Advancing explanatory and tonal dialectometry

Sung, H.W.M.

Citation

Sung, H. W. M. (2026, February 13). *Advancing explanatory and tonal dialectometry. LOT dissertation series*. LOT, Amsterdam. Retrieved from https://hdl.handle.net/1887/4291801

CHAPTER 7

Comparison of the Existing Tone Distance Calculation Methods[1]

7.1 Introduction

Traditional dialectology (also known as dialect geography, Chambers and Trudgill 1998: 14) originated in Europe, pioneered in Germany and France, has primarily been centering around phonetics and lexicon, while syntax has entered the scene more recently. The study of phonetic variation has been heavily focused on segments, and not tones, despite the fact that 60-70% of the world's languages also include lexical tones (Yip 2002). This focus is not surprising, however, since European languages mostly do not use pitch to differentiate word meaning. Although the traditional methodology of dialect geography spread to different corners of the world, the application of computational mapping and later computational processing of dialectal data have mostly been developed and applied to European languages, such as Dutch (Heeringa 2004), French (Goebl 1984), Swiss German (Scherrer and Stoeckle 2016) and Bulgarian (Prokić 2010). For the same reason, these computational methods

---

[1] This chapter is based on Sung et al. (2025).

do not concern tonal variation[2], and whether these methods are suitable for the tonal data is a current issue yet to be solved.

In Chinese dialectology, for example, there are numerous studies on dialects spoken in China (e.g. Chao's (1928) survey on Wu dialects, Grootaers's (2003) work in Shanxi, Zhan's (2002) *Introduction to Yue dialects in Guangdong*), and the field has a century-long tradition. However, most studies on tonal variation are descriptive. This means that studies usually report the tonal inventory of a dialect after a fieldwork investigation, and/or tones are analysed in terms of how they correspond to historical tone categories (from the Middle Chinese period, based on the rhyme dictionary descriptions) and how they split or merged. An example of such studies is Zhan (2002). Although there is a huge amount of dialect data available for Chinese (in the form of IPA transcriptions, including tones), currently the field is still at the exploration stage in finding a methodology which allows dialectologists to measure tone distances for the purpose of dialect classification. The lack of analytic tools is not only a problem for dialectologists working with Chinese, but also a problem for dialectologists in general, as well as linguists, around the world. While most of the world's languages are tonal, we do not have proper methods which would allow us to investigate dialectal variation of tones, the most important aspect of tonal languages by which they differ from the non-tonal languages.

In this chapter, four existing tone distance calculation methods are compared in order to assess their adequacy for the purpose of dialectometry. The structure of this chapter is as follows: Section 7.2 briefly sketches what tonal languages are and their relative isolation in dialectometry. Section 7.3 gives an overview of the transcription notations and formal representations of tones. A sketch of the existing methodologies used in measuring tone distances is provided in Section 7.4, followed by the comparison of the tone distance calculation methods in Section 7.5. The chapter ends with a summary and discussion of the results in Section 7.6 and a conclusion in Section 7.7.

---

[2] There are studies, such as Gooskens and Heeringa (2006), which explored prosodic variation in pitch accent languages. However, for other tonal languages (see Section 7.2), the methodology in dealing with more complex tones is still very much under-explored.

7.2 Tonal Languages

What makes tonal languages different from, e.g. most European languages, is the use of pitch to distinguish meaning of a word from another, together with segments. Hyman (2006:229) defines tone as the use of "pitch… [in] the lexical realization of at least some morphemes". Around the world, it has been estimated that around 60-70% of the languages are tonal (Yip 2002:1). Some relatively well-studied tonal languages can be found in the Sino-Tibetan language family and the Niger-Congo family, e.g. Cantonese and Yoruba.

In addition, there is a group of languages which are in between tonal and non-tonal ('stress') languages, namely the accentual languages (Yip 2002:4). Yip classifies accentual languages, including Japanese, Serbo-Croatian and Limburgish, as a subclass of tone languages. Accentual languages tend to have "a small number of contrasting tones (usually only one or two), and these are sparsely distributed or even absent on some words and usually belong to specific syllables" (Yip 2002:4).

In fact, one can place tonal languages on a continuum by tonal density (Gussenhoven 2004:35). On one end of the continuum, languages like Sikaritai have a tone marked in every syllable in every word; on the opposite end of the continuum, we have languages like Dagaara, which contrasts only H(igh) and L(ow) in stressed syllables. Comparing to Yip's (2002) description of accentual languages, Gussenhoven's (2004) continuum roughly matches Yip's tonal vs. accentual languages, and Yue is a strong tonal language in the continuum. In the current thesis, 'tonal languages' refer to languages which the pitch of the word can change its meaning (Yip 2002:1), with a bigger number of contrasting tones (more than one or two) and are not sparsely distributed or absent on some words (Yip 2002:4), opposed to accentual languages.

The main types of tones are Level, Falling, Rising, as well as complex tones, which consist of Concave (Falling-Rising) and Convex (Rising-Falling) tones. A tone is a level tone when the contours are kept flat throughout the whole tone production[3], and a rising tone is when the pitch of the offset is higher than the onset, whereas a falling tone is the opposite. Concave and Convex tones are the combinations of rising and falling tones. Languages and dialects may have more than one tone of

---

[3] In reality, some of the level tones also have a slight falling or rising F0 trajectory, but this chapter focuses on impressionistic transcription data, which does not always capture such fine trajectories.

the same shape. Take Cantonese for example, it has three level tones, two rising tones and one falling tone. It has been found that across 737 Sinitic varieties (but over 500 are Mandarin dialects), falling tone is the most common tone (1125 tokens), closely followed by level tones (1086 tokens). Rising (790 tokens) and Concave (352 tokens) tones are less common and Convex (80 tokens) tones are the least common tones (Cheng 1973).

7.3 Transcription notations and formal representations

The following introduction of the transcription notations of tones (found in dialect surveys) and formal tone representations (used in phonology) serves as a background for tone representations, which is extremely important for tone distance calculations.

Transcription notations and formal representations of tones should be treated separately. By transcription notations, I refer to the notations fieldworkers or scholars use to describe and represent the phonetic realisation of tones in their publications. On the other hand, formal representations are notations in which phonologists use to represent the mental representation of tones in the speakers' minds. The following subsections will give an overview of these notations for modern tone languages, with an emphasis on Chinese dialectology (based on You 2016).

7.3.1 Transcription Notations

There are two types of transcription notations for tones in Chinese. The first being the systems of notations which transcribe the tone fully (phonetically transparent) and the second type is much more (phonetically) opaque. I shall address the former as *transparent notation systems* and the latter *opaque notation systems*. The differences between the two systems are explained below.

The transparent notation system aims to show the full tone contour in the transcription. The earliest transparent notation can be found in Chao's (1928) survey on Wu dialects, where tone contours were documented with two notations. The first notation is a tone contour diagram (Chao 1928:76), which is a time series diagram with time on the x-axis and pitch on the y-axis. A mid-point pitch is indicated in the middle of the diagram and the relative pitch to the mid-point was then plotted. In addition, the tone is also transcribed in the Sol-fa system (a musical notation, in the form of digits) and it contains *sharps* and *flats* from the music notation. Chao's (1930) *tone letters* is a system for tone transcription which consists of 5 digits, 1, 2, 3, 4, 5, representing different (possible) contour levels in a tone. 1 represents the lowest contour level and 5 represents the highest, while the rest sit in between. When combined (as two-digit or three-digit sequences), they can indicate a change in the contour, which represents the shape of the tone. For example, 53 is a falling tone, whereas 213 is a convex or dipping tone (a falling contour followed by a rising contour). Chao's (1930) system also has a graphic counterpart, which consists of the contour and a reference height bar on its side. This graphical system is also adopted by the IPA (see the IPA chart, International Phonetic Association 2005).

It should be noted that there is another notation which is very similar to Chao's (1930) tone letters, but the pitch which each digit represents is reversed. In the Central American tradition, tone contours are also represented by the digits 1 to 5, but 5 represents a low tone whereas 1 represents the high tone (Yip 2002:21). To indicate a change in the contour, a hyphen '-' is added, e.g. si^{1} is a high-level tone, and si$^{2\text{-}3}$ is a high falling tone (adapted from Yip 2002:21).

The IPA chart has one other tone notation system, other than Chao's. This system uses diacritics to convey tonal information. In this system, pitch levels are differentiated through the acute and grave accents, such as á and à (representing High and Low tones respectively, Yip 2002:21). In addition, a bar indicates a Mid tone, but it is sometimes unmarked. To represent extra High or Extra Low tones, the accent is doubled. Contour shapes are also possible to be represented in this notation. A fall from high to low is represented by a circumflex and a rise from low to high is represented by a caron. This diacritic notation is common in the African tradition (Yip 2002:19).

The opaque notation systems do not directly convey the phonetic information of the tones, and they are usually dialect-/ language-specific. Often, these systems show the tone categories from a reference system, which one can then convert back to the actual phonetic values by referring to the inventory table. For instance, in Chinese dialectology, a historical reference system is used, namely the Middle Chinese (MC hereafter) tone categories, which are *Ping*, *Shang*, *Qu* and *Ru* (tone category found in checked syllables), and they can be further divided into high and low registers, which are represented by *Yin* and *Yang* categories respectively.[4] These categories can be marked as symbols (You 2016:42-44):

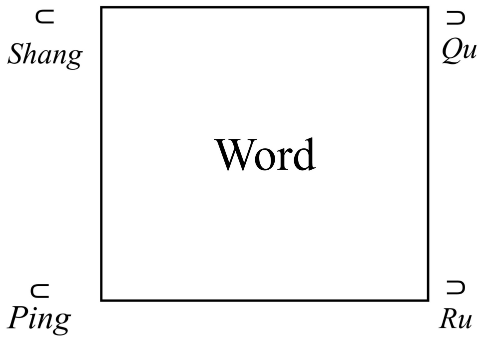


Figure 7.1: Tone symbols for the Middle Chinese tone categories in the Yin register

A subset symbol '⊂' placed in different corners of the word (transcription) indicates different MC tone categories for the respective word, as illustrated in Figure 7.1. Furthermore, the Yang register is marked by adding a bar under the subset symbol, and the Yin register is unmarked. Since most Chinese dialect surveys follow the wordlist which is based on MC rhyme dictionaries *Qieyun/ Guangyun*, the tone inventories are often listed in terms of MC tone categories. Therefore, to retrieve the phonetic tonal information of a word, one has to find the corresponding tone value from the inventory by matching the MC tone category represented by the symbol in the transcription.

A variant of the MC system uses numbers instead of symbols. Odd numbers (1, 3, 5, 7) represent *Ping*, *Shang*, *Qu* and *Ru* tones in the *Yin* category, and even numbers (2, 4, 6, 8) for the *Yang* categories. For some Chinese dialects, not all MC categories are reflected in the contemporary tonal inventory (due to mergers). In these cases, some numbers (tone categories which merged) are omitted in the inventory. In Shanghainese, for example, *yin ping*, *yin qu*, *yang qu*, *yin ru*, *yang*

---

[4] Yin and Yang categories generally correspond to the tones which occur in syllables with a voiceless and voiced (obstruent and sonorant) respectively. However, there is variation in the correspondences among the dialects.

ru would be marked as 1, 5, 6, 7 and 8. Since this is a system based on MC categories, I will call this number system the *Diachronic Numbering System.*

	*A	*B	*C	*DS	*DL
Voiceless Fricatives	A1	B1	C1	DS1	DL1
Voiceless Unaspirated Stops	A2	B2	C2	DS2	DL2
Pre-glottalised Onsets	A3	B3	C3	DS3	DL3
Voiced Onsets	A4	B4	C4	DS4	DL4

Figure 7.2: Gedney's (1989) Tone Box

In the study of Tai varieties, there is a similar diachronic numbering system to Chinese, since both language families have similar historical developments in tones (Zhang et al. 1999:24). Gedney (1989) has proposed a reference system of tones based on the Proto-Tai tonal system, which Gedney calls the *Tone Box.* The Proto-Tai tone system consists of four tones, A, B, C and lastly D, which is reserved for checked syllables. Each tone category is further split into four subcategories, based on the types of the onset of the syllable, namely 1) voiceless fricatives, 2) voiceless unaspirated stops, 3) preglottalised consonants and 4) voiced consonants. Tone D is also split into D-short and D-long, because historically, tones could split within Tone D based on the vowel length. Gedney's reference system has been used widely in comparative Tai dialectology for investigating tone splits in numerous Tai varieties. The Tone Box can be found in Figure 7.2. The use of this alphanumeric notation for tones can also be found in works of other languages, such as Vietnamese (e.g. Kirby 2011) and Hmong-Mien (e.g. Ratliff 2010). This will be labeled as the *Diachronic Alphanumeric System.*

There is another system very similar to the diachronic numbering system, which instead of omitting numbers (which makes the tone numbers look incoherent), uses a chronological numbering system for tones, even though not all reflexes of the tones are present. The tone marking system of Mandarin (Pinyin) is a representative example of this system.

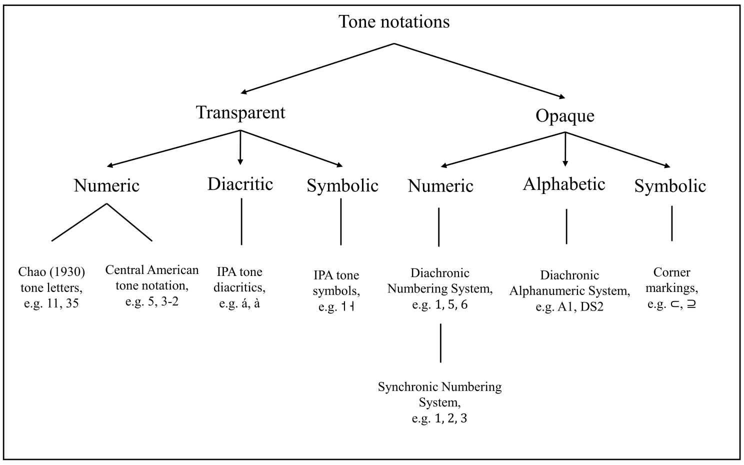


Figure 7.3: Typology of tone notations across traditions

Mandarin has *yin ping*, *yang ping*, *shang* and *qu* tones (in the traditional analysis), but these tones are referred to as Tone 1, 2, 3 and 4. The number still follows the order from the previous number system, except it replaces the incohesive numbers with the next digit. This discards the historical dimension and bases the numbering on synchronic tones. Therefore, I will call this numbering system the *Synchronic Numbering System.*

In summary, one can divide the tone notation systems into two families, transparent vs. opaque, and within each family, one can further differentiate between numeric, symbolic, diacritic and alphabetic representations of tones. The typology of tone notations is visualized in Figure 7.3.

7.3.2 Formal representations

Formal representations of tones attempt to model tone representations in the speakers' grammars. A detailed account of the formal approaches (in the 20th century) can be found in Bao (1999). This section will briefly highlight the representations introduced in Bao (1999) and point out the potential problems of these representations in the context of cross-dialectal comparison.

The earliest formal tone representation, which was proposed by Wang

(1967), relies on distinctive features. Wang proposed to use seven distinctive features to account for thirteen tones, which are listed below in Figure 7.4.

	1	2	3	4	5	6	7	8	9	10	11	12	13
	˥	˩	˦	˨	˧	˧˥	˩˧	˥˧	˧˩	˥˧˥	˧˩˧	˧˥˧	˩˧˩
Contour	-	-	-	-	-	+	+	+	+	+	+	+	+
High	+	-	+	-	-	+	-	+	-	+	-	+	-
Central	-	-	+	+	+	-	-	-	-	-	-	-	-
Mid	-	-	-	-	+	-	-	-	-	-	-	-	-
Rising	-	-	-	-	-	+	+	-	-	+	+	+	+
Falling	-	-	-	-	-	-	-	+	+	+	+	+	+
Convex	-	-	-	-	-	-	-	-	-	-	-	+	+

Figure 7.4: Wang's (1967) tone features (replicated)

Wang's tone features consist of height and contour parameters. The pitch height of the tone is determined by [High], [Mid] and [Central], whereas the shape of the tone is determined by [Contour], [Rising], [Falling] and [Convex]. Wang's proposal was critised by Sampson (1969), however, for a number of reasons. [Mid] was said to be too non-economical (we can see from Figure 7.4 that only one tone, Mid-Level, activates this feature). The purpose of establishing hypothesised universal features is to be used as widely as possible, which is not the case for [Mid] here. Sampson therefore made a small modification of Wang's features, namely replacing [Mid] with [Low], so that different combinations of the height features can differentiate five pitch levels. This is illustrated in Figure 7.5, from the highest pitch level to the lowest starting from the top.

More criticisms of Wang's tone features came from Woo (1969). Woo objects to the use of the contour features. Firstly, [Rising] and [Falling] are like instructions of a series of feature matrices, e.g. [-High] [+high] for [Rising]. Secondly, these two features are just the opposite feature of each other. This makes Woo question the validity of these features as distinctive features like the ones we find for segments. Furthermore,

[+ HIGH, - CENTRAL, - LOW]
[+ HIGH, + CENTRAL, - LOW]
[- HIGH, + CENTRAL, - LOW]
[- HIGH, + CENTRAL, + LOW]
[- HIGH, - CENTRAL, + LOW]

Figure 7.5: Sampson's (1969) modification of Wang's (1967) tone height features (remade)

[Contour] is only used to distinguish static and dynamic tones, but it is at the same time not dependent from [Rising] and [Falling]. [Convex] serves no purpose unless both [Rising] and [Falling] are activated, and it is a feature which only informs us the order of [Rising] and [Falling].

Rather than using features to represent a pitch transition, Woo proposed to use distinctive features for pitch height only, and contour tones are represented as sequences of these features. [High] and [Low] are kept from Wang's system, like Sampson (1969), but the intermediate levels are represented using [modify], against Wang's [Mid]. Woo's proposal is illustrated in Figure 7.6.

From 1970 onwards, tone features started to develop towards minimalism, meaning there was an attempt to represent tones as simplistically as possible. These representations do not convey as much information about the phonetic tone itself unlike Wang (1967), Sampson (1969) and Woo's (1969) features, but rather abstract representations which can be used to explore phonological theories.

Halle and Stevens (1971) proposed to use binary features which characterize the laryngeal effects of tones. These features include [Stiff], [Slack] and [Voice]. Unfortunately, this system only distinguishes three levels, H, M, L, which is not enough to account for complex tones which exist in Chinese dialects. Yip (1980) proposed the use of a register feature [upper] together with [raised] (represented as H and L) to distinguish tones. While Yip's data were mostly drawn from Chinese, reflected in the parallelisms with [upper] and *Yin* vs. *Yang* categories (see Section 7.3.1), the system can only differentiate four distinct levels (for level tones). For contour tones, each shape can have 2 variants, one for each

	˥	˦	˧	˨	˩	˥˥		˩˩		˧˧		˧˥		˩˥		˦˥		˩˨	
High tone	+	+	-	-	-	+	+	-	-	-	-	-	+	-	+	+	+	-	-
Low tone	-	-	-	+	+	-	-	+	+	-	-	-	-	+	-	-	-	+	-
Modify	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-

	˧˥		˥˧		˧˩		˥˦		˥˩		˧˩˥			˩˩˥		
High tone	-	+	+	-	-	-	+	+	+	-	-	-	-	-	-	-
Low tone	-	-	-	-	-	+	-	-	-	+	-	+	-	+	+	-
Modify	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-

Figure 7.6: Woo's (1969) proposal of tone features (remade)

register (see [upper]). This highly restrictive representation of tones only allows twelve distinct tones, which is not suitable for a cross-dialectal comparison.

Clements (1983) represents tones without the use of binary features, unlike previous representations. Instead, he uses tone matrices which consist of elements h, l, or 0 with number of rows, depending on how many levels the tone system has. The four-level tone matrix is similar to Yip's (1980) feature representation for a four-level (level) tone system. This shows that Clement's representation is only suitable for non-contour tonal languages. Clements's (1989) improvement of his tone representation expands on its theoretical implications on tones, but it does not improve the ability to distinguish enough tones for the purpose of dialect comparison of tones in Yue dialects.

The later developments of formal tone representations are similar to Yip (1980) and Clement (1983, 1989) and even Wang's (1967) representations, in the way that they also cannot distinguish enough types of tones for the purpose of tone distance calculation for cross-dialectal comparison. These representations include Hyman's (1986) autosegmental representation of tones, Shih's (1986) hierarchy, Inkelas's (1987) autosegmental approach based on African languages, Hyman's (1993) tonal geometry, Yip's (1989) model of tone features, Duanmu's (1990, 1994) structure of tone, Snider's (1990) model of tone, Chang's (1992) prosodic hierarchy and Tsay's (1994) proposal of tone classes.

7.3.3 Optimal representation of tones for measuring tone distances

In addition to the problems in defining the early phonological features of tones, formal representations in general are not suitable for cross-dialectal comparisons for the following reasons. No single language variety has the need for speakers to differentiate a huge number of tones (e.g. over 50). This has led phonologists to find feature systems which can only distinguish a handful of tones. In the later proposals, the features were based on African languages, where the tone contour does not vary as much as the tones found in Southeast Asia. These representations are even more unsuitable for a cross-dialectal comparison of tones for dialects of Southeast Asian languages.

An optimal representation (notation) for tones should aim to fit cross-dialectal (or cross-linguistic) comparisons. This is similar to finding a tone representation which is universal. This is because the goal here is to serve linguistic comparison, which means that we need to have a maximal amount of distinctions in order to account for all possible tones across the dialects in the data. Such a representation should also be as similar to the F0 trajectory as possible (because we have no recordings for the data, this kind of representation is the best we can have).

At the same time, such a representation should also make computation efficient. Opaque notations can be converted back to the transparent notations, but it creates an extra step for the computation of tone distances, which is not ideal.

To satisfy the maximal distinction and transparency criteria, Chao's tone letters seem to be the closest to what an ideal representation is out of the existing tone representations for the current purpose. Firstly, Chao's notation is transparent, and this representation can reflect the F0 trajectories the best. Secondly, it can distinguish much more tones than other notations. It is able to distinguish 5 contour levels (opposed to H, M and L), and it can reflect the contour by using a combination of digits. In terms of computing tone distances, the combination of digits would be more useful than a single character or diacritic representation of contour, since the more fine differences between two tones with the similar contours can be distinguished. For instance, 13 and 15 are both rising tones, but the steepness of the slope is reflected through the differences between the digits (contours). For the reasons listed above, the computation of tone distances in the following sections all starts from Chao's notation.

7.4 Previous approaches to measuring tone distances

Many studies in Southeast Asian tonal languages, such as Sinitic languages, are rather descriptive, i.e., listing the tonal inventory of a dialect or showing diachronic correspondences with historical tone categories, as in Zhan's (2002) *Introduction to the Yue dialects in Guangdong*, for example.

In dialectometry, to date, there are not many large-scale dialectometric studies on tones either. A number of studies looked at no more than 30 dialects (e.g. Yang and Castro 2008, Tang 2009). In some dialectometric studies, tones were neglected (e.g. Wichmann and Ran 2019), while others used a rather simplified method (e.g. Stanford 2012). In addition, there are studies on the correlation between phonetic distance and the perception of tones (e.g. Onset-Contour-Offset (OCO), Yang and Castro 2008), which do not focus on the application of these tone distance calculation methods on dialect classification.

In the following subsections, four existing tone distance calculation methods are introduced and their procedures are explained.

7.4.1 Onset-Contour-Offset

Onset-Contour-Offset (OCO hereafter) is a representation of tones proposed by Yang and Castro (2008). This representation gives a more phonetic representation of tones, instead of an abstract one, as its purpose is to approximate multiple cues of tones in the distance calculation methods in order to generate a more accurate prediction for intelligibility between dialects.

OCO involves a transformation of the tone letters (5-level transcription, Chao 1930) into a representation which consists of three components: *Onset*, *Contour* and *Offset*, each represented with one character, except for Contour, which can have up to two characters. Onset and Offset are the starting and ending contour levels of the tone, and the Contour is the shape of the tone. For the contour levels, the original 5-level transcription is converted into three categories, which are *H(igh)*, *M(id)* and *L(ow)*. H represents levels 4 and 5, M represents 3 and L represents 1 and 2. For contours, the basic shapes include *R(ising)*, *F(alling)*, *L(evel)*, and the complex tones are represented by the combination of the basic shapes, hence it has up to two characters.[5] In addition, phonetic short tones are not distinguished. Examples of the Contour representations can be found in Table 7.1.

Representation	Contour	Examples
L	Level	11, 33
R	Rising	12, 35
F	Falling	31, 52
RF	Convex	131, 253
FR	Concave	213, 424

Table 7.1: Contours in OCO representation with examples

As an example, the OCO representation of 221 would be LLFL, and for 24, it would be LRH. To calculate tone distances, Yang and Castro (2008) applied the Levenshtein distance algorithm on the OCO representation. This is illustrated in Table 7.2.

When two tones with different lengths are compared (length of three and four, like in Table 7.2, the Onset (Slot 1) and Offset (Slot 4) are always aligned together. In this example, we can find two substitutions and one deletion out of four alignment slots. This yields a (3 / 4 =) 0.75 difference between the tone pair.

Last but not least, the aggregated distance is calculated by summing

---

[5] The complex tones are additions to Yang and Castro's original proposal, as the original article did not account for these types of tone, but they are present in the dataset.

Slot 1	Slot 2	Slot 3	Slot 4	Operations	Distance
L	L	F	L	-	-
L	R	F	L	Substitution of L > R	1
L	R	_	L	Deletion of F	1
L	R		H	Substitution of L > H	1
Sum					3

Table 7.2: Calculation of Levenshtein Distance between 221 and 24 in OCO representation

the tone distances of the tone pairs and dividing the sum by the number of words compared in pairwise comparison between two dialects. This is done for any dialect pair in the dataset. For the rest of the dissertation, 'OCO' (or the 'OCO method') is used to refer to the aggregate distance calculation using OCO as the tone representation, combined with the application of Levenshtein distance.

OCO has shown high correlation with mutual intelligibility for both Zhuang and Bai in Yang and Castro's (2008) study. Do and Lai's (2021) results basically agree with Yang and Castro. However, Tang (2009:122-125) found that OCO fails to yield a classification which can differentiate Mandarin from other Sinitic languages. Based on this result, Tang (2009:125) argued that this representation does not "provide a handle on dialect affinity".

It should be mentioned that in Yang and Castro (2008), a related representation, Onset-Contour, correlates significantly higher than OCO with intelligibility as a predictor for Zhuang. However, this representation can differentiate even fewer tones than OCO, which makes it even more unsuitable for dialectometry. Hence, in this chapter, the focus is on the OCO representation.

7.4.2 Tone-to-string

The tone-to-string method (Tang 2009) applies the Levenshtein distance algorithm directly on Chao (1930) tone letters. The differences in the digits are not accounted in this method, i.e. a substitution from 2 to 1 costs the same distance as from 4 to 1. In addition, when a two-digit tone is compared with a three-digit tone, the first digit of the two-digit tone aligns with the second digit of the three-digit tone[6], as shown in the example below. Note that in this approach, short tones are not distinguished from the other tones in the dataset.

The example in Table 7.3 illustrates how tones 325 and 15 are aligned and the distance between them is calculated using the Levenshtein algorithm. In this example, one substitution and deletion are required to convert 325 to 15, which yields distance of (2 / 3 =) 0.67 between two tones.

Slot 1	Slot 2	Slot 3	Operations	Distance
3	2	5	-	-
__	2	5	Deletion of 3	1
	1	5	Substitution of 2 > 1	1
Sum				2

Table 7.3: Calculation of Levenshtein Distance between 325 and 15 with tone-to-string method

The aggregated distance is again calculated as the sum of the tone distances of the word pairs, divided by the number of words compared in the pairwise comparison between two varieties. Thus, the aggregated distance is calculated for any pair of varieties in the dataset.

For the rest of the dissertation, 'Tone-to-string' is used to refer to the aggregate distance calculation using Chao's (1930) tone letters as the tone representation, combined with the application of Levenshtein distance.

Yang and Castro (2008) found that applying Levenshtein distance directly on Chao (1930) tone letters has a lower correlation to mutual intelligibility than OCO. Tang's (2009) evaluation also shows that this method yields quite a high number of misclassifications when compared to the traditional classification (separating Mandarin from other Sinitic languages).

7.4.3 Binary comparison

Cheng (1997:53) proposed to calculate the similarities between dialects by measuring the ratio of shared items to all the items (segments and tones) that are occurring in both varieties in a pairwise comparison.

---

[6] This alignment is based on Tang (2009).

The data matrix consists of the binary encoding of the items, with *1* (present) and *0* (absent), and shared items are features encoded with 1 in both dialects. This is similar to Goebl's (1984) *Relative Identity Value* (see Chapter 3), except the data matrix is comprised using the bag-of-words approach (based on the sound inventory of all dialects in the data). Tang (2009:105-106) has found that this method yields a number of misclassifications when compared to the traditional taxonomy (Mandarin vs. other Sinitic languages).

In addition to the binary bag-of-words approach, Tang (2009:114-115) weighted the tone inventories by lexical frequency (out of 764 items in the database). This method yields one misclassification more than the binary inventory comparison. Another related approach by Cheng (1991:88-89) includes Middle Chinese sound categories (as indicated by the ancient rhyme dictionaries *Guangyun*[7]) in the data matrix, in addition to adding lexical weighting and comparing tone inventories, using more than 2700 words. The compared elements are not simply the synchronic sound segments in the inventories, but reflexes of certain Middle Chinese sound categories. For example, instead of stating Dialect A has 79 tokens of [p^h], Dialect A now has 40 tokens of [p^h] as a reflex of *p^h and and 39 tokens of [p^h] as a reflex of *b. Tang's (2009:131-132) evaluation shows that the result from Cheng's method highly resembles the traditional classification. Due to the low number of items (which makes it not reliable with Cheng's method) and a lack of Proto-Yue reconstruction for a number of items in the current dataset, the weighted lexical frequency method will not be used in the analysis. Instead, the binary comparison approach is used.

Comparing the tonal inventory (bag-of-words) between dialects might be too simplistic, since dialects could share the same tonal inventory (with the same phonetic tone values), but the lexical distribution (Wells 1982:78) of the tones could be different. For example, the Taishan and Kaiping dialects are members of the Siyi dialects which have an identical tonal inventory, with 5 tonemes which are phonetically the same (Zhan and Cheung 1990: 85-87). However, not all lexical items share the same tones. Table 7.4 has listed three examples where each item has a different tone in each dialect.

In order to account for the lexical distribution differences between

---

[7] *Guangyun* is a rhyme dictionary from the Middle Chinese period. In Chinese dialectology, it is often used as a proto-system for modern Chinese dialects (You 2016:85-86).

Item	Taishan	Kaiping
First syllable of ‘lychee’ 荔	lai^{31}	lai^{21}
‘inner’ 内	^{n}dui^{31}	lui^{22}
‘caldron’ 釜	fu^{21}	fu^{55}

Table 7.4: Examples of lexical distribution differences in tones of two Siyi dialects

the dialects, the binary comparison has been modified from the inventory level (comparing differences in the tonemic inventory only) to the lexical level (comparing tone differences in word pairs). Instead of comparing the proportion of elements (tones) shared, the tone distance for each lexical item is calculated by identifying whether the tone is identical (distance = 0) or not (distance = 1). An overall aggregate distance for each dialect pair is then calculated by summing all the items with a ‘1’, then dividing the sum by the total number of items compared.

The binary method is a harsh measure which implies that there is no difference in the distance between similar and very different tones, i.e. 11 vs. 22 and 11 vs. 523 are equally distant.

For the rest of the dissertation, ‘Binary comparison’ is used to refer to the aggregate distance calculation using Chao’s (1930) tone letters as the tone representation, followed by the binary comparison.

7.4.4 Gandour-Harshman-Tang tone distance measurement

Unlike the previous methods in Sections 7.4.1-7.4.3, the Gandour-Harshman-Tang tone distance measurement is based on the perceptual experiment of tone distances published in Gandour and Harshman (1978). Their stimuli included a range of tones, which differ in contour, pitch and duration, and they asked listeners to rate their distances. Based on the perceived distances of the listeners (with Thai, Yoruba and English backgrounds), five (MDS) dimensions have been extracted and interpreted as *Average Height*, *Direction* (Rising, Level, Falling), *Length*, *Extreme End Point* and *Slope*. These dimensions are underlying ‘cues’ which can explain the variance found in the perceptual tone distance matrix and have been implemented as multivalued features by Tang (2009:125-126) as a way to measure objective tone distances. This is the tone distance measure which accounts for perception, unlike other

Cue	Features	Feature value	Maximal difference
Average Height	Lower than 2.5	1	4
	Between 2.5 and 3.5	3	
	Higher than 3.5	5	
Direction[8]	Falling	0	2
	No change/ one digit tone	1	
	Rising	2	
Duration	1-digit (or short 2-digit) tone	1	2
	2-digit (or short 3-digit) tone	2	
	3-digit tone	3	
Slope[9]	Difference smaller than 3	0	1
	Difference bigger than 3	1	
Extreme endpoint[10]	Ends with 2, 3 or 4	0	1
	Ends with 1 or 5	1	

Table 7.5: Cues and their relative feature values in GH-T

methods, which have been criticized as “very crude and unrealistic”, since any substitution or changes in the pitch and contour have equal importance (Tang 2009:125). However, a method as sophisticated as this still gave “highly unsatisfactory results”according to Tang’s (2009: 127) evaluation method, since it also cannot differentiate Mandarin from other Sinitic languages, just like OCO.

Based on the contour of the tone, a feature value is assigned for each cue under the criteria listed in Table 7.5, following Tang’s (2009: 125) implementation. It should be noted that in Tang’s dissertation, only the last two digits of the tone are taken into account for Direction and Slope, which means that complex tones and their counterparts without their onset element will share the same values for these two cues.

It should also be noted that a slight modification has been made to Tang’s (2009) implementation in this thesis. For Direction, instead of assigning the values *0*, *1*, *2* to Level, Falling, and Rising, respectively, I have made Level the middle value (i.e. *1*). This change can make

---

[5] Last two digits.

[6] Last two digits.

[10] Last digit only.

Tone string	Height	Direction	Duration	Slope	Extreme end point	Total
551	5	0	3	1	1	
12#	1	2	1	0	0	
Difference	4	2	2	1	1	10

Table 7.6: Example of tone distance calculation using GH-T

the distances between the directions more continuous (Rising-Level-Falling forming a continuum), rather than being arbitrary categorical values. Furthermore, since this method is Tang's (2009) implementation of the perceptual dimensions found by Gandour and Harshman (1978), this method will be addressed as the GH-T hereafter, which stands for Gandour-Harshman-Tang.

Here is an example how tone distance is calculated with GH-T, which is illustrated in Table 7.6. The maximum difference two tones can have is 10, based on sum of the maximal difference each cue can have ($4 + 2 + 2 + 1 + 1$). Take 551 and 12# as an example (# indicates a shorter duration of the tone, often found in checked syllables), their feature specifications of these tones are listed in Table 7.6. After deducting the absolute difference of each feature value and summing the differences, this value is the distance between the tone pair. 551 and 12# is one of the possible pairs of tones with the greatest distance in Chao's (1930) transcription system.

For the computation of GH-T, the Manhattan distance has been used. Manhattan distance in this context is the sum of absolute differences of the values for all the cues shown in Table 7.5 (maximum distance is 10). In addition, all the differences are divided by 10 so that the range of differences are kept between 0 and 1, just like the other methods in the previous subsections. Again, the overall aggregated difference is calculated between all dialect pairs in the data.

For the rest of the dissertation, 'GH-T' is used to refer to the aggregate distance calculation using Chao's (1930) tone letters as the tone representation, followed by the calculation of the Manhattan distance of the feature values these tone letters represent according to GH-T.

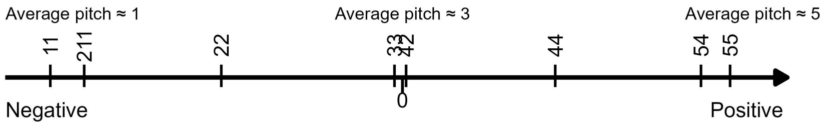
7.5 Comparison of the tone distance calculation methods

In the following subsections, a comparison is made with the four tone distance calculation methods introduced in the previous section based on i) the number of tones distinguished, ii) comparison with perceptual dimensions (Gandour and Harshman 1978), iii) local incoherence, iv) comparison with traditional classification, and v) comparison with segmental classification (see Chapter 5).

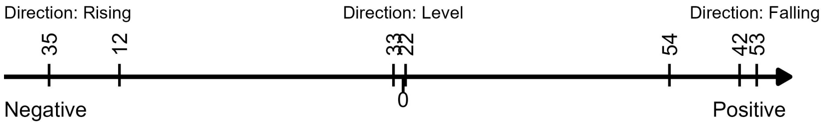
The purpose of the comparison is to use exploratory techniques to assess the current state-of-the-art approaches and see how adequate these methods are at dealing with a larger dialect dataset (with a bigger diversity of tones), as well as seeking areas which need to be improved for the purpose of dialectometry, instead of directly applying any of these methods to a dataset which is largely different from previous studies.

An ideal tone distance measure is able to distinguish all the (attested) tones, including those found in the data. By computing the distances between all the tones in the Yue dataset using the methods introduced in Section 7.4, we can visualize these distances on an MDS plot and inspect how much overlap there is between the converted tone representations, i.e. how much do tones (Chao's tone letters) share the same representation under each method. In addition, we can interpret the first two dimensions of the MDS plot and check whether they match Gandour and Harshman (1978) perceptual dimensions of tones. Previous studies in dialectometry have validated their methods through comparing the linguistic distances with perceptual distances (e.g. Gooskens and Heeringa 2004). This study follows the same logic. For tones, Gandour and Harshman (1978) and Gandour (1983) have repeatedly found average pitch and direction are the most important dimensions in tone distance perception of native speakers from several tone languages. An ideal tone distance measurement method should also yield similar dimensions in the tone distances that each method produces. Figure 7.7 illustrates what the MDS dimensions would look like if the tested method matches the perceptual dimensions.

Local incoherence is another way to compare different distance calculation methods used in dialectometry. Previous studies have shown that dialects which are geographically closer to each other tend to be more similar than distant ones (the *Fundamental Dialectological Postu-*



(a) *Average pitch*



(b) *Direction*

Figure 7.7: Illustration of the two most important perceptual dimensions of tones

late, Nerbonne and Kleiweg 2007). This pattern has been found not only within dialectal variation, but also in other domains (in Geography, it is known as *Tobler's First Law of Geography*, Tobler 1970). The idea behind local incoherence is to measure how much dialectal variation (in terms of the distances between the nearest dialects from any locality) matches the tendency stated above. The optimal score is of local incoherence is 0. According to Nerbonne and Kleiweg (2007), a method that gets a small local incoherence score is considered to be more suitable.

Lastly, it is useful to compare how the classifications generated by each method differ from the traditional (see Chapter 2) and segmental classifications (from segmental dialectometry, see Chapter 5). The comparison can inform us whether tones behave similarly to segments as well as whether tones were considered in the classification of dialects in the LAC (Chinese Academy of Social Sciences (CASS) 2012), which may not have been explicitly stated.

The comparison between the tonal classifications and the traditional classification or the segmental classification requires additional steps. Firstly, cluster analysis (Ward's method, Ward 1963) was performed on the segmental classification as well as the classification for each tone distance measure. The traditional classification in the *Language Atlas of China* originally gives 10 different groups. However, in the current dataset, one of them consist of only two dialects (Suixi and Litang under 'Unclassified'[11]); the other two groups consist of one dialect (Huazhou (Hexijie) under 'Wuhua' and Lingui under 'Guibei Pinghua'). To avoid

---

[11] The LAC did not include this locality in their classification.

having clusters (in the traditional classification) consisting of only one element, some modifications of the LAC grouping have been made. Suixi and Huazhou (Hexijie) dialects are now classified under the Gaoyang dialect group, whereas Litang and Lingui are now labelled as Guinan Pinghua, based on their geographical proximity.

To compare the similarity of the classifications (dialect groups) generated by the different approaches, the *Adjusted Rand Index* is used as an indicator to measure how much each cluster overlaps between a Reference Classification and an Observed Solution.[12] The Adjusted Rand Index (ARI hereafter) is a method used for comparing two different clustering solutions (with chance correction, Hubert and Arabie 1985), which derived from the Rand Index (Rand 1971). This analysis assumes that the traditional classification and the segmental classification are Reference Classifications, and the classification generated with each tone distance measure are the Observed Solutions, which are compared against the Reference Classification. If two classifications completely overlap, the score is 1, if they only overlap on the chance level, the score is 0. Furthermore, the ARI can also measure the overlapping of clusters among the tone distance calculation methods.

7.5.1 Tone overlaps

In this and the next subsection, I will present MDS plots to visualize the distances between all the tones in the Yue data under different tone distance measures introduced in Section 7.4. MDS plots are used to represent the distances between tones (represented by labelled points in the plot), and the further the points are from each other, the more different they are. The MDS plot for each method[13] is shown from 7.8 to 7.10.[14] These plots are very insightful in showing tone overlaps and are suitable for making comparisons with perceptual dimensions (Gandour and Harshman 1978; Gandour 1983). Note that unless specified, classical

---

[12] In the literature, the reference classification is known as the Gold Standard. However, since the classifications used are not the absolute correct solution in this context (since there is no one 'correct' solution in dialectology), an alternate name is used instead.

[13] Except for Binary method, since it is not possible to visualise the kind of distances measured using this approach. See below.

[14] Plots and explained variances were produced and calculated with *LED-A* (Heeringa et al. 2024), except Figure 7.10.

Method	Tones differentiated	Distinction rate
Binary	73	100%
Tone-to-string	52	71.2%
OCO	32	43.8%
GH-T	40	54.8%

Table 7.7: Differentiation of tones under each tone distance measure

MDS has been used in the analysis.[15]

The first evaluation on the tone distance calculation methods is based on the number of overlapping representations of tones in the dataset. Tone overlaps are calculated by dividing the number of unique tone representations generated from each method by the total number of tones in the data. Additionally, some tone overlaps can also be identified through the lines connecting to the tone labels to the dots on the MDS plots (see Figure 7.8-7.10). Table 7.7 summarizes the number of tones each method distinguishes from the original dataset, in both raw token and percentage. The total number of distinctive tones in the data is 73.

In Table 7.7, we can see that the Binary method yields full a distinction between all the tones in the data, and Tone-to-string can differentiate slightly over 70% of the tones in the data. For GH-T, only 54.8% of the tones can be differentiated, and lastly, OCO performs the poorest out of all methods, with under half of the tones in the data only.

The overlap of tones for the Tone-to-string method mainly occurs in cases where there is long and short distinction (indicated by the '#' in Figure 7.8). Both Yang and Castro (2008) and Tang (2009) did not specify how they dealt with short phonetic tones, therefore, the current analysis has removed the short tone marker in the data. For the OCO method (Figure 7.9), in addition to the lack of length distinction, the contrasts between several tones with similar contour have been collapsed. For example, we can see a cluster of six tones on the right consisting of 41, 42, 42#, 51, 52, 52# sharing the same OCO representation. This overlap is the result of collapsing a five-level contour transcription system into a three-level representation as well as not representing length distinction.

---

[15] Kruskal's non-metric MDS sees the distances in the distance matrix in an ordinal fashion (as rankings), whereas 'distances' are meaningful to classical MDS. Since dialect distances are not merely rankings in the analysis, classical MDS has been used.

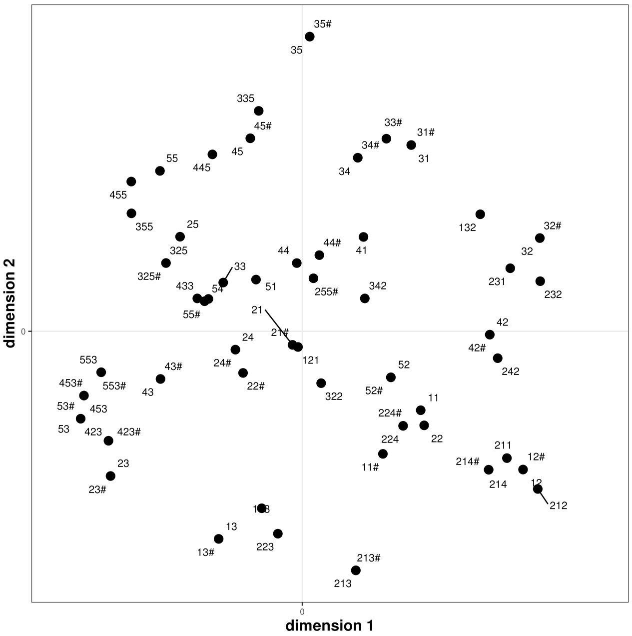


Figure 7.8: MDS plot for Tone-to-string, $r^2 = 0.26$

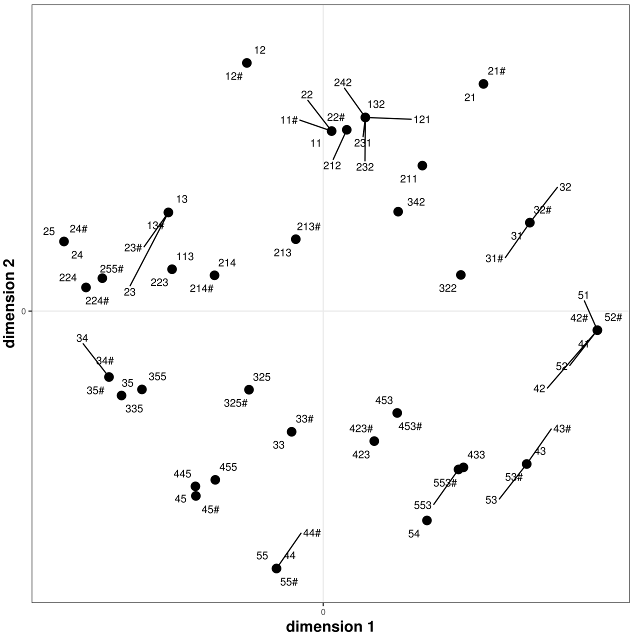


Figure 7.9: MDS plot for OCO, $r^2 = 0.59$

Lastly, with the GH-T method (Figure 7.10), there are also overlapping tones with similar contours (e.g. the cluster at the top, consisting of 42, 32, 43) or tones which share the same latter two digits (e.g. the cluster at the bottom consisting of 23, 24, 224#, 423#). This is due to how tones are encoded with their features as well as how tones with two and three digits align (see 7.4.4).

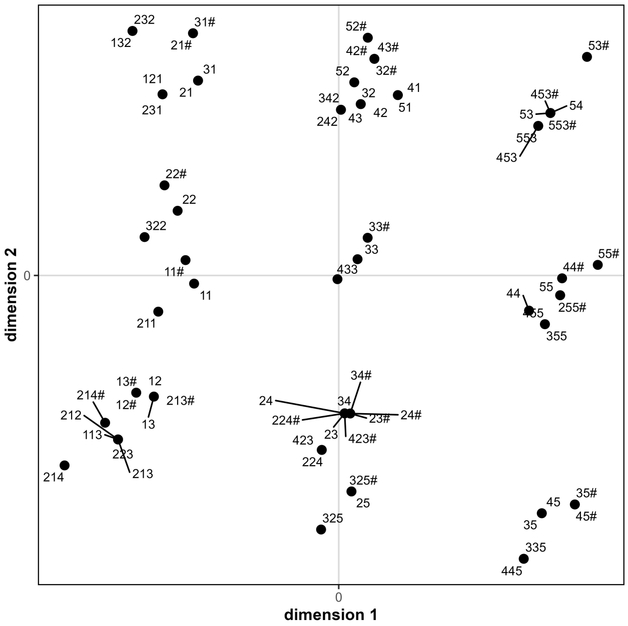


Figure 7.10: MDS plot for GH-T, $r^2 = 0.74$

7.5.2 Comparison with the perceptual dimensions

Gandour and Harshman (1978) and Gandour (1983) have consistently found that cross-linguistically, the two major perceptual dimensions in tones are average pitch and direction (see Figure 7.7a and 7.7b as references). A comparison between Gandour and Harshman's perceptual dimensions and the first two dimensions extracted from the MDS plots are summarised in Table 7.8 below.

Firstly, for the Binary method, it is not possible to visualise the distances between the tones. This is because the distances between all the pairs of tones are the same (all of them being 1), and it is impossible

Method	Dimension 1	Dimension 2
Gandour and Harshman (1978), Gandour (1983)	Average Pitch	Direction
Binary	Uninterpretable	Uninterpretable
Tone-to-string	Uninterpretable	Uninterpretable
OCO	Direction	Average Pitch
GH-T	Average Pitch	Direction

Table 7.8: Interpretation of the first two dimensions in the MDS plots of each tone distance measure

to project them in 2-dimenional space (i.e. an MDS plot). Therefore, it is also impossible to extract and interpret the first two dimensions of the distances and make a comparison with the perceptual dimensions. The MDS plot for Tone-to-string (Figure 7.8) is also uninterpretable if we only focus on Dimensions 1 and 2.[16] However, tones in Figure 7.8 are clustered together loosely by the pitch level that the tone ends in. This suggests the plot is a product of the operations and calculations on the tone letters, instead of the actual linguistic components (pitch, contour) of the tones.

OCO (Figure 7.9) shows a clearer picture than the previous two methods. In Dimension 1, we can see that on the right of the y-axis, the tone contours tend to be falling (e.g. 32, 52 and 53), and on the left, the contours tend to be rising (e.g. 12, 25, 45). For the level tones, they are situated near the middle (y-axis). In Dimension 2, we can see that the tones above the x-axis tend to have a low onset, and their average pitches are mostly below 2.5. On the other hand, the tones under the x-axis are the opposite. Lastly, the GH-T plot (Figure 7.10) shows nine clear clusters across the plot. The first dimension also shows average pitch differences; if we look at the dots around the x-axis, we can easily identify that the low-level tones are clustered on the left, mid-level tones are clustered in the middle, and high-level tones on the right. The second

---

[16] In terms of the procedures, it is possible to extract dimensions beyond dimension 2. With three and four dimensions, the cumulative explained variances are 0.40 and 0.55 respectively. However, if the first dimension, which has the highest variance explained for the distance matrix, is not even interpretable, the higher dimensions, which have lower explained variances, probably will not yield more insights into the validity of the method.

dimension also quite clearly suggests direction. If we focus on the dots around the y-axis, we will discover that the clusters of tones go from falling to level to rising, starting from the top.

7.5.3 Local incoherence

Local incoherence measures how different the geographical distances of the physical neighbours and the geographical distances of the 'neighbours' sorted by linguistic distances are, in order to see how much linguistic distance conforms to the *Fundamental Dialectological Postulate* (Nerbonne and Kleiweg 2007).[17] The local incoherence scores of each method are calculated using *Gabmap* (Nerbonne et al. 2011; Leinonen et al. 2016), with an optimal score being 0. It is generally considered that the lower the local incoherence score, the more suitable the distance calculation method is (Nerbonne and Kleiweg 2007). The results are shown in Table 7.9 below.

Distance measure	Local incoherence
OCO	4.51
Tone-to-string	3.63
Binary comparison	3.74
GH-T	4.36
Segments (classic Levenshtein)	2.10

Table 7.9: Local Incoherence scores for each tone distance measure and segments. Lower values represent better fit to the Fundamental Dialectological Postulate.

In Table 7.9, we see that none of the tone distance calculation methods gets closer to 0 than the segmental level. Amongst the methods, Tone-to-string gets the lowest score, although it is not much lower than Binary comparison. OCO appears to be the method that produces the highest local incoherence, compared to the more linguistically naïve methods, while GH-T sits in between Binary and OCO.

In general, all tone distance measures have higher local incoherence than segments. This suggests two possibilities. The first possibility is

---

[17] There are other ways to calculate local incoherence, e.g. Lameli and Schönberg (2023), where the degree of dissimilarity a reference location to its neighbours is measured.

that none of these measures are currently good enough to capture geographical variation of tones. The second possibility is that tones behave differently from segments, i.e. tones do not follow the *Fundamental Dialectological Postulate*. An additional contribution to a high local incoherence may come from similar varieties (the Guangfu and Yongxun dialects, see Chapter 5) that are distributed a long distance apart, with the Yongxun dialects being surrounded by dissimilar dialects in the west. In order to explore these possibilities further, the tone distance calculation method should first be refined, so that inherent problems of each method can be eliminated from the investigation. In addition, one can focus on a smaller region only (e.g. Guangdong), in order to avoid the irregularity of distance introduced by the Yongxun dialects in Guangxi.

7.5.4 Adjusted Rand Index (ARI)

The cluster solutions of each of the tone distance measures (performed in *Gabmap* (Nerbonne et al. 2011; Leinonen et al. 2016)) are compared to the traditional taxonomy from the *Language Atlas of China* as well as the segmental distances (calculated with classical Levenshtein distance with *Gabmap*) by using the ARI. ARIs indicate how much cluster solutions overlap with the Reference Classification (see Chapter 3). Results are shown in Table 7.10.

	LAC	Segments	Binary	Tone-to-string	OCO	GH-T
LAC	1	0.46	0.09	0.20	0.23	0.16
Segments	0.46	1	0.13	0.17	0.19	0.14
Binary	0.09	0.13	1	0.42	0.32	0.34
Tone-to-string	0.20	0.17	0.42	1	0.44	0.42
OCO	0.23	0.19	0.32	0.44	1	0.5
GH-T	0.16	0.14	0.34	0.42	0.5	1

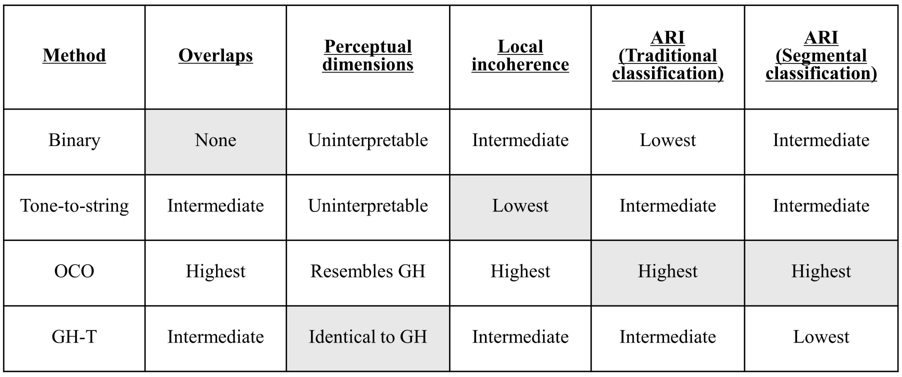
Table 7.10: ARI Similarity matrix between tone distance measures, traditional classification and segmental classification. Higher scores represent more overlap.

The ARI score of 1 represents a complete overlap of cluster solutions between two classifications. In Table 7.10, the highest ARI score belongs to the pair of OCO and GH-T. In fact, OCO, GH-T and the Tone-to-string method all share high resemblance in terms of the classifications they yield. In fact, OCO and GH-T have even higher ARI scores than the traditional classification and the segmental classifications.

When the Reference Classification is set to LAC, the closest classification to the traditional taxonomy is the classical Levenshtein segmental classification, with a score of 0.46. In terms of tonal classifications, OCO had a score of 0.23, followed by Tone-to-string with 0.20. Between the segmental classification and the tone distance calculation methods, OCO resembles segments the most, with an ARI score of 0.19, while Tone-to-string follows with an ARI of 0.17. Overall, the Binary method has the lowest ARI score with almost all classifications. Although OCO and Tone-to-string resemble the traditional and segmental classifications the most respectively, their ARI scores are rather low, suggesting rather dissimilar classifications.

7.6 Summary of the results and discussion

The overall results are summarized in Figure 7.11



Method	**Overlaps**	**Perceptual dimensions**	**Local incoherence**	**ARI (Traditional classification)**	**ARI (Segmental classification)**
Binary	None	Uninterpretable	Intermediate	Lowest	Intermediate
Tone-to-string	Intermediate	Uninterpretable	Lowest	Intermediate	Intermediate
OCO	Highest	Resembles GH	Highest	Highest	Highest
GH-T	Intermediate	Identical to GH	Intermediate	Intermediate	Lowest

Figure 7.11: Overall results of evaluation across 4 tone distance calculation methods

Which method to choose?

Each method seems to be more suitable for one specific task than other methods. OCO has shown the best result with respect to the ARI scores, while others performed best in one task each. This does not automatically make OCO the go-to method for the dialect classification of tones, though. OCO suffers the most from the overlapping representations, whilst the Binary method could distinguish all tones. The Binary method, however, cannot differentiate similar vs. very distinct tones, which makes the whole distance calculation rather arbitrary. For this reason, the Binary method is not suggested to be used in dialectometry.

In terms of perceptual dimensions, OCO shows resemblance to Gandour and Harshman (1978)'s findings. Although the relative importance of the dimensions is not in the same order, as reflected in the MDS plot (see 7.9), OCO still has some potentials in being a more suitable tone distance measurement method than others.

Looking at local incoherence, the best performing method is Tone-to-string, but like the Binary method, this method is also linguistically naïve. In addition, its first two dimensions on the MDS are uninterpretable, hence this method is also not recommended in dialectometry. Furthermore, readers should be reminded that local incoherence should be taken with a grain of salt here, as mentioned in Section 7.5.3: The possibility that tones do not behave like segments and the dialect islands (Yongxun dialects) in the Guangxi province surrounded by dissimilar dialects both might contribute to the higher local incoherence.

As suggested in Section 7.5.4 and based on the observation that there is low resemblance between the traditional, segmental and tonal (OCO, GH-T) classifications, there is a possibility that the nature of tonal variation is different from segments. Therefore, we cannot simply use the ARI scores as a determining factor for choosing one method over another. OCO shares the highest ARI score with GH-T (and to some extent, the perceptual dimensions), which makes both methods valid to be used in dialectometry, provided that some improvements are made (e.g. increase the distinction rate). However, there is one consideration which makes us favour one method over another – the ability to combine tonal and segmental analysis. The segmental analysis is done using Levenshtein Distance, as is the OCO method. Ideally, the method for calculating tone distances is not limited to a tone-only analysis, but also a combined analysis with segments. This will not be possible with GH-T, since it requires a separate specification of tone features. For this reason, I argue that the OCO method should be chosen and further improved.

Possible improvements

The most fundamental problem OCO has is not being able to distinguish the tones in the dataset (not even 50%). The first thing one could do is to find a way to take duration into consideration. Secondly, one could modify the three-level distinction into five, since five seems to be the maximum number of levels used by any tone language (Yip 2002:20).

Another issue which has been pointed out is that some tones, despite having different contours, are still yielding the same distance using OCO. The above can be illustrated with two tone pairs, 53 vs. 31 and 12 vs. 31. The former pairs are Falling tones (with the same slope), only differing in pitch levels. The latter pair are Rising and Falling tones respectively, and they have different degrees of slope. Many linguists would agree that one tone should be more different from 31 than the other. If one uses the OCO representation to calculate the tone distances for these two tone pairs, the distances OCO yields will be identical. This is illustrated in Figure 7.12.

	53 vs. 31				12 vs. 31		
	Onset	**Contour**	**Offset**		**Onset**	**Contour**	**Offset**
53	H	F	M	12	L	R	L
1 subs.	**M**	F	M	1 subs.	**M**	R	L
1 subs.	M	F	**L**	1 subs.	M	**F**	L
31	M	F	L	31	M	F	L
Total: 2/3				Total: 2/3			

Figure 7.12: Comparison of tones 53 and 12 with 31

In order to distinguish differences in these cases, perhaps weighting (onset, offset and/or contour) should be tested on each parameter.[18]

---

[18] A reviewer has suggested introducing *steps* when dealing with substitution in the OCO representation. For example, if Onset or Offset has a High, then a substitution to Mid would be 1, but to Low would be 2. Similarly, for Contour, a substitution from

Alternatively, other features (based on perceptual cues, see Gandour and Harshman 1978) could be explored.

Tonal variation in Yue and Pinghua

Lastly, I would like to make a remark on the patterns of tonal variation in the Yue and Pinghua-speaking so far. Although in general the ARI scores of the classifications from all the tone distance calculation methods and the segmental classification are relatively low, by visual inspection of the cluster maps (see Appendix C), Tone-to-string, OCO and GH-T yield some clusters which strongly resemble the segmental clusters, as well as the LAC. For instance, Siyi dialects (consisting of 5 localities, Taishan, Kaiping, Enping, Xinhui and Doumen) can be identified, which align with both segmental and traditional classification. Another notable observation for the same methods is that we can see the East-most cluster (traditional Guangfu dialects) spread westward towards Guangxi. This pattern matches the segmental analysis, and can be explained by historical migration around 150 years ago (de Sousa 2020: 268). The other patterns are yet to be explained.

To further understand the tonal variation of Yue and Pinghua (and any other tonal languages), therefore, one first needs an improved tone distance calculation method. That will allow us to find out how tones differ between dialects (e.g. is it gradual or abrupt?), to explore whether tones may behave differently from segments (as I have suggested in 7.5.3 with local incoherence) as well as to explore the possibility of combining segments and tones in the same analysis.

7.7 Conclusion

Traditionally dialectology has focused on segments, and not much attention has been paid to tones, despite the wealth of data linguists have on tone languages, such as Sinitic languages. The limited research in exploring tone distance measures has only been tested on small datasets. Additionally, there is a lack of systematic comparisons on the existing methods for the purpose of dialect classification. This chapter offers such systematic comparison, with a dataset which comprises of 104 dialects.

Rising to Level would be 1, but to Falling would be 2.

The results show that the OCO and GH-T representations make meaningful contributions to dialectometry, but they are not yet ready for the application in measuring tonal distances. A few possible improvements have been suggested and the coherence problem was discussed in relation to the tone distance calculation methods and ways which tones might vary.

Throughout the comparison, it was repeatedly seen that tones do not follow the same variation patterns as segments. Perhaps tonal variation really has a different nature in how it varies geographically. This is a rather unexplored area which awaits more research and improved methodology, and I hope that this chapter is the first step in that direction.