Bilingualism: Language and Cognition

cambridge.org/bil

Research Note

Cite this article: Pelzl E, Carlson MT, Guo T, Jackson CN, van Hell JG (2020). Tuning out tone errors? Native listeners do not down-weight tones when hearing unsystematic tone errors in foreign-accented Mandarin. *Bilingualism: Language and Cognition* 1–8. https://doi.org/10.1017/S1366728920000280

Received: 11 September 2019
Revised: 13 March 2020
Accepted: 1 April 2020

Keywords:
foreign-accented speech; adaptation; cross-modal priming; lexical tones; Mandarin

Address for correspondence: Eric Pelzl, E-mail: pelzlea@gmail.com

Tuning out tone errors? Native listeners do not down-weight tones when hearing unsystematic tone errors in foreign-accented Mandarin

Eric Pelzl[1], Matthew T. Carlson[1], Taomei Guo[2], Carrie N. Jackson[1] and Janet G. van Hell[1]

[1]The Pennsylvania State University and [2]Beijing Normal University

Abstract

Listeners can adapt to errors in foreign-accented speech, but not all errors are alike. We investigated whether exposure to unsystematic tone errors in second language Mandarin impacts responses to accurately produced words. Native Mandarin speakers completed a cross-modal priming task with words produced by foreign-accented talkers who either produced consistently correct tones, or frequent tone errors. Facilitation from primes bearing correct tones was unaffected by the presence of tone errors elsewhere in the talker's speech. However, primes bearing tone errors inhibited recognition of real words and elicited stronger accentedness ratings. We consider theoretical implications for tone in foreign-accent adaptation.

Introduction

Listeners can adapt to foreign-accented pronunciation (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Reinisch & Holt, 2014; Xie, Weatherholtz, Bainton, Rowe, Burchill, Liu & Jaeger, 2018). They can even adapt to outright syntactic, semantic, or pronunciation errors produced by second language (L2) speakers (Brehm, Jackson & Miller, 2018; Grey & Van Hell, 2017; Hanulíková, van Alphen, van Goch & Weber, 2012; Lev-Ari, 2015; Samuel & Larraza, 2015). However, adaptation to foreign-accented speech is not always a given. Inconsistent pronunciation patterns – within or across speakers – can prevent or inhibit listener adaptation (Baese-Berk, Bradlow & Wright, 2013; Grohe, Poarch, Hanulíková & Weber, 2015; Reinisch & Holt, 2014; Witteman, Weber & McQueen, 2014; Xie & Myers, 2017). For example, Witteman et al. (2014) found that adaptation to foreign-accented Dutch vowels was delayed when the speaker switched between foreign and nativelike pronunciation. Listeners are also sensitive to the information value of specific acoustic cues (e.g., F0), and will quickly down-weight those that stop being informative for word recognition (Idemaru & Holt, 2011).

The present study considers what happens when an L2 speaker produces frequent pronunciation errors that mislead the listener due to a lack of any underlying pattern – what we call UNSYSTEMATIC ERROR. This occurs in the context of L2 Mandarin speech, where categorical tone errors are common (Chen, Wee, Tong, Ma & Li, 2016).

Accented speech, systematic errors, and unsystematic errors

Nonnative pronunciation takes different forms. Figure 1 illustrates distinctions between ACCENTED PRONUNCIATION and PRONUNCIATION ERROR, and between SYSTEMATIC and UNSYSTEMATIC ERROR. As we assume all speech is probabilistic (Kleinschmidt & Jaeger, 2015), categories are pictured as distributions. The left panel compares native speaker productions for a phonological category (A) with those of an L2 speaker (A′). Compared to the idealized native speaker, the L2 speaker produces shifted approximations of the target category, sometimes nativelike (where distributions overlap), but generally a bit outside of native norms. An illustration might be a speaker who produces the vowel /ɪ/ (as in 'ship') with a sound somewhere between /ɪ/ and /i/ (as in 'sheep') – but not so similar to /i/ that it misleads the listener. Importantly, the accented shift is highly systematic; though probabilistic, it forms a predictable pattern. Research has shown that listeners can quickly learn this type of accented pattern (e.g., Clarke & Garrett, 2004; Xie et al., 2018).

Sometimes L2 pronunciation moves beyond accent into the realm of error (as defined from the listener's perspective). In contrast to accent-shifted pronunciation, errors are categorically inappropriate. They are not just odd-sounding, but potentially MISLEADING as lexical cues. Figure 1 (middle and right panel) indicates such errors as inappropriate categories B′, C′, or D′.

We can also distinguish types of categorical error. The middle panel depicts systematic pronunciation errors: though not appropriate from the listener's perspective, the L2 category is

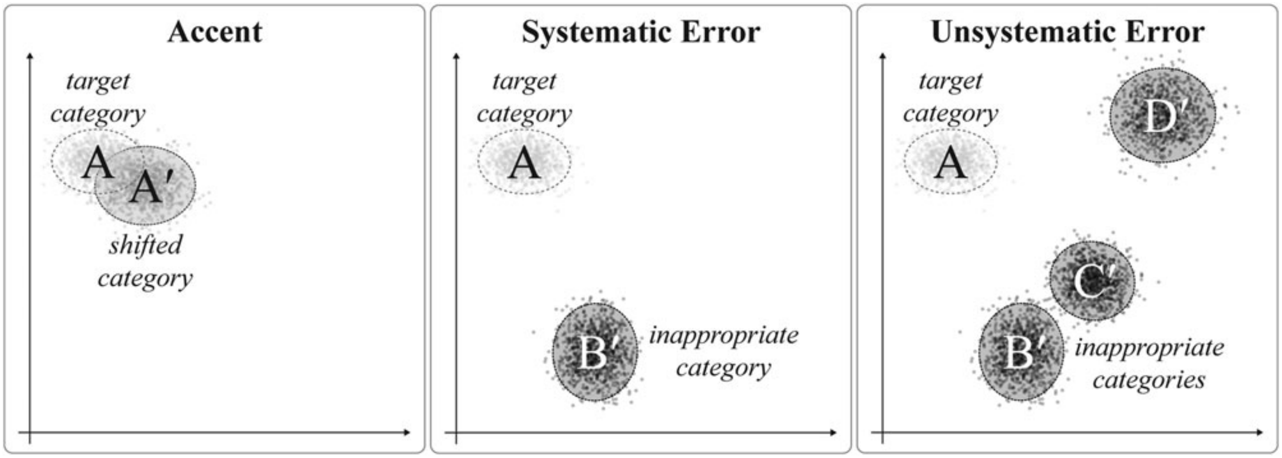

Fig. 1. Illustration of pronunciation error types in L2 speech. Distributions are mapped along two undefined dimensions in phonetic space. The left panel depicts an accent-shifted category (A′) with realizations that approach and sometimes overlap with the native phonological category (A). The middle panel depicts systematic error, where B′ is realized as an inappropriate but consistent category. The right panel depicts unsystematic errors realized variably and unpredictably as belonging to multiple inappropriate categories (B′, C′, D′).

still consistent and predictable. As an illustration, the vowel /ɪ/ might always be pronounced so similarly to /i/ that it creates lexical ambiguity (e.g., 'ship' vs. 'sheep'). This ambiguity might initially confuse listeners, but given enough experience, they can adapt (Samuel & Larraza, 2015). The right panel illustrates unsystematic categorical errors. Now the expected category is sometimes realized as B′, C′, or D′, without any underlying pattern. This implies that listeners have nothing to learn except that the speaker makes frequent errors. This would be akin to hearing /ɪ/ pronounced sometimes as /i/, sometimes as /e/ ('shape'), and sometimes as /ɛ/ ('shep', a nonword).

These distinctions are theoretically important because they have consequences for models of listener adaptation (e.g., Kleinschmidt & Jaeger, 2015). They are practically important because they have implications for what L2 speakers can and cannot do to influence listener behavior. Critically, although unsystematic pronunciation errors may not be the norm in all L2 contexts, they are typical for L2 Mandarin tone.

Mandarin tones and L2 tone errors

Modern Standard Mandarin has four lexical tones, conventionally numbered 1–4 (Figure 2). Tone 1 has a high-level pitch (indicated by an iconic level diacritic over a vowel in Pinyin romanization: *ā*); Tone 2 has a rising pitch (*á*); Tone 3 has a low (sometimes dipping) pitch (*ǎ*); Tone 4 has a falling pitch (*à*). Additionally, sometimes syllables bear a so-called 'neutral tone', with their pitch determined primarily by the tone of the preceding syllable (cf. Chen & Xu, 2006; Lee & Zee, 2014). While tones often disambiguate monosyllabic words (*tāng* 'soup' vs. *táng* 'candy'), for disyllabic words a tone deviation will likely produce a nonword (*tāngchí* 'soup spoon' vs. nonword *tángchí*) (Pelzl, 2018).

L2 Mandarin learners often struggle to produce tones accurately. They have been reported to produce what we would classify as accent-shifted tones, e.g., slightly too high or low in onset or overall pitch, but without necessarily becoming categorically inappropriate (cf. Miracle, 1989; Shen, 1989; Wang, Jongman & Sereno, 2003). They also produce categorical tone errors (Chen et al., 2016; Zhang, 2010). Chen et al. (2016) report that novice L2 speakers made tone errors on 32% of all syllables. This was the case despite the fact that participants were reading words

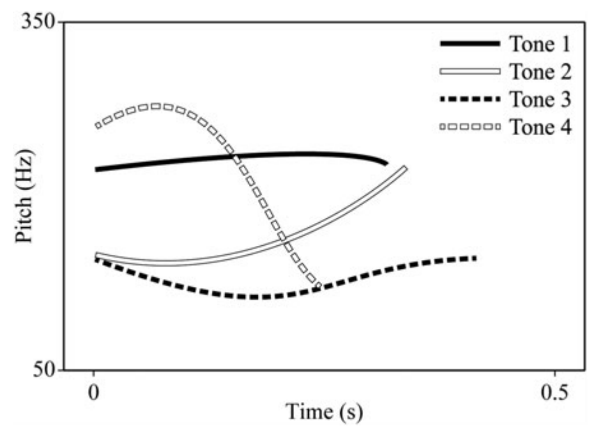

Fig. 2. Pitch contours of the four Mandarin tones produced in isolation

with tones marked explicitly. No previous studies have noted the distinction between systematic and unsystematic errors, but recent work indicates that even advanced L2 learners often have incorrect, uncertain, or incomplete knowledge of tones for known vocabulary (Pelzl, 2018). As these gaps in knowledge are unique to each learner, the resulting errors are largely unsystematic.

In the context of tones, a systematic error occurs when a speaker consistently produces one tone when another is appropriate. For example, if a speaker consistently produced Tone 1 in place of Tone 4, the word *èmèng* 'nightmare' would be produced as nonword *ēmēng*. In contrast, unsystematic tone errors occur when a speaker randomly substitutes one tone for another. Rather than *èmèng*, this speaker might produce the nonword *émēng*, with Tone 4 replaced by multiple different tones. While there is some evidence that listeners might adapt to systematic tone errors in native speech (Mitterer, Chen & Zhou, 2011), we are unaware of any studies examining effects of unsystematic tone errors.

Present study

In a cross-modal priming experiment, native Mandarin listeners heard an auditory prime, followed by a visually presented target, and decided if the target was a Chinese word or not. They

https://doi.org/10.1017/S1366728920000280

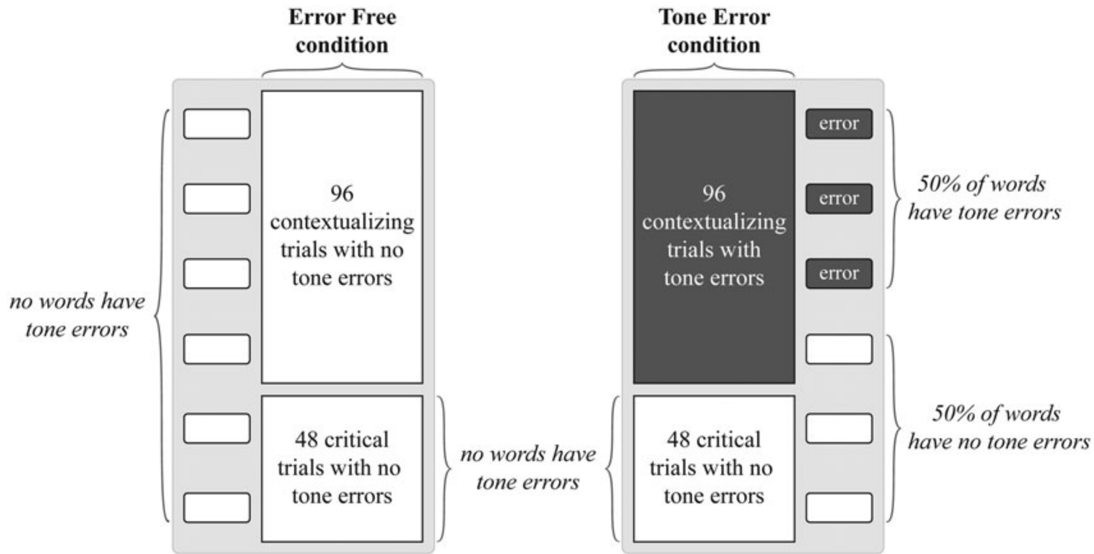

Fig. 3. Overview of trials for the two contextualizing conditions

completed two trial blocks. In both blocks, critical trials were always error free, but contextualizing filler trials differed. In one block, listeners heard an L2 speaker who made no tone errors (Error Free condition). In the other block, a second L2 speaker made unsystematic tone errors in contextualizing trials (Tone Error condition).

Our primary research question was: *Does the presence of frequent unsystematic tone errors impact Mandarin listeners' recognition of foreign-accented speech when tone is produced accurately?* To answer this question, we analyzed response times (RTs) for critical trials (all error free), when the prime either matched (identity priming) or did not match the target word. We compared the indirect effects of contextualizing trials in the Error Free and the Tone Error condition. We call these effects indirect as they reflect the sustained influence of previously encountered tone errors (or lack of errors) on subsequently encountered words that do not contain tone errors. In addition to the down-weighting effects observed by Idemaru and Holt (2011), other recent studies have also observed this type of indirect influence. McQueen and Huettig (2012) found listeners responding more slowly to phonetic cues in clearly produced critical words when context around the words contained intermittent radio static. Similarly, Hopp (2016, Experiment 2) found that German listeners stopped using grammatical gender cues predictively when a speaker made frequent gender errors. Considering such results, we hypothesized that, if listeners learn to expect frequent tone errors from a speaker, they will become uncertain for all tones, and thus slower overall, even on items that the speaker has produced accurately. In other words, they will down-weight tones. Alternatively, it is possible that, despite the demonstrated lack of control of the L2 speaker, listeners will still use whatever tone cues are available, resulting in equivalent RTs for the accurately produced words, regardless of contextualizing condition.

Methods

Participants

We recruited 80 native Mandarin speakers in Beijing, China (26 male, 53 female, 1 other; age: $m = 22.7$, $sd = 3.1$). All were highly educated (2 high school; 40 college; 38 grad school), identified Mandarin as their native language, and reported no history of language or neurological disorder. On a post-experiment survey, most (91.25%) indicated they had little experience speaking to L2 Mandarin speakers. Participants gave informed consent and were compensated for their time. (See online supplementary materials for additional details on all methods, Supplementary Materials.)

Table 1. Prime types (* indicates a syllable with a tone error)

Words	*Translation*	*Tone Error*	*Error location*	*% occurrence*
nénglì	'ability'	**nèng***lì	1st syllable	25%
shíyóu	'oil'	shí**yòu***	2nd syllable	25%
yífàn	'criminal suspect'	**yǐ*fán***	both syllables	25%
yóutǐng	'yacht'	—	none	25%

Materials

Stimulus words were selected from the SUBTLEX-CH corpus (Cai & Brysbaert, 2010). Auditory primes (both critical primes and contextualizing primes) were disyllabic Mandarin words. Visual targets were displayed as Chinese characters, with half of the targets being real words and half nonwords. Half of all primes were identical to the targets, half were unrelated.

Primes

Critical primes (96 total) were high frequency nouns. They were divided into two sets of 48 words, matched for (log) frequency (Set A: $m = 2.82$; $sd = .23$; Set B: $m = 2.82$; $sd = .23$). Two sets of 96 contextualizing filler primes (192 total) were created, with word frequencies balanced between them (set 1: $m = 2.65$, $sd = .33$; set 2: $m = 2.69$, $sd = .34$). To control any set-specific effects, the

https://doi.org/10.1017/S1366728920000280

Table 2. Examples of target items types and their relations to prime words. For Nonword trials, the Prime is a real word and the Target is a homophonous written form that is not a word.

	Trial type	*Prime*	*Tone Error*	*Target*	*Pinyin/Translation*
Real word	identical	nénglì 'ability'	**nèng***lì	能力	nénglì 'ability'
	unrelated	shíyóu 'oil'	shí**yòu***	幻想	huànxiǎng 'illusion'
Nonword	identical	zīyuán 'resources'	**zì***yuán	兹园	zīyuán [nonword]
	unrelated	hēibāng 'gang'	**hèi*bàng***	井申	jǐngshēn [nonword]

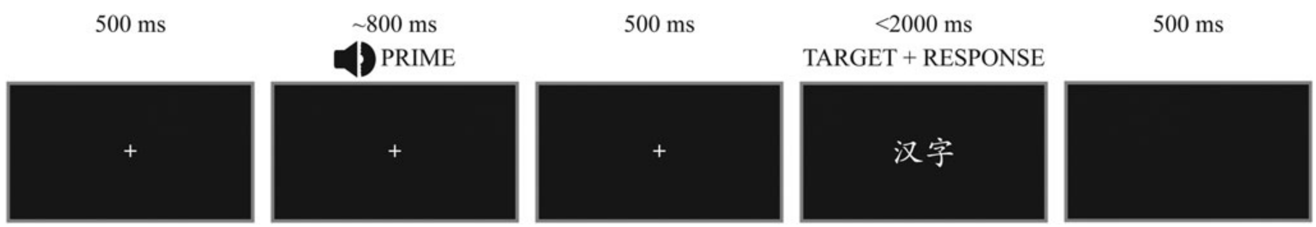

Fig. 4. Timing parameters for trials in cross-modal priming experiment

Table 3. Stimuli examples for the two conditions (Error Free/ Tone Error). Note: 25% of contextualizing trials in the Tone Error condition were free of tone errors.

		Prime				
Stimuli type	*Trial type*	*Error Free*	*Tone Error*	*Target type*	*Target*	*Trials per block*
Critical	*identical*	xīnwén	xīnwén	*real word*	新闻	24
	unrelated	xiāngcūn	xiāngcūn	*real word*	嘴巴	24
Contextualizing	*identical*	nénglì	**nèng***lì	*real word*	能力	12
	unrelated	shíyóu	shí**yòu***	*real word*	幻想	12
	identical	zīyuán	**zǐ*yuǎn***	*nonword*	兹园	36
	unrelated	hēibāng	***hèi*bàng****	*nonword*	井申	36
Total trials						144

pairing of contextualizing primes and critical primes was rotated across participants.

All spoken critical and filler primes in the Error Free condition contained accurately produced tones (Figure 3). In the Tone Error condition, three-quarters of all contextualizing filler primes were produced with a categorical tone error on one or both syllables (Table 1). This meant that, over all trials in the Tone Error condition, an error occurred on 50% of words (72 trials).

Two female L2 speakers of Mandarin (both native English speakers) recorded all auditory stimuli. As these are the only speakers in this experiment, all stimuli are foreign-accented. To avoid the influence of either speaker's specific segmental pronunciation features on outcomes, the combination of speaker, condition (contextualizing stimuli), and critical stimuli were all counterbalanced across participants.

Targets

Targets consisted of two Chinese characters. Half of the targets (72 trials) were real words meant to elicit 'yes' responses, half were nonwords meant to elicit 'no' responses. Half of the real words (36 trials) were identical to the primes, half were unrelated. Nonwords utilized real Chinese characters, but inappropriate combinations, so that participants needed to process the targets before rejecting them. Example stimuli are shown in Table 2.

Nonword targets were evenly distributed across identical and unrelated trials. For the identical nonword trials (36 per condition), nonwords were homophonous with the prime, but infelicitous. For example, the real word prime *zīyuán* ('natural resources') is written 资源. By combining the characters 兹 *zī* ('now, present') and 园 *yuán* ('garden, park'), we created the homophonous nonword 兹园. Importantly, homophonous nonwords provide cues about the accuracy of L2 tones even for nonword trials.

Procedure

The experiment was conducted with a computer in a quiet room in the lab at Beijing Normal University. Participants were

https://doi.org/10.1017/S1366728920000280

Table 4. Overview of RTs and Error Rates in the cross-modal priming task, by experiment block half and overall

		Mean RT (ms)		*Error Rate (%)*	
Condition	*Trial Type*	identical	unrelated	identical	unrelated
Error Free	first half	553 (139)	657 (155)	1.2	6.4
	second half	547 (132)	640 (140)	0.9	7.0
	overall	**550 (136)**	**649 (148)**	**1.0**	**6.7**
Tone Error	first half	552 (129)	649 (143)	0.7	6.0
	second half	547 (136)	643 (132)	0.6	5.0
	overall	**550 (133)**	**646 (138)**	**0.7**	**5.5**

Table 5. Model results (simple effects) for analysis of indirect effect of tone errors

Fixed Effects	*Estimate*	*Std.Error*	*df*	*t-value*	*Pr(>|t|)*
(Intercept: *Error Free/identical*)	550.19	10.46	129.38	52.60	<.001
Tone Error prime	−0.54	5.35	131.51	−0.10	.920
unrelated trial	99.48	8.77	133.69	11.34	<.001
Tone Error prime × unrelated trial	-1.06	5.24	7082.57	-0.20	.840

instructed to respond as quickly and accurately as possible by pressing the "J" key for "YES" (是), "F" for "NO" (否). Timing parameters, illustrated in Figure 4, were modelled after Witteman et al. (2014).

After completing 20 practice trials with feedback (correct/incorrect), participants completed two blocks of trials, one presented in the Error Free condition and the other in the Tone Error condition. In each block, a different L2 speaker produced the stimuli. Speakers were rotated across participants so that each speaker sometimes produced tone errors and sometimes not. Block order was counterbalanced across participants.

Each block contained 144 prime-target trials, with 48 critical trials and 96 contextualizing filler trials (Table 3). Within each block, trials were presented in two sub-blocks of 72 trials, with half of the critical trials in the first sub-block and half in the second sub-block (order of sub-blocks was counterbalanced across participants). The order of presentation was pseudo-randomized uniquely for each participant using *Mix* (van Casteren & Davis, 2006), with the restriction that at least one contextualizing trial had to intervene between critical trials.

The entire task took less than 20 minutes to complete, after which participants answered questions about the accentedness of the L2 speakers, and filled out a language history questionnaire.

Data analysis

Data were processed and analyzed using *R* (R Core Team, 2019) and the *lme4* package (Bates, Mächler, Bolker & Walker, 2015). Analyses reported here used raw RTs. Supplementary materials provide model details as well as alternative analyses, and exploratory analyses of adaptation over halves and trials (Supplementary Materials).

RT results

Average RTs and Error Rates are summarized in Table 4. Error Rates were low overall, though slightly higher for unrelated trials.

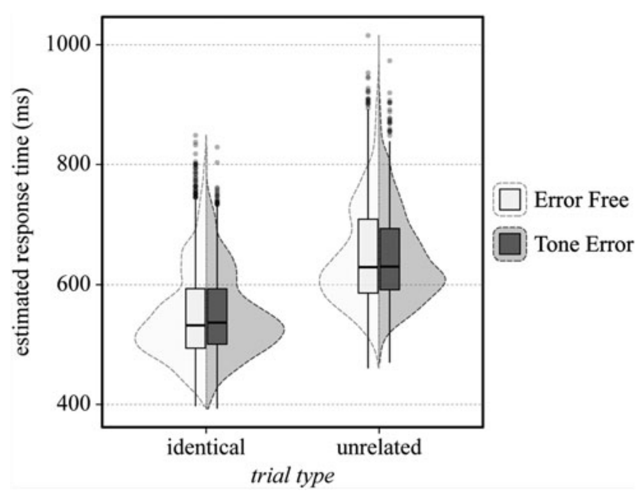

Fig. 5. Boxplots of model estimates for the indirect effect of tone errors. Shaded areas behind boxplots indicate the estimated distribution of responses.

Incorrect trials were removed before further analysis (3.5% of data, 267 trials). RTs show little change from the first to second half of the experiment, suggesting little adaptation occurred.

RTs were submitted to a linear mixed effects model (Table 5). Results revealed a statistically significant effect of trial type, with unrelated trials 99 ms slower than identical trials. The effect of condition and the interaction between condition and trial type were not significant and were very small (about 1 ms each). Model estimates are depicted visually in Figure 5.

Exploratory analysis: The direct effect of tone errors (in contextualizing filler trials)

Although we found no evidence of an indirect effect of contextualizing tone errors on recognition of foreign-accented words,

https://doi.org/10.1017/S1366728920000280

Table 6. Model results (simple effects) for analysis of direct effect of tone errors

Fixed Effects	*Estimate*	*Std.Error*	*df*	*t-value*	*Pr(>|t|)*
(Intercept: *no error/identical*)	549.68	9.77	140.07	56.25	<.001
tone error prime	52.59	11.61	122.15	4.53	<.001
unrelated trial	98.48	9.04	138.67	10.89	<.001
tone error prime × unrelated trial	−57.42	16.46	121.23	−3.49	.001

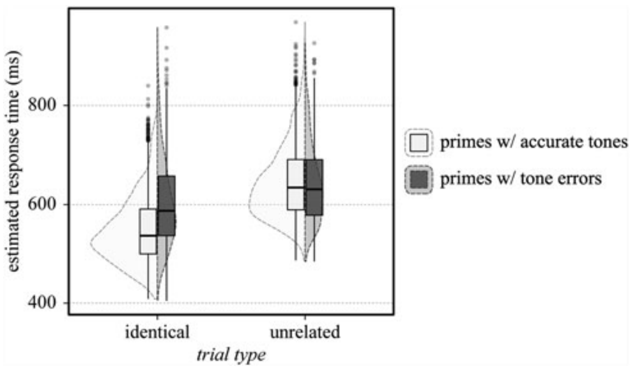

Fig. 6. Boxplots of model estimates for the direct effect of tone errors. Shaded areas behind boxplots indicate the estimated distribution of responses.

we wondered whether there might be a direct effect: that is, whether a prime containing a tone error might inhibit recognition of the visual target that immediately followed. Contextualizing filler stimuli contained nine real word trials with overt tone errors in the Tone Error condition (e.g., the prime *zǐdàn* 'bullet' was mispro-duced as ***zī*****dàn* followed by identical real word target 子弹). By comparing RTs for these trials with RTs for accurately produced words in the critical trials, we explored whether there might be a direct impact of tone errors on RTs. Model results revealed an inhibitory effect of about 53 ms for tone errors (Table 6, Figure 6). As expected, there was also an interaction, indicating that responses were slower for unrelated trials with tone error primes than for related trials with tone error primes.

Post-experiment questions

After the experiment, participants answered four questions about the accentedness of the two L2 speakers. Responses suggest a clear impact of tone errors on listener impressions. In response to the question "*Do you think the speaker is a foreigner?*" 90% of participants identified a speaker as foreign when she had made tone errors, compared to 60% when there were no tone errors. Listeners also tended to rate the Error Free speaker as having a mild accent, and the Tone Error speaker as having a strong accent (Figure 7).

General discussion

We asked whether the presence of frequent unsystematic tone errors would have an indirect effect on Mandarin listeners' recognition of foreign-accented speech when tone is produced accurately. We found typical identity priming effects, but failed to find indirect effects of contextualizing tone errors. However, this does not mean that listeners were insensitive to L2 tone errors. Post-hoc analyses provided evidence of direct inhibitory effects on target word recognition when primes contained tone errors. This aligns with previous studies examining tones in native Mandarin word recognition, which suggest that – relative to identical words – words with mismatched tones are recognized more slowly, though still faster than unrelated words (Lee, 2007; Sereno & Lee, 2014). This inhibition is evidence that tones played an essential role in word recognition during our experiment, that is, they were not just ignored. Additionally, listeners assigned stronger accentedness ratings to the speaker who made tone errors, again indicating that they were not simply tuning out tones altogether.

Why did we fail to find evidence of accent adaptation while previous studies found it? Limitations in statistical power cannot be entirely ruled out – though compared to many previous studies, we used a simpler, within-participants design, more trials, and more participants. For this reason, we do not think this is the best explanation for our results. A more theoretically motivated explanation is that the type of accent/error targeted in previous studies differs qualitatively from the unsystematic errors we investigated. Previous studies tested adaptation to what we would classify as accent-shifted pronunciation or systematic errors (see Figure 1), and found that, while inconsistencies slowed listeners down, they could still adapt to foreign-accented pronunciation (Grohe et al., 2015; Witteman et al., 2014). Critically, such adaptation results in more efficient word recognition for the listener. In contrast, unsystematic tone errors (as commonly found in L2 Mandarin speech) provide no useful cues for adaptation. Even if a listener learns to anticipate the tone errors, they cannot anticipate the specific direction of future deviations. The only adaptation available is global down-weighting of tone cues. This is a negative type of adaptation – avoiding misleading lexical cues – rather than learning to more efficiently recognize words. Our results may simply reflect that listeners are much more resistant to this type of adaptation, or perhaps that priming effects such as those measured here are not sensitive enough to detect it – though a shorter ISI (e.g., 100 ms) or other measures, such as eye-tracking, might be (e.g., Hopp, 2016; McQueen & Huettig, 2012).

Another explanation would be that, under the specific conditions of the present experiment, listeners responded optimally. This aligns with the ideal adapter framework (Kleinschmidt & Jaeger, 2015), which posits that listeners are highly sensitive to probabilistic statistical patterns in speech, and will adapt in a computationally rational way. Within our experiment, the evidence available to listeners indicated that tones – though categorically wrong in 50% of words – were still more often informative than they were misleading (66% accurate overall at the level of syllables). Furthermore, participants generally had little contact with L2 speakers of Mandarin, so, for them, previous experience had consistently shown tonal accuracy to be the norm. Listeners who have experienced more foreign-accented speech might be

https://doi.org/10.1017/S1366728920000280

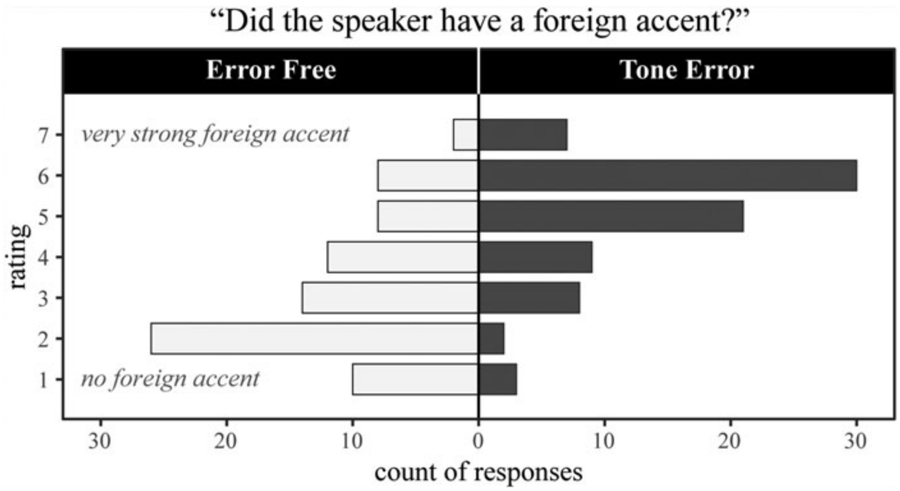

Fig. 7. Accentedness ratings for the speakers without tone errors (left) and with tone errors (right)

expected to adapt more readily under these experimental conditions.

A simple way to probe this further would be to increase the ratio of tone errors to non-errors in contextualizing filler stimuli. For the present study, we chose a moderate frequency of errors in order to approximate what seems typical in L2 production (Chen et al., 2016). A higher error rate in the contextualizing stimuli might be a stronger test of whether there is any indirect effect of unsystematic errors on responses to accurate L2 Mandarin speech. More complex designs might also incorporate a contrast with systematic errors to test whether the presence of unsystematic tone error interferes with adaptation to otherwise learnable accented patterns.

Finally, a theoretically important way in which the current study differs from previous work on foreign-accented speech processing is the linguistic level at which adaptation was targeted. While previous studies tested whether listeners could adapt to a single phonological or acoustic cue (Idemaru & Holt, 2011; Witteman et al., 2014), we tested tone as a phonological class – any given tone could be mistakenly substituted for any other tone. This reflects the fact that, for non-tonal native language speakers, it is not any specific tone contrast, but the entire class of functional tone cues that is novel. Errors at such a level may behave quite differently from more typically examined foreign-accented speech errors that affect only specific or closely related segments. Just as models of speech comprehension are starting to make room for tones (Shuai & Malins, 2017), models of foreign accent adaptation also need to consider potential impacts of tone that do not arise in the context of more commonly studied languages. Can humans adapt at the level of phonological class?

The current study found robust priming when the correct tone was present, and there was no evidence that the size of this effect was diminished when the talker produced tone errors on other words. This research raises important questions about the nature of L2 pronunciation errors, as well as theoretically important issues that arise in the context of lexical tone languages.

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728920000280

Acknowledgements. This research was supported in part by NSF OISE grant 1545900 to Paola Dussias, Judith Kroll, John Lipski, and Janet van Hell; NSF BCS grant 1349110, NSF grant 1561660, and NSF grant 1726811 to Janet van Hell. We are grateful to Arthur Samuel and two anonymous reviewers for their invaluable feedback. We also thank Kunning Yang and her colleagues at the University of Kansas who shared the materials we modified for use as our Chinese language history questionnaire, three Penn State students who agreed to record the Mandarin stimuli, and students at Beijing Normal University for help navigating a busy lab smoothly during data collection.

References

Baese-Berk MM, Bradlow AR and Wright BA (2013) Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America* **133**, EL174–EL180. https://doi.org/10.1121/1.4789864

Bates D, Mächler M, Bolker B and Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48. https://doi.org/10.18637/jss.v067.i01

Bradlow AR and Bent T (2008) Perceptual adaptation to non-native speech. *Cognition* **106**, 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Brehm L, Jackson CN and Miller KL (2019) Speaker-specific processing of anomalous utterances. *Quarterly Journal of Experimental Psychology* **72**, 764–778, 174702181876554. https://doi.org/10.1177/1747021818765547

Cai Q and Brysbaert M (2010) SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One* **5**, e10729.

Chen NF, Wee D, Tong R, Ma B and Li H (2016) Large-scale characterization of non-native Mandarin Chinese spoken by speakers of European origin: Analysis on iCALL. *Speech Communication* **84**, 46–56. https://doi.org/10.1016/j.specom.2016.07.005

Chen Y and Xu Y (2006) Production of Weak Elements in Speech – Evidence from F_0 Patterns of Neutral Tone in Standard Chinese. *Phonetica* **63**, 47–75. https://doi.org/10.1159/000091406

Clarke CM and Garrett MF (2004) Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America* **116**, 3647. https://doi.org/10.1121/1.1815131

Grey S and Van Hell JG (2017) Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics* **42**, 93–108. https://doi.org/10.1016/j.jneuroling.2016.12.001

Grohe A.-K, Poarch GJ, Hanulíková A and Weber A (2015) Production inconsistencies delay adaptation to foreign accents. *Sixteenth Annual Conference of the International Speech Communication Association*. https://www.researchgate.net/profile/Ann_Kathrin_Grohe/publication/281741688_Production_inconsistencies_delay_adaptation_to_foreign_accents/links/55f68fed08ae1d9803976fc7.pdf

Hanulíková A, van Alphen PM, van Goch MM and Weber A (2012) When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience* **24**, 878–887.

Hopp H (2016) Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research* **32**, 277–307. https://doi.org/10.1177/0267658315624960

Idemaru K and Holt LL (2011) Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance* **37**, 1939–1956. https://doi.org/10.1037/a0025641

Kleinschmidt DF and Jaeger TF (2015) Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review* **122**, 148–203. https://doi.org/10.1037/a0038695

Lee C-Y (2007) Does horse activate mother? Processing lexical tone in form priming. *Language and Speech* **50**, 101–123. https://doi.org/10.1177/00238309070500010501

Lee W-S and Zee E (2014) Chinese phonetics. In Huang C-TJ, Li Y-HA and Simpson A (eds), *The handbook of Chinese linguistics*. Wiley Blackwell, pp. 369–399.

Lev-Ari S (2015) Comprehending non-native speakers: Theory and evidence for adjustment in manner of processing. *Frontiers in Psychology* **5**. https://doi.org/10.3389/fpsyg.2014.01546

McQueen JM and Huettig F (2012) Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America* **131**, 509–517. https://doi.org/10.1121/1.3664087

Miracle WC (1989) Tone production of American students of Chinese: A preliminary acoustic study. *Journal of the Chinese Language Teachers Association* **24**, 49–65.

Mitterer H, Chen Y and Zhou X (2011) Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science* **35**, 184–197. https://doi.org/10.1111/j.1551-6709.2010.01140.x

Pelzl E (2018) *Second language lexical representation and processing of Mandarin Chinese tones*. Ph.D. dissertation, University of Maryland, College Park.

R Core Team. (2019) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Reinisch E and Holt LL (2014) Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance* **40**, 539–555. https://doi.org/10.1037/a0034409

Samuel AG and Larraza S (2015) Does listening to non-native speech impair speech perception? *Journal of Memory and Language* **81**, 51–71. https://doi.org/10.1016/j.jml.2015.01.003

Sereno JA and Lee H (2014) The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech* **58**, 131–151. https://doi.org/10.1177/0023830914522956

Shen XS (1989) Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association* **24**, 27–47.

Shuai L and Malins JG (2017) Encoding lexical tones in jTRACE: A simulation of monosyllabic spoken word recognition in Mandarin Chinese. *Behavior Research Methods* **49**, 230–241. https://doi.org/10.3758/s13428-015-0690-0

van Casteren M and Davis MH (2006) Mix, a program for pseudorandomization. *Behavior Research Methods* **38**, 584–589. https://doi.org/10.3758/BF03193889

Wang Y, Jongman A and Sereno JA (2003) Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America* **113**, 1033. https://doi.org/10.1121/1.1531176

Witteman MJ, Weber A and McQueen JM (2014) Tolerance for inconsistency in foreign-accented speech. *Psychonomic Bulletin & Review* **21**, 512–519. https://doi.org/10.3758/s13423-013-0519-8

Xie X and Myers EB (2017) Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers. *Journal of Memory and Language* **97**, 30–46. https://doi.org/10.1016/j.jml.2017.07.005

Xie X, Weatherholtz K, Bainton L, Rowe E, Burchill Z, Liu L and Jaeger TF (2018) Rapid adaptation to foreign-accented speech and its transfer to an unfamiliar talker. *The Journal of the Acoustical Society of America* **143**, 2013–2031. https://doi.org/10.1121/1.5027410

Zhang H (2010) Phonological universals and tonal acquisition. *Journal of the Chinese Language Teachers Association* **45**, 39–65.