SEPARABILITY OF TONES AND RHYMES IN CHINESE SPEECH PERCEPTION: EVIDENCE FROM PERCEPTUAL MIGRATION

Biao Zeng[a] *& Sven L. Mattys*[b]

[a]School of Education, Communication and Language Sciences, Newcastle University, UK;
[b]School of Experimental Psychology, University of Bristol, UK
`b.zeng@ncl.ac.uk; sven.mattys@bris.ac.uk`

ABSTRACT

The present study uses the perceptual-migration paradigm to explore whether Chinese tones and rhymes are encoded separately during Mandarin speech perception. Following the logic of illusory conjunctions, we calculated the cross-ear migration rates of tones, rhymes, and their combination in Chinese and English listeners. For Chinese listeners, we found that tones migrated more than rhymes. The opposite pattern was found for English listeners. These results suggest that tones are registered separately from their tone-bearing rhyme in Chinese listeners, but they have no independent representation in speakers of non-tonal languages. The results lend support to autosegmental theory, which claims separability and mobility between tonal and segmental representations.

Keywords: speech perception, Chinese, tones, migration paradigm, autosegmental theory

1. INTRODUCTION

In the psycholinguistic literature on the structure and content of the mental lexicon, the representational separability of lexical tones and segments has been relatively under-studied [14, 15]. From a phonological perspective, however, Goldsmith's autosegmental theory [3] has received widespread attention for its claim that tones are represented in a separate tier from segments, even though both types of representations are co-registered at the phonetic level. How the alleged separability of tones and segments translates to speech perception is unclear.

The migration paradigm [5, 7, 8], based on the illusory-conjunction phenomenon [13], is a valuable tool for exploring the perceptual separability and representationality of speech units. For example, [5] showed that French listeners are more likely to erroneously recombine the syllables of dichotically-presented French disyllables than any other units (vowels, consonants, voicing, and place of articulation). That is, relative to a control condition, participants reported hearing a target word, e.g., "bijou" (/biʒu/, "jewel"), more often when this target was embedded in a dichotic pair inducing syllable migration, e.g., /bitɔ̃-kɔʒu/ (underlined is the information forming the target word), than in one inducing the migration of the first vowel, e.g., /bɔʒu-kitɔ̃/, or first consonant, e.g., /kiʒu-bɔtɔ̃/. Under the assumption that the migration of linguistic units reflects the involvement of these units as independent building blocks during speech perception [4, 12], the predominance of syllable migration suggests that French listeners rely predominantly on a syllabic code to perceive speech.

From a methodological perspective, insight into perceptual mechanisms and mental representations can be better gained by using tasks that bypass conscious access to knowledge [1, 6, 9]. Migrations suit this category quite well. For instance, illiterate Portuguese speakers have no conscious awareness of phonemes, as measured by various phonemic tasks [10], but they experience phoneme migration to the same extent as literate speakers do [11]. Thus, migrations involve speech properties that do not need to be accessible to conscious experience. This property makes the paradigm ideal for studying the perceptual reality of phonological constructs such as tones and segments, whose phenomenology is highly variable across listeners.

The present study uses the migration paradigm to test the separability of Mandarin rhymes and tones. Listeners were asked to report whether a pre-specified target syllable (e.g., $b\acute{a}i^1$) was present in a pair of dichotically presented syllables. In the key trials, the dichotic syllables allowed the target to be (erroneously) perceived through rhyme migration, e.g., $b\acute{\imath}m\grave{a}i$, tone migration, e.g., $b\grave{a}i$-$m\acute{\imath}$ or rhyme+tone migration (R+T), e.g., $b\grave{\imath}m\acute{a}i$. Following the logic of illusory conjunctions, if tones are represented separately from their tone-

bearing rhyme, as predicted by autosegmental theory, they should migrate independently from rhymes. Alternatively, if tones and rhymes are represented as undividable wholes (a unitary view), they should either not migrate at all when manipulated independently or migrate just as much as the R+T units. Moreover, if representational separability is the result of long-term exposure to a language with tonal phonology rather than a language-general perceptual property, tone migration should be more prevalent among speakers of a tonal language (Mandarin) than a non-tonal language (English).

2. METHODS

2.1. Participants

Twenty-two native Mandarin speakers and 25 native English speakers participated in the experiment. The Mandarin speakers were from Beijing or northern China where Mandarin is spoken. The English speakers had no knowledge or experience with Mandarin or other tonal languages. Participants had no known hearing problems. They received payment or course credits for their participation.

2.2. Materials

Ten monosyllabic Mandarin morphemes were selected as target syllables, all of which were composed of an onset and a rhyme. Because of the large number of morphemes required for each target (see Table 1 for an example), the rhyme of some of the morphemes contained a coda whereas the rhyme of others did not. This distinction was not taken into account in the analyses.

Table 1: Pairs of dichotic syllables for the target *bái*.

	Target-absent		Target-present		
	Exp.	**Ctr.**			
Rhyme	bímǎi	bímèng	bá-bí	bá-mǎi	bá-mèng
Tone	bǎi-mí	bǎi-mǐ	bá-bǎi	bá-mí	bá-mǐ
R+T	bìmái	tì-mái	bá-bì	bá-mái	bá-tì

The syllables bore one of the four Mandarin tones. For each of the 10 target syllables, three pairs of experimental stimuli were constructed. These pairs assessed the migration of rhymes, tones and the combination of rhymes and tones (R+T), respectively. Each experimental pair contained the components of the target syllable distributed over the two syllables of the pair. The distribution of information across the two syllables

depended on the specific unit under study (see Table 1 for an example with the target *bái*.). In order to ascertain that erroneous target detection, when it occurred, reflected genuine cross-ear migration of the key unit rather than mere perceptual confusion due to the complexity of the listening situation, each experimental pair was matched with a control pair. The control pair was derived from the experimental pair by substituting the critical migrating unit with an unrelated counterpart. Furthermore, in order to compute the discriminability score d' for each of the three units under study, nine target-present pairs were included for each target syllable, three for each of the three migration units, i.e., rhyme, tone and R+T. These pairs consisted of the target syllable itself and the syllable that was used in the corresponding experimental or control pair. How the migration rate for each unit was derived from these stimuli is described in the Results section.

In all, the experiment included 300 pairs: 10 target syllables x 3 migration units (rhyme, tone, R+T) x 5 types (target-absent experimental, target-absent control, 3 target-present) x 2 ear assignments (left/right, right/left).

2.3. Procedure

The target syllables were recorded by a female Mandarin speaker and the syllables for the dichotic pairs were recorded by a male Mandarin speaker. The voice contrast was intended to avoid detection responses being purely acoustically driven. The two syllables within a pair were edited to have the same duration and their onset and offset were synchronized. The experiment was preceded by 20 practice trials unused in the main experiment.

Trials were presented quasi-randomly so that no more than three target-absent or target-present trials and no trials sharing the same rhyme were presented consecutively. Each participant received a different quasi-random order. The left/right ear assignment was randomized for every pair, but each pair was presented in both a left/right and right/left formats. Headphone ear assignment was counterbalanced between participants. Participants were told that, on each trial, they would first hear a target pronounced by a female voice followed by two syllables pronounced by a male voice played simultaneously, one in each ear. They were asked to pay attention to both syllables (i.e., to both ears) in order to decide whether the target had been presented or not. They gave their response through two response keys

labeled "Yes" (target present) and "No" (target absent). Participants were encouraged to respond as quickly and accurately as possible.

Within each trial, the target and the dichotic pair were separated by 500 ms. Participants had up to 2.5 s after the end of the dichotic pair to respond. Upon button press or at the end of the 2.5-s period, there was a 1-s interval before the next target was played.

3. RESULTS

Hit rates and false-alarm rates, as well as the *d'* scores derived from them, were calculated separately for the rhyme, tone, and R+T units (Table 2). The migration rate, plotted in Figure 1, was defined as the difference between *d'* on control trials and *d'* on experimental trials. This index corresponds to the difference in discriminability between trials in which not all of the components of the target syllable are present in the stimuli (the control trials) and trials in which all components are present, albeit in a distributed way (the experimental trials).

Table 2: Hit rate (Hit), false alarm rate (FA), and *d'* (calculated as the average across participants' individual *d'*) for all the conditions of the design.

	Chinese Listeners					
	Experimental			Control		
	Hit	**FA**	***d'***	**Hit**	**FA**	***d'***
Rhyme	.77	.08	2.75	.83	.10	2.78
Tone	.79	.12	2.58	.78	.07	2.94
R+T	.79	.26	1.73	.81	.17	2.27
	English Listeners					
	Experimental			Control		
	Hit	**FA**	***d'***	**Hit**	**FA**	***d'***
Rhyme	.59	.35	.63	.57	.24	1.06
Tone	.65	.56	.39	.64	.56	.33
R+T	.59	.37	.64	.57	.24	1.12

From the hypothesis that perceptual units may erroneously recombine during dichotic listening, experimental trials should constitute a situation of lesser discriminability than control trials [4, 13] because the former contain all the information necessary for (mis)perceiving the targets while the latter do not. Thus, lower discriminability (*d'*) in experimental trials than in control trials, resulting in a positive migration rate, is a sign that migration occurred.

A two-way mixed ANOVA was performed on the migration rate, with the within-subject factor Unit (rhyme, tone, R+T) and the between-subject factor Listener (Mandarin, English). The main effect of Unit was significant, $F(2, 90) = 6.42$, $p < .005$, but the main effect of Listener was not, $F(1, 45) < 1$. However, a significant interaction between Unit and Listener, $F(2, 90) = 7.43$, $p < .005$, showed that units migrated to different extents in the two language groups. In particular, a clear 2-by-2 cross-over interaction appeared between Unit (restricted to Rhyme and Tone) and Language group, $F(1, 45) = 3.86$, $p = .001$: Rhyme migration was larger for the English than Chinese listeners, $t(45) = 2.19$, $p < .05$ and the opposite was observed for tones, $t(45) = -2.79$, $p < .01$. There was no difference between the two groups in the R+T condition, $t(45) = -.39$, $p = .70$. For the Mandarin listeners, while Tone migration was numerically larger than Rhyme migration, the difference only showed a trend toward significance, $t(21) = 1.71$, $p = .10$. R+T migrated more than Rhymes, $t(21) = 3.24$, $p < .005$, and there was no difference between R+T and Tones, $t(21) = 1.33$, $p = .20$. For the English listeners, Rhymes migrated more than Tones, $t(21) = 3.28$, $p < .005$, and less than R+T, $t(21) = 3.74$, $p < .001$. No difference was observed between Rhymes and R+T, $t(21) = .48$, $p = .64$.

Figure 1: Migration rate (and error bars) as a function of Unit and Listener. Significance of migration rate, i.e., departure from .00, is shown as: ns = $p > .05$; ** = $p < .01$; *** = $p < .001$.
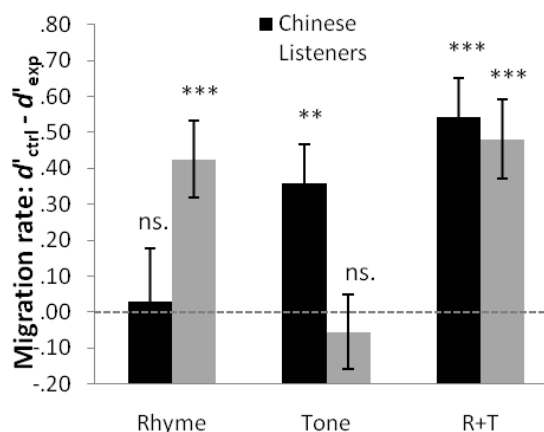
4. DISCUSSION

These results lead to several conclusions. First, contrary to a unitary view of the phonological system, whereby tones and rhymes are represented in a combined rather than independent format, our results revealed distinct migration patterns for the two units. Following the migration rationale, it can thus be inferred that tones are likely to have

perceptual and representational distinctiveness from their tone-bearing rhyme. This conclusion is in keeping with the autosegmental theory, which posits that tones and their corresponding segments should indeed be distinguished on a representational level. Support from the present study is all the stronger because the illusory nature of the phenomenon makes a meta-linguistic account of the results unlikely. A potential question for future studies is whether this pattern is equally strong for single-segment rhymes (nucleus only) and multi-segment rhymes (nucleus + coda), and hence, the degree to which autosegmental theory might have to be modified accordingly.

Second, the distinct patterns we found for the Mandarin and English listeners confirm the effect of language exposure on perceptual processes and mental representations. Tones and R+T migrated to the same degree in Mandarin listeners, but rhymes did not migrate at all. This suggests that, for those speakers at least, tones have greater representational flexibility (or *mobility*, cf. [3]) than strings of segments. In contrast, English listeners experienced comparable migration rates for rhymes and R+T, but no migration of tones. The latter probably indicates that English speakers, having no experience with linguistic tones in their language, have no stored representations for them and are therefore likely to ignore tonal information altogether when listening to Mandarin stimuli. The equally high R+T migration rate for Mandarin and English listeners confirms the primacy of the rhyme for its tonal content in Mandarin and for its segmental content in English. A factor worth exploring for further research is the extent to which the pattern of tone migration in Chinese listeners is modulated by the laterality of the ear in which the target tone is played. Evidence for left lateralization of lexical-tone processing in Chinese listeners as opposed to right lateralization for low-level tone processing [2] suggests that the strength of the migration pattern could very well be asymmetrical as well.

In summary, the present study provides empirical support for the autosegmental assumption that tones and segments have independent representations, and does so by using a methodology that minimizes strategic control and recruitment of meta-linguistic knowledge. The study also shows that such representations are deeply shaped by the listeners' linguistic experience.

5. REFERENCES

[1] Fodor, J., Pylyshyn, Z. 1981. How direct is visual perception? Some reflection on Gibson's "Ecological Approach". *Cognition* 9, 139-196.

[2] Gandour, J., et al. 2004. Hemispheric roles in the perception of speech prosody. *Neuroimage* 23, 344-357.

[3] Goldsmith, J. 1979. *Autosegmental Phonology*. New York: Garland Publishing.

[4] Kolinsky, R. 1992. Conjunction errors as a tool for the study of perceptual processes. In Alegria, J., Holender, D., Morais, J., Radeau, M. (eds.), *Analytic Approaches to Human Cognition*. Amsterdam: North-Holland, 133-149.

[5] Kolinsky, R., Morais, J., Cluytens, M. 1995. Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language* 34, 19-40.

[6] Marcel, A.J. 1983. Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology* 15, 238-300.

[7] Mattys, S.L., Melhorn, J.F. 2005. How do syllables contribute to the perception of spoken English? Evidence from the migration paradigm. *Language and Speech* 48, 223-253.

[8] Mattys, S.L., Samuel, A.G. 1997. How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory and Language* 36, 87-116.

[9] Morais, J. 1985. Literacy and awareness of the units of speech: Implications for research on the units of perception. *Linguistics* 23, 707 -721.

[10] Morais, J., Cary, L., Alegria, J., Bertelson, P. 1979. Does awareness of speech as a sequence of phones arise spontaneously? *Cognition* 7, 323-331.

[11] Morais, J., Kolinsky, R. 1994. Perception and awareness in phonological processing: The case of the phoneme. *Cognition* 50, 287-297.

[12] Treisman, A., Paterson, R. 1984. Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance* 10, 12-32.

[13] Treisman, A., Schmidt, H. 1982. Illusory conjunctions in the perception of objects. *Cognitive Psychology* 14, 107-141.

[14] Ye, Y., Connine, C.M. 1999. Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes* 14, 609-630.

[15] Zhou, X., Marslen-Wilson, W. 1994. Words, Morphemes and syllables in the Chinese mental lexicon. *Language and Cognitive Processing* 9, 393-422.

---

[1] Diacritics for tones: ī = Tone 1; í= Tone 2; ǐ = Tone 3; ì= Tone 4