

This is a repository copy of *Separability of tones and rhymes in Chinese speech perception: Evidence from perceptual migrations*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/105532/>

Version: Accepted Version

Article:

Zeng, Biao and Mattys, Sven orcid.org/0000-0001-6542-585X (Accepted: 2016)

Separability of tones and rhymes in Chinese speech perception: Evidence from perceptual migrations. *Language and Speech*. ISSN: 0023-8309 (In Press)

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

**Separability of tones and rhymes in Chinese speech perception:
Evidence from perceptual migrations**

Biao Zeng^a and Sven L. Mattys^b

^a Department of Psychology, Bournemouth University, UK

^b Department of Psychology, University of York, UK

Corresponding authors: Biao Zeng, Department of Psychology, Bournemouth University, Poole, BH12 5BB, UK, bzeng@bournemouth.ac.uk, and Sven Mattys, Department of Psychology, University of York, York, YO10 5DD, UK, sven.mattys@york.ac.uk.

ABSTRACT

This study used the perceptual-migration paradigm to explore whether Mandarin tones and syllable rhymes are processed separately during Mandarin speech perception.

Following the logic of illusory conjunctions, we calculated the cross-ear migration of tones, rhymes, and their combination in Chinese and English listeners. For Chinese listeners, tones migrated more than rhymes. The opposite pattern was found for English listeners. The results lend empirical support to autosegmental theory, which claims separability and mobility between tonal and segmental representations. They also provide evidence that such representations and their involvement in perception are deeply shaped by a listener's linguistic experience.

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

1. Introduction

Whether tones and segments are represented separately or integrally has been a long-standing debate in speech-perception research (e.g., Lee & Nusbaum, 1993; Ye & Connine, 1999; Zhou & Marslen-Wilson, 1994). From a phonological perspective, Goldsmith's autosegmental theory (Goldsmith, 1979) has received widespread attention for its claim that tones are represented in a separate tier from segments, even though both are co-registered at the phonetic level. This claim is supported by the concept of tone mobility: Depending on context (e.g., citation form vs. connected speech), tones can change their affiliation from one segment to another (Yip, 2002). In perceptual terms, tone mobility means that the interpretation of tones is subject to contextual analysis, and hence, that it might be delayed relative to the analysis of the tone-bearing segments. Indeed, Cutler and Chen (1997) found that tones take more time to process than their host segments (see also Speer, Shih, & Slowiczek, 1989; see Schirmer, Tang, Penney, Gunter, & Chen, 2005, for a decisional rather than perceptual account), which led Ye and Connine (1999) to suggest that tones should be represented separately from segments in computational models. Evidence in partial support of that claim was provided recently by Sereno and Lee (2015) through auditory priming. They showed that, while a joint segmental and tonal overlap between prime and target led to substantial priming, an overlap in tone only did not. The authors interpreted their results as showing that tones are less constraining than segments in Mandarin speech recognition.

Although the evidence for independent contributions of segments and tones to on-line speech recognition is compelling, the issue of representational separability is comparatively poorly documented. For exception, word games in some African languages offer indirect, though thought-provoking evidence that a distinction ought to

be made between segmental and suprasegmental information at a representational level. For instance, in Mangbetu, a language spoken in the northeast of Congo, a word game called *nekondi* involves swapping syllables without swapping their tones (Demolin, 1991). Thus, in that game, the tonal envelope of a word is unaffected by, and independent of the segmental reorganization imposed by the rules of the game—an example of *segment mobility*, the corollary of tone mobility. Interestingly, Hombert (1986) showed that receptivity to this kind of game when taught to adult speakers, was substantially influenced by the status of tones in the native language of the player, suggesting that the psychological representation of tone may be constrained by the native language.

The goal of this study is to address the issue of representational independence between tones and segments using the migration paradigm (Kolinsky & Morais, 1996; Kolinsky, Morais, & Cluytens, 1995; Mattys & Melhorn, 2005; Mattys & Samuel, 1997). The migration paradigm is based on the illusory-conjunction phenomenon originally reported in vision research to uncover the primitive features of visual object perception (Treisman & Schmidt, 1982). In a typical migration experiment in the speech domain, participants hear two spoken stimuli played dichotically and are asked to report whether a pre-specified target was present in either ear. The paired stimuli are manipulated so that accidental cross-ear migration of portions of the stimuli can lead to the illusory perception of the target.

Using this technique, Kolinsky et al. (1995) showed that French listeners were more likely to erroneously recombine the syllables of dichotically-presented French disyllables than any other units (vowels, consonants, voicing, and place of articulation). For instance, relative to a control condition, participants reported hearing a target word, e.g., “bijou” (/bi·zu/, “jewel,” with · representing a syllable boundary),

more often in a dichotic pair inducing syllable migration, e.g., /bi:tõ/-/kɔ:zu/ (underlined is the information forming the target word /bi:zu/), in a pair inducing the migration of the first vowel, e.g., /bɔ:zu/-/ki:tõ/, or the first consonant, e.g., /ki:zu/-/bɔ:tõ/. Note that in all three conditions, the information making up the target was always present. However, it was distributed across the ears in three different ways. Under the assumption that the migration of linguistic units reflects the involvement of these units as independent building blocks during speech perception (Kolinsky, 1992; Treisman & Paterson, 1984), the predominance of syllable migration was taken to suggest that French listeners rely on a syllabic code to perceive their native language. By extension, the logic of the migration paradigm is that the predominant perceptual units of a language should be the ones with the highest migration rate.

The migration paradigm is well suited to the study of underlying linguistic representations because, being an illusory phenomenon, it bypasses conscious access to knowledge and metalinguistic analysis (e.g., Marcel, 1983). For example, although illiterate Portuguese speakers have no conscious awareness of phonemes as measured by explicit phoneme-manipulation tasks (Morais, Cary, Alegria, & Bertelson, 1979), they experience phoneme migration to the same extent as literate speakers do (Morais & Kolinsky, 1994; see also Morais, Castro, Scliar-Cabral, Kolinsky, & Content, 1987, for feature blending). Importantly, too, the migration paradigm was recently used to show that the perceptual segmentation of Japanese stimuli into morae, syllables, and phonemes was relatively unaffected by orthographic representations (Nakamura & Kolinsky, 2014). Thus, the paradigm's ability to measure the involvement of speech properties that do not need to be accessible to conscious experience or written representations makes it ideal for investigating the perceptual reality of tones and segments.

In Experiment 1, native Chinese speakers were asked to report whether a pre-specified target syllable (e.g., bai2¹) was present in a pair of dichotic syllables. In the key trials, the dichotic syllables allowed the target to be (erroneously) perceived through rhyme migration (e.g., bi2-mai4), tone migration (e.g., bai4-mi2), or rhyme+tone migration (e.g., bi4-mai2). Note that the label rhyme+tone is somewhat arbitrary since evidence for migration in that condition could also be described as resulting from the migration of the initial consonant. However, critically, in both cases, the illusory percept leaves the rhyme and tone undivided.

Following the logic of illusory conjunctions, if tones are represented separately from their tone-bearing rhyme, as predicted by autosegmental theory, both units should migrate independently from each other. Additionally, greater migration of one unit than the other would indicate the units' respective degrees of mobility. On the other hand, if tones and rhymes are stored as single representations, they should either not migrate at all when manipulated independently or both migrate to the same extent as the rhyme+tone units. Finally, to test whether the pattern observed in Experiment 1 was due to the listeners' linguistic experience with tones (as suggested by Hombert, 1986) rather than language-general perceptual processes, the same experiment was run with native English speakers in Experiment 2.

2. Experiment 1

2.1 Method

2.1.1 Participants

Thirty native Mandarin speakers from Beijing or northern China received payment or course credits to participate in the study. They had no self-reported hearing problems.

¹ Mandarin tones are indicated by the number following a syllable: 1 = Tone 1; 2 = Tone 2; 3 = Tone 3; 4 = Tone 4.

2.1.2 Materials

Ten monosyllabic Mandarin morphemes were selected as target syllables, all of which were composed of an onset and a rhyme. To have a diversity of rhyme structures, the rhyme of some of the morphemes contained a coda whereas the rhyme of others did not. One of the target morphemes had tone 1, three had tone 2, two had tone 3, and four had tone 4. Because of the large number of stimuli required for each target and because of the strict phonological constraints imposed by the design (see Table 1 for an example), not all stimuli could be nonwords. Therefore, our stimuli were a mixture of words and nonwords. However, these were randomly mixed across all stimulus categories.

	Target-Absent		Target-Present		
	Exptl.	Ctrl.	Exptl. 1	Exptl. 2	Ctrl.
Rhyme	<u>bi</u> 2-mai4	<u>bi</u> 2-meng4	<u>bai</u> 2- <u>bi</u> 2	<u>bai</u> 2-mai4	<u>bai</u> 2-meng4
Tone	bai4- <u>mi</u> 2	bai4-mi3	<u>bai</u> 2-bai4	<u>bai</u> 2-mi2	<u>bai</u> 2-mi3
R+T	<u>bi</u> 4-mai2	<u>bi</u> 4-meng1	<u>bai</u> 2- <u>bi</u> 4	<u>bai</u> 2-mai2	<u>bai</u> 2-meng1

Table 1: Examples of dichotic pairs for the target bai2. Underlined are the components of the syllables making up the target (bai2).

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. For example, to assess rhyme migration in the target bai2, bi2 was paired with mai4 (see Table 1 for the full example). In order to ascertain that erroneous target detection, when it occurred, reflected genuine cross-ear migration of the test unit rather than mere perceptual confusion due to the complexity of the listening task, 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. For example, the control pair for the experimental pair bi2-mai4 was bi2-meng4, in which everything but the migrating rhyme was identical (ai was substituted with eng). The same logic was applied to the tone and R+T conditions.

Target-present pairs were created following the practice used in previous migration studies (e.g., Mattys & Melhorn, 2005). Nine target-present pairs were created for each target—three for each unit. These pairs consisted of the target itself (e.g., bai2) and the syllable that was used in the corresponding experimental or control pair. To take into account the fact that the illusory perception of a target in the target-absent experimental trials could result from bidirectional migration, there were two target-present pairs (e.g., bai2-bi2 and bai2-mai4) associated with each experimental target-absent pair (bi2-mai4).

In all, the experiment included 300 pairs: 10 target sets 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).

2.1.3. Procedure

The target stimuli were recorded by a female Mandarin speaker and the stimuli for the dichotic pairs by a male Mandarin speaker. The voice contrast was intended to avoid detection responses being based on a simple acoustic match. All stimuli, recorded in a sound-attenuating chamber, were digitized at a 10kHz sampling rate (12 bit A/D) and their intensity normalized. The two syllables within a pair were edited to have the same duration such that their onset and offset were synchronized.

The pairs were played over headphones at approximately 70dB SPL. 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). 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.

2.2 Results and Discussion

Hit rates and false-alarm rates, as well as the d' scores derived from them, were calculated separately for the experimental and control trials in the rhyme, tone, and R+T conditions (Table 2). The hit rates for experimental trials were calculated by averaging the two target-present experimental trials. The migration rate of each unit, 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). From the hypothesis that perceptual units

may erroneously recombine during dichotic listening, experimental trials should constitute a situation of poorer discriminability than control trials (Kolinsky, 1992; Kolinsky et al., 1995), because the former contain all the information necessary for (mis)perceiving the target whereas the latter do not. Thus, lower discriminability (d') in experimental than control trials, resulting in a positive migration rate, is taken as evidence that migration has occurred.

Experiment 1: Chinese Listeners						
	Experimental			Control		
	Hit	FA	d'	Hit	FA	d'
Rhyme	.76	.09	2.63	.81	.11	2.64
Tone	.77	.13	2.47	.77	.08	2.85
R+T	.77	.27	1.68	.79	.18	2.13

Experiment 2: 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

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 in Chinese listeners (Experiment 1) and English listeners (Experiment 2).

A two-way ANOVA was performed on the d' scores by participants², with Trial Type (Experimental, Control) and Unit (Rhyme, Tone, R+T) as repeated-measure factors. The significant effect of Trial Type, $F(1, 29) = 17.45$, $p < .001$, $\eta_p^2 = .376$,

² Analyses are reported by participants only. In migration experiments, the large number of syllables required per stimulus set (as illustrated in Table 1) means that it is difficult to generate a sufficiently large number of sets without recycling syllables across sets, which we wanted to avoid. Therefore, the relatively small number of stimulus sets (10) that resulted from that constraint made it difficult to statistically confirm generalization of all the reported by-participant patterns across sets, although many of the key comparisons reached significance.

showed that d' was lower for experimental than control trials, which confirms that the design was successful in eliciting migrations. A Unit effect, $F(2, 58) = 21.59, p < .001, \eta_p^2 = .427$, indicated generally better discrimination in the rhyme or tone conditions than when rhymes and tones were combined, $F(1, 29) = 46.11, p < .001, \eta_p^2 = .614$ and $F(1, 29) = 22.30, p < .001, \eta_p^2 = .435$, respectively, with no difference between rhymes and tones, $F(1, 29) < 1$.

Critically, Trial Type and Unit interacted, $F(2, 58) = 4.66, p = .01, \eta_p^2 = .139$, which means that migration rates differed between units. A significant Trial Type effect (i.e., evidence for migration) was found in the Tone condition, $F(1, 29) = 15.86, p < .001, \eta_p^2 = .354$, and in the R+T condition, $F(1, 29) = 15.73, p < .001, \eta_p^2 = .352$, but not in the Rhyme condition, $F(1, 29) < 1$. A series of 2-by-2 interaction tests confirmed that the Trial Type effect did not differ between the Tone and R+T conditions, $F(1, 29) < 1$, but that it was larger for the Tone condition than the Rhyme condition, $F(1, 29) = 4.87, p = .03, \eta_p^2 = .144$, and for the R+T condition than the Rhyme condition, $F(1, 29) = 6.77, p = .01, \eta_p^2 = .189$.

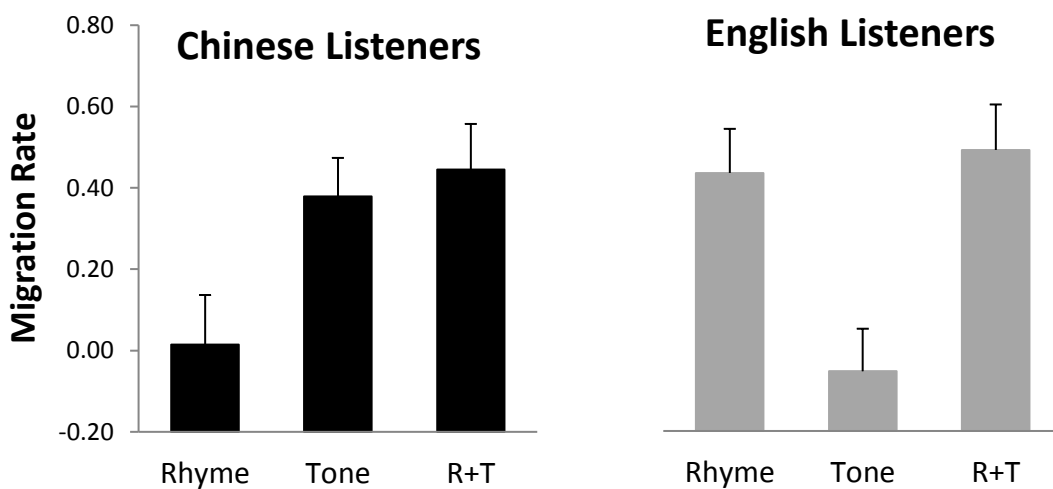


Figure 1: Migration rate (and standard error of the mean by participants) as a function of migrating unit for Chinese listeners (Experiment 1) and English listeners (Experiment 2). The migration rate is calculated as d' control - d' experimental.

These results show that tones can migrate independently from the rhymes they modify. However, there was no indication that rhymes migrated on their own or that the combination of tones and rhymes migrated more than tones alone. Migration of tones is consistent with their presumed mobility across tone-bearing units (Yip, 2002). The absence of rhyme migration suggests that rhymes are rigidly anchored into their segmental frame, and hence, show less mobility than tones. This finding is somewhat unexpected, especially in the context of Demolin's (1991) and Hombert's (1986) report of segment mobility in words games in some African languages where the rhymes of disyllables are swapped while the tonal structure remains the same. We come back to this point in the General Discussion.

A critical question is whether this perceptual pattern is constrained by the listeners' years of experience with tonal phonology and, in particular, the knowledge that tones are formally independent from the segments they modify. If the migration pattern we found reflects perceptual processes associated with the listeners' familiarity with tonal phonology, as predicted by Hombert (1986), we expect it to be attenuated in speakers of a non-tonal language. Specifically, if tonal information is treated as extraneous "noise" by those speakers, then whether the tone associated with the target syllable is heard in the correct ear, the incorrect ear, or neither, should make little difference to the perceiver. The difference between control and experimental trials—our migration index—should therefore be smaller or non-existent. Alternatively, if the migration pattern in Experiment 1 reflects either language-general perceptual properties or stimulus/design properties, speakers of a non-tonal language should exhibit migration patterns similar to those of Chinese speakers. Experiment 2 addressed this question by running Experiment 1 with native English speakers.

3. Experiment 2

3.1 Method

All methodological details were the same as in Experiment 1, except that the participants were 25 native English speakers with no knowledge or experience with Mandarin or other tonal languages. Participants were told that they were going to hear Chinese stimuli and that they should try to spot whether the target stimulus was played in either ear, listening for the closest match with the target.

3.2 Results and Discussion

Data coding and analyses were the same as in Experiment 1. Trial Type was significant, $F(1, 24) = 15.22, p = .001, \eta_p^2 = .388$, and interacted with Unit, $F(2, 48) = 9.47, p < .001, \eta_p^2 = .283$. The Trial Type effect was significant in the Rhyme condition, $F(1, 24) = 15.84, p = .001, \eta_p^2 = .398$, and in the R+T condition, $F(1, 24) = 19.00, p < .001, \eta_p^2 = .442$, but not in the Tone condition, $F(1, 24) < 1$. The Trial Type effect did not differ between the Rhyme and R+T conditions, $F(1, 24) < 1$, but it was larger for the Rhyme condition than Tone condition, $F(1, 24) = 10.79, p = .003, \eta_p^2 = .310$, and for the R+T condition than the Tone condition, $F(1, 24) = 14.03, p = .001, \eta_p^2 = .339$.

These results, which are in sharp contrast with those of Experiment 1, demonstrate that the migration of tones in Chinese listeners is the result of experience with tonal phonology rather than a language-general property or a property of the migration paradigm. Indeed, a cross-experiment analysis showed a significant 3-way interaction between Experiment, Trial Type, and Unit, $F(2, 106) = 8.34, p < .001, \eta_p^2 = .136$, confirming the difference in migration patterns across the two groups. Additionally, a main effect of Experiment showed that the Chinese listeners had larger d' values than English listeners, $F(1, 53) = 61.24, p < .001, \eta_p^2 = .536$. As the group

difference was noticeable in both hit and false alarm rates, the lower performance for the English listeners highlights the distracting nature of tones in establishing a distinct representation of the target. Yet, even at that relatively low discrimination level, the unit contrast was clear and the data showed control-vs-experimental d' differences similar to those of the Chinese group.

4. General Discussion

This study set out to test whether the claim that Mandarin tones and rhymes are represented separately on a phonological level (Goldsmith, 1979; Yip, 2002) can be demonstrated empirically. Using the migration paradigm, an experimental procedure known to tap into pre-attentive stages of speech processing (Kolinsky, 1992; Morais & Kolinsky, 1994), we found that tones are perceptually separable from the rhymes they modify, and that this pattern is tightly linked to the listener's experience with tonal phonology. Chinese listeners experienced tone migration, but not migration of tone-bearing rhymes, whereas English listeners showed the exact opposite pattern. These results lead to two conclusions.

First, contrary to a unitary view of the phonological system whereby tone is considered just another segmental feature (e.g., Duanmu, 2002; Woo, 1970), our results suggest that tones are perceptually and representationally distinct from their tone-bearing rhyme. This conclusion is in keeping with autosegmental theory (Goldsmith, 1979), which posits that tones and their corresponding segments are represented in separate tiers, a proposal endorsed by Ye and Connine (1999). Tone migration also provides empirical support for the phonological notion of tone mobility (Yip, 2002): Not only can tones move temporally from one syllable to another within an utterance, they can also migrate from one spatial sound source to another, as in the

case of dichotic listening—a form of *spatial* mobility. The extent to which our migration patterns would be constrained by top-down factors is unknown, however. In particular, a question for future research is whether tone migration is modulated by sentence-level intonational contour (prosodic deviation from intended tone) and semantic context (perceptual anticipation of intended tone). Given the extensive evidence that on-line tone perception is influenced by both dimensions (e.g., Kung, Chwilla, & Schriefers, 2014; Liu & Samuel, 2007; Ma, Ciocca, & Whitehill, 2006), one would expect perceptual migrations to be affected accordingly, unless the migrating elements reflect abstract rather than perceived representations.

Second, the distinct migration patterns between Mandarin and English listeners confirm the effect of language exposure on perceptual processes and underlying representations. The fact that, in the Chinese group, rhymes were less easily misallocated to their sound source than were tones is consistent with the view that segments take interpretive and temporal precedence over tones during speech perception (e.g., Cutler & Chen, 1997; Speer, Shih, & Slowiaczek, 1989; Tong, Francis, & Gandour, 2008). Greater mobility for tones could be partly the consequence of their weaker constraining power for word identity compared to segments (Serenio & Lee, 2015). Accordingly, the lack of rhyme migration among Chinese listeners might indicate that the dominant and earlier processing of segmental information (Cutler & Chen, 1997) protects segments from instability, perhaps by anchoring attention to the segmental structure first and to the prosodic contour second. Interestingly, listeners might have some control over the dominance of segmental over tonal information. Indeed, the African word games documented by Hombert (1986) and Demolin (1991) showed that rhymes could be moved within a fixed tonal frame, an instance of rhyme

migration. However, the very fact that this type of transposition was instantiated in a game probably reflects its challenging and unusual nature.

Following the principle underlying illusory conjunctions (Treisman & Schmidt, 1982), the absence of tone migration in Experiment 2 shows that tones do not have stand-alone representations for English listeners, or that if they do, those representations are not considered when perceiving speech. Tones only migrated when they were attached to their host rhyme. The absence of a need to integrate segments and tones for English listeners might make it less necessary for rhymes to serve as perceptual anchors, and hence, more likely to migrate (i.e., to be subject to spatial mobility). Rhyme migration in English has not been studied as such before. However, Mattys and Melhorn (2005) showed that vowels in English words migrated significantly, but only if they belonged to stressed syllables, and they migrated more in trisyllables than disyllables. In contrast, entire syllables were far more likely to migrate. Thus, it seems like vowels (or rhymes) constitute units of moderate instability for English listeners, whereas syllables are highly unstable (mobile), possibly because they are the basis of metrical contrasts, and tones are unlikely to migrate on their own because they lack the necessary representations to do so. In contrast, for Chinese listeners, rhymes would be stable and would serve as perceptual anchors for word recognition, whereas tones would be highly unstable (mobile), probably because of their secondary, modifying status.

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

5. References

- Cutler, A. & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, 59, 165-179.
- Demolin, D. (1991). L'analyse des segments, de la syllabe et des tons dans un jeu de langage mangbetu. *Langages*, 101, 30-50.
- Duanmu, S. (2002). *The Phonology of Standard Chinese*. Oxford University Press.
- Goldsmith, J. (1979). *Autosegmental Phonology*. New York: Garland Publishing.
- Hombert, J. M. (1986). Word games: Some implications for analysis of tone and other phonological constructs. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 175-186). Orlando, FL: Academic Press.
- 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.
- Kolinsky, R. & Morais, J. (1996). Migrations in speech recognition. *Language and Cognitive Processes*, 11, 611-620.
- 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.
- Kung, C., Chwilla, D. J., & Schriefers, H. (2014). The interaction of lexical tone, intonation and semantic context in on-line spoken word recognition: An ERP study on Cantonese Chinese. *Neuropsychologia*, 53, 293-309.
- Lee, L. & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics*, 53, 157-165.

Liu, S. & Samuel, A. (2007). The role of Mandarin tones in lexical access under different contextual conditions. *Language and Cognitive Processes*, 22, 566-594.

Ma, J. K., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *The Journal of the Acoustical Society of America*, 120, 3978-3987.

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

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.

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.

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

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

Morais, J., Castro, S. L., Scliar-Cabral, L., Kolinsky, R., & Content, A. (1987). The effects of literacy on the recognition of dichotic words. *The Quarterly Journal of Experimental Psychology*, 39, 451-465.

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

- Nakamura, M., & Kolinsky, R. (2014). Multiple functional units in the preattentive segmentation of speech in Japanese: Evidence from word illusions. *Language and speech*, 57, 513-543.
- Schirmer, A., Tang, S. L., Penney, T. B., Gunter, T. C., & Chen, H. C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience*, 17, 1-12.
- Speer, S. R., Shih, C. L., & Slowiczek, M. L. (1989). Prosodic structure in language understanding: evidence from tone sandhi in Mandarin. *Language and Speech*, 32, 337-354.
- Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*, 23, 698-708.
- Treisman, A. & Paterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance* 10, 12-32.
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107-141.
- Woo, N. (1970). Tone in Northern Tepehuan. *International Journal of American Linguistics*, 36, 18-30.
- Ye, Y. & Connine, C.M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, 14, 609-630.
- Yip, M. (2002). *Tone*. Cambridge University Press.
- Zhou, X. & Marslen-Wilson, W. (1994). Words, morphemes and syllables in the Chinese mental lexicon. *Language and Cognitive Processing*, 9, 393-422.

6. Acknowledgments

This study was made possible thanks to a studentship from the Overseas Research Scholarship (ORS) scheme to Biao Zeng.