

TONE AS A PREDICTOR OF MUTUAL INTELLIGIBILITY OF CHINESE DIALECTS

Chaoju Tang^a & Vincent J. van Heuven^b

^aSchool of Foreign Languages, University of Electronic Science and Technology of China, Chengdu, China;

^bPhonetics Laboratory, Leiden University Centre for Linguistics (LUCL), the Netherlands
chaoju.tang@gmail.com; V.J.J.P.van.Heuven@hum.leidenuniv.nl

ABSTRACT

This paper attempts to establish the importance of differences in the tone inventories for the mutual intelligibility of 15 Chinese dialects. Tone influence for mutual intelligibility was tested in two studies. In the first study mutual intelligibility was measured through opinion testing. The fable “The North Wind and the Sun” was read by one male speaker in each of 15 Chinese dialects with melody and was subsequently monotonized. Both versions were presented to groups of 24 listeners for each of the 15 dialects. In this way we obtained intelligibility judgments for all 225 combinations of speaker and listener dialects. The results show that the absence of pitch in the monotonized version only marginally affected the intelligibility judgments. In the second study we determined the mutual intelligibility of the same 15 dialects (again 225 combinations of speaker and listener dialects) using functional intelligibility tests (recognition of words in isolation and in sentences). The intelligibility scores were then correlated with tonal distance measures computed in three different ways on representative word lists for the 15 dialects: (i) Levenshtein string edit distance on 3-digit tone transcriptions, (ii) Levenshtein distance with symbols for starting level and tonal change, and (iii) a perceptually weighted tone distance measure. None of the distance measures correlated with the intelligibility scores, so that our overall conclusion is that differences in lexical tones contribute little to the mutual intelligibility of Chinese dialects.

Keywords: tone, mutual intelligibility, Chinese dialects, Levenshtein distance

1. INTRODUCTION

Mutual intelligibility of dialects can be measured experimentally through opinion tests and functional tests [8, 9, 10]. Opinion tests measure

how well the hearer thinks s/he understands the other dialects. The testing methods are to collect the opinion scores via designed experiments and these methods are often proposed as a shortcut for functional testing. Functional tests measure how well the hearer actually understands the other dialect, in terms of percentage of correctly recognized or translated words between pairs of dialects. The intelligibility scores obtained can be correlated with various structural distance measures between dialects, e.g. lexical distance (percentage of cognates shared between two languages/dialects), degree of phonological/phonetic similarity between cognates (e.g. by Levenshtein distance), etc., to see to what extent the mutual intelligibility (subjective measures) can be predicted from various structural distances (objective measures).

In earlier experiments we determined the mutual intelligibility of the following 15 Chinese dialects: Beijing, Chengdu, Jinan, Xi’an, Taiyuan, Hankou (Mandarin dialects), Suzhou, Wenzhou (Wu dialects), Nanchang (Gan dialect), Meixian (Hakka dialect), Xiamen, Fuzhou, Chaozhou (Min dialects), Changsha (Xiang dialect), and Guangzhou (Yue dialect). All Chinese dialects have a lexical tone system but the complexity of the tones differs. Mandarin dialects typically have four lexical tones whereas the non-Mandarin (Southern) dialects have five tones or more [2, 11].

This paper aims to find out whether the tone information contributes to mutual intelligibility of Chinese dialects and how well the mutual intelligibility can be predicted from tonal differences. The following questions will be targeted. (1) Do tone differences play a crucial role in the judgment of mutual intelligibility between Chinese dialects? (2) How can we compute an objective measure to express the difference between tones? (3) How well can mutual intelligibility can be predicted from the objective tone difference?

2. TONE AS A PREDICTOR FOR MUTUAL INTELLIGIBILITY

2.1. Effect of tone on judged intelligibility

Like vowels and consonants, tone is a distinguishing factor in the identity of word forms. Taking the inventory of Mandarin as an example, there are 22 onset consonants (including zero), 35 rhymes and 4 tones. These onsets and rhymes can form various meaningful homo-syllables. In this case, the tones play very crucial roles for distinguishing these homo-syllables or homophonic syllables. The typical example is the different syllables with the same initial /m/ and final /a/, illustrated in the following table:

Table 1: The four tones of Mandarin.

Pinyin	Tone name	transcription	Character	Meaning
mā	Tone "1"	55	妈	mother
má	Tone "2"	35	麻	hemp
mǎ	Tone "3"	214	马	horse
mà	Tone "4"	51	骂	scold

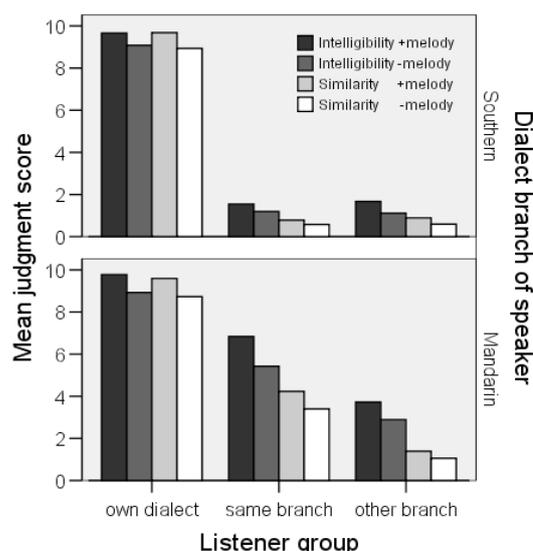
(from <http://mandarin.about.com/od/pronunciation/a/tones.htm>)

The tones are used to determine the meaning of a Mandarin word. So /mǎ/ (horse) is very different from /mā/ (mother). Our presupposition is that the tones would contribute a lot to mutual intelligibility.

In order to test this hypothesis, we did an experiment to test the tone's effect for mutual intelligibility of 15 Chinese dialects. We used recordings of the fable "the North wind and the sun" with melodic and (artificially) monotonized versions read by one male speaker in different 15 Chinese dialects (6 Mandarin dialects and 9 Southern dialects). We obtained judgment scores on mutual intelligibility (and similarity) from 360 listeners for each of the melodic and monotonized fables (12 female and 12 male listeners in each of 15 dialects, for more details, see [7, 8]). Judgments were made on an 11-point scale where '10' represented 'perfect intelligibility' (or 'complete similarity with the listener's own dialect') and '0' stood for 'no intelligibility whatsoever' (or 'no similarity at all'). We always presented the monotonized version at an earlier point in the stimulus order than the corresponding version with full melodic information. The results show that judged intelligibility is consistently lower when tone is absent, but the effect is rather small (Figure 1). However, there are much larger effects due to

judged similarity between speaker and hearer dialects. The interim conclusion is that as a distinguishing factor to identifying the word form, tone can contribute to mutual intelligibility between pairs of dialects to some extent, but we should not overestimate the importance of tone for mutual intelligibility of Chinese dialects.

Figure 1: Effect of absence vs. presence of pitch information on judged intelligibility (and judged similarity) of Chinese dialects within and between the Mandarin and Southern branches.



2.2. Tone difference and mutual intelligibility

The number of lexical tones in Chinese dialects may range from 4 to 10. Tone differences are potentially important for distinguishing lexical units, especially between homophonic words, although all tones are evolved from the same Middle Chinese tone system (see Table 2).

Also, the classification of Mandarin versus Southern dialects is mostly based on tone evolution. Normally, when the number of tones is less than or equal to 5, they are classified as Mandarin dialects; when the number of tones is larger than 5, they are called Southern or non-Mandarin dialects [7, 11].

Table 2: Tone splitting of Chinese dialects.

Register	Tone (Sheng)			
	Level (Ping)	Rising (Shang)	Departing (Qu)	Entering (Ru)
Upper (Yin)	Yin Ping	Yin Shang	Yin Qu	Yin Ru
Lower (Yang)	Yang Ping	Yang Shang	Yang Qu	Yang Ru

2.2.1. Measuring the tone difference

We are interested to know how much tone difference or similarity affects the mutual intelligibility between pairs of Chinese dialects (within and between the Mandarin and Southern branches). We firstly need to determine the size of difference between tone systems and then correlate the measure with the intelligibility scores collected.

For any pair of lexical transcriptions of cognate words in two dialects, we may use the Levenshtein Distance (LD) to measure their difference. The LD measure is a string edit distance based on the number of string operations (insertion, deletion, substitution) needed to convert the phonetic transcription of a word in language A to its counterpart in language B (or vice versa) [4]. For words with tone information, we need to solve how to quantify distance between tone systems. The principle is to firstly collect tone transcriptions for comparable word lists and then to compute their LD.

We computed the Levenshtein distance (LD) of tone differences between 15 dialects based on phonetic transcriptions of 764 cognates a database of 40 Chinese dialects compiled by the Linguistic Institute of the Chinese Academy of Social Sciences (CASS) [5]. The tones in the CASS database were transcribed in three-digit tone marks based on Chao [1]. The solution for this computing is to count the minimal number of edit operations when converting string A to string B. Theoretically, the tone digit sequences can be treated as strings with a maximum length of three, on which LD can be computed. When one member of a pair of tone strings is a single digit, this digit will be matched with the leftmost digit of a two-digit tone, and with the second digit of a three-digit tone. When a three-digit tone is compared with a shorter tone sequence, the second digit of the triplet (three-digit tone) will be matched with the first digit of the shorter string. We counted substitutions as 1 unit of distance and insertions/deletions (indels) 0.5 unit. We then normalized for string length by dividing the summed distance by the number of alignment slots (Table 3).

The results show that this tonal LD has no meaningful correlation with the mutual intelligibility between the 15 target dialects (Pearson's $r = .100$), when intelligibility was functionally measured at the sentence level (for details on this intelligibility test see [6, 7, 10]). The correlation was even poorer, $r = .030$, when judged

intelligibility (§2.1, version with melody) was used as the criterion.

Table 3: Example: Levenshtein distance (LD) computed for all pairs of Mandarin tones.

Tone pair	Members	String operations	Relative LD	
(1-2)	55	1 substitution / 2 alignments	1 / 2	0.50
	35			
(1-3)	55	1 indel, 2 substitutions / 3 alignments	2.5 / 3	0.83
	214			
(1-4)	55	1 substitution / 2 alignments	1 / 2	0.50
	51			
(2-3)	35	1 indel, 2 substitutions / 3 alignments	2.5 / 3	0.83
	214			
(2-4)	35	2 substitutions / 2 alignments	2 / 2	1.00
	51			
(3-4)	214	1 indel, 2 substitutions / 3 alignments	2.5 / 3	0.83
	51			

Next, in a second tonal-distance measure, we used the number of string-edit operations as the distance measure but now the symbols in the strings were chosen so as to reflect some of the auditory characteristics of tone-language listeners. Following [12], who showed that results obtained with this method correlated best with mutual intelligibility of Tibeto-Burman and Tai-Kadai languages, we transformed the three-digit tone strings to sequences of two symbols. The first symbol (letter) represents the onset of the tone, the second the contour shape. We assume that Sinitic languages can be adequately described with three onset tone levels, viz. high (H), mid (M) and low (L). We further distinguished five contour types, viz. Level (L): 55 = HL, Rising (R): 35 = MR, Falling (F): 51 = HF, Dipping (D): 214 = LD, Peaking (P): 241 = LP. The correlation between the onset+shape LD and mutual intelligibility was slightly better than before but still insignificant ($r = .150$ for functional intelligibility and $r = .160$ for judged intelligibility).

The third solution we attempted was to compute a distance metric after perceptually weighting the various dimensions underlying the tonal space. Taking our cue from [3] we used five dimensions (tone features) and assigned values as follows. (1) We computed a value h (height, pitch) for a tone as the mean of the (maximally) three tone digits. If $h > 3.5$ height was set to 5 (=high), if $h < 2.5$ it was set to 1 (=low); all values between (and including) 2.5 and 3.5 were set to 3 (=mid). (2) Duration (0=1 mora, 1=2 morae, 2=3 morae): Depending on the number of tone digits present in

the string, duration was 1, 2 or 3 timing units (morae). Three-morae tones are always of the complex contour type (peaked or dipping), so that this feature covers more than just duration. (3) Direction of pitch change (0=level, 1=down, 2=up): Direction was defined on the last two digits in the tone string. Direction was set to 0 if the string contained just one digit or if there was no change in pitch level on the last two digits. Any falling pitch (on the last two digits) was given the value 1, and any rising pitch 2. (4) Slope of change (0=gradual, 1=steep): Steep slopes are found on tone strings with a difference of 3 or more tone levels (either up or down) on the last two digits. Steep slopes were specified as '1', all non-steep slopes as '0'. (5) Extreme endpoint (0=no, 1=yes): (see Table 4). It was specified as '1' if the final digit was either 1 or 5, and as '0' for any other final digit. (Table 4)

Table 4: Example of computation of perceptual distance between two tones.

Tone string	height	direction	duration	slope	extremity	total
5	5	0	1	0	1	
214	1	2	3	1	0	
△	4	2	2	1	1	10

We defined the perceptual distance between any two tone strings as the sum of the (implicitly weighted) feature differences divided by 10. As a result the perceptual distance between any two tones is a fraction between 0 (no difference) and 1 (maximally different).

This time correlation between tonal distance and mutual intelligibility was as low as $r = .140$ (ins.) when determined functionally and $r = .040$ (ins.) for judged intelligibility.

3. CONCLUSION AND DISCUSSION

The results obtained in §2.1-2 show that differences in the tones system have little effect on the mutual intelligibility of Chinese dialects, even across the Mandarin-Southern divide. In [7] we found clearly better correlations between segmentally defined LD and mutual intelligibility scores, with r-values around .500 for both onsets and rhymes, whether established through functional listening tests or by judgments.

Much better correlations between tone distance and intelligibility were obtained by [12] for South Chinese and Vietnamese languages, even for crude string edit distances. There is no immediate explanation for the discrepancy between their results

and ours. Possibly, the relationships between the tones in the various Sinitic dialects that developed from Middle Chinese are so arbitrary that listeners do not use the tones when listening to other dialects. Obviously, additional research is called for in order to better understand of the role of tonal differences in cross-dialect intelligibility in China.

4. REFERENCES

- [1] Chao, Y.R. 1928. *Studies in the Modern Wu Dialects*. Tsinghua College Research Institute Monograph 4, Beijing.
- [2] Cheng, C.C. 1997. Measuring relationship among dialects: DOC and related resources. *Computational Linguistics & Chinese Language Processing* 2, 41-72.
- [3] Gandour, J., Harshman, R. 1978. Crosslanguage differences in tone perception: A multidimensional scaling investigation. *J. Acoust. Soc. Am.* 62, 693-707.
- [4] Gooskens, C., Heeringa, W. 2004. Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Lang. Var. Change* 16, 189-207.
- [5] Hou, J. 2003. *Xiandai Hanyu Fangyan Yiku [The Sound Databank of Modern Chinese]* (CD-ROM version, in Chinese). Shanghai: Shanghai Education Press.
- [6] Kalikow, D.N., Stevens, K.N., Elliott, L.L. 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J. Acoust. Soc. Am.* 61, 1337-1351.
- [7] Tang, C. 2009. *Mutual Intelligibility of Chinese Dialects: An Experimental Approach*. LOT dissertation series 228. Utrecht: LOT.
- [8] Tang, C., van Heuven, V.J. 2007. Mutual intelligibility and similarity of Chinese dialects. In Los, B., van Koppen, M. (eds.), *Linguistics in the Netherlands*. Amsterdam: John Benjamins, 223-234.
- [9] Tang, C., van Heuven, V.J. 2008. Mutual intelligibility of Chinese dialects tested functionally. In van Koppen, M., Botma, B. (eds.), *Linguistics in the Netherlands*. Amsterdam: John Benjamins, 145-156.
- [10] Tang, C., van Heuven, V.J. 2009. Mutual intelligibility of Chinese dialects experimentally tested. *Lingua* 119, 709-732.
- [11] Yan, M.M. 2006. *Introduction to Chinese Dialectology*. LINCOM Studies in Asian Linguistics. München: LINCOM.
- [12] Yang, C., Castro, A. 2008. Representing tone in Levenshtein distance. *Int. J. Hum. Arts Computing* 2, 205-219.