VAT of the lexical tones in Mandarin Chinese

Authors: Jiangping Kong, Ruifeng Zhang, Jiangping Kong

Date: 2019-06-20T00:00:00+00:00

Abstract

The purpose of this research was to investigate the association of vocal attack time (VAT) and tones in speakers of Mandarin Chinese, and to explore how tones initiated at different pitch levels affected VAT. SP and EGG signals were synchronously recorded from 72 young undergraduates or postgraduates (42 females and 30 males) while they were reading aloud a wordlist of 50 disyllabic words at their most comfortable pitch, loudness and rate. VAT measures revealed three findings. (1) Vocal attack time shows no significant difference between the common yangping and the yangping derived from shangsheng. This, from a physiological perspective, supports the argument that the tone sequence 3 3 in Mandarin is indeed converted into 2 3, nothing else. (2) The tones of Mandarin Chinese that start from low pitch levels (35, 21) tend to present significantly different VAT values from those that start from high pitch levels (55, 51), with mean VATs of the former being much longer than those of the latter. This embodies the nonlinear contravariant relationship between VAT and F0 at vowel onsets. (3) There are deviations or individual differences: a small number of people do not follow this pattern.

Full Text

Preamble

Vocal Attack Time of Lexical Tones in Mandarin Chinese
Jiangping Kong[1], Ruifeng Zhang[1]
[1]Center for Chinese Linguistics, Peking University
Department of Chinese Language and Literature, Peking University
Joint Research Center for Language and Human Complexity

ABSTRACT

This study investigated the association between vocal attack time (VAT) and lexical tones in Mandarin Chinese speakers, and explored how tones initiated at

different pitch levels affected VAT. SP and EGG signals were synchronously recorded from 72 undergraduate and graduate students (42 females and 30 males) while they read aloud a wordlist of 50 disyllabic words at their most comfortable pitch, loudness, and rate. VAT measures revealed three key findings. First, vocal attack time shows no significant difference between common yangping and yangping derived from shangsheng. This finding provides physiological support for the argument that the tone sequence 3-3 in Mandarin is indeed converted into 2-3, and nothing else. Second, Mandarin tones that start from low pitch levels (35, 21) exhibit significantly different VAT values from those that start from high pitch levels (55, 51), with mean VATs of the former being much longer than those of the latter. This demonstrates the nonlinear contra-variant relationship between VAT and F0 at vowel onsets. Third, there are deviations and individual differences: a small number of speakers do not follow this pattern.

Subject Keywords: Vocal attack time, Lexical tones, Phonation onset, Nonlinear contra-variant relationship

1.1 VOCAL ATTACK TIME

Vocal attack time (VAT) is a concept proposed by Baken et al. (1998a, 1998b) based on the time delay between the rise of the sound pressure (SP) signal and the appearance of a clear electroglottographic (EGG) signal when SP and EGG signals are recorded simultaneously. In the presence of transglottal airflow, the vocal folds oscillate with small amplitudes before reaching the midline of the glottis. Upon arriving at the glottal midline with periodic contact achieved and stabilized, the amplitude of their oscillations grows rapidly. Consequently, the SP signal begins its growth to large magnitude well before vocal-fold contact occurs.

However, the EGG signal, as a record of vocal-fold contact area, has nearly no amplitude until vocal-fold contact is achieved, and only after that does its amplitude grow rapidly. The EGG and SP signals are thus offset with respect to each other, and VAT is taken to be the time lag between the rise of these signals measured at the onset of phonation. Positive VAT values indicate that the initiation of the SP signal leads that of the EGG signal, while negative values signify the latter preceding the former. When the two types of signals rise at the same time, VAT equals zero. Therefore, VAT provides a potentially useful measure that varies with vocal attack characteristics. Orlikoff et al. (2009), for example, have reported negative VATs for all attempts by their subjects to produce a hard glottal attack. A computer program was developed to automatically extract VAT measures from simultaneously recorded EGG and SP signals, and the validity of this measurement was experimentally demonstrated by Orlikoff et al. (2009). In 2012, Roark et al. proposed a figure of merit (FOM) for VAT measurement, which was actually Pearson' s correlation coefficient determined from the amplitude features of SP and EGG signals (Roark et al. 2012). VAT measurement has been used in nonlinguistic research by Roark et al. (2012) to

acquire normative data for healthy young adults. In 2012, VAT was also measured for linguistically constrained voice onsets during the production of the six Cantonese tones (Ma et al. 2012).

1.2 LEXICAL TONES IN MANDARIN CHINESE

As a well-known tone language in Asia, Mandarin Chinese has four distinctive lexical tones. The first, named yinping, is a high-level tone with pitch sustained high on level 5. The second, named yangping, is a mid-rise with pitch climbing from level 3 to level 5. The third, shangsheng, is a fall-rise that dips first from level 2 to 1 and then rises to level 4. The last, qusheng, is a full fall that starts from level 5 and glides down to level 1. The values of these tones are consequently recorded as yinping (55), yangping (35), shangsheng (214), and qusheng (51). Because these lexical tones distinguish meanings in Mandarin Chinese, the same morpheme may have different meanings when adopting different pitch contours. For example, /mi/ with a fall-rise (214) signifies "rice," but means "honey" when its pitch contour is altered to a full fall (51).

Unlike English, where phonemic variation occurs frequently, Mandarin Chinese exhibits frequent Tone Sandhi. The fall-rise pitch contour of shangsheng (214) mentioned above appears only on syllables before pauses or in citation form. However, when two such contours are juxtaposed in speech flow, the first is always altered into a mid-rise (35), the pitch contour that yangping always adopts. Furthermore, in connected speech, the fall-rise of shangsheng preceding yinping, yangping, or qusheng is almost invariably modified into a low-fall (21), with pitch dipping slightly from level 2 to 1. These processes make the original contour of Tone 3 (214) very rare in connected speech. Tone Sandhi can also be optional. The original tone of "一" (a Chinese word meaning "one") is 55, a high-level pitch pattern, when in citation form or at the end of a sentence. However, in connected speech, this pattern can be modified to 35 when it precedes qusheng morphemes (e.g., "一样 55+51→35+51"), or to 51 when it occurs before yinping, yangping, or shangsheng (e.g., "一般 55+55→51+55", "一直 55+35→51+35", "一起 55+214→51+214"). Not all Mandarin speakers follow these rules, and some still pronounce "一般" as 55+55.

1.3 PURPOSE

The aforementioned linguistic features of Mandarin tones render the most frequent contours in connected speech as four types: yinping (55), yangping (35), shangsheng (21), and qusheng (51). By comparing the pitch levels from which they start, these four types can be distinguished as two categories. The first and fourth tones both start with the highest pitch and belong to one category, while the second and third tones, which start from low levels 3 and 2, belong to the other. As is well known, pitch is an important perceptual correlate of F0, which is associated with the rate of vocal-fold oscillation. Since tones with a high-pitch onset have a higher rate of vocal-fold oscillation than those with a

low-pitch onset during the initial stage of phonation, they may adopt different mechanisms of laryngeal adjustment and present dissimilar characteristics of vocal attack. The purpose of the present investigation is to examine the association between VAT and tone in Mandarin Chinese speakers and to explore how tones initiated at different pitch levels affect vocal attack time. This represents an attempt to measure VAT for linguistically constrained voice onsets.

2.1 WORDLIST

Three considerations determined the selection of disyllabic words for this study. First, the first syllable of each word should start with a head vowel (Chao 1970), meaning no initial consonant or semivowel medial should appear at the beginning. This is because the computer program designed for VAT extraction works efficiently only on syllables beginning with vowels, and a large portion of a Chinese tone contour is spread across the head vowel of a syllable. Second, the three vertex vowels of Mandarin Chinese (/a/, /i/, /u/) should serve as the head vowels of the first syllables of words because they occupy the extreme points on the vowel chart and represent the entire scope of tongue movement during speech. Third, each final of the first syllable should adopt four distinctive tone patterns (yinping, yangping, shangsheng, and qusheng), with each pattern being further followed by yinping, yangping, shangsheng, and qusheng respectively as the second syllables of words.

Since VAT measurement was taken only for the voice onset of the first syllables, there was no requirement regarding the composition of the second syllables. These criteria resulted in a wordlist of 50 disyllabic words with /ai/, /an/, /aŋ/, /au/, /i/, /u/ chosen as the finals of the first syllables, as shown in Table 1 .

Table 1. 50 disyllabic words used in the present study

2.2 SUBJECTS AND INSTRUMENTATION

Recordings of the wordlist were obtained from 42 females (mean age = 24.0 years, standard deviation (SD) = 2.1) and 30 males (mean age = 22.7 years, SD = 1.9), all of whom were undergraduate or graduate students. They were able to speak standard Mandarin Chinese and use it freely for daily communication. None had any voice or hearing problems, and all were in good health at the time of testing. The recording process was conducted in a sound-treated booth at the Language Lab of the Chinese Department, Peking University, where background noise was below 25 dBA. During recording, Adobe Audition 1.5 was set to a double-channel interface with a sampling rate of 44100 Hz and a resolution of 16 bits for each channel. The electroglottograph (Model 6103) used for collecting EGG signals and the microphone and sound card (Creative Labs Model No. sb1095) used to acquire SP signals were synchronously connected to a personal computer through a sound console (Behringer XENYX502). With their lips approximately 10 cm from the microphone, subjects were asked to read the disyllabic words aloud using their most comfortable pitch, loudness,

and rate. Each subject read the wordlist twice, and the reading with better quality was selected for analysis.

Excluding 13 recordings of poor quality, the total number of speech samples eventually acquired was 3587, with 2095 from females and 1492 from males. Among the 3587 first syllables of words, 1036 were spoken with yinping (55), 1075 with yangping (35), 640 with shangsheng (21), and 836 with qusheng (51). The original tone contour 214 was absent from the database, and the number of shangsheng tokens was smaller than others due to Tone Sandhi: 214+55/35/51→21+55/35/51, 214+214→35+214. Additionally, some subjects produced the first syllable of "一般 (55+55)" as 51, while others produced it as 55.

2.3 PARAMETER EXTRACTION AND DATA PREPROCESSING

VAT measures were extracted largely automatically from the EGG and SP signals using the computer program developed by Roark et al. (2012). The process consisted of four components. The second component automatically identified a 600-millisecond segment of the SP and EGG signals centered at the approximate time of vocal onset of the first syllable of the disyllabic word. This identification was based on two criteria that had to be simultaneously satisfied for the band-pass filtered EGG signal: local energy had to exceed 15% of the maximum energy, and local cycle length had to show less than 15% variation. However, observation revealed that for 107 speech samples from our database, the 600-millisecond segments thus identified were not centered at the voice onset of the first syllables but elsewhere (e.g., at the onset of the second syllables), suggesting that inadequate EGG signal quality in these samples failed to meet the two criteria. VAT measures for these samples were marked "WS" (wrongly segmented) in the comments column of Excel sheets.

From the 3587 VAT values obtained, the 107 "WS" measures were first removed, leaving 3480 values that were divided into two groups: 2025 measures for females and 1455 for males. Each group was then processed separately in the same way. Due to the large number of outliers among VAT values, measures beyond ±2 standard deviations from the mean VAT were eliminated. Eventually, another 100 measures were deleted from the database, leaving 3380 values (skewness = -0.32, kurtosis = 3.397) with VAT ranging from -40.4 ms to 37.1 ms. The average and standard deviation were -0.32 ms and 7.16 ms, respectively. SPSS 13.0 (SPSS Inc., USA) was used for all analyses below.

3 RESULTS

Among the preprocessed database, there were 1000 speech samples whose first syllables carried mid-rise pitch contours of yangping (35). Of these contours, 195 were derived from shangsheng (214) through the Tone Sandhi pattern 214+214→35+214. A one-sample t-test showed that the VAT values of these

195 derived yangping contours were not significantly different from those of common yangping contours ($t(194) = 1.486$, $p = 0.139 > 0.05$). Similarly, 43 subjects pronounced the first syllable of "一般 (55+55)" as 51, a full fall, and according to another one-sample t-test, the VAT values of these 43 derived qusheng pitch contours (51) were not significantly different from those of common qusheng contours either ($t(42) = 0.822$, $p = 0.416 > 0.05$). It is therefore reasonable to regard these derived pitch patterns of 35 and 51 as belonging to common yangping and qusheng categories, respectively.

Since the final data corpus comprised 1961 VAT measures for females and 1419 for males, with 992 for yinping (55), 1000 for yangping (35) (including 195 derived ones), 599 for shangsheng (21), and 789 for qusheng (51) (including 43 derived ones), a two-way mixed-measures ANOVA was conducted with speaker gender (female vs. male) as the between-subject factor and tone (the four tones) as the within-subject factor. Results revealed a significant main effect of tone ($F(2.444) = 33.59$, $p < 0.05$, $R^2 = 0.324$), a non-significant main effect of gender ($F(1,70) = 0.179$, $p = 0.673 > 0.05$, $R^2 = 0.003$), and a non-significant tone-by-gender interaction effect ($F(2.444) = 0.262$, $p = 0.813 > 0.05$, $R^2 = 0.004$). Means and standard deviations of VAT calculated across the 72 subjects are listed in Table 2 and plotted in Figure 1 [Figure 1: see original paper].

Table 2. Means and SDs of VAT for different tones and genders across all 72 subjects

Figure 1 shows that average VATs for three of the four tones are smaller for females than for males, except for tone 1, suggesting the need for simple effects analyses. A paired-samples t-test further demonstrated that for both males and females, VATs between yinping (55) and yangping (35), and between shangsheng (21) and qusheng (51), are significantly different (for males: 55 vs. 35: $t = -5.421$, $p = 0.00 < 0.05$; 21 vs. 51: $t = 4.479$, $p = 0.00 < 0.05$; for females: 55 vs. 35: $t = -4.526$, $p = 0.00 < 0.05$; 21 vs. 51: $t = 4.291$, $p = 0.00 < 0.05$), while those between yangping (35) and shangsheng (21), and between yinping (55) and qusheng (51), are not (for males: 35 vs. 21: $t = 1.007$, $p = 0.322 > 0.05$; 55 vs. 51: $t = 0.702$, $p = 0.488 > 0.05$; for females: 35 vs. 21: $t = 1.222$, $p = 0.229 > 0.05$; 55 vs. 51: $t = 1.34$, $p = 0.188 > 0.05$). Specifically, mean VATs are much longer for yangping and shangsheng than for yinping and qusheng in both genders. This is why a cluster analysis grouped yinping and qusheng into one category and yangping and shangsheng into another based on VAT measures of their voice onsets (see Figure 2 [Figure 2: see original paper]).

Figure 2. Result of a hierarchical cluster analysis

However, close inspection of the mean VATs for the four tones (55, 35, 21, and 51) for each subject revealed that the 72 subjects (including males and females) could be divided into two groups: 46 subjects (63.89%) had mean VATs for both yangping and shangsheng that were longer than those for yinping and qusheng (see Figure 3 [Figure 3: see original paper]), while 26 subjects (36.11%) displayed various other patterns (see Figure 4 [Figure 4: see original paper]).

Figure 3. 46 of the 72 subjects displayed mean VATs of four tones in the same pattern. Each line indicates the average VATs of one person

Figure 4. 26 of the 72 subjects displayed mean VATs of four tones in miscellaneous patterns. Each line indicates the average VATs of one person

A two-way mixed-measures ANOVA conducted on the 46 subjects still showed a significant main effect of tone ($F(3) = 87.644$, $p < 0.05$, $R^2 = 0.666$), a non-significant main effect of gender ($F(1,44) = 2.163$, $p = 0.149 > 0.05$, $R^2 = 0.047$), and a non-significant tone-by-gender interaction effect ($F(3) = 0.632$, $p = 0.595 > 0.05$, $R^2 = 0.014$). Table 3 and Figure 5 [Figure 5: see original paper] present the means and SDs of VAT for these 46 subjects as a function of tone and gender. The exception observed in Figure 1 no longer appears in Figure 5, suggesting that individual differences may slightly affect results for the whole population.

Table 3. Means and SDs of VAT for different tones and genders across the 46 subjects

Figure 5. Means of VAT as a function of tone and gender (46 subjects)

4.1 TONE SANDHI

One debate concerning tone sandhi in Mandarin Chinese has been whether the tone sequence 3-3 is homophonous with the sequence 2-3. This issue was eventually settled by Wang et al. (1967, 2006) through a perception experiment. The 130 pairs of test items used in their research were designed such that the two members of each pair shared the same phonological features except for pitch contour; in other words, one member carried the tone sequence 2–3 while the other carried 3–3. These items were recorded in random order and then presented randomly for native Mandarin speakers to identify as either sequence 3-3 or 2-3. None of the listeners achieved accuracy above 55%, and even the speaker from whom the test items were recorded could not correctly identify more than 60% of them, suggesting that the yangping pitch contour (35) derived from shangsheng (214) was perceptually no different from that of common yangping (35). The first one-sample t-test in the present study supports this argument from a physiological perspective: VAT values of the second tone derived from the third tone are not significantly different from those of the common second tone, suggesting that all yangping contours, whether original or derived, share similar laryngeal adjustments at their voice onsets and therefore display comparable features of vocal attack. In short, both perception and physiology point to one conclusion: the pitch contour 214 before another 214 is indeed phonemically identical to yangping. The second one-sample t-test leads to a similar judgment: the onsets of phonation for all qusheng contours, whether the original 51 or those derived from 55, are physiologically alike, and the two types should be perceptually indistinguishable.

4.2 VAT AND LEXICAL TONES

The second finding from the analyses above can be summarized as follows. In a large group of Mandarin speakers, the two lexical tones with high-pitch onsets, yinping (55) and qusheng (51), display smaller VAT values, while the other two with low-pitch onsets, yangping (35) and shangsheng (21), show much larger values (see Tables 1 and 2 and Figures 1, 2, 3, and 5). In other words, a higher rate of vocal-fold oscillation tends to be associated with a shorter VAT value, and vice versa. This negative VAT-F0 correlation at linguistically constrained voice onsets is also observed in the three level tones of Cantonese (Ma et al., 2012). In females, mean VATs for high, mid, and low level tones are 0.72 ms, 1.70 ms, and 1.78 ms, respectively; in males, mean VATs for high, mid, and low level tones, although longer than in females, follow the same pattern: 3.99 ms, 4.64 ms, and 4.69 ms. However, Tables 2 and 3 also indicate counterexamples. For both males and females, mean VATs for shangsheng (21) should always be larger than those for yangping (35) because the former has a lower initial pitch than the latter, but this is not actually the case. These findings align with the VAT study on five linguistically unconstrained pitch levels in Mandarin (Zhang et al.). In a large group of speakers, as pitch levels shift from one to five, there is a linear increase in pitch but a nonlinear decrease in VAT: from Levels Two to Five, each mean VAT value is not always larger than the one that follows. However, the average VAT at Level One is always the largest among the five pitch levels and is much larger than all others. Therefore, for both linguistically constrained and unconstrained vocal onsets, VAT and pitch tend to exhibit a nonlinear contra-variant relationship in most Mandarin speakers.

4.3 INDIVIDUAL DIFFERENCES

Forty-six of the 72 subjects produced low-pitch onsets for the second and third tones (35, 21) with longer mean VATs than for high-pitch onsets of the first and fourth tones (55, 51), while 26 subjects displayed inconsistent VAT patterns across the four tones. This appears to support the findings by Zhang et al. that as pitch means increase linearly from Levels One to Five, mean VATs decrease nonlinearly in a large group of speakers but increase nonlinearly in a small group, and that different speakers tend to use different strategies for increasing pitch height. However, among the 26 subjects observed in the present study, mean VATs for the four tones were ordered as yangping (35) = 1.736 ms > yinping (55) = 1.586 ms > qusheng (51) = 0.697 ms > shangsheng (21) = 0.375 ms, and a positive VAT-F0 correlation was not observed at the phonation onsets of the four tones. The causes of these individual differences require further investigation.

5 CONCLUSION

First, vocal attack time, as a measure of vocal fold phonatory function, shows no significant difference between common yangping and yangping derived from

shangsheng, nor between common qusheng and qusheng derived from yinping. This provides physiological support for the argument that the tone sequence 3-3 in Mandarin is indeed converted into 2-3, and nothing else. Second, Mandarin tones starting from low pitch levels (35, 21) tend to exhibit significantly different VAT values from those starting from high pitch levels (55, 51), with mean VATs of the former being much longer than those of the latter. This demonstrates the nonlinear contra-variant relationship between VAT and F0 at vowel onsets. Third, there are deviations and individual differences: a small number of speakers do not follow this pattern.

NOTES

1. This research was funded by the National Social Sciences Foundation of China. Grant No: 10&ZD125.

REFERENCES

- Baken, R.J. and Orlikoff, R.F. 1998a. Vocal fold adduction time estimated from glottographic signals. Presented at: the 25th Mid-Winter Meeting of the Association for Research in Otolaryngology; February 1998; St. Petersburg, FL.
- _____. 1998b. Estimating vocal fold adduction time from EGG and acoustic records. In: Schutte HK, Dejonckere P, Leezenberg H, Mondelaers B, HF, eds. Programme and Abstract Book: 24th IALP Congress, Amsterdam; 1998:15.
- Ma, E., Baken, R.J., and Roark, R.M. 2012. Effect of Tones on Vocal Attack Time in Cantonese Speakers. *Journal of Voice*, Vol. 26, issue 5, pp. 670.e1-670.e6.
- Roark, R.M., Watson, B.C., and Baken, R.J. 2012. A Figure of Merit for Vocal Attack Time Measurement. *Journal of Voice*. Vol. 26, issue 1, pp. 8-11.
- _____, Watson, B.C., Baken, R.J., Brown, D.J., et al. 2012. Measures of Vocal Attack Time for Healthy Young Adults. *Journal of Voice*. Vol. 26, issue 1, pp. 12-17.
- Orlikoff, R.F., Deliyski, D.D., Baken, R.J., and Watson, B.C. 2009. Validation of a Glottographic Measure of Vocal Attack. *Journal of Voice*. Vol 23, issue 2, pp. 164-168.
- Wang, W. S-Y., and Li, K.P. 1967. Tone 3 in Pekinese. *Journal of Speech Hearing Research*. 10 (3), pp. 629-36.
- _____, and Peng Gang. 2006. Language, Speech and Technology. Shanghai Educational Publishing House. pp. 120-121.
- Chao, Yuen Ren. 1970. *A Grammar of Spoken Chinese*. University of California Press, Berkeley, Los Angeles, London. pp. 18-25.
- Zhang, Ruifeng, Baken, R.J., and Kong, Jiangping. Vocal Attack Time of Different Pitch Levels and Vowels in Mandarin. *Journal of Voice* (in press).

Chinese Abstract

Title: A Study of Vocal Attack Time of the Four Tones in Mandarin Chinese

Authors: Kong Jiangping, Zhang Ruifeng

Affiliations: Department of Chinese Language and Literature, Peking University; Center for Chinese Linguistics; Joint Research Center for Language and Human Complexity

Abstract: As a typical tone language, Chinese has four meaning-distinguishing lexical tones: yinping (55), yangping (35), shangsheng (214), and qusheng (51). However, due to tone sandhi in connected speech, the four tone values that frequently occur are: yinping (55), yangping (35), shangsheng (21), and qusheng (51). This study investigates the relationship between vocal attack time (VAT) and tones in Mandarin Chinese speakers, and explores how tones initiated at different pitch levels affect VAT. We simultaneously recorded SP and EGG signals for 50 disyllabic words from 72 speakers (30 males and 42 females), all university students or graduate students in their twenties. One-sample t-tests showed no significant difference in VAT values between yangping and yangping derived from shangsheng, nor between qusheng and qusheng derived from yinping. This finding provides physiological support for the view that when two shangsheng tones occur consecutively, the first changes to yangping. A two-way repeated measures ANOVA revealed that VAT values for tones starting from low pitch levels were significantly different from those starting from high pitch levels, with the former being significantly larger than the latter. However, individual differences existed, with 26 of the 72 speakers not following this pattern.

Keywords: Vocal attack time; Lexical tones; Phonation onset; Nonlinear contra-variant relationship

Note: Figure translations are in progress. See original paper for figures.