

# Acoustic Analysis of English Codas by Mandarin Learners of English

*Nan Xu and Katherine Demuth*

Child Language Lab, Linguistics Department,  
ARC Centre of Excellence in Cognition and its Disorders, Macquarie University

nan.xu@mq.edu.au

## Abstract

Mandarin has a much more limited segmental inventory than English, and permits only nasals in coda position, presenting a challenge for learners of English. However, previous studies have mainly explored this issue using perceptual transcription. This study provides an acoustic analysis of coda consonant productions by Mandarin L2 learners of Australian English. The results indicate that they produced voiceless stop and fricative codas well, but exhibit considerable difficulty with voicing contrasts and coda clusters. These findings and their theoretical implications for current models of L2 learning are discussed.

**Index Terms:** Mandarin, English, second language learning, coda consonants, voicing contrasts

## 1. Introduction

Adult speakers learning a second language (L2) often have difficulties with L2 phonology, especially when these differ in the segments and syllable structures allowed in the native language (L1). For example, Mandarin only permits /n, ɲ/ in word-final (coda) position, whereas English permits a large number of singleton codas and coda clusters. It is therefore of interest to know more about how Mandarin speakers acquire L2 English coda consonants. This is all the more critical given the high functional load of coda consonants in English words and inflectional morphemes (e.g., plural *-s*, past tense *-ed*). This is also an important group to study as there is a large population of Mandarin speakers in Australia.

According to the perceptual assimilation model (PAM) [1], how well an L2 sound can be perceived depends on how similar or different it is to the L1 phonology. L2 contrasts that are equivalent to L1 contrasts should be discriminated at close to ceiling levels while L2 contrasts that resemble good versus poor exemplars of a single L1 phoneme should be discriminated poorly. Similarly, the speech learning model (SLM) predicts that “new” sounds not in the L1 phonology can be acquired with little problem, but L2 sounds that are “similar” to those in the L1 phonology cannot be mastered and continue to be identified as the L1 counterpart [6]. Although the PAM does not make explicit predictions about L2 production, the SLM does, suggesting that L2 sounds that are difficult to perceive will also be difficult to produce.

Indeed, several studies have shown that Mandarin speakers do not find English codas universally difficult to acquire [3, 6, 7]. However, to date few studies have examined the predictions of the SLM for Mandarin speakers learning L2 English for different coda types. One study examining Mandarin speakers’ production of the English alveolar coda voicing /d/ – /t/ contrast did not find support for the predictions made by SLM [6]. The authors predicted that Mandarin speakers should initially have problems producing

this coda voicing contrast as Mandarin does not allow for stop codas, but with increased exposure to native English, their production should match more closely that of native English speakers [6]. Production studies of native English-speaking adults show that there are many potential cues to stop coda voicing, including the duration of the preceding vowel, presence of a voice bar (i.e., low-frequency periodicity indicating continued vocal fold vibration after oral closure), closure duration and the presence of aspiration noise produced after the oral release [5, 9, 11]. However, the results from [6] showed that, regardless of length of exposure to English, Mandarin speakers had non-native acoustic measures for coda voicing contrasts (closure duration, vowel duration, and changes in F1) and used voice bar to indicate voicing on less than 20% of the target /d/ codas.

One explanation offered for the pattern of voicing results reported in [6] was that, compared to velar and bilabial stops, alveolar stops are more variable in L1 English productions, and are often produced as glottal stops or flaps. While all three stops /b, d, g/ were examined in another study [3], place distinctions were pooled in the results so it is impossible to tease apart the voicing characteristics found for alveolar vs. velar and bilabial stops. In addition, codas were coded perceptually so acoustic measures such as vowel and closure duration and voice bar as reported in [6] are not available. In the present study, we carried out acoustic measures for both alveolar /d, t/ and velar /g, k/ codas to determine if voicing contrasts were being made and whether voicing contrasts were more readily observed in velar than alveolar stops.

Another explanation offered for the lack of support for SLM was that, while single word utterances are assumed to be the optimal condition for pronunciation of English, in practice, this may not be the case [6]. Indeed inconsistencies in results can be observed between studies using perceptually coded transcriptions of English codas in free speech [7] vs. words in isolation [3]. The accuracy of voiced coda stops /b, d, g/ was rated better in words produced in free speech (100%, 76% and 100% correct respectively) [7] than for words produced in isolation (0-20% correct) [3]. Furthermore, for words in free speech, voiced stops /b, d, g/ are rated as more accurate than voiceless stops /p, t, k/ [7]. One possibility for these discrepancies is that different acoustic cues for codas may vary depending on the prosodic context in which the word appears. In our study we examined codas in two sentence positions, sentence medial and final, to observe any contextual effects on coda productions.

Apart from [6], no other studies have used acoustic analysis to examine the predictions of the SLM on Mandarin speakers’ acquisition English codas. The aim of the present study was therefore to examine how well the SLM can be used to explain Mandarin speakers’ productions on a range of English codas. The SLM would predict that Mandarin speakers will have problems with all English coda consonants

that are not permitted in Mandarin, at least initially. Since Mandarin only allows for nasal codas, all other English codas, including stops /d, t, g, k/, fricatives /s/ and clusters /ts/, should present a challenge. However, the nasal coda /n/ should be produced close to target-like since it is permitted in Mandarin. We included the cluster /ts/ to test for any effects of phonological complexity. If the documented variability in the realization of alveolar stops leads to difficulties with learning this phonemic contrast, we would expect the acoustic cues to alveolar stop voicing contrasts should be less accurately produced than for velar codas. Finally, any model that assumes L1 to L2 influence could also predict that L2 speakers can apply their knowledge of L1 word initial consonants (onsets) to the production of L2 codas. If so, we would predict native like production of voiceless stops /t, k/ and fricative /s/ since all four are permitted as initial consonants in Mandarin. In contrast, poor productions for the voiced stops /d, g/ and the /ts/ cluster were predicted as none of these are permitted in Mandarin).

## 2. Method

### 2.1. Participants

A total of eight female L1 speakers of Mandarin living in Sydney, Australia participated in the study as part of their undergraduate course requirement. All participants indicated Mandarin as their only native language with no knowledge of any other dialects; seven spoke the Mainland northern dialect of Mandarin and one was from Taiwan. The speaker from Taiwan did not show a different pattern of results from that of the speakers from mainland China and was therefore included in the analysis. The mean age of participants was 23 yrs (range 19-30), mean length of residency in Australia was three and a half years (range 1-9 years), mean length of formal English instruction is 11.83 years (range 7-17 years) conducted in China, where English has recently become part of the mainstream primary and secondary curriculum. The percent of daily English use was 39% (range 20%-50%).

### 2.2. Stimuli

Each target coda was embedded in four high frequency monosyllabic CVC words with half of the words for each coda having high vowels /i, u, e/ and half low vowels /æ, ɛ, ɔ/. The codas examined included (1) voiced and voiceless alveolar stops /d/ - /t/ (*kid*, *bed*, *bud*, *mud*, *cat*, *bat*, *pet*, *net*), (2) voiced and voiceless velar stops /g/ - /k/ (*pig*, *peg*, *dog*, *bug*, *neck*, *book*, *back*, *duck*), (3) the voiceless fricative /s/ (*bus*, *kiss*, *gas*, *mess*), (4) the stop+fricative cluster /ts/ (*cats*, *bats*, *pets*, *nets*), and (5) the nasal /n/ (*bun*, *gun*, *bin*, *pin*). A total of 56 sentences we used, half with the target words in sentence medial position (e.g., *her pets bark*) and half in sentence final position (e.g., *she has pets*). All target words in sentence medial position were followed by a high frequency verb beginning with /b/. All stimuli were recorded in a child-directed speech register by a female monolingual speaker of Australian English. This register is slower in pace and more carefully articulated than speech directed to adults, providing optimal acoustic cues. The order of the sentences was randomized across participants.

### 2.3. Procedure

The data were collected in a sound attenuated room. An elicited imitation task was used where participants sat in front

of a computer and saw a picture accompanied by a semantically related spoken sentence. Participants were asked to repeat each sentence exactly as they had heard it. If participants were unable to produce the sentence after the initial presentation, the audio was repeated a maximum of two more times before moving to the next sentence. Most participants required only one presentation and none continued to have problems after three repetitions.

Productions were recorded via a directional microphone onto a MAC computer using the Protools software, with sampling rate set to 44.1kHz at 16bits per sample recorded on one channel. Uncompressed WAV files were exported for acoustic analysis using the Praat software [2].

#### 2.3.1. Acoustic Analyses

The presence vs. absence of different acoustic cues was coded for each coda. For coda stops, this included the following measures: (1) vowel duration, period from the onset to clear offset of F2 energy, (2) voice bar, period of low frequency periodicity with voicing during closure after the abrupt drop in vowel amplitude, (3) closure, period of abrupt stop in amplitude at the end of vowel in the absence of voice bar before the coda burst or post-release noise, and (4) coda burst and post-release noise, period of aperiodic noise following closure.

For the coda fricative /s/, the acoustic measures included (1) vowel duration, and (2) duration of the frication noise following vowel offset. For the coda cluster /ts/, measures included both those for stops and fricatives.

For the coda nasal /n/ the following measures were taken: (1) vowel duration, and (2) nasal duration, entailing the voiced period of reduced amplitude with either absence of F2 or downward movement for high vowels and upward movement for low vowels.

## 3. Results

The analyses for the number of coda productions are presented first, followed by analyses of the acoustic events for the stop codas.

### 3.1. Coda Productions

Stops were coded as ‘produced’ where there was the presence of at least closure. Fricatives were coded as ‘produced’ where frication was observed, and for cluster /ts/ where both closure and post-release frication noise were observed. Production of /n/ was assumed where nasal cues were observed. Percentage of codas produced is illustrated in Figure 1.

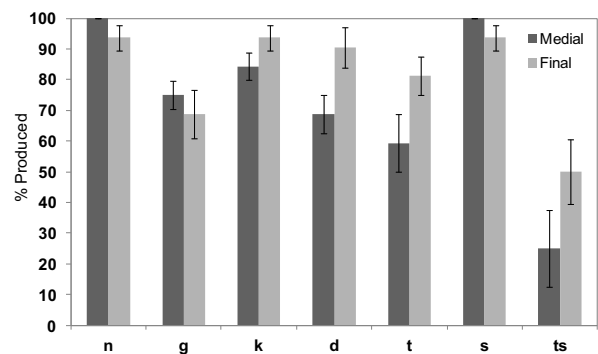


Figure 1: Percentage of codas produced as a function of utterance position (medial, final) with standard error bars

A two-way chi-square test revealed no significant relationship between the production of coda types and utterance position. However, collapsed across positions, a one-way chi-square test with alpha set at .05 revealed a significant difference in accuracy of productions across coda types,  $\chi^2(6, N = 347) = 21.256, p = .002$ . In descending order, the nasal /n/ and fricative /s/ were produced near ceiling at 97% and 98% followed by stops /g, k, d, t/ (87%, 89%, 85% and 74%) and then the cluster /ts/ 38% (medial = 25% & final = 50%). Thus, the stops were produced reasonably well, whereas the cluster was much more problematic.

Figure 2 shows the percentage by which the /ts/ cluster was realized either as /ts/, with the cluster reduced (this always resulted in only the /s/ being preserved), or with total omission of the cluster. To examine cluster realization as a function of sentence position, a two-way chi-squared test (without omissions) was conducted. A significant relationship was found,  $\chi^2(1, N = 58) = 7.895, p = .005$ , with more cluster simplification to /s/ in sentence medial position (75%) but /ts/ tending to be preserved in sentence final position (50%). This pattern is commonly found in studies of typically developing children, suggesting increased perceptual/production challenges in medial position [10].

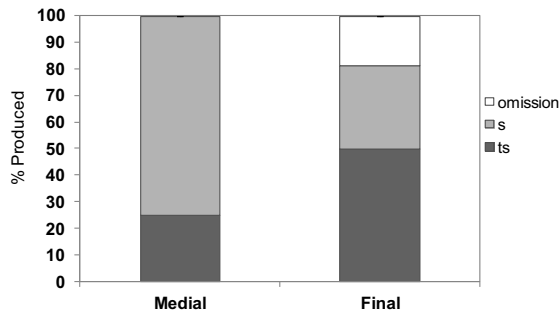


Figure 2: Types of productions made for the /ts/ cluster as a percentage of total /ts/ productions with standard error bars

### 3.2. Acoustic Analyses

Several acoustic cues to stop coda voicing in English were examined including voice bar, vowel and closure duration. The nature of these cues in the Mandarin speakers' L2 English is discussed below.

#### 3.2.1. Voice Bar

Consistent with previous findings, voice bar was produced on less than 20% of target voiced stops, the same rate as for voiceless stops. A two-way chi-squared test of place (alveolar vs. velar) by sentence position (medial vs. final) with alpha set at .05 was not significant. This suggests that voice bar was not a good cue for distinguishing voiced from voiceless consonants as a function of either place of articulation (alveolar or velar stops) or position in the sentence.

#### 3.2.2. Vowel Duration

If Mandarin speakers produced target like voicing contrasts then we should see longer durations in vowels preceding voiced than voiceless stops [5, 8]. See Figure 3 for vowel durations on vowels preceding each stop coda as a function of sentence position. A repeated measures analysis of variance (ANOVA) on voicing (voiced vs. voiceless) by place (alveolar vs. velar) by sentence position (medial vs. final) was conducted with alpha set at .05. Significant main effects were

found for place,  $F(1, 7) = 14.304, p = .007, \eta^2 = .671$ , and position,  $F(1, 7) = 5.181, p = .057, \eta^2 = .425$ , but not for voicing. This suggests that vowel duration was longer for alveolar /d, t/ ( $M = .181$ ) than velar /g, k/ stops ( $M = .167$ ), and longer in sentence final ( $M = .181$ ) than medial positions ( $M = .165$ ). However no differences were found between voiced and voiceless stops.

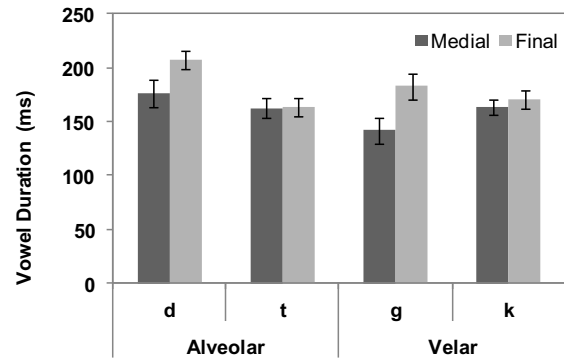


Figure 3: Duration (ms) of vowels preceding Alveolar and Velar voiced vs. voiceless stops as a function of utterance position (medial, final) with standard error bars.

Significant two-way interactions were found for voicing by place of articulation,  $F(1, 7) = 10.367, p = .015, \eta^2 = .597$ , and for voicing by sentence position,  $F(1, 7) = 22.077, p = .002, \eta^2 = .767$ , but place by sentence position was not significant nor was the three way voicing by place by position interaction. This suggests that vowel durations were longer for alveolar voiced /d/ than voiceless /t/ stops but vowel durations did not differ significantly between voiced and voiceless velar stops /g, k/. In addition, vowel duration was longer for voiced than voiceless stops in sentence final positions only.

#### 3.2.3. Closure Duration

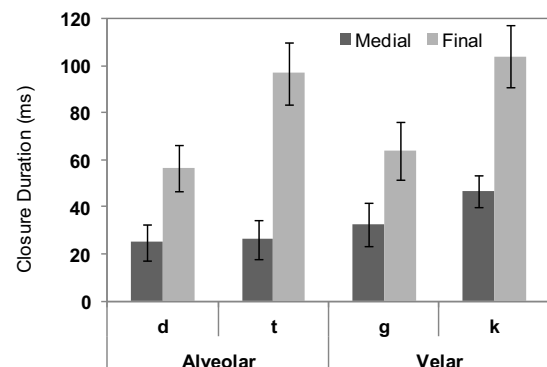


Figure 4: Closure duration (ms) for Alveolar and Velar voiced vs. voiceless stops as a function of utterance position (medial, final) with standard error bars.

Target-like voicing distinctions would predict longer stop closure duration for voiceless than voiced vowels. See Figure 4 for closure durations produced for each stop coda by sentence positions. A repeated measures ANOVA was performed using the same factors as above. Significant main effects were found for all three factors; voicing,  $F(1, 7) = 7.300, p = .031, \eta^2 = .510$ , place,  $F(1, 7) = 16.153, p = .005, \eta^2 = .698$ , and position,  $F(1, 7) = 13.442, p = .008, \eta^2 = .658$ , but no significant two-way or three-way interactions were found. This suggests that closure duration was longer for voiceless /t, k/ ( $M = .072$ ) than voiced /d, g/ ( $M$

= .050) stops, longer for velar /g, k/ ( $M = .067$ ) than alveolar /d, t/ ( $M = .055$ ) stops, and longer in sentence final ( $M = .051$ ) than medial positions ( $M = .039$ ).

#### 4. Discussion

Our results provide varying degrees of support for the SLM's predictions. As expected, Mandarin speakers learning L2 English show little difficulty producing nasal codas, the only type of coda that is permitted in Mandarin. For the English codas not permitted in Mandarin, speakers were able to produce the fricative /s/ well, followed by the alveolar /d, t/ and velar /g, k/ stops which were produced reasonably well. Performance was poor for the cluster /ts/, produced less than 50% of the time. There was no general position effect, codas in sentence medial positions were not harder to produce than in final positions, except for the cluster /ts/.

Several points can be made from this pattern of results. Firstly, how well a coda was produced did not simply rely on whether L1 codas are available for L2 comparison. There appears to be a coda complexity effect, with singleton codas produced better than the cluster. This suggests that syllable structure may influence coda production. If it is a straightforward case of L1 to L2 influence at a segmental level then Mandarin having no codas could conceivably be an ideal situation where "new" L2 coda categories could be created with ease. This does not seem to be the case, nor does the opposite, i.e., that the lack of L1 comparisons makes all L2 codas universally difficult. The alternative, that L2 codas may be influenced by the segmental inventory of L1, even if only found in onsets, seems to handle the segmental coda data presented here.

Another interesting finding is the pattern of /ts/ coda cluster reduction to /s/, especially in sentence medial environments. One possible explanation for this pattern is that /s/ codas are easier to produce (and/or to perceive) than /t/ codas (98% vs. 75% accuracy). Thus, perhaps Mandarin speakers prefer more sonorant codas, as is the case in Mandarin and cross-linguistically [4]. Lastly, all words in our stimuli set ending in /ts/ were plurals (*cats, bats, pets, nets*). It is possible that participants realized that /s/ in this context is an inflectional morpheme which carries more linguistic content than /t/, so /s/ was preserved in their production. In the present study, it is not possible to tease these issues apart. However, given that Mandarin is a language without plural marking, it is unclear how plurals are treated in the L2 phonology of these speakers. A follow up study comparing final /s/ clusters that are either morphemic or not, would answer this question. Whatever the reasons for this pattern of simplification, it is different from L2 onset productions where vowel epenthesis prevails over deletions [8]. Clearly there are many further issues to be investigated comparing the learning of complex onsets and codas in Mandarin speakers' L2 English.

There was little support for the claim that variability in English production of alveolar stops should lead to less target like productions in voicing contrasts for /d, t/ than /g, k/. Of the range of cues examined (presence of voice bar, vowel duration, closure duration), only closure duration was consistently used to signal voicing contrasts, with longer closure durations produced for voiceless than voiced stops regardless of place of articulation. Vowel durations were also longer for voiced alveolar stop /d/ than the voiceless /t/ but no differences were observed for velar stops. This suggests that Mandarin speakers may not be sensitive to the entire set of

cues used in English to signal voicing contrasts. However it is unclear whether Mandarin speakers will acquire a fuller set of cues over time. This would be an interesting issue for future research.

#### 5. Conclusions

Mandarin learners of Australian English produce nasal and voiceless fricative codas well, voiceless stops relatively well, but show difficulty with producing the entire range of cues necessary for stop coda voicing contrasts and for cluster /ts/ codas. Our results suggest that coda complexity influences L2 coda acquisition rather than just presence or absence in L1. In addition, Mandarin speakers are able to produce some cues to voicing contrasts, i.e., closure and vowel duration, so it is possible that increased exposure to English can lead to acquisition of the full range of acoustic cues for producing voicing contrasts. Current models of L2 must be expanded to account for these results.

#### 6. Acknowledgements

We thank Felicity Cox for comments and assistance. This research was supported in part by funding from Macquarie University, ARC DP110102479, ARC CE110001021, and NIH R01 HD057606.

#### 7. References

- [1] Best, C. T., "A direct realist view of cross-language speech perception", in W. Strange (Ed.), *Speech perception and linguistic experience: Issue in cross-language research*, York Press, Baltimore, MD, pp. 171–204, 1995.
- [2] Boersma, P., "Praat, a system for doing phonetics by computer", *Glott International*, 5:9/10, 341-345, 2001.
- [3] Broselow, E. and Zheng, X., "Differential difficulty in the acquisition of second language phonology", *International Journal of English Studies*, 4, 135-163, 2004.
- [4] Clements, G. N., "the sonority cycle and syllable organization" in W. Dressler, H. Luschutzky, O. Pfeiffer, and J. Rennison (eds.): *Phonologica 1988: Proceedings of the 6<sup>th</sup> International Phonology Meeting*, Cambridge: Cambridge University Press, 1992.
- [5] Cole, J., Kim, H., Choi, H., and Hasegawa-Johnson, M., "Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech", *J. Phonetics* 35, 180–209, 2007.
- [6] Flege, J. E., Munro, M. J., & Skelton, L., "Production of the word-final English /t-/d/ contrast by native speakers of English, Mandarin, and Spanish", *The Journal of the Acoustical Society of America*, 92, 128–143, 1992.
- [7] Hansen, J., "Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese", *Applied Linguistics*, 22, 338–365, 2001.
- [8] Chen, S. W., "Acquisition of English onset clusters by Chinese learners in Taiwan", *The University of Edinburgh, Postgraduate Conference*, 2003.
- [9] Song, J. Y., Demuth, K. and Shattuck-Hufnagel, S., "The development of acoustic cues to coda contrasts in young children learning American English", *Journal of the Acoustical Society of America*. 131(4), 3036-3050, 2012.
- [10] Sundara, M., Demuth, K., & Kuhl, P., "Sentence-position effects on children's perception and production of English 3rd person singular -s." *Journal of Speech, Language, and Hearing Research*, 54, 55-71, 2011.
- [11] Wright, R., "A review of perceptual cues and cue robustness", in *Phonetically-Based Phonology*, edited by B. Hayes, R. Kirchner, and D. Steriade (Cambridge University Press, New York, NY), pp. 34–57, 2004.