

# Perception of English Codas in Various Phonological and Morphological Contexts by Mandarin Learners of English

Valeria Peretokina<sup>1</sup>, Catherine T. Best<sup>1,2</sup>, Michael D. Tyler<sup>1,3</sup>, Jason A. Shaw<sup>1,2</sup>, Bruno Di Biase<sup>2</sup>

<sup>1</sup>MARCS Institute, <sup>2</sup>School of Humanities and Communication Arts, <sup>3</sup>School of Social Sciences and Psychology, University of Western Sydney, Sydney, Australia

[V.Peretokina@uws.edu.au](mailto:V.Peretokina@uws.edu.au), [C.Best@uws.edu.au](mailto:C.Best@uws.edu.au), [M.Tyler@uws.edu.au](mailto:M.Tyler@uws.edu.au), [J.Shaw@uws.edu.au](mailto:J.Shaw@uws.edu.au), [B.Dibiase@uws.edu.au](mailto:B.Dibiase@uws.edu.au)

## ABSTRACT

The present study aims to explore how phonological, morphological and prosodic features of an utterance affect perception of English codas by Mandarin listeners. A phoneme detection task was conducted with the word-final target phoneme /s/ appearing utterance-medially or finally in various phonological (consonant cluster vs. singleton) and morphological (plural morpheme vs. coda of a stem word) contexts. The results indicate that the patterns of second language (L2) perception of /s/ are influenced by the utterance position and phonological complexity of the coda and cannot be explained exclusively by presence or absence of plurality in Mandarin.

**Keywords:** speech perception, phoneme detection, phonology, morphology, sentential prosody.

## 1. INTRODUCTION

Adult L2 learners often find it challenging to perceive and produce certain phonological and morphological features of their target language. These difficulties are especially evident if the same features do not exist in the learners' native language (L1) or if they create phonotactic violations. For instance, English and Mandarin are typologically distant languages that differ greatly in their phonological and morphological structures. Mandarin has a phonotactic restriction against consonant clusters, and only nasals /n, ŋ/ are allowed in syllable-final (coda) position, whereas English permits a variety of singleton and cluster codas [2, 4]. Furthermore, Mandarin contains almost none of the inflectional morphemes that have a high functional load in English (e.g., -s indicates plurality, possessiveness and third person singular) [2]. Due to these phonological and morphological dissimilarities, Mandarin-speaking children, adolescents, and adults learning L2-English are reported to experience persistent difficulties with reliable production of coda consonants that represent English inflectional morphology [6, 10, 11, 16].

Researchers have taken various approaches to account for learner difficulties. For instance, the Failed Functional Features hypothesis [8] posits that

morphological inflections that are expressed lexically rather than grammatically in the L1 will be hard to acquire in L2.

Alternatively, the Consonant Cluster Reduction hypothesis [12] proposes that failure to consistently supply the L2-English past tense inflection -ed cannot be explained by the absence of grammatical tense-marking in Mandarin, as errors only occur when verbs end in a consonant cluster (e.g., *stopped*), and therefore must instead result from the L1 phonotactic constraint against consonant clusters in coda position in Mandarin.

[14] took a different approach and focused on L2-English speech *perception* by Mandarin listeners. The researchers tested automated processing of the phoneme /n/ in a phoneme monitoring task [3]. The target phoneme always appeared in coda position of the syllable [\_Cən], which is phonologically permitted in Mandarin. However, it encoded different morphological structures, some that are also used in Mandarin and some that are not. The syllable represented part of a word stem in nouns (e.g., *kitten*), and non-nouns (e.g., *often*), and a grammatical inflection of past participles in passive voice and perfective aspect (e.g., *taken*). These morphological features are considered congruent in English and Mandarin. It was also used as a derivational inflection in polymorphemic words (e.g., *straighten*), which is a novel morphological feature for Mandarin listeners. Their findings demonstrated that L2 listeners were able to process /n/ successfully across congruent and incongruent word types. Also, like English listeners, they even responded faster and more accurately to morphologically complex than to simple targets. The results indicate that inflections that are not represented in L1 do not pose significant problems for L2 speech processing and confirm that L2 listeners can make use of L1 phonological information to aid them in L2 speech perception.

Following from L1 acquisition research, which shows that children's perception and production of morphophonological features in English as a native language are significantly affected by sentential prosody [15], the researchers also examined the influence of prosodic factors on L2 speech perception. Manipulation of utterance positions of

the target phoneme uncovered that, similarly to L1 acquisition, the targets placed utterance-finally were perceived faster and more accurately than the ones embedded utterance-medially.

The present study sought to explore L2-English speech perception by Mandarin listeners further and to investigate how phonological, morphological and prosodic features of an utterance influence processing of incongruent aspects of English and Mandarin. To achieve that, a phoneme detection study [3] was conducted with a target phoneme /s/ appearing in coda position of monosyllabic words. /s/ exists in Mandarin but is phonotactically restricted to a syllable-initial position; moreover, it does not have a high morphological load that it possesses in English [2]. Therefore in the present study /s/ was embedded in various structures, allowing us to determine whether it is phonological context (singleton vs. cluster) or morphological complexity (singular vs. plural) that hinders L2 perception the most. /s/ encoded:

1. part of a word stem in singular nouns in [Vs] phonological context (e.g., *house, face*);
2. part of a word stem in singular nouns in [Cs] consonant cluster context (e.g., *dance, box*);
3. plural marker -s in [Cs] consonant cluster context (e.g., *lights, parks*).

If L2 listeners rely on morphological knowledge available in their native language, it is hypothesised that Mandarin listeners' processing of /s/ would be slowest when it is a plural marker, as plurality is not expressed grammatically in Mandarin. This prediction is supported by the Failed Functional Features hypothesis [8]. Following the Consonant Cluster Reduction hypothesis [12] we propose that between the two singular word types, automated processing of /s/ would be hindered the most when /s/ appears in [Cs] phonological context as Mandarin has no consonant clusters in coda position.

Since omission of grammatical inflection is reported to be a common error among Mandarin learners of English in speech production [7, 9] this study also aimed to test whether L2 listeners are

sensitive to those errors in perception. Therefore, plural nouns with omitted inflections were included in the stimulus set (e.g., She had many *cat*).

To probe whether the effect of utterance position can be generalised to processing of an unfamiliar morphophonological feature, or if it is significant for congruent structures only, the target phoneme /s/ appeared in the context of five-syllable intonational phrases either utterance-finally (coda of the fifth syllable) or medially (coda of the third syllable).

## 2. METHOD

### 2.1. Subjects

The participants in this study comprised a test group of 16 Mandarin-Chinese (ManC) speakers (9 females;  $M_{age} = 24$  years, Range: 18-32 years) and a control group of 16 Australian-English (AusE) speakers (9 females;  $M_{age} = 26$  years, Range: 18-38 years) living in Sydney, Australia. All participants reported having normal hearing and no speech or language impairments.

AusE participants were born and raised in Australia and reported little to no knowledge of languages other than English. ManC participants had been in Australia for less than one year and were classified as intermediate learners of English, as determined by IELTS scores. They had similar educational backgrounds; previous English exposure was mostly limited to formal instruction in China.

### 2.2. Stimuli

A 26-year-old female monolingual speaker of AusE produced 272 stimulus sentences in a neutral voice. The sentences were divided into three categories: experimental, control, and fillers (see Table 1). Frequency of the targets was controlled for using the CELEX database [1]. Only high-to-medium frequency words were selected as targets so that their semantic difficulty should have minimal effect on participants' performance. Stimuli were recorded

**Table 1:** Stimulus sentence types.

Stimulus Sentences	Quantity	Subgroups	Examples
Experimental	16 (target words) x 2 (utterance positions)	/s/ in a word stem in singular monosyllabic nouns in [Vs] context, medially or finally	We bought a large <i>house</i> . Her new <i>house</i> looked clean.
	8 x 2	/s/ in a word stem in singular monosyllabic nouns in [Cs] context, medially or finally	Buy that candy <i>box</i> . Get a <i>box</i> of fruit.
	16 x 2	/s/ as a plural inflection in monosyllabic nouns in [Cs] context, medially or finally	I don't like cold <i>nights</i> . Here the <i>nights</i> are short.
	16 x 2	/s/ as a plural inflection in nouns in [Cs] with a cue (numerals or quantifiers) preceding the target word, medially or finally	She bought two green <i>plants</i> . All my <i>plants</i> dried out.
Control	16 x 2	plural nouns with omitted inflection in [Cs] context with a cue preceding the target word, medially or finally	I've read many <i>book</i> . Many <i>book</i> lay there.
Filler	48	No /s/	She can ride a bike.
	80	/s/ in open-class words, in consonant cluster or singleton, any non-final position	I drink <i>strong</i> coffee. I had a <i>sore</i> throat.

in a soundproof booth using a Shure SM10A-CN headset microphone, a MOTU Ultralite-mk3 audio interface and Cool Edit Pro 2.1 software at a sample rate of 44.1 kHz.

### 2.3. Procedure

Participants completed a phoneme detection task with stimulus sentences presented in random order with 1-second inter-trial intervals (ITI). Each was tested individually in a sound attenuated room using an Acer TravelMate P653 laptop, Sennheiser HD 650 headphones, an Edirol UA-25EX sound card, and DMDX 4.3.0.0 software [5]. The task was to listen to the utterances and to press a space bar every time they heard /s/, as fast as they could.

Accuracy and reaction time (RT) for correct responses and false alarm rates for control targets with omitted inflections were recorded and analysed. RTs were calculated from the onset of /s/.

## 3. RESULTS

A 2 (*language group*: AusE, ManC) x 4 (*word type*: singular nouns ending in [Vs], singular nouns ending in [Cs] clusters, plural nouns ending in [Cs], plural nouns ending in [Cs] with a cue preceding the target) x 2 (*utterance position*: medial, final) mixed ANOVA was conducted on accuracy (% correct), and another on RT for correct responses to experimental targets. Language group was a between-subjects factor, word type and utterance position were within-subjects factors. Separate ANOVAs were conducted by subjects ( $F_1$ ) and items ( $F_2$ ), and only the results that reached significance in both analyses are reported. The following planned contrasts on the word type factor were tested:

1. Morphological complexity (plural nouns vs. singular nouns);
2. Phonological context (singular nouns with [Vs] vs. singular nouns with [Cs] clusters);
3. Cue provision (plural nouns vs. plural nouns preceded by a cue).

### 3.1. Accuracy for experimental targets

Analyses revealed a main effect of language group,  $F_1(1, 30) = 6.74, p = .014, F_2(1, 104) = 50.51, p < .001$ . As expected, AusE listeners responded more accurately ( $M = 87%$ ) than ManC listeners ( $M = 75%$ ). The main effect of utterance position was also significant,  $F_1(1, 30) = 8.04, p = .008, F_2(1, 104) = 11.13, p = .001$ , with overall more accurate responses to targets in final ( $M = 84%$ ) than in medial position ( $M = 78%$ ). There was a significant effect of phonological context such that participants responded more accurately to singular nouns ending

in [Vs] ( $M = 83%$ ) than in [Cs] ( $M = 78%$ ),  $F_1(1, 30) = 5.91, p = .021, F_2(1, 104) = 4.10, p = .045$ .

A two-way interaction between utterance position and morphological complexity,  $F_1(1, 30) = 21.33, p < .001, F_2(1, 104) = 11.47, p = .001$ , showed that difficulty with the medial position was more pronounced for singular nouns than plural nouns. A two-way interaction between utterance position and phonological context was observed,  $F_1(1, 30) = 14.79, p = .001, F_2(1, 104) = 6.78, p = .011$ , such that more pronounced differences in accuracy between medial and final position were observed for targets ending in consonant cluster than for those ending in a singleton (see Figure 1).

### 3.1. RT for experimental targets

Analyses of the RT data corroborated the main effects of language group,  $F_1(1, 30) = 17.01, p < .001, F_2(1, 104) = 516.96, p < .001$ , and of utterance position,  $F_1(1, 30) = 44.51, p < .001, F_2(1, 104) = 233.22, p < .001$ . Native listeners responded faster ( $M = 525$  ms) than L2 listeners ( $M = 859$  ms), and both groups responded more rapidly to targets in final ( $M = 563$  ms) than in medial position ( $M = 821$  ms). A significant main effect of morphological complexity was also observed,  $F_1(1, 30) = 8.25, p = .007, F_2(1, 104) = 5.80, p = .018$ . Both groups responded more rapidly to plural targets ( $M = 711$  ms) than to singular targets ( $M = 673$  ms). This main effect did not interact with language group, suggesting the pattern was similar for both groups.

A two-way interaction between language group and utterance position  $F_1(1, 30) = 5.92, p = .021, F_2(1, 104) = 24.05, p < .001$ , showed that medial position hindered L2 listeners more than native listeners, as the difference in RT between final and medial position is larger for ManC than AusE (see Figure 2).

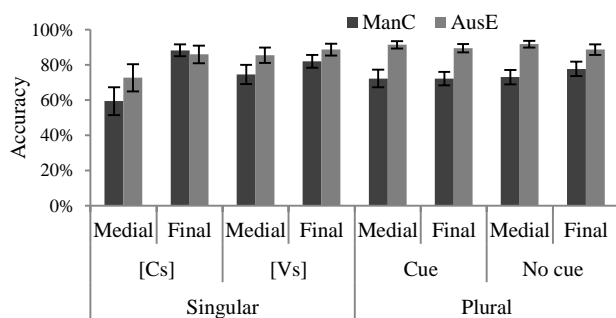
### 3.1. False alarm rate for control targets

A 2 (*language group*: AusE, ManC) x 2 (*utterance position*: medial, final) mixed ANOVA was conducted with language group as a between-subjects factor, utterance position as a within-subjects factor, and with false alarm rate for control targets as the dependent variable. A main effect of language group was found,  $F_1(1, 30) = 7.51, p = .01, F_2(1, 30) = 5.91, p = .021$ , with AusE participants making fewer false alarms ( $M = 15%$ ) than ManC participants ( $M = 25%$ ).

## 4. DISCUSSION

The results from this study provide further insights into how phonological, morphological, and prosodic

**Figure 1:** Language group difference in accuracy as a function of utterance position and word type. Error bars represent standard error of the mean.

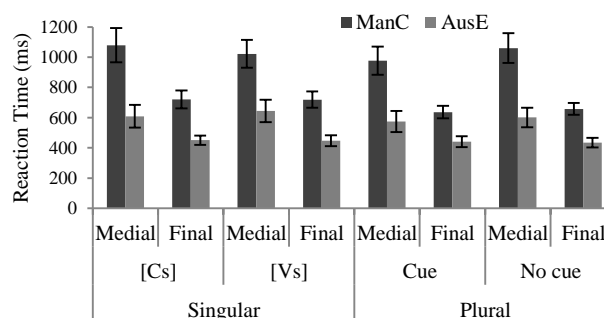


features influence phoneme detection in connected speech in L1 and L2 speech perception.

As anticipated, native listeners were able to react faster and more accurately than ManC listeners to the target phoneme /s/ and they also made significantly fewer false alarms in the control condition with omitted inflections. However, the results indicate that both L1 and L2 participant groups were affected by the utterance position of the target, performing more accurately and rapidly when /s/ appeared utterance-finally than medially. Phoneme detection was more accurate in final than medial position for singular targets, especially for those ending in consonant clusters, which appeared to be difficult for both participant groups to process. Therefore, we posit that final position can aid L1 and L2 speech perception. It is interesting to note that the RT of L2-listeners was influenced by utterance position even more than for native listeners, as their phoneme detection was markedly hindered when the targets appeared medially. Thus, we propose that the effect of sentential prosody extends beyond L1 acquisition and can be generalised to perception of novel L2 aspects.

However, the question as to why targets in medial position are more difficult to process in general, and even more so for L2 learners in particular, remains unresolved. Words in the middle of an utterance are typically not emphasised by pitch accent and are shorter in duration due to the absence of phrase-final lengthening [13]. Moreover, information following the word in medial position causes additional acoustic and cognitive load and, thus, hinders its processing (i.e., backward masking effect). In future studies we will attempt to test the effect of backward masking by conducting the same experiment with no inter-trial intervals. This would ensure auditory stimulation after each final word and thus increase acoustic masking and cognitive load in final position. Another direction to take would be to test a group of “immersed” ManC learners of English, who have lived in Australia for more than

**Figure 2:** Language group difference in RT as a function of utterance position and word type. Error bars represent standard error of the mean.



two years, to examine whether their perceptual patterns are closer to those of native or L2 listeners.

Both language groups responded more accurately to singular nouns ending in singleton [s] rather than in consonant clusters. This finding may seem to support the Consonant Cluster Reduction hypothesis [12] as extended to perception, not just L2 production. However, this interpretation is complicated by the fact that we found phonological context affects both L1 and L2 perception. That is, this perceptual effect is not restricted to the L2 group like cluster reduction in *production*, which is seen in young L1 children but not in L1 adults.

Similarly to the results reported in [14], both participant groups responded faster to morphologically complex targets (plural nouns) rather than to morphologically simple targets (singular nouns), therefore indicating that the patterns of L2 coda perception could not be explained exclusively by grammaticalisation of a certain morphological feature in the L1. If L2-Mandarin listeners were hindered by the absence of grammatical plurality in their native language, they should have exhibited longer RT to /s/ in plural targets, particularly in the condition with no cues.

In summary, our results suggest that utterance position influences both L1 and L2 speech perception, but has a more pronounced effect on non-native listeners, with medial position being harder to process for language learners than for native listeners. At the same time, final position was observed to significantly facilitate processing even of novel L2 structures. The findings also indicate that phonological context of the coda affects processing by native and L2 listeners more than its morphological complexity. Thus, we propose that difficulties in perception of English codas by L2-Mandarin listeners cannot be attributed exclusively to absence of grammatical plurality in their native language and that they need to be examined in conjunction with phonological and phonotactic factors.

## 5. REFERENCES

- [1] Baayen, R. H., Piepenbrock, R., Van Rijn, H. 1993. *The CELEX lexical database (CD-ROM)* Philadelphia: University of Pennsylvania.
- [2] Catford, J., Palmer, J. D., Dew, J. E., Barry, R. G., Cheng, H. L., Hsu, V. L., Li, Y. 1974. *A contrastive study of English and Mandarin Chinese*. Monterey: Defence Language Institute.
- [3] Connine, C. M., Titone, D. 1996. Phoneme monitoring. *Language and Cognitive Processes* 11, 635–646.
- [4] Duanmu, S. 2007. *The phonology of standard Chinese* New York: Oxford University Press.
- [5] Forster, K. I., Forster, J. C. 2003. DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers* 35, 116–124.
- [6] Goad, H., White, L., Steele, J. 2003. Missing inflection in L2 acquisition: defective syntax or L1-constrained prosodic representations? *Canadian Journal of Linguistics* 48, 243–263.
- [7] Hansen, J. 2001. Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics* 22, 338–365.
- [8] Hawkins, R., Liszka, S. 2003. Locating the source of defective past tense marking in advanced L2 English speakers. In: van Hout, R., Hulk, A., Kuiken, F., Towell, R. (eds), *The Interface Between Syntax and Lexicon in Second Language Acquisition*, Amsterdam: John Benjamins, 21–44.
- [9] Jia, G. 2003. The acquisition of the English plural morpheme by native Mandarin Chinese-speaking children. *Journal of Speech, Language, and Hearing Research* 46, 1297–1311.
- [10] Jia, G., Fuse, A. 2007. Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences. *Journal of Speech, Language, and Hearing Research* 50, 1280–1299.
- [11] Ku, Y. M., Anderson, R. C. 2003. Development of morphological awareness in Chinese and English. *Reading and Writing* 16, 399–422.
- [12] Lardiere, D. 2003. Second Language Knowledge of [ $\pm$ Past] vs. [ $\pm$  Finite]. *Proc. 6th Generative Approaches to Second Language Acquisition Conference* Ottawa, 176–189.
- [13] Oller, D. K. 1973. The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America* 54, 1235–1247.
- [14] Peretokina, V., Tyler, M. D., Best, C. T. 2014. Influence of phonological, morphological, and prosodic factors on phoneme detection by native and second-language adults. *Proc. 15th Australasian International Conference on Speech Science and Technology* Christchurch, New Zealand, 211–214
- [15] Sundara, M., Demuth, K., Kuhl, P. K. 2011. Sentence position effects on children's perception and production of English third person singular -s. *Journal of Speech, Language, and Hearing Research* 54, 55–71.
- [16] Xu, N., Demuth, K. 2012. Acquisition of English codas by Mandarin-speaking adult learners of English. *Proc. 14th Australasian International Conference on Speech Science and Technology* Sydney, Australia, 41–44.