénmíng* 'ancient civilization') and 眼 *yǎn* (in three expressions, 眼镜蛇 *yǎnjìngshé* 'cobra', 眼中钉 *yǎn zhōng dīng* 'nail in the eye', and 小心眼 *xiǎoxīnyǎn* 'narrow-minded'). Whatever the explanation, there is no support for the view that the complexity of the onset has an impact on syllable duration.

Wu and Kenstowicz (2015: 91) also report some variation in syllable duration, based on 48 monosyllabic Mandarin words. Each word (type) was read 20 times (tokens),

twice in isolation and twice in a carrier sentence, by five speakers. The mean values are shown in 37, estimated from a histogram. The onset C varies; the medial glide is always [w] (if any); the vowel is always [a]; and the coda is always [n] (if any). I have added the column on the percentage of onsets that are stops, whose role is discussed below.

(37) Variation in syllable duration in Mandarin (Wu and Kenstowicz 2015: 91, Fig. 2)

Structure	Type	Token	Mean (ms)	C = stop (%)
C[aː]	12	240	300	100
C[an]	11	220	345	73
C[waː]	13	260	350	62
C[wan]	12	240	380	58

Since all the syllables have an onset and a full rime, the CVX analysis predicts them to be similar in duration, which is mostly confirmed, as Wu and Kenstowicz (2015, p. 90) report that most of the differences among the means are statistically insignificant. Nevertheless, there is one significant difference, which is between C[aː] and C[wan], and it requires an explanation. I suggest that it is due to the under-measurement of stop onsets in C[aː]. Specifically, the test words in Wu and Kenstowicz's study were read twice in a carrier sentence and twice more in isolation. In the latter case, the closure duration of a stop onset is not included, because there is no visible boundary for the start of oral closure (a fact confirmed by Michael Kenstowicz, personal communications). In contrast, non-stop onsets (specifically, [x m l]) have a clear starting point and are unlikely to be under-measured. As seen in 37, C[aː] had the highest percentage of stop onsets and C[wan] had the lowest. This means that C[aː] was the most under-measured and C[wan] the least under-measured. Indeed, we could explain the durational difference among all syllable structures, in that the ranking in the percentage of stop onsets is the inverse of the reported ranking in mean duration, namely, C[aː] > C[an] > C[waː] > C[wan] in percentage of stop onsets, and C[aː] < C[an] < C[waː] < C[wan] in mean duration.

In summary, the CVX analysis makes the best predictions on syllable duration. There is no evidence that the complexity of the onset has a noticeable effect on syllable duration.

5.2 Lexical access and reaction time

Neergaard and Huang (2016) offer a new approach to the non-uniqueness problem through an experiment that involves an auditory shadowing task. In the experiment, subjects were asked to repeat monosyllabic Chinese words that they hear, and for each word, the reaction time of a subject was recorded. The idea is as follows. The auditory shadowing task presumably involves lexical access, in that the subject has to perceive words and produce words. It has been reported that a number of factors could affect lexical access. Some are shown in 38.

(38) Factors that could influence lexical access
    a. Word recognition is faster when a word has a high frequency.
    b. Auditory lexical decision is slower when a word has greater homophone
     density (Wang et al. 2012).

    c. Word recognition is slower when a word has greater phonological
       neighborhood density (PND) (Luce and Pisoni 1998).

    d. Word production (as in picture naming) is slower when a word has
       many interconnected neighbors (i.e., high clustering coefficient)
       (Chan and Vitevitch 2010).

38c and 38d are dependent on phonemic analysis. PND refers to how many "phonological neighbors" a word has. Two words A and B are phonological neighbors if we can derive B from A by adding, deleting, or substituting a single phoneme (or vice versa). The clustering coefficient of a word W refers to the percentage of W's phonological neighbors that are also phonological neighbors of each other. For example, let N be the set of phonological neighbors of W. If N has 20 members, and if 10 of them is a phonological neighbor of another member in N, then the clustering coefficient of W is 10/20 = 50%.

Now, the point of interest is that different phonemic analyses will yield different PND values, which in turn could yield different clustering coefficient values. As an example, consider the word 关 *guān* 'close'. The CG_VX analysis is shown in 39, where CG is a phoneme and VX is another. The C_G_V_X analysis is shown in 40, where C, G, V, and X are four separate phonemes.

(39) Phonological neighbors in the CG_VX analysis of 关 [kʷ_an] 'close'

Word	Phoneme	Neighbor?	Comment
干 'dry'	[k_an]	Yes	Change [kʷ] to [k]
碗 'bowel'	[w_an]	Yes	Change [kʷ] to [w]
男 'male'	[n_an]	Yes	Change [kʷ] to [n]
安 'peace'	[an]	Yes	Delete [kʷ]
瓜 'melon'	[kʷ_a]	Yes	Change [an] to [a]
龟 'turtle'	[kʷ_ei]	Yes	Change [an] to [ei]
锅 'wok'	[kʷ_o]	Yes	Change [an] to [o]
…	…	…	…

(40) Phonological neighbors in the C_G_V_X analysis of 关 [k_w_a_n] 'close'

Word	Phoneme	Neighbor?	Comment
干 'dry'	[k_a_n]	Yes	Delete [w]
碗 'bowel'	[w_a_n]	Yes	Delete [k]
男 'male'	[n_a_n]	No	Change [k] to [n]; delete [w]
安 'peace'	[a_n]	No	Delete [k]; delete [w]
瓜 'melon'	[kʷ_a]	Yes	Delete [n]
龟 'turtle'	[kʷ_e_i]	No	Change [a] to [e]; change [n] to [i]
锅 'wok'	[kʷ_o]	No	Change [a] to [o]; delete [n]
…	…	…	…

In general, finer segmentation leads to fewer neighbors. In the CG_VX analysis, CG is a single phoneme, and all onset changes yield neighbors. In contrast, in the C_G_V_X analysis, the onset is split into C_G; of the four onset changes (first four

lines), just two yield neighbors. In addition, the CG_VX analysis does not split the rime, and all the rime changes yield neighbors. In contrast, the C_G_V_X splits the rime into V_X, and of the three rime changes, just one yields a neighbor.

Neergaard and Huang (2016) compared seven ways of segmentation (ignoring tones), shown in 41, with two sample words, where the notations are as in the original.

(41) Seven ways of phonemic segmentation considered by Neergaard and Huang (2016)

Segmentation	联 *lián* 'connect'	拿 *ná* 'hold'
C_V_C	[l_iɛ_n]	[n_a]
C_V_V_X	[l_i_ɛ_n]	[n_a]
C_V_VX	[l_i_ɛn]	[n_a]
C_VVX	[l_iɛn]	[n_a]
CV_V_X	[li_ɛ_n]	[n_a]
CV_VX	[li_ɛn]	[n_a]
CVVX	[liɛn]	[na]

For each segmentation, PND is calculated for each word, so is its clustering coefficient. Values of other factors are gathered, too, such as word frequency and homophone density. Then statistics and model selection were performed. It was found that the C_V_C segmentation yields the best statistics to fit the reaction time data in the auditory shadowing task.

The C_V_C segmentation is the same as the "CV segmentation" discussed in Section 2, in which diphthongs and triphthongs are not segmented. We have shown in Section 4 that it fails most of the criteria we have proposed. How, then, do we reconcile the present conclusion with the result of Neergaard and Huang (2016)?

I would like to point out an oversight in the analysis of Neergaard and Huang (2016), which is the failure to represent vowel length in monophthongs. For example, in all the analyses in 41, the vowel in "hold" should be long vowel [aː] or [aa], not a short vowel [a]. This has two consequences. First, it will affect the calculation of PND, which is illustrated in 42 and 43

(42) Phonological neighbors in the C_V_C analysis of 路 *lù* 'road'

Word	C_V [l_u] 'road'	C_V [l_uu] 'road'
老 'old'	C_V [l_au] yes	C_V [l_au] yes
拉 'pull'	C_V [l_a] yes	C_V [l_aa] yes
龙 'dragon'	C_V_C [l_u_ŋ] yes	C_V_C [l_u_ŋ] no
…	…	…

(43) Phonological neighbors in the C_V_X analysis of 路 *lù* 'road'

Word	C_V [l_u] 'road'	C_V_X [l_u_u] 'road'
老 'old'	C_V_X [l_a_u] no	C_V_X [l_a_u] yes
拉 'pull'	C_V [l_a] yes	C_V_X [l_a_a] no
龙 'dragon'	C_V_X [l_u_ŋ] yes	C_V_X [l_u_ŋ] yes
…	…	…

In the C_V_C analysis, 'road' is C_V (with no final C), for which both 'old' and 'pull' are neighbors, each involving the change of V. If vowel length is ignored, 'dragon' is also a neighbor (by adding C), but if vowel length is represented, 'dragon' is not a neighbor, because it involves two changes, replacing V from [uu] to [u] and adding C.

In the C_V_X analysis, if vowel length is ignored, 'road' is C_V, for which 'pull' is a neighbor (by changing V), so is 'dragon' (by adding X), but 'old' is not a neighbor, because it involves two changes, replacing V from [u] to [a] and adding X. If vowel length is represented, 'road' is C_V_X, for which 'old' is a neighbor (by changing V), so is 'dragon' (by changing X), but 'pull' is not a neighbor, because it involves two changes, replacing both V and X.

Representing vowel length can also change the calculation of PND and homophone density in polysyllabic words. It is worth pointing out that PND is based on the entire lexicon. The lexicon that Neergaard and Huang (2016) used has some 70,000 phonologically distinct items (excluding tones), in which 0.6% are monosyllabic, 43.8% disyllabic, 34.4% trisyllabic, and 21.2% quadrisyllabic or longer. Therefore, polysyllabic items (mostly compounds or names) contribute heavily to the calculation of PND and homophone density. Now, consider the examples in 44.

(44) Effect of vowel length on PND and homophone density in polysyllabic items

	Word	Length ignored	Length represented
a.	聊 'chat'	[liau]	[liau]
b.	李敖 (name)	[li][au] ➔[liau]	[lii][au] ➔[liiau]
c.	李瑶 (name)	[li][iau] ➔[liiau]	[lii][iau] ➔[liiiau]
d.	李一敖 (name)	[li][i][au] ➔[liiau]	[lii][ii][au] ➔[liiiiau]

If vowel length is ignored, 44a–b are homophones, so are 44c–d. If vowel length is represented, 44a–d are all distinct.

In the 70,000 phonologically distinct lexical items (excluding tones) that Neergaard and Huang (2016) used, there are nearly 200,000 syllables, 37% of which are annotated as CV, which ought to be CVV instead. The effect of such syllables on the calculation of PND and homophone density could be substantial. Therefore, it seems premature to accept the conclusion that the C_V_C segmentation offers the best explanation of reaction time data in the auditory shadowing task.

6 Conclusions

The non-uniqueness theory of Chao (1934) has had a profound influence on phonemic analysis. According to the theory, there is no best solution in phonemic analysis. In-stead, competing solutions of the same language may co-exist, each having its own merit and all being valid.

The lack of a clear solution makes phonemic analysis seem not very useful, which may explain the fact that the Chinese tradition of phonological descriptions rarely offers a phonemic inventory; instead, inventories of onsets, rimes, and tones are offered, and oftentimes an inventory of syllables as well, evidently because such inventories are usually unambiguous.

One might wonder why, despite its ambiguity, phonemic analysis is nevertheless widely used in the Western tradition of phonological descriptions. A plausible answer, suggested by Ladefoged (2001), is that the Western tradition is influenced by the accidental fact that most Western languages are spelled alphabetically, where consonants and vowels seem to be the basic units. Another answer, and in my view a more reasonable one, is that in Western languages, most words are polysyllabic, where syllable boundaries are unclear. Therefore, it is hardly feasible to come up with an inventory of onsets and rimes, or an inventory of syllable, and a phonemic analysis is the only option. Nevertheless, the lack of rigor in phonemic analysis has led some scholars to doubt its validity, such as Ladefoged (2001) and Fowler (2015).

A fundamental shortcoming in the non-uniqueness theory is the assumption that there is no common set of criteria that applies to all solutions. I have proposed instead that such a set of criteria can be established, which include riming properties, rime structure, constraints on syllable gaps, phonetic facts (syllable duration and articulation), syllable complexity, and feature theory. I have illustrated the proposal with an in-depth examination of Chengdu, whose phonemic analysis has not been offered before. Four proposals are compared, based on the "CGV" segmentation, the "CV" segmentation, the "finest" segmentation, and the "CVX" segmentation, and the CVX analysis is shown to be unambiguously the only viable solution and the best one. The present study removes a long-standing problem in phonology and offers an example of how to determine phonemes in other languages.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
Ao, Benjamin XP. 1992. The non-uniqueness condition and the segmentation of the Chinese syllable. In *Ohio State University working papers in linguistics 42: Papers in phonology*, ed. Elizabeth Home, 1–25. Columbus: Department of Linguistics, Ohio State University.
Archangeli, Diana. 1988. Aspects of underspecification theory. *Phonology* 5(2): 183–207.
Browman, Catherine P., and Louis M. Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6(2): 201–251.
Browman, Catherine P., and Louis M. Goldstein. 1995. Gestural syllable position effects in American English. In *Producing speech: Contemporary issues*, ed. Fredericka Bell-Berti and Lawrence J. Raphael, 19–33. New York: AIP Press (For Katherine Safford Harris).
Chan, Kit Ying, and Michael S. Vitevitch. 2010. Network structure influences speech production. *Cognitive Science* 34(4): 685–697.
Chao, Yuen-Ren. 1934. The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica* 4(4): 363–398.
Chao, Yuen-Ren 赵元任. 1923. *A new vocabulary of rimes* 国音新诗韵. Shanghai: Commercial Press.
Chinese Academy of Social Sciences (Dictionary Office, Institute of Linguistics) 中国社会科学院 (语言研究所词典编辑室). 2005. *Modern Chinese Dictionary* 现代汉语词典 (5th edition). Beijing: Commercial Press.
Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.

Duanmu, San. 1990. A formal study of syllable, tone, stress and domain in Chinese languages. Doctoral dissertation. Cambridge: MIT.

Duanmu, San. 1994. Against contour tone units. *Linguistic Inquiry* 25(4): 555–608.

Duanmu, San. 2007. *The phonology of standard Chinese (2nd edition)*. New York: Oxford University Press.

Duanmu, San. 2008. *Syllable structure: The limits of variation*. Oxford: Oxford University Press.

Duanmu, San. 2016. *A theory of phonological features*. New York: Oxford University Press.

Feng, Long 冯隆. 1985. Duration of initials, finals, and tones in Beijing dialect 北京话语流中声韵调的时长. In *Working papers in experimental phonetics* 北京语音实验录, ed. Tao Lin and Lijia Wang 林焘, 王理嘉, 131–195. Beijing: Peking University Press.

Fowler, Carol A. 2015. The segment in articulatory phonology. In *The segment in phonetics and phonology*, ed. Eric Raimy and Charles E. Cairns, 25–43. Oxford: Wiley-Blackwell.

Gao, Mingkai, and Anshi Shi 高名凯, 石安石. 1963. *Introduction to linguistics* 语言学概论. Beijing: Zhonghua Press.

Goldsmith, John. 2011. Syllables. In *Handbook of phonological theory*, ed. John A. Goldsmith, Jason Riggle, and Alan C. L. Yu, 164–196. Malden: Wiley-Blackwell.

He, Wan, and Dongmei Rao 何婉, 饶冬梅. 2014. *A survey study of the phonology and vocabulary of Chengdu speech in Sichuan* 四川成都话音系词汇调查研究. Chengdu: Sichuan University Press.

Hoard, James E. 1971. The new phonological paradigm. *Glossa* 5: 222–268.

Hockett, Charles F. 1960. The origin of speech. *Scientific American* 203(September): 88–96.

House, Arthur S. 1961. On vowel duration in English. *Journal of the Acoustical Society of America* 33(9): 1174–1178.

Kohler, Klaus. 1999. German. In *Handbook of the international phonetic association: A guide to the use of the international phonetic alphabet*, ed. International Phonetic Association, 86–89. Cambridge: Cambridge University Press.

Ladefoged, Peter. 2001. *Vowels and consonants: An introduction to the sounds of languages*. Malden: Blackwell.

Lee, Wai-Sum, and Eric Zee. 2003. Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33(1):109–112.

Lin, Maocan, and Jingzhu Yan 林茂灿, 颜景助. 1980. Acoustic characteristics of neutral tone in Beijing Mandarin 北京话轻声的声学性质. *Dialects* 方言 3: 166–178.

Luce, Paul A., and David B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19(1): 1–36.

Ma, Qiuwu. 2003. *Optimality theory and Mandarin syllable structure*. Tianjin: Nankai University Press.

Neergaard, Karl, and Chu-Ren Huang. 2016. Graph theoretic approach to Mandarin syllable segmentation. Paper presented at *The Fifteen International Symposium on Chinese Language and Linguistics (IsCLL)*. Hsinchu: Hsinchu University of Education.

Öhman, Sven EG. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* 39(1): 151–168.

Peterson, Gordon E., and Ilse Lehiste. 1960. Duration of syllabic nuclei in English. *Journal of the Acoustical Society of America* 32(6): 693–703.

Pike, Kenneth. 1947a. *Phonemics: A technique for reducing languages to writing*. Ann Arbor: University of Michigan Press.

Pike, Kenneth. 1947b. On the phonemic status of English diphthongs. *Language* 23(2): 151–159.

Ren, Hongmo. 1983. A linguistic model for duration in Chinese. M.A. thesis. Los Angeles: UCLA.

Shih, Chi-lin, and Benjamin Ao. 1997. Duration study for the Bell Laboratories Mandarin text-to-speech system. In *Progress in speech synthesis*, ed. Jan van Santen, Richard Sproat, Joseph Olive, and Julia Hirschberg, 383–399. New York: Springer-Verlag.

Steriade, Donca. 1987. Redundant values. In *Papers from the 23rd Annual Regional Meeting of the Chicago Linguistic Society, part 2: Parasession on autosegmental and metrical phonology*, ed. Eric Schiller, Barbara Need, and Anna Bosch, 339–362. Chicago: Chicago Linguistic Society.

Swadesh, Morris. 1935. The vowels of Chicago English. *Language* 11(2): 148–151.

Trager, George L, and Henry L. Smith. 1957. *An outline of English structure (5th edition)*. Washington, DC: American Council of Learned Societies.

Wang, Wenna, Xiaojian Li, Ning Ning, and John X. Zhang. 2012. The nature of the homophone density effect: An ERP study with Chinese spoken monosyllable homophones. *Neuroscience Letters* 516(1): 67–71.

Wiese, Richard. 1996. *The phonology of German*. New York: Oxford University Press.

Wu, Di. 2017. Cross-regional word duration patterns in Mandarin. Doctoral dissertation. Champaign: University of Illinois Urbana-Champaign.

Wu, Fei, and Michael Kenstowicz. 2015. Duration reflexes of syllable structure in Mandarin. *Lingua* 164: 87–99.

Xu, Yi. 2017. Syllable as a synchronization mechanism that makes human speech possible. Manuscript, University College London. Available at http://www.homepages.ucl.ac.uk/~uclyyix/Syllable_manuscript.pdf. Accessed 13 Nov 2017.

Xu, Yi, and Fang Liu. 2006. Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Rivista di Linguistica* 18(1): 125–159.

Yi, Li, and San Duanmu. 2015. Phonemes, features, and syllables: Converting onset and rime inventories to consonants and vowels. *Language and Linguistics* 16(6): 819–842.

You, Rujie, Nairong Qian, and Zhengxia Gao 游汝杰, 钱乃荣, 高钲夏. 1980. On the phonological system of Putonghua 论普通话的音位系统. *Chinese Philology* 中国语文 5 (158): 328–334.

Zhu, Xiaonong. 1995. Shanghai tonetics. Doctoral dissertation. Canberra: Australian National University.