YIDDISH Vol 12 Issue 02March 2023
ISSN NO: 0364-4308

# The Perception of Second-Language English Emotional Onsets by Native Mandarin Speakers

**K. Ganga Bhavani Gowri (M.A)[1],D. Mamatha (M.A)[2],L. Sudha (M.A)[3]**

**Assistant Professor[1,2,3]**

## ABSTRACT

***This article presents the results of an experiment that tested the ability of native Mandarin speakers to distinguish between the /t/ and /t/ affricate onset contrasts in two vowel contexts (/I/ and /u/) in English. Specifically for the lengthy voice of onset (VOT) contrast, /t/-/t/, the results imply that vowel quality increases discriminating accuracy, and that /u/ generates a challenging situation for Mandarin listeners. Rich information regarding online processing throughout the identification method was revealed by mouse-tracking data, and various metrics demonstrated a substantial influence of vowel context, which was in accordance with the findings from the discriminating challenge. VOT may have a more subtle function than vowel context, as shown by the fact that it was not observed to influence cursor movements during identification, in contrast to the findings of discriminating.***

**Keywords:***affricate, perception, Mandarin, mouse- tracking.*

## INTRODUCTION

The English /l/-// contrast, for example, is notoriously difficult for native Japanese listeners to accurately perceive [1]. Perceiving unfamiliar non-native and second language (L2) consonants can be difficult for listeners, especially when certain consonants form phonemic contrasts in the L2 but not in the listener's native language (L1). Dissimilarities between L1 and L2 inventories are the root cause of many events of this kind [2]. Non-native category confusions can be influenced by phonological and phonotactic contexts, as shown by additional research [3, 4]; for example, L1 English and L1 French listeners are unable to accurately perceive */to/-/kl/ and */dl/-/l/ contrasts because both */to/ and */dl/ are unattested sequences according to English or French phonology. Comparing the relative ease of */dl/-/l/ and */to/-/kl/, a thorough examination of such patterns suggests that voice onset time (VOT) may play a role in the former. In line with this observation, a recent study [5] demonstrated that L1 Japanese listeners' perception of the English /s/-// contrast is conditioned by the nucleus vowel context: whereas Japanese listeners perform well in discriminating */us/-/u/, they perform significantly worse in discriminating */is/-/I/, where the non-native sequence is unattested in the L1. These investigations show that vowel context, phonotactic quality of the sequence (unattested vs. attested), and maybe voicing specification of the plosives of the contrast (long- vs. short-lag) all have a role in how L2 consonant onsets are perceived. Since English /t/-/t/ and /d/-/d/ diphone contrasts have been reported to be difficult for Mandarin listeners [6, 7], the current research explores how L1 Mandarin speakers understand these contrasts. However, the role that phonological context plays in the processing of these two contrasts is still unclear. Even though the phonological realizations of English /t/ and /d/ are stop-rhotic sequences, their phonetic realizations are more akin to affricate-rhotic sequences [8]. The Chinese language has a rhotic category (//) and both long- and short-lag affricates (/t, d/). Affricate-/w/ sequences are allowed in Mandarin phonology, whereas affricate-rhotic sequences are not [9].Therefore, Mandarin speakers may hear the English phonetic affricate categories /t, d/ as foreign, while they hear the English /t, d/ as native. Understanding how sensitive Mandarin listeners are to labial (rounding) and lingual (narrowing) motions connected with the // segment in unknown phones is the focus of the current investigation. If Mandarin speakers are unable to hear the rhotic segment at the start, they may confuse the sounds /t, d/ with the English sounds /t, d/ since they are phonologically and phototactically equivalent. As an alternative, Mandarin listeners may use some (but not all) of the gestural clues in perception, such as focusing on the labial motion and mistaking English /t/ for the Mandarin sequence /two/. This is confirmed by examples found in the adaptation patterns of Mandarin loanwords, such as the adaptation of the English name Trump as " /twan-pu/. If this is the case, then the perception of /t/-/t/ in English will be mapped to /tw/ in Mandarin, allowing for correct discrimination of unknown categories even when just auditory and gestural clues are available. Mandarin speakers who heavily rely on 2.2. Stimuli the labial gesture (i.e., replacing // with /w/) may still have trouble perceiving /t/-/t/ when the following segment also has the [+labial] feature, such as a rounded vowel (/u/) due to anticipatory coarticulation; in this case, it would be difficult to differentiate between the two affricate categories. In contrast, the labial motion is supposed to be at its most noticeable when an unrounded vowel like /i/ is used (because only /ti/ will be formed with the labial gesture). Accordingly, we anticipate that the vowel context will affect Mandarin listeners' ability to discriminate between and correctly identify English /t/ and /d/, with /u/ resulting in poor discrimination and identification and /i/ resulting in more accurate perception.

Here, we provide an AXB discrimination task and a mouse-tracking identification test to examine how listeners in Mandarin perceive the /t/-/t/ and /d/-/d/ contrasts. Analysis of discrimination accuracy, which is reflective of perceptual outcomes, is used in the discrimination task to examine the pairwise discriminability of /t/-/t/ and /d/-/d/ in English [10]. Because vision, cognition, and hand motion are tightly coupled, and goal-approaching movement is a valid index of cognitive conflicts, a mouse-tracking identification task is used to investigate online processing patterns, i.e. how phonetic-phonological information is integrated during the decision-making process. [11]–[14

# METHODS

## Participants

Twenty native Mandarin speakers (eighteen females and two men) with a dominant right ear took part in the research. No one mentioned having any kind of articulation or hearing issues. All were non-native English speakers who studied abroad in Australia (Mage = 24.3). Eleven of them spoke a regional Mandarin dialect in addition to Standard Mandarin, and two spoke Cantonese. No one was proficient in a third language. On average, they had been learning English as a second language for 13.3 years before coming to Australia, and they had been living there for 2.2 years. On average, they were 22.1 years old when they arrived, and 6.6 years old when they first started showing signs of acquiring new skills. The average vocabulary size test (VST) score for all participants was 8075 [15]. These results indicate that the participants had high levels of proficiency in English as a second language.

## Stimuli:

## The stimuli were eight English CVCV pseudowords,

Phonetically trained male Australian English native speaker: /tu-ti, du-ti, tu-ti, du-ti, ti-ti, di-ti, ti-ti, di-ti/. This provocation were used to generate a total of six contrasts: four essential ones (/ti/-/ti, /tu/-/tu/, /di/-/di/, and /du/-/du/) and two supplementary ones (/du/-/tu/, /du/-/di/).Different vowel contexts (/u/ vs. /i/), phonological structures (real affricate vs. stop-rhotic sequence), and VOT (short- vs. long-lag, or voiced vs. voiceless) characterize the target syllables. To provide a consistent phonological environment across all stimuli, the second syllable /ti/ was introduced. In order to maximize the acoustic variations, the speaker repeated each pseudoword three times in a clear speaking style. Each stimulus word had its first syllable emphasized.

## Procedures

On separate days, individuals performed the discriminating task and the mouse-tracking experiment. Activities included PsyToolkit [16], [17], and PsychoPy [18] were used to gather data while the test was provided online. Participants were given a total of 144 trials (6 contrasts, 4 triplets, and 6 repeats) to complete the discriminating task. Each trial presented the listener with a trio of stimuli (A, X, B) separated by a 1.0 s interstimulus interval (ISI), with X being phonologically similar to either A or B. Phonological processing was boosted by the extended ISI [19]. Within three seconds, the participant was asked to type "F" (X = A) or "J" (X = B) depending on whether they thought the first two or final two stimuli were more comparable.The canvas size in the mouse-tracking identification challenge was normalized to 2 units by 2 units. The "start" box was placed in the center bottom [0, 0] of the screen, as per the standard mouse-tracking paradigm, and the two answer labels were shown at the top left [-1, 2] and top right [1, 2] of the screen, respectively. The listener began each trial by clicking the "start" button, which played the auditory stimuli, and then used the mouse to choose either the "CH" or "TR" or "J" or "DR" category label, which corresponded to the phonemes /t, t, d, d/. After each trial, the "start" box would be reprinted and the user would have to click on it to proceed to the next round of testing. This method insured that the mouse pointer would always be positioned in about the same place. Both the presentation sequence of the stimuli and the direction of correct answers (left vs. right) were randomized. In all, there were 288 possible outcomes throughout the task's 4 consonants, 2 vowels, 3 tokens per combination, 2 orientations, and 6 iterations.During the response process, the mouse movements were tracked. Each user's mouse movements were recorded at a rate of 60 frames per second (FPS), meaning that any two consecutive mouse positions within 17 milliseconds of one another indicate the same cursor movement. All rightward trajectories were converted to leftward trajectories before being used in the statistical analysis. We anticipate that the mouse tracking trajectory would seem like a straight line between the "start" button with the right answer when making an easy choice, but trajectories may be more or less curved when facing cognitive difficulties. Based on the literature [11–13], we analyzed the latency of mouse movements and the complexity of their curves. In specifically, we tracked how long it took for individuals to respond by monitoring their identification RT and

motor pauses (the amount of time spent still after beginning a movement, also measured in seconds). Total trajectory distance and maximum deviation from a straight line were used to quantify the complexity of trajectory curvature (in standard units).

# RESULTS

## AXB discrimination

Participants' performance on the filler contrasts was excellent (/du/-/tu/, 95%; /du/-/di/, 97%), and the filler trials were well-designed. eliminated methodically in the statistical study. To assess the impact of the onset, the vowel, and their interaction while controlling for participants as a random factor, we developed a generalized linear mixed-effects model (GLMM, binomial link) for AXB accuracy (Table 1). Tukey-adjusted post hoc tests showed that the accuracy of the /tu/-/tu/ contrast was considerably worse than that of the /tie/-/tie/ contrast (p =.013), the /di/-/di/ contrast (p =.017), and the /du/-/du/ contrast (p =.027). There was no significant difference between mono- and poly-dialectal Mandarin speakers and those who spoke another language (all t-test p-values >.05). No significant relationships were found (p's >.05, Pearson's r) between accuracy data and the speakers' vocabulary sizes, indicating that the cohort of participants evaluated here was homogeneous.

## Mouse-tracking identification

Table 2 summarizes the identification task accuracy statistics. Accuracy in the /a/ segment ranged from 93% to 97% among the participants. context; however, the /u/ context resulted in much lower performance (57-85%). We constructed a GLMM (binomial link) to describe the interplay between vowel condition (/u/ vs. /I/), phonological structure (affricate vs. sequence), and VOT (short vs. long) for statistical analysis. The vowel condition [2 = 479, p .0001], the structure [2 = 149, p .0001], and the VOT [2 = 16, p .0001] all had significant main effects, as determined by the Wald Chi-squared test. Vowel-VOT interaction effect was also significant [2 = 13.0, p =.0003], as was vowel-structure interaction effect [2 = 9.1, p =.0026], and vowel-structure-VOT interaction [2 = 7.3, p =.0070]. Furthermore, real affricates (/to, du/) exhibited poorer accuracy than the comparable sequence categories (/too, du/), and a shorter VOT led to greater identification accuracy in the /u/ context.

| CV-CV | Condition | Acc. (SD) |
|---|---|---|
| /tʃu/-/tɯ/ | Long VOT | 88 (13) |
| /dʒu/-/dɯ/ | Short VOT | 95 (6) |
| /tʃi/-/tɯi/ | Control | 98 (4) |
| /dʒi/-/dɯi/ | Control | 97 (9) |

***Table 1: AXB discrimination accuracy***

| CV | Acc. (SD) | CV | Acc. (SD) |
|---|---|---|---|
| /tʃu/ | 57 (31) | /tʃi/ | 97 (8) |
| /dʒu/ | 66 (30) | /dʒi/ | 93 (20) |
| /tɯ/ | 78 (26) | /tɯi/ | 96 (6) |
| /dɯ/ | 85 (18) | /dɯi/ | 97 (8) |

***Table 2: Identification accuracy*.**

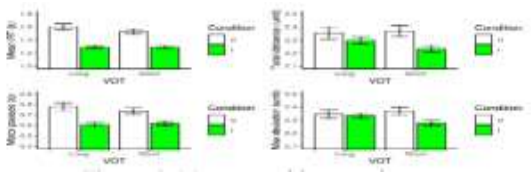

***Figure 1: Mouse-tracking metrics.***

The proper answers to the target responses of "/t/" and "/d/" were the subject of our mouse trajectory study. corresponding distractors, since our listeners had a harder time recognizing the genuine affricate categories compared to the sequence categories, as shown by the accuracy results. As shown in Figure 1, we first analysed

two latency measures in the responses: the mean RT and motor pauses, which we defined as the period the mouse was inactive after the commencement of a motion. To assess the impact of vowel conditions and VOT, we used logarithmic transformations and constructed two linear mixed-effects models (LMMs).A Wald Chi-squared test revealed that vowel condition affected response time (RT) but VOT condition did not (2 = 1.9, p =.1686; p =.2223); thus, vowel condition but not VOT condition affected RT. Listeners took significantly more time to make decisions about identification in the /u/ context. Again, we discovered a main effect of the vowel [2 = 32.8, p .0001] for motor pauses, but we found no significant effect of VOT [2 = 0.6, p =.4533], or vowel-VOT interaction [2 = 1.7, p =.1935], indicating that listeners' pauses were significantly longer in the /u/ context than in the /I/ context. Then, we looked at the overall distance (length) and the maximum deviation (curvature) of mouse trajectories, both of which can be shown in Figure 1. LMMs were also used for analysis of these indicators. Vowel had a significant influence on overall distance [2 = 18.7, p .0001], but neither VOT nor the vowel-VOT interaction did [2 = 0.8, p =.3637, 0.9, p =.3474]. Participants' mouse trajectories were considerably longer in the /u/ context compared to the /I/ context, indicating that vowel condition but not VOT impacted the trajectory lengths.

The results for maximum deviation were similar, with vowel having a significant impact [2 = 15.0, p =.0001], VOT having no effect [2 = 1.6, p =.2025], and the vowel-VOT interaction having no effect [2 = 1.5, p =.2184].

# GENERAL DISCUSSION

Both the AXB and identification tasks supported our hypothesis that Mandarin /u/ might provide a more difficult situation than /I/. English (phonetic) affricate onsets were perceived similarly in both tasks, although there were also some subtle differences. Accuracy information gathered from the AXB assignment revealed that the lengthy VOT condition was the only one in which /u/ posed a challenge. Although accuracy was better in the short VOT condition than the long VOT condition, the identification test showed that the /u/ context created perceptual uncertainty in both cases. It's interesting to note that the disparities between short- and long-lag obstruent's in Mandarin and English are based on comparable VOTs [20]. Our results show that VOT has a more subtle interfering effect than vowel context, and that this effect becomes more pronounced with increasing task complexity, since performing the identification task, but not the AXB task, necessitates additional knowledge to draw the correspondence between orthographic and phonological representations. However, these results are consistent with previous research showing that phonological and phonotactic context influence the difficulty level of non-native consonant perception [5], and that VOT may also play a role in the perceptual ease of unfamiliar onset categories [3, 4]. VOT differences may influence the temporal structure and phasing relations between the articulatory gestures, reducing the perceptual salience of other gestures (such as the lingual and labial gestures for producing the rhotic sound) in favour of aspiration (wide laryngeal). The gesture signals may also be diminished and more challenging for L2 listeners to respond to if high ambition causes partial devoicing of the subsequent Sonoran's. As for why our two experiments yielded such different findings, it's possible that auditory learning comes first, followed by orthography; alternatively, perhaps the explicit metalinguistic knowledge creates another layer of phonological representations beyond perception itself [21], causing the discrimination and identification tasks to draw on distinct L2 phonological systems. See also Table 2 for evidence that listeners may be biased towards the genuine affricate classes when deciding how to pronounce /u/.Finally, the VOT impact was not significant, but the vowel effect was constant across all four online-processing measures for mouse-tracking. Our results showed that Mandarin listeners do indeed use the labial cue, but not the lingual cue, when classifying English /t, d/ sounds. More generally, this study suggests that L2 segment perception is context-dependent; for example, the // segment in /t, d/ is likely to be replaced as a /w/ in the /I/ context, but may be considered as perceptually 'removed' in the /u/ context. Future research could recruit a group of L2 listeners who are relatively inexperienced with the target language to determine if and to what extent English /du/-/du/ can cause perceptual confusion at the beginning of L2 learning to further explore the nuanced effect of VOT. However, the mouse-tracking method offers a plethora of further measures for learning about the mental processes involved in selecting choices [11]- [14]. We believe that mouse-tracking may supplement keystroke paradigms (e.g., AXB/AX tasks or identification by key-pressing) by providing new insights into the live processing of L2 speech input, such as the change of mind end route as shown by the curved mouse trajectories. The sorts of mouse trajectories and the frequency with which they occur under different experimental settings should be analysed in a future research.

# ACKNOWLEDGEMENT

Twenty people took part in the study, and we'd like to give special thanks to Alexander Kilpatrick for his assistance in designing the experiments' stimuli.

# REFERENCES

*[1] A. Sheldon and W. Strange, "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception," Appl. Psycholinguist., vol. 3, no. 3, pp. 243–261, 1982.*

*[2] C. T. Best, "A direct realist view of cross-language speech perception," in Speech perception and linguistic experience: Issues in cross-language research, W. Strange, Ed. Timonium, MD: York Press, 1995, pp. 171–204.*

*[3] C. T. Best and P. A. Hallé, "Perception of initial obstruent voicing is influenced by gestural organization," J. Phon., vol. 38, no. 1, pp. 109–126,*

*[4] P. A. Hallé and C. T. Best, "Dental-to-velar perceptualassimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters," J. Acoustic. Soc. Am., vol. 121, no. 5, pp. 2899–2914, 2007.*

*[5] A. Kilpatrick, R. L. Bundgaard-Nielsen, and B. J. Baker,"Japanese co-occurrence restrictions influence second language perception," Appl. Psycholinguist. Vol. 40, no. 2, pp. 585–611, 2019.*

*[6] Y. Lan, "Perception of English fricatives and affricatesby advanced Chinese learners of English," in Proceedings of INTERSPEECH 2020, H. Meng, B. Xu, and T. Zheng, Eds. Shanghai, China: International Speech Communication Association, 2020, pp.4467–*

*[7] Y. Lan, "Vowel effects on L2 perception of English consonants by advanced learners of English," in Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation, M. Le Nguyen, M. C. Luong, and S. Song, Eds. Hanoi, Vietnam: University of Science, Vietnam NationalUniversity, 2020, pp. 149–157.*

*[8] L. MacLaughlin, /to/ and /ds/ in North American English: phonologization of a coarticulatory effect. Unpublished doctoral thesis: University of Ottawa,2018.*

*[9] S. Danum, The phonology of Standard Chinese. Oxford, UK: Oxford University Press, 2007.*

*[10]W. Strange and V. L. Shafer, "Speech perception in second language learners," in Phonology and second language acquisition, J. G. Hansen-Edwards and M. L. Zampini, Eds. Amsterdam: Benjamins, 2008, pp. 153– 192.*

*[11]M. J. Spivey, M. Grosjean, and G. Knoblich, "Continuous attraction toward phonological competitors," Proc. Natl. Acad. Sci. U. S. A., vol. 102, no. 29, pp. 10393–10398, 2005.*

*[12]P. E. Stillman, X. Shen, and M. J. Ferguson, "How mouse-tracking can advance social cognitive theory," Trends Cong. Sci., vol. 22, no. 6, pp. 531–543, 2018.*

*[13]D. U. Wulff et al., "Movement tracking of cognitive processes: A tutorial using mousetrap," PsyArXiv,2021.*

*[14]Y. Wang, R. L. Bundgaard-Nielsen, B. J. Baker, and O. Maxwell, "Native phonotactic interference in L2 vowel processing: Mouse-tracking reveals cognitive conflicts during identification," in Proceedings of INTERSPEECH 2022, H. Ko and J. H. L. Hansen, Eds.*