Tonal Aspects of Languages 2016
24-27 May 2016, Buffalo, New York

The Timing of Acoustic vs. Perceptual Availability of Segmental and Suprasegmental Information

Katrina Connell, Annie Tremblay, Jie Zhang

Department of Linguistics, University of Kansas, Lawrence, KS, USA

katconnell213@ku.edu, atrembla@ku.edu, zhang@ku.edu

Abstract

This study investigates the timing of perception of tonal and segmental information. Its purpose is to determine whether the apparent delay that previous priming studies have reported for the processing of tonal information (relative to the processing of segmental information) may stem from low-level speech perception. Native Chinese listeners and native English listeners without knowledge of tone languages completed a gated AX-discrimination task where they heard increasingly large fragments of Chinese word pairs that differed only in tones or only in segments. The gates where listeners perceived the contrasts were compared after factoring out differences in when the contrasts became reliably present acoustically. The results of both groups show that the perception of tonal information is delayed compared to that of segmental information, with English speakers showing a larger delay than Chinese speakers.

Index Terms: tone, intonation, production, perception

1. Introduction

Segmental cues and suprasegmental cues can both signal lexical differences in language. To illustrate, in Mandarin Chinese (henceforth Chinese), both tonal and segmental information contribute to lexical identity: The word *ma* can have four meanings depending on its tone (e.g., T1 *mā* 'mother' [level tone] vs. T2 *má* 'hemp' [rising tone] vs. T3 *mǎ* 'horse' [dipping tone] vs. T4 *mà* 'to scold' [falling tone]). How segmental and suprasegmental information are integrated in the word recognition system is poorly understood, however. The importance of both segmental and suprasegmental cues in languages such as Chinese raise crucial questions about the relationship between these two types of information in lexical access, specifically their relative weighting and their relative timing. The current paper focuses on the issue of timing.

Although models of lexical access have attempted to incorporate the use tonal information in word recognition [1,2], no model exists that incorporates the use of suprasegmental information in general. Research has shown that suprasegmental information, including not only tonal information, but also stress and prosody, constrains lexical access in important ways [3,4,5]. It is thus critical for theories of auditory word recognition to incorporate the use of suprasegmental information alongside the use of segmental information in lexical access. Chinese, being a tone language, provides an ideal test case for examining questions about the timing of use of segmental and suprasegmental information.

The literature on lexical tone reveals a debate about when tone is used in relation to segments in lexical access. On the basis of results obtained from priming, gating and error detection tasks, some researchers have argued that tone is used at a later stage of lexical access than segments [6,7,8]. Lee [6], in particular, used several priming experiments to examine the timing of use of tonal information in Chinese lexical access. He compared primes with a direct semantic relationship to the target (e.g., *lóu* 'hall' – *jiànzhu* 'building') and primes that were segmentally identical to, but differed in tones from, the primes in the previous condition (e.g., *lǒu* 'hug' – *jiànzhu* 'building'. At a 250-ms interstimulus interval (ISI) (Experiment 3), only the primes in the semantically related condition yielded a significant priming effect. By contrast, at an ISI of 50 ms (Experiment 4), there was a significant priming effect for both the semantically related and the non-semantically related pairs, indicating that the segmentally identical word that differed in tone was also activated. In light of these results, the author proposed that at early stages of lexical access, tone is not used to constrain the word search, and thus there is priming between *lǒu* 'hug' and *jiànzhu* 'building' since all words with the segmental structure of *lou* have been activated. At later stages (as shown by the 250-ms ISI results), however, tone is used to select among the segmental candidates already active.

Contrary to Lee's [6] claims, however, more recent, eye-tracking and electrophysiological studies in Chinese have shown that tone is used at the same stage of lexical access as segments [2,9,10]. Although these time-course studies suggest that tone and segments are processed in parallel, they leave Lee's [6] finding (i.e., that tonal information did not constrain lexical access at a 50-ms ISI) unexplained. One possibility is that the apparent delay of use of tonal information relative to the processing of segmental information in Lee's [6] study stemmed from low-level speech perception.

Not all information is perceived on the same time scale, and the timing by which different types of information can be perceived, and thus become available for use in lexical access, could differ greatly. If tonal information is perceived later than segmental information, it will be used later to constrain the lexical search, not necessarily because it is used at a later stage of lexical access than segments, but because the information becomes available to the processor later. Hence, a delay in perception could potentially explain Lee's [6] results without necessarily entailing that tonal information is used at a later stage of lexical access than segments.

The present study aims to investigate potential timing differences between the perception of segmental and tonal information. It does so by examining the perceptual availability of tonal and segmental information in relation to when this information becomes reliably present in the acoustic signal. By doing so, this research will shed light on whether tonal information is perceived later than segmental information; even when potential differences in the acoustics have been factored out.

2. Present Study

We conducted a gated AX discrimination task to find the point in time where listeners could hear the difference between word pairs that differed only in tone or only in segments. We also examined the point in time where the tonal

http://dx.doi.org/10.21437/TAL.2016-21

and segmental pairs selected for the stimuli became acoustically different. The perceptual results were analyzed in relation to these acoustic analyses. In addition to native Chinese listeners, participants included native English listeners who had no experience with Chinese or any tone language. If tonal information is perceptually available later than segmental information, then this perceptual delay should be present for both language groups.

2.1. Method

The present study used a gated AX discrimination experiment, that is, an AX discrimination task that compared auditory stimuli consisting of increasingly larger word fragments. Participants heard fragments (of equal length) of two words, and decided if the two fragments were the same or different. Because participants heard fragments rather than complete words, and because their task was to discriminate between the fragments rather than identify them, this gated AX discrimination experiment has the ability to by-pass lexical access, which would be typical of a classic gating task. Given the gated nature of the stimuli, this task is ideal to test for the timing of disambiguation in perception. An ISI of 250 ms was used in order to target the phonetic level of perception and to reduce native-language effects [11,12].

2.2. Participants

Participants were native Mandarin Chinese listeners (*n*=20) and native English listeners with no experience learning Chinese (*n*=27). Three native English listeners were excluded from the analysis for not following the instructions and not paying close enough attention to the task. The number of native English listeners was thus 24. Participants received payment or course credit in return for their time.

2.3. Materials and Design

The task included two conditions: A tonal contrast condition and a segmental contrast condition. The tone pairs included in the tonal stimuli were T1-T2, T1-T3, T4-T2, T4-T3, and T1-T4.[1] Four different monosyllabic word pairs were selected for each of the five tone pair types, creating five tonal sets. All words began with voiceless initial consonants to control the timing of the tonal information. The words within a given tonal pair were identical segmentally.

The segmental condition included words with four hypothesized timings of disambiguation of the segments, ranging from early disambiguation in monophthong vowels (*bi-ba*) to late disambiguation in the change of a vowel to a nasal coda (*tao-tang*), as illustrated in Table 1. As with the tonal condition, the segmental condition included four word pairs in each of these four timing categories, creating four segmental sets. The words within pairs were identical tonally.

Table 1: *Example segmental contrasts*

Discrimination Expected		Early ⟷		Late
Set Name	Vowel	Allophonic Onglide	Offglide	Nasal Coda
IPA	pi4 – pa4	tʰai2 - tʰɑʊ2	tʰiɑʊ1 - tʰiɛ1	sɑʊ3 – sɑŋ3

[1] None of the stimuli tested the pair T2-T3. In the task, participants heard no more than the first half of the syllable; thus, differences between T2 and T3 would likely not be present acoustically in any of the gates, yielding no useful perception data.

The 36 word pairs were recorded by a male native speaker of Mandarin Chinese in the Anechoic Chamber of the University of Kansas with a cardioid microphone (Electrovoice, model N/D767a) and a digital solid-state recorder (Marantz, model PMD671) at a sampling rate of 22,050 Hz. All words were then normalized for duration. The consonant portions were normed to 117 ms and the rhyme portions were normed to 407 ms, as found from the average durations from all productions.

After duration was normalized, the initial consonant and the first half of the rhyme portion of each word together was divided into 12 gates. The first gate was the initial consonant (117 ms) and all 11 subsequent gates in the rhyme included 18 ms more information than the previous gate. All items were also normalized for intensity. Filler trials were added to balance the number of 'same' and 'different' trials.

2.4. Acoustic Analyses

In order to ascertain when the tonal and segmental information became reliably present in the acoustic signal, we analyzed each word pair for F0 (tonal condition) or F1 and F2 (segmental condition) for the time window that corresponded to new information in each gate, based on the analysis used by Malins and Joanisse [2]. Gate 1, which included only the initial consonant, was analyzed as a separate time window. From there, each window for analysis was the new information present in each gate that was not in the preceding gate. Since each gate added approximately 18 ms of the rhyme to the preceding gate, each window (except for the consonant in Gate 1) was 18 ms.

The point of disambiguation (POD) for the tonal condition was defined as the first of three consecutive time windows where F0 differed significantly over the four pairs in a set. The POD for the segmental condition was defined at the first of three consecutive time windows where *either* F1 or F2 (whichever came first) differed significantly over the four pairs in a set. For each set and in each window, the POD was established with paired-samples *t*-tests. Table 2 shows the results of this analysis. A *t*-test comparing all tonal pairs to all segmental pairs reveals no significant difference between the acoustic PODs of the tonal and segmental conditions ($p>.05$).

Table 2: *POD analysis of stimuli*

Tonal		Segmental	
Set	POD	Set	POD
T1-T3	2	Vowel	3
T1-T2	2	Allophonic	2
T4-T3	2	Offglide	5
T4-T2	2	Nasal Coda	5
T4-T1	8		
Mean	3.2	Mean	3.75

Surprisingly, the vowel set had a later POD than the allophonic set. Investigation into this result revealed that a single word pair in the vowel set caused the lack of significance at Gate 2: The consonant in the pair *sha-shu* had a heavy effect on the F2 of the vowel for several gates, resulting in later disambiguation of the two words. A second analysis was conducted without this word pair, and the two words in the vowel set became marginally different ($p<.052$) at Gate 2, as expected. The problematic word pair was therefore excluded from all analyses, and accordingly, the vowel set was analyzed at having a POD of 2.

Center of gravity (CoG), variance, skewness, and kurtosis measurements were also taken for the whole consonant and for the last 10 ms of the consonant (Gate 1) for all segmental items. The initial F2 of the vowel was also

measured for all segmental items. Paired-sample t-tests revealed no differences for any pair under any measure.

2.5. Results

The perceptual POD was calculated for each item perceived by each participant. The perceptual POD was defined as the first of three consecutive gates within a single item where the participant responded "different" for all three gates. From these values, difference scores were calculated by subtracting the acoustic POD for that set (see Table 2) from the perceptual POD of that item. This value gives the delay between when the information was reliably present in the acoustic signal and when the information was perceived. Table 3 presents the POD means and the difference score means by set and by group, with the acoustic PODs presented again for reference.

Table 3. *Averages of POD and difference scores*

	Set	POD		Acoustics	Difference	
		Chinese	English		Chinese	English
Tonal	T1_T3	4.26	4.27	2	2.26	2.27
	T1_T2	4.60	6.18	2	2.60	4.18
	T4_T3	4.08	4.21	2	2.08	2.21
	T4_T2	4.00	5.53	2	2.00	3.53
	T1_T4	9.70	9.35	8	1.70	1.35
	Mean	5.33	5.91	3.20	2.13	2.71
Segmental	Vowel	1.61	1.67	2	-0.39	-0.33
	Allophon.	3.01	2.79	2	1.01	0.79
	Offglide	4.45	4.09	5	-0.55	-0.91
	Nasal	7.33	7.49	5	2.33	2.49
	Mean	4.10	4.01	3.50	0.60	0.51

These values were entered into linear mixed-effects models in R using the lme4 package. A first and second model examined the effect of the hypothesized point of disambiguation (i.e., set) on participants' *perceptual POD* scores separately for, respectively, the tonal and the segmental conditions. These two models ascertained whether Chinese and English listeners showed the same effect of acoustic disambiguation of the tonal and segmental contrasts. For the tonal model, the sets were contrast coded to reflect the acoustic timings of disambiguation. Therefore, all early sets (T1-T3, T1-T2, T4-T3 and T4-T2) were coded as –0.5 and the late set (T1-T4) was coded as 0.5. For the segmental model, the sets were also contrast coded to reflect the acoustic timings of disambiguation. This means that all early sets (vowel, allophonic) were coded as –0.5 and the late sets (offglide, nasal coda) were coded as 0.5. The two models each included set ("early" vs. "late," labeled as such for the sake of convenience), group (Chinese vs. English), and their interaction as fixed effects, and participant and item as crossed random effects. For both models, the Chinese listeners' performance in the early sets was the baseline against which the effects of set and group and the interaction between set and group were measured. For all models reported, the fixed effects and their interaction were added individually to the model, and the models were compared using log-likelihood ratio tests (i.e., the ANOVA function).

For the tonal model, the best model included set, group, and their interaction. For the segmental model, the best model included set and group, but not the interaction. The model estimates are provided in Table 4.

Table 4. *Linear mixed-effects model on perceptual POD scores*

		Estimate	Std. Error	DF	*t* value	Pr(>F)
Tonal	Intercept	4.23	0.52	62.70	8.20	.001
	Set	5.47	0.60	26.80	9.10	.001
	Group	1.39	0.61	48.50	2.29	.027
	Set × Group	–1.46	0.40	873.90	–3.66	.001
Segmental	Intercept	2.33	0.62	23.71	3.74	.001
	Set	3.54	0.79	15.98	4.46	.001
	Group	0.16	0.37	46.32	0.44	.663

Table 4 shows that, as expected, for the tone model the effect of set was significant and had a positive estimate, indicating that the Chinese listeners showed longer perceptual PODs in the late set as compared to the early sets (9.70 vs. 4.23). For the segmental model, Table 4 shows that effect of set was also significant and had a positive estimate, indicating that the Chinese listeners showed longer perceptual PODs in the late sets as compared to the early sets (5.89 vs. 2.31). This confirms that the acoustic PODs of words in the tonal and segmental conditions affect Chinese listeners' perception.

A third and fourth model examined the effect of the hypothesized point of disambiguation (i.e., item set), this time on participants' *difference* scores. Recall that the difference scores were calculated by subtracting the acoustic timing from the perceptual timing; hence, if the acoustic analyses correctly captured the disambiguation point in the speech signal, these models should no longer reveal an effect of set on perceptual PODs. The models included the same baseline and fixed and random variables as the first two models.

For the tonal model, the best model included set, group, and their interaction, whereas for the segmental model, the best model included only the intercept. The model estimates are provided in Table 5.

Table 5. *Linear mixed-effects model on difference scores*

		Estimate	Std. Error	DF	*t* value	Pr(>F)
Tonal	Intercept	2.23	0.52	62.70	4.33	.001
	Set	–0.53	0.60	26.80	–0.89	.382
	Group	1.39	0.61	48.50	2.29	.027
	Set × Group	–1.46	0.40	873.90	–3.66	.001
Segmental.	Intercept	0.69	0.44	21.77	1.59	.127

Table 5 shows that for the tonal model, the effect of set was not significant, as predicted. The effect of group was significant and had a positive estimate, indicating that English listeners had longer delays in the early tonal sets than did Chinese listeners (3.63 vs. 2.23). For the segmental model, no fixed factor or interaction improved the fit of the model.

As can be seen from these results, Chinese listeners no longer showed an effect of set in the difference scores. This lack of effect in the difference models suggests that the analysis using the difference scores was able to effectively control for the hypothesized timing differences for Chinese listeners. English listeners, by contrast, differed from Chinese listeners in their perception of the tone contrasts even after controlling for disambiguation in the signal, a finding that is likely due to their native language not being a tonal language. Since the effect of acoustic disambiguation was well controlled for by using the difference scores, as evidenced in the Chinese listeners' data, the final, and primary, analysis was conducted on difference scores to control for the effect of the acoustic timings on perception.

A final model was therefore conducted on participants' difference scores in the tonal and segmental conditions, with condition (tonal vs. segmental), group (Chinese vs. English),

and their interaction as fixed effects, and participant and item as crossed random effects. For this model, the Chinese listeners' performance in the segmental condition was the baseline against which the effects of condition, group, and the interaction between condition and group were measured. The best model included the interaction between condition and group. The model estimates are provided in Table 6 below.

Table 6. *Linear mixed-effects model on difference scores in both the tonal and segmental conditions*

	Estimate	Std. Error	DF	*t* value	P(>F)
Intercept	0.60	0.46	77.80	1.31	.194
Condition	1.53	0.47	41.50	3.28	.002
Group	–0.09	0.44	52.00	–0.20	.839
Condition × Group	0.67	0.44	1504.90	2.94	.003

Table 6 shows that the effect of condition was significant and had a positive estimate, indicating that the Chinese listeners showed a larger delay in the tonal condition than in the segmental condition (2.13 vs. 0.60). This means that tonal contrasts were perceived approximately 1.5 gates, or 28 ms, later than segmental contrasts. The effect of group was not significant, indicating that English listeners were not significantly different from Chinese listeners in the segmental condition (.60 vs. .51). The interaction between condition and group indicates that English listeners had a longer delay in the tonal condition than did Chinese listeners (see Table 3), with Chinese listeners showing a delay of 2.13 gates (approx. 38 ms) and English listeners a delay of 2.71 gates (approx. 48 ms). This means that the relative delay of tonal information to segmental information is approximately 28 ms for Chinese listeners and approximately 40 ms for English listeners.

3. Discussion

This study used a gated AX discrimination task to determine whether native Chinese and native English listeners would show a delay in perceiving tonal information as compared to segmental information. Analyses were conducted on the different hypothesized timing sets within each condition for both perceptual PODs and difference scores.

Crucially, the comparison of the perceptual POD models to the difference score models showed that for Chinese listeners, the effect of the disambiguation timings within the conditions were effectively neutralized for both tonal and segmental conditions by controlling for the acoustics and using difference scores. This suggests that using difference scores between when the information is reliably present in the acoustics and when the information is perceived is a suitable method for controlling for these acoustic differences in perception tasks.

The results from the primary analysis on the difference scores further indicated that tonal information was perceived approximately 28 ms later than segmental information for native listeners and 40 ms later for English listeners, even after controlling for acoustic differences between when tonal and segmental information disambiguated in the speech signal. Although English listeners showed a longer delay than Chinese listeners, the fact that both groups showed the delay is indicative of a timing difference between the use of tonal and segmental information in perception.

These results can explain why previous work such as Lee [6] found delayed use of tone, when none may actually be present, according to more recent work. Tonal information being delayed compared to segmental information in perception would cause the relevant information to arrive to the lexical access system later, and thereby create the appearance of a tonal delay in lexical access. One question that arises, then, is why recent time-course studies [2,9,10] did not find such a delay. After all, if tonal information takes longer to be perceived than segmental information, this delay should also impact lexical access. One possibility is that these studies did not provide a sufficiently tight control of when the tonal and segmental pairs disambiguated acoustically, resulting in no difference between tonal and segmental pairs.

The present results have begun to clear up the discrepancy between previous and more current research on the timing of the use of tonal information in lexical access, but further work is necessary to investigate why no delay is found in recent time-course studies.

4. Conclusions

The present study showed that tonal and segmental contrasts are not perceived on the same time scale, even after controlling for when the contrasts were acoustically present in the speech signal. Tonal information was found to be perceived later than segmental information for native speakers of Chinese, a tone language, but also for native speakers of English, a language without lexically contrastive tones. This difference must be taken into account in future research on the processing of tonal and segmental information.

5. Acknowledgements

We would like to thank Dr. Allard Jongman, Dr. Joan Sereno, and all the students in LING 850 for their valuable feedback on every stage of this work. We are also thankful to those who participated in the experiment.

6. References

[1] Ye, Y., & Connine, C. M. (1999). Processing Spoken Chinese: The Role of Tone Information. *Language and Cognitive Processes, 14*(5-6), 609-630.

[2] Malins, J., & Joanisse, M. (2012). Towards a Model of Tonal Processing During Mandarin Spoken Word Recognition. In *Tonal Aspects of Languages-Third International Symposium*

[3] Cutler, A., & Chen, H. (1995). Phonological similarity effects in Cantonese word recognition. In *Proceedings of the Thirteenth International Congress of Phonetic Sciences*, 1, 106-109.

[4] Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and speech*, *45*(3), 207-228.

[5] Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Vowel use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *The Quarterly Journal of Experimental Psychology*, *63*(4), 772-783.

[6] Lee, C. (2007). Does Horse Activate Mother? Processing Lexical Tone in Form Priming. *Language and Speech, 50*(1), 101-123.

[7] Cutler, A., & Chen, H. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics, 59* (2), 165-179.

[8] Taft, M., & Chen, H.C. (1992). Judging homophony in Chinese: The influence of tones. *Advances in psychology, 90*, 151-172.

[9] Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language, 62*(4), 407-420.

[10] Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: evidence from ERP analyses. *Neuropsychologia, 49*(7), 1761-1770.

[11] Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*(2), 253-260.

[12] Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35-44.