Influence of Phonological, Morphological, and Prosodic Factors on Phoneme Detection by Native and Second-Language Adults

Valeria Peretokina[1], *Michael D. Tyler*[1,2], *Catherine T. Best*[1,3]

[1]MARCS Institute, [2]School of Social Sciences and Psychology, [3]School of Humanities and Communication Arts, University of Western Sydney, Sydney, Australia

V.Peretokina@uws.edu.au, M.Tyler@uws.edu.au, C.Best@uws.edu.au

Abstract

This study investigates how phonology, morphology, and prosody affect phoneme identification within sentences. The influence of these factors was examined in a phoneme monitoring task with Australian-English and Second-Language-English Mandarin-Chinese participants monitoring for /n/ in four morphological contexts (part of a word stem, correct/altered grammatical inflection, or derivational inflection) and two prosodic contexts (medial vs. final utterance position). Analysis of accuracy and reaction time to targets revealed that final position is more beneficial for phoneme detection across both groups, and that non-native listeners profited from first-language-permissible phonological information rather than suffering from absence of morphological features in their native language.

Index Terms: speech perception, phoneme monitoring, phonology, morphology, sentential prosody.

1. Introduction

Acquisition of phonological, morphological, and prosodic features are all important but challenging steps in mastering a second language (L2), especially for adult L2 learners. So far, difficulties in L2 acquisition have been addressed exclusively from either a phonological or a morphological perspective.

It has been shown that processing of a non-native language by adults is significantly influenced by their experience with the phonology of their native language (L1) [1, 2]. Therefore, it is especially difficult for L2-listeners to perceive inflectional morphological structures added to a word stem if they employ phonemes that do not exist in the L1, or produce sound combinations that are illicit in a native phonological system (i.e., phonotactically impermissible). For instance, the Consonant Cluster Reduction hypothesis [3] suggests that failure to reliably produce the L2-English regular past tense inflection –ed by L1-Chinese speakers cannot be attributed to the absence of tense-marking in Chinese, as they only experience difficulty when producing phonologically complex verbs ending in a consonant cluster (e.g., “stopped”). The deficits in acquisition of the regular past tense inflection are therefore thought to be caused by an L1 phonotactic constraint against consonant clusters in coda position.

Another possibility is that morphological features themselves may not be grammaticalised in the L1, thus hindering their successful perception and production in the L2. Failed Functional Features [4] and Morphological Congruency [5] hypotheses support this approach, suggesting that morphological structures that are not activated in learners’ L1 grammar will be hard to master in L2.

The prosodic features of an utterance present an additional factor that may influence L2 processing, but this has received little attention even though prosody is one of the key aspects of linguistic structure for speech perception [6]. Thus, in the present study we aim to explore the extent to which phoneme identification is affected by phonological, morphological, and prosodic factors. To achieve this, we conducted a phoneme monitoring task [7] as it is known to reflect automatic processing of a targeted phoneme in continuous speech.

Two language groups were selected to participate in the study, Australian-English monolingual adults (AusE) as a control group and L2-English-learning Mandarin-Chinese adults (ManC) as a test group. The reason for this choice lies in the fact that English and Mandarin are typologically distant languages. The dissimilarities between them allow us to examine how ManC listeners apply native language properties to L2 acquisition (i.e., *L1 transfer*). This should be apparent at various levels of language structure that are considered to be the fundamental building blocks of any language: phonology, morphology, and prosody.

Even though the Chinese family of languages, in general, does not possess overt morphology, it should be noted that certain morphological features do correspond in English and Mandarin. For instance, in both languages perfective aspect and passive voice are grammaticalised by morphemes attached to word stems rather than lexically by separate words [5].

Therefore, the phoneme /n/ was chosen for this study as a target that participants had to detect while listening to the sentences. This phoneme always appears in coda position of the syllable [ən], which is phonologically and phonotactically allowed in Mandarin [8], and which acts as part of several different morphological structures in English:

1. part of the word stem in monomorphemic nouns and non-nouns (e.g., kitten, often);
2. a grammatical inflection in past participles in either passive voice or perfective aspect, or as an adjectival attribute (e.g., spoken, fallen);
3. a derivational inflection in polymorphemic words (e.g., wooden, soften).

The AusE control group was predicted to perform at ceiling across target word types as they are familiar with morphological structures used in their native language, so words of various morphological complexity should not affect accuracy of their automatic processing of the target phoneme /n/. However, using all of these English target word types was crucial as it allowed us to compare L2-listeners’ perception of /n/ in monomorphemic versus L2-only polymorphemic words and grammatical versus derivational inflections.

If L2-listeners predominantly rely on their L1 morphological knowledge, we hypothesise that ManC listeners’ processing of /n/ should be the slowest when it is a part of derivational inflections. This prediction is supported by the fact that L2 learners are generally reported to show poor knowledge of L2 derivational morphology regardless of their L1 background and L2 proficiency [9]. Following Failed

Functional Features [4] and Morphological Congruency [5] hypotheses, however, we expect no differences between processing of /n/ in grammatical inflections or word stems, as passive voice and perfective aspect are both grammaticalised in Mandarin and should not pose any additional difficulties as compared to uninflected monomorphemic words.

Alternatively, no difference in /n/ detection for those word types will be observed if L2-listeners rely on phonological information available in their native language. Despite [ən] being used in L2 words of varying morphological complexity, it is phonologically and phonotactically permissible in Mandarin. Therefore, if performance of L2-listeners in phoneme detection is driven simply by familiarity with this phonological feature, no additional processing will be required, even for novel derivational morphology.

To address the question of whether prosodic features of an utterance can affect automatic phoneme detection, the target phoneme appeared in the context of a five-syllable intonational phrase either finally (coda of the fifth syllable) or medially (coda of the third syllable). Even though the influence of sentential prosody has not been touched upon in second language acquisition (SLA) literature, similar manipulations with utterance position have been done in L1 acquisition research [10]. Those researchers demonstrated that children are able to produce 3rd person singular –s in verbs with higher accuracy in final position than in medial position and also are better able to perceive presence/absence of this inflectional morpheme utterance-finally rather than utterance-medially. This suggests that children's perception and production of morphophonological features in English as a native language are significantly affected by sentential prosody. In this study, we test whether the perceptual finding generalises to adult L2 learners.

2. Method

2.1. Subjects

Data from 24 AusE monolingual participants recruited from the student body at the University of Western Sydney (14 females; M_{age} = 21 years, Range = 18-29 years,) and 24 ManC participants recruited from Sydney, Australia (14 females; M_{age} = 23 years, Range = 18-30 years), with no language or hearing impairments, were included in the final analysis.

To ensure the strength of any L1-transfer effect in the ManC group we only recruited participants who had been in Australia for less than one year, with prior English language exposure limited to classroom experience in China. English proficiency level varied from pre-intermediate to intermediate, as determined by IELTS, TOEFL, or similar language proficiency tests scores.

2.2. Stimulus materials

256 five-syllable stimulus sentences were produced by a 26-year-old female monolingual native speaker of AusE as intonational phrases in a neutral voice. The recording was made in a soundproof room using a Shure SM10A-CN headset microphone. The sentences were recorded using a MOTU Ultralite-mk3 audio interface and Cool Edit Pro 2.1 software. Stimulus sentences fell into three categories: experimental, control and filler sentences, as shown in Table 1.

Frequency of the target words was checked using the CELEX database [11] as a control measure to ensure that they do not pose an additional lexical and semantic challenge for the participants. Most of the target words revealed to be high-to-medium frequency with the exception of 13 low-frequency polymorphemic words and one low-frequency word in other categories. Participants indicated their familiarity with these words in a lexical checklist test. Moreover, base frequencies of polymorphemic words (e.g., frequency of the word 'straight' instead of 'straighten') were high which makes them comparable to words in other groups. Two control-word types were selected for the stimulus set according to the most common types of errors made among L1-Chinese learners of English in speech production [12] to test whether ManC listeners are sensitive to those errors in perception.

2.3. Design and procedure

All participants completed a phoneme monitoring task in which stimulus sentences were presented one after another with 1-second inter-trial intervals (ITI), in random order. Participants were instructed to listen to the sentences and to press a space bar as quickly as they could every time they heard /n/. Participants were tested individually in a quiet room using an Acer TravelMate P653 laptop, Sennheiser HD 650 headphones, an Edirol UA-25EX sound card, and DMDX 4.3.0.0 software [13].

Accuracy and reaction time (RT) for correct responses were recorded. False alarm rates for control targets with altered/omitted inflections were also analysed, as increased false alarm rates in those conditions may suggest phonological, morphological or prosodic influences on both native and L2 speech perception. For instance, extrapolating

Table 1. *Stimulus sentence types.*

Stimulus Sentences	Number	Subgroups	Example
Experimental	16 (target words) x 2 (utterance positions)	/n/ in a word stem in monomorphemic nouns, medially or finally	This *kitten* is white. – We bought a *kitten*.
	8 x 2	/n/ in a word stem in monomorphemic non-nouns, medially or finally	She *often* wears shorts. – She wears shorts *often*.
	16 x 2	/n/ in grammatical inflection in past participle, medially or finally	The *fallen* star gleamed. – The star has *fallen*.
	16 x 2	/n/ in derivational inflection in polymorphemic words, medially or finally	I *straighten* my hair. – That's how to *straighten*.
Control	16 x 2	past participles with mispronounced inflection –ed, medially or finally	The *falled* star gleamed. – The star has *falled*. [ˈfɔːləd]
	16 x 2	past participles with omitted inflection, medially or finally	The small *fall* star gleamed. – The star has just *fall*.
Filler	16	No /n/	It is her birthday.
	16	/n/ in closed-class words, any position	Birds sat *on* the roof.
	48	/n/ in open-class words, in consonant cluster or singleton, any position	*Snowflakes* look pretty. It's a *funny* book.

from PAM-L2 [2], which explores the probability of successful discrimination and learning of L2 contrasts based on their assimilation to L1 categories, we predict that increased false alarm rate by native ManC listeners in the mispronunciation with –ed condition may be attributed to the assimilation of the final /d/, which is not allowed in this position in their L1, to their permissible word-final /n/.

For statistical analysis, RTs were calculated from the onset of the consonant preceding [ən]. This cut-off point was chosen because in English the vowel preceding a nasal consonant becomes nasalized and thus, provides important acoustic information about the upcoming nasal. At the same time, nasalization in the vowel has shown to be extremely variable, which is why measuring RT to /n/ from a preceding consonant provides a more reliable landmark for the analysis. Moreover, this approach did not result in high variance in RT.

3. Results

3.1. Accuracy

A 2 (*language group*: AusE, ManC) x 4 (*word type*: monomorphemic nouns, monomorphemic non-nouns, past participles, polymorphemic words) x 2 (*utterance position*: medial, final) mixed ANOVA was conducted on accuracy (% correct), with language group as a between-subject factor, word type and utterance position as within-subject factors, and the following planned contrasts: poly- versus monomorphemic word types (past participles with grammatical inflections and polymorphemic words with derivational inflections versus nouns and non-nouns), grammatical versus derivational inflections, and nouns versus non-nouns. Separate ANOVAs were conducted by subjects (F_1) and items (F_2).

Analyses revealed a main effect of language group $F_1(1, 46) = 24.72$, $p < .001$, $F_2(1, 104) = 281.31$, $p < .001$. AusE listeners responded more accurately (M = 93%) than ManC listeners (M = 72%). The main effect of utterance position was significant, $F_1(1, 46) = 18$, $p < .001$, $F_2(1, 104) = 14.90$, $p < .001$, with overall more accurate responses to targets in final (M = 85%) than in medial position (M = 80%). A two-way interaction between language group and utterance position $F_1(1, 46) = 8.85$, $p = .005$, $F_2(1, 104) = 8.55$, $p = .004$ showed that the position effect was more pronounced for ManC listeners (M_{medial} = 68%, M_{final} = 76%) than AusE listeners (M_{medial} = 92% , M_{final} = 94%). A two-way interaction between utterance position and inflection type was observed $F_1(1, 46) = 5.79$, $p = .014$, $F_2(1, 104) = 4.98$, $p = .028$, with more pronounced differences in accuracy between medial and final position for targets with grammatical inflections (M_{medial} = 78%, M_{final} = 90%) than for those with derivational inflections (M_{medial} = 82%, M_{final} = 86%).

Both participant groups responded more accurately to targets in non-nouns than nouns across positions, but this main effect only reached significance by subjects, $F_1(1, 46) = 7.85$, $p = .007$, but not items, suggesting participants may have found a subset of nouns to be particularly difficult.

3.2. RT

We employed the same ANOVA design to analyse correct-response RT data as we used for accuracy. Main effects and interactions found in the RT analysis mirrored those observed for accuracy. Analyses verified main effects of language group $F_1(1, 46) = 31.37$, $p < .001$, $F_2(1, 104) = 445$, $p < .001$, and of utterance position $F_1(1, 46) = 145.57$, $p < .001$, $F_2(1, 104) = 353.70$, $p < .001$. Native listeners responded more rapidly (M = 810 ms) than L2-listeners (M = 1212 ms), and both groups responded more quickly to targets in final (M = 836 ms) than in medial position (M = 1186 ms). A two-way interaction between language group and utterance position $F_1(1, 46) = 7.71$, $p = .008$, $F_2(1, 104) = 17$, $p < .001$, showed that medial position hindered L2-listeners more than native listeners, as can be seen in Figure 1.

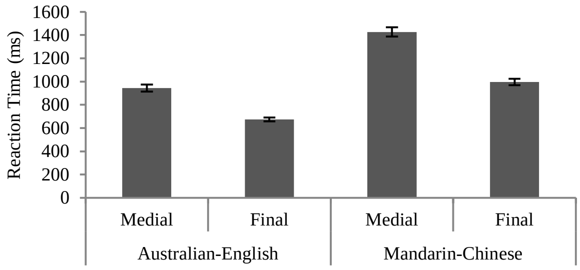

Figure 1. *Language group difference in RT to targets in medial vs. final positions, collapsed across word type.*

Overall, participants responded more rapidly to polymorphemic than to monomorphemic words $F_1(1, 46) = 8.32$, $p = .006$, $F_2(1, 104) = 9.54$, $p = .003$. This main effect did not interact with language group, suggesting that the pattern was similar for both groups. Although the contrast comparing RT to targets in grammatical versus derivational inflections was not significant, a two-way interaction between that contrast and utterance position was observed $F_1(1, 46) = 22.89$, $p < .001$, $F_2(1, 104) = 8.90$, $p = .004$. As shown in Figure 2, the difference in RT for medial versus final position was greater for targets with derivational inflections than for those with grammatical inflections.

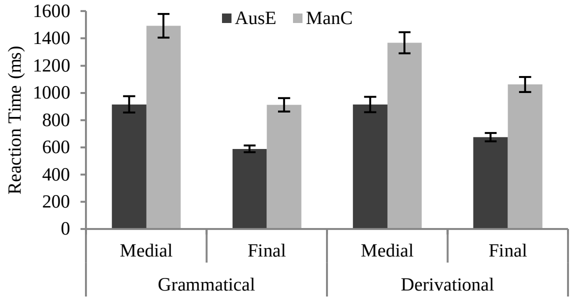

Figure 2. *Difference in RT to grammatical vs. derivational inflections embedded medially or finally.*

As can be seen in Figure 2, the by-subjects analysis also uncovered a three-way interaction among inflection type, utterance position and language group $F_1(1, 46) = 4.80$, $p = .034$. The relatively larger differences in RT between medial and final position for targets with grammatical inflections than those with derivational inflections were even more prominent for ManC than for AusE listeners. However, this interaction was not found in the by-items analysis, so it appears premature to draw firm conclusions about this pattern yet.

3.3. False alarm rate

A 2 (*language group:* AusE, ManC) x 2 (*error type:* past participles with mispronounced inflection, past participles with omitted inflection) x 2 (*utterance position:* medial, final) mixed ANOVA was conducted with language group as a between-subject factor, error type and utterance position as within-subject factors, and with false alarm rate as the dependent variable. A main effect of error type (mispronunciation vs. omission) was found, $F_1(1, 46) = 36.10$, $p < .001$,

$F_2(1, 60) = 39.63$, $p < .001$. Both language groups were more prone to errors when they heard mispronounced (/d/) than omitted inflections. The main effect of utterance position reached significance only by subjects $F_1(1, 46) = 5.1$, $p = .029$, and language group was significant only by items $F_2(1, 60) = 27.96$, $p < .001$. No significant interactions were observed.

4. Discussion

Results obtained from this study demonstrate how phonological, morphological, and prosodic features affect phoneme detection in continuous speech in both L1 and L2 speech perception. Somewhat counterintuitively, both participant groups responded faster and more accurately to polymorphemic than to monomorphemic targets, however, no differences were observed depending on inflection type. L1-listeners are familiar with native morphological structures of various complexities and might be predicting the appearance of grammatical or derivational inflection while they are still listening to a word stem. As for L2 speech perception, this finding supports the Morphological Congruency hypothesis [3] which posits that grammatical inflections represented in a native language do not pose significant problems for L2 speech processing, which indeed is what we observed. The finding also provides evidence of L2-listeners benefiting from phonological knowledge of their native language rather than being hindered by the absence of similar derivational patterns in it. Since final [ən], in which the target phoneme occurred, is phonologically and phonotactically congruent in English and Mandarin, but the derivational inflection –en exists only in English, these results allow us to infer that L2-listeners can make use of L1 phonological information to aid them in phoneme detection. If L2-listeners were unable to do so, they would have exhibited lower accuracy and longer RT to /n/ in polymorphemic words, particularly in derivational inflections.

It is also interesting to note that the shortest RT across language groups was observed for past participles embedded utterance-finally. That finding might be due to the context of the stimulus sentences, which makes the past participle in final position more predictable than other word types (e.g., The car was *driven* vs. We bought a *kitten*).

Analysis of false alarm rate demonstrated that both groups were more likely to respond to mispronounced than omitted inflections as the correct target /n/. This can be explained by the fact that the words' syllabic structure was preserved in the mispronunciation but not the omission condition. Therefore, listeners' expectations of morphophonologically correct inflections overpowered their perception of an ungrammatical stimulus. According to PAM-L2, these findings in L2-learners may be due to assimilation of the mispronounced inflections that contained /d/, which is impermissible in coda position in Mandarin, to /n/, one of the few coda consonants that is permissible in that language.

Following from L1 acquisition research that found a significant effect of utterance position on perception and production of morphophonological features in children [10], we can conclude that the same is true for both L1 and L2 adult listeners. Even though native listeners' accuracy remained unaffected by utterance position, analysis of RT, a more sensitive measure of speech processing, shows that sentential prosody still plays a significant role in native speech perception. Both accuracy and RT of L2-listeners were influenced by utterance position more than for native listeners, as their phoneme detection was noticeably hindered when the targets appeared medially. Final position may be more beneficial for speech perception due to: 1) utterance-final lengthening of the word allowing more time for morphophonological processing; 2) utterance-final pitch accent emphasising morphophonological properties of an utterance-final word, or; 3) a backward masking effect from information following the word in medial position causing additional acoustic and cognitive load. Therefore, a future direction of this study will be to disentangle those possible explanations by providing or eliminating backward masking effects, as the first step. To achieve that, phoneme monitoring will be conducted with the same stimulus set, but presented with 0 ms between utterances to place additional auditory stimulation directly following each final word and thereby increase acoustic masking and cognitive load in final position.

Findings from this study will be further compared with results from a phoneme monitoring task examining morphologically *incongruent* aspects of English and Mandarin, specifically in pluralisation. Even though the notion of plurality exists in Mandarin it is not grammaticalised. Furthermore, /z/, one of allophones of the English plural marker –s, is not present in the Mandarin phonological inventory, and /s/, the other allophone, is phonotactically constrained to the syllable initial position. Comparison of the two tasks will allow us to investigate the extent to which a novel L2 morphophonological feature affects L2 perception.

5. References

[1] 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. Baltimore: York Press, 1995, pp. 171–204.

[2] Best, C.T. and M.D. Tyler, Nonnative and second-language speech perception: Commonalities and complementarities, in Second language speech learning: The role of language experience in speech perception and production, M.J. Munro and O.-S. Bohn, Editors. 2007, John Benjamins: Amsterdam. p. 13-34.

[3] D. Lardiere, "Second Language Knowledge of [±Past] vs. [± Finite]," in *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002)*, 2003, pp. 176–189.

[4] R. Hawkins and S. Liszka, "Locating the source of defective past tense marking in advanced L2 English speakers," in *The Interface between Syntax and Lexicon in Second Language Acquisition*, R. van Hout, H. Aafke, F. Kuiken, and R. Towell, Eds. Amsterdam: Benjamins, 2003, pp. 21–44.

[5] N. Jiang, E. Novokshanova, K. Masuda, and X. Wang, "Morphological congruency and the acquisition of L2 morphemes," *Lang. Learn.*, vol. 61, no. 3, pp. 940–967, 2011.

[6] A. Cutler, *Native listening: Language experience and the recognition of spoken words*. Cambridge, Massachusetts: The MIT Press, 2012.

[7] C. M. Connine and D. Titone, "Phoneme monitoring," *Lang. Cogn. Process.*, vol. 11, no. 6, pp. 635–646, 1996.

[8] S. Duanmu, *The phonology of Standard Chinese*, 2nd ed. Oxford: OUP Oxford, 2007.

[9] B. E. Friedline, "Challenges in the second language acquisition of derivational morphology: From theory to practice," 2011.

[10] M. Sundara, K. Demuth, and P. Kuhl, "Sentence-position effects on children's perception and production of English third person singular -s," *J. Speech, Lang. Hear. Res.*, vol. 54, no. February, pp. 55–72, 2011.

[11] H. Baayen, R. Piepenbrock, and H. Van Rijn, "The CELEX lexical database (CD-ROM)." Linguistic Data Consortium, University of Pennsylvania, Philadelphia, 1993.

[12] G. Jia and A. Fuse, "Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences," *J. Speech, Lang. Hear. Res.*, vol. 50, no. 5, pp. 1280–99, 2007.

[13] K. I. Forster and J. C. Forster, "DMDX: A Windows display program with millisecond accuracy," *Behavior Research Methods, Instruments & Computers*, vol. 35, pp. 116–124, 2003.