L2 phonological category formation and discrimination in learners varying in L2 experience

Mona M. Faris[1]*, Catherine T. Best*[2]*, & Michael D. Tyler*[3]

[1]The MARCS Institute, Western Sydney University, Australia
[2]School of Humanities and Communication Arts, Western Sydney University, Australia
[3]School of Social Sciences and Psychology, Western Sydney University, Australia
m.faris@westernsydney.edu.au, c.best@westernsydney.edu.au, m.tyler@westernsydney.edu.au

Abstract

Non-native phones that are perceived as speech-like, but do not closely resemble any single first-language (L1) category, are assimilated as uncategorised. The Perceptual Assimilation Model for Second-Language (L2) Speech Learning [1] predicts that such phones are likely to be acquired as L2 categories, which should result in improvements in discrimination. This study investigated the acquisition of uncategorised L2 phones and discrimination performance in L1 Egyptian-Arabic learners varying in L2 Australian-English experience. While no firm conclusions can yet be drawn about L2 category formation, age of acquisition and L2 vocabulary size predicted discrimination accuracy, but this was dependent upon assimilation type.

Index Terms: vowel perception, L2 immersion, individual variability

1. Introduction

Research on L2 speech perception has demonstrated that, unlike young children, adult learners experience difficulty in discriminating and acquiring certain L2 phones [2]. It is well known, for example, that the discrimination of L2 phones varies as a function of both L1 attunement and the specific contrasts tested. For instance, German and Spanish learners of English differ on their discrimination of the English vowels /ɛ/ and /æ/ [3]. The Spanish listeners were able to successfully discriminate between these two English vowels, which they perceived as two contrasting L1 vowel categories (Spanish /e/ and /a/, respectively). The German listeners, on the other hand, discriminated this English contrast poorly, which they may have perceived as instances of a single L1 vowel category (German /ɛ/ or /ɛː/). Similarly, Japanese learners experience varying degrees of perceptual difficulty in discriminating between certain Australian-English vowel contrasts [4]. Good discrimination was observed for the L2 contrast /iː/-/ɪ/, which they perceived as two separate L1 vowel categories, namely, /iː/ and /i/, respectively. The L2 contrast /iː/-/ɪə/ was discriminated poorly, however, since they perceived both phones within the contrast as the single L1 phoneme /iː/.

Several models on L2 speech perception have been developed to account for the variations in discrimination performance on L2 contrasts by L2 learners. One of the most prominent models of L2 speech acquisition is the Perceptual Assimilation Model for L2 Speech Learning [PAM-L2; 1]. PAM-L2 predicts the likelihood of establishing new L2 categories, which would result in changes in discrimination performance over L2 learning time. According to PAM-L2, L1 attunement shapes L2 speech acquisition in ways that imply a shared L1-L2 phonological system. PAM-L2 makes various predictions about L2 speech learning based on the way in which L2 phones are initially assimilated to the L1 phonological system. An L2 phone that is perceived as somewhat similar to an existing L1 phoneme will be assimilated as categorised and may vary in its goodness-of-fit to the native ideal. If, however, it is perceived as speech-like, but does not closely resemble any of the L1 categories, then it will be assimilated within the L1 phonological space as uncategorised. However, phones that are perceived as non-speech will not be assimilated within the phonological space. When considered as pairs of phones, various predictions about L2 learning are possible. Consider assimilation types where one or both phones are uncategorised, namely, Uncategorised-Categorised and Uncategorised-Uncategorised assimilations. According to PAM-L2, a new L2 category is likely to be formed for phones assimilated as uncategorised. This is predicted to result in improvements in discrimination over L2 learning time. It is the discrimination and acquisition of those contrasts that are of interest in the present study.

New L2 category formation is influenced not only by the different ways in which pairs of L2 phones assimilate to the L1 phonological system, but there are also a number of factors that have been shown to affect L2 speech acquisition. High discrimination accuracy is associated with a longer length of residence [LOR; 5], a younger age of acquisition [AOA; 6] and age of immersion in the L2 speaking environment [AOI; 7], use of the L2 is relatively more than the L1 [proportion L2 use; 2], a higher L2 vocabulary size [L2 VS; 4], and a longer period of L2 training prior to L2 immersion [EFL; 8]. Given the variability among L2 learners, the effect of these factors on L2 perception will be considered in the current study. While the PAM-L2 predictions are based on beginner learners who are immersed in the L2 speaking environment, its principles are also applicable to other learning situations.

This study is part of a larger project that aims to test the PAM-L2 predictions of new L2 phonological category acquisition and to track the changes in discrimination performance over L2 learning time in learners varying in L2 experience for contrasts assimilated as Uncategorised-Categorised and Uncategorised-Uncategorised. Here we present data from the first testing session of the longitudinal study to investigate how variations in L2 experience influence L2 category acquisition and discrimination, and to test PAM contrast assimilation predictions. The PAM-L2 predictions were examined in Egyptian-Arabic (EA) learners of Australian-English (AusE) who had been exposed to the L2 prior to immersion. Based on the perceptual assimilation results in [9], two Uncategorised-Categorised (/ʊ/-/ɔ/, /ɪ/-/e/) and seven Uncategorised-Uncategorised (/ɪə/-/iː/, /æɪ/-/ɑe/, /əʉ/-/ʉː/, /oː/-/ʉː/, /oː/-/əʉ/, /æ/-/ɐ/, /eː/-/ɜː/) AusE vowel contrasts were selected for the current study.

2. Method

2.1. Participants

Thirty-eight native adult EA speakers participated in the study (20 females, M_{age} = 41 years, age range: 17 – 73 years). They were recruited from the Greater Western Sydney community and through snowball sampling. Participants varied to some extent on each of the six factors (see Table 1). They indicated that they were native-born speakers of EA, with no hearing or language impairments, and normal or corrected-to-normal vision. None of the participants had had any extended stay in an English-speaking country prior to immersion in Australia. English instruction in Egypt was typically from non-native speakers of British or American English. They received monetary reimbursement for their participation.

Table 1. *Characteristics (means and ranges) of the learners on each of the six factors.*

Variable	Mean	Range	
		Min.	Max.
Length of residence	1.41 years	0.03 years	6.34 years
Age of acquisition	13 y/o	2 y/o	52 y/o
Age of immersion	40 y/o	16 y/o	71 y/o
Proportion of L2 use	0.37	0.06	0.61
L2 vocabulary size	9200 words	4600 words	14200 words
English as a foreign language	10.03 years	0 years	23 years

2.2. Stimulus and Apparatus

The stimuli were the same as those used in [9]. Briefly, the auditory stimuli were produced in a sound-attenuated booth at the Western Sydney University by two female speakers of AusE (34 and 44 years old) recruited from the Greater Western Sydney region. All AusE monophthongs /ɐ, e, ɪ, ɔ, ʊ, æ, ɐː, eː, iː, oː, ʉː, ɜː/, diphthongs /ɑe, æɔ, æɪ, əʉ, ɪə, oɪ/, and /ə/ [10] were produced in /ˈhVbə/ nonsense words. Selected tokens were those produced with a falling intonation and spoken with a consistent speaking rate across talkers. The tokens containing the vowels /æ, ɐ, ɪ, e, ʊ, ɔ, eː, ɜː, iː, oː, ʉː, æɪ, ɑe, ɪə, əʉ/ were selected for the current study.

The stimuli were recorded at a 44.1 kHz sampling rate using a Shure SM10A headset microphone connected to an Edirol UA-25EX external USB sound card. The recordings were high-pass filtered at 70 Hz to attenuate low-frequency noise and to correct for the DC component. Tokens were ramped such that the onset and offset of each token had a 10 ms fade-in and 20 ms fade-out, respectively. Four tokens per vowel category were selected from both speakers resulting in a total of 120 tokens (15 vowels x 4 repetitions x 2 speakers). Any audible clicks detected in the tokens were excised.

L2 vocabulary size was assessed using a bilingual version of the Nation and Beglar L2-English Vocabulary Size Test [11]. It is an assessment of decontextualised knowledge of written receptive vocabulary presented in a multiple-choice format. As the English version of the test requires grammatical knowledge and fair reading abilities, a bilingual version of the test was developed by the first author. The bilingual version of the test required participants to select one out of four translated definitions that best match the test word or phrase. They were given one of two equivalent versions of the test, each containing 100 multiple-choice questions. The readability of the test and the accuracy of the translations were checked by native EA speakers prior to the administration of the test.

2.3. Procedure

Participants first completed an AXB categorial discrimination task for each of the nine AusE vowel contrasts. Participants indicated whether the vowel in the middle token (X) belonged to the same phonemic category as the vowel in either the first (A) or last (B) token. To encourage phonological perception, all three tokens per trial were physically different, with tokens A and B produced by a different speaker than token X. The interstimulus interval was 1 s. Participants were asked to attend to the first vowel in the nonsense syllable and to select one of two keys on a computer keyboard. If a response was not collected within 2 s, the trial was repeated a random number of trials later. No feedback was provided. Participants were first familiarised with the procedure on three practice trials with feedback, and the tokens were produced by a different female AusE speaker than those from the experimental trials. For each AXB task, there were 64 trials, which were randomised for each participant. All four trial types (i.e., AAB, ABB, BAA, BBA) were presented an equal number of times per contrast. As there were three tokens per speaker, using a Latin-square design, each token was presented an equal number of times in each position (i.e., A, X, B). The order of presentation of each AXB contrast was pseudorandomised.

Participants then completed an L1 perceptual assimilation task with goodness-of-fit ratings. On a given trial, they were presented with a /ˈhVbə/ nonsense syllable over headphones and were instructed to attend to the target vowel. A grid was then presented containing all L1 core phonemic (/a, i, u, aː, iː, uː, eː, oː, aw, aj/) and allophonic ([æ, æː, ɑ, ɑː, ɛː, ɛ̆ː, e, o, ɪ, ɪː, ʊ, ʊː, ə]) vowel categories, and /ʔ/ presented in Arabic CVC or CV keywords, with the vowels highlighted in red. Using a computer mouse, participants selected an L1 keyword containing the vowel closest to the auditorily presented AusE vowel. After the token was presented again, they rated its goodness-of-fit to their chosen EA vowel using a scale from 1 (strange) to 7 (perfect). No feedback was provided. A keyword selection and rating response were required to be made within 6 s and 3.5 s, respectively, otherwise the entire trial was reinserted into the random sequence. There were 120 trials (15 vowels x 2 speakers x 4 repetitions), the intertrial interval was 500 ms, and the presentation order of the trials was randomised for each participant.

In addition to the L1 perceptual assimilation task, an L2 perceptual assimilation task was administered in order to allow for inferences to be made about new L2 phonological category formation. The procedure was similar to that of the L1 task except that participants categorised the L2 vowels to L2 AusE vowel category labels. All 18 AusE vowels were presented in CVC or CV English keywords, with the vowels highlighted in red. The order of presentation of the two perceptual assimilation tasks was counterbalanced. Stimulus presentation and response collection for the AXB task and both perceptual assimilation tasks were controlled using PsyScope X B57 on a MacBook laptop, Sennheiser HD 650 headphones, and an Edirol UA-25EX external USB sound card.

Participants were given the vocabulary size test, and a language background information questionnaire in order to collect information on the participants' AOA, AOI, LOR, EFL, and proportion of L2 usage.

3. Results

3.1.1. New L2 phonological category formation

Inferences about new L2 phonological category formation were made based on the perceptual assimilation patterns from both L1 and L2 perceptual assimilation tasks. For a given L2 phone, if it was uncategorised in the L1 but categorised in the L2, then this was taken as indirect evidence that a new L2 phonological category had been formed. Phones were deemed categorised if an L1 category label was consistently selected more than 50% of the time, otherwise it was deemed uncategorised. No systematic differences were found in whether an AusE vowel was categorised to an L1 core phonemic versus allophonic category, so the allophonic vowel categories were collapsed into the appropriate main phonemic categories [see 9]. The only two AusE vowels that were categorised to an L1 category were /æ/ and /ɐ/, which were categorised to the EA /aː/ 53% and 54% of the time, respectively, while none of the AusE vowels were categorised to an L2 vowel category label.

The individual perceptual assimilation patterns revealed a high degree of variability in terms of whether a given L2 phone was categorised or uncategorised, both in the L1 and L2 tasks. Given the high degree of interindividual variability, for each individual participant instances were identified where an L2 AusE vowel was uncategorised in the L1, but categorised in the L2. A binomial logistic regression was conducted with each of the six variables (i.e., AOA, AOI, LOR, L2 VS, proportion of L2 usage, and EFL) to determine whether any of those factors are related to the likelihood of forming a new L2 phonological category. An L2 phone was uncategorised in the L1 but categorised in the L2 in 21% of instances. The logistic regression was not statistically significant, $\chi^2(6) = 3.339$, $p > .05$, suggesting that none of the six factors reliably predicted the likelihood of new L2 phonological category acquisition for this first testing session of the longitudinal project.

3.1.2. Assessing PAM's predictions of discrimination

PAM assimilation types were determined in the same way as in [12]. When the L2 phones were considered as contrasts, there was a high degree of interindividual variability in the PAM assimilation patterns. For example, while the contrast /oː/-/əʉ/ was assimilated as Uncategorised-Uncategorised in the L1 perceptual assimilation task at the group level, individual participants assimilated it either as Uncategorised-Uncategorised or Uncategorised-Categorised. Given this high degree of variability, to analyse the discrimination results, we adopted the same approach as in [12]. Specifically, each individual's assimilation type for each of the nine contrasts was determined. The mean discrimination accuracy scores were then grouped according to individual assimilation type rather than on vowel contrast. For example, the discrimination accuracy scores for Uncategorised-Uncategorised assimilations were grouped together, regardless of the contrast in which they occurred. Individual assimilation patterns were determined per individual for both the L1 and L2 tasks.

Individual assimilation types were compared across the L1 and L2. There were cases where a contrast was assimilated as a Single-Category in the L1, but as a Two-Category contrast in the L2, suggesting that the participant was able to discern phonological differences between the pair of contrasting L2 phones, *and* that they had learned the new L2 contrast. Therefore, it may be more meaningful to consider both L1 and L2 perceptual assimilation patterns than either one alone. Taking into account both L1 and L2 assimilation types, we created a composite L1-L2 assimilation type by selecting the L1 or L2 assimilation type that was predicted to result in the more accurate discrimination across the two. For instance, for an individual participant, if a given contrast was assimilated as Uncategorised-Uncategorised in the L1, but as a Two-Category in the L2, then the L2 perceptual assimilation type was selected. Similarly, if a contrast was Uncategorised-Categorised in the L1, but Single-Category in the L2, then the L1 perceptual assimilation type was selected. Eight percent of cases were Single-Category assimilations, another 8% were Two-Category, 40% were Uncategorised-Categorised, and 42% were Uncategorised-Uncategorised. Since only 1% of cases were of Category-Goodness assimilations (comprised of two data points), they were excluded from further analyses.

A one-way between-subjects analysis of variance was conducted to determine if the discrimination accuracy scores vary as a function of assimilation type. There was a significant difference in the discrimination accuracy scores among the assimilation types, $F(3, 336) = 5.446$, $p = .001$. A Bonferroni post-hoc comparison revealed Two-Category assimilations were discriminated more accurately than Single-Category, $M_{\text{diff}} = 15.92\%$, $p = .001$, $SE = 4.27\%$. Uncategorised-Categorised assimilations were discriminated more accurately than Single-Category assimilations, $M_{\text{diff}} = 10.68\%$, $p = .009$, $SE = 3.34\%$. The results are displayed in Figure 1.

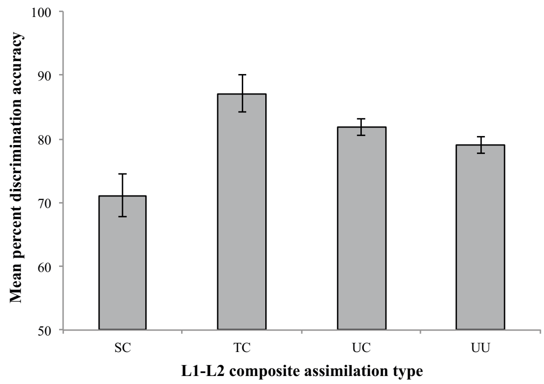

Figure 1: *Mean percent discrimination accuracy for the composite L1-L2 assimilations. Error bars represent standard error of the mean*

3.1.3. Effect of the six factors on discrimination accuracy

The relationship between the six predictors on discrimination accuracy was assessed using the composite L1-L2 assimilation types. Bivariate Pearson correlations were conducted between the mean percent discrimination accuracy score for a given assimilation type with each of the six factors. The correlations are presented in Table 2, with significant bivariate Pearson correlation coefficients presented in bold, which ranged from .202 to -.490. A younger age of acquisition was associated with more accurate discrimination for all assimilation types, except for Single-Category. Uncategorised-Categorised and Uncategorised-Uncategorised contrast assimilations each yielded a larger number of significant correlations than Single-Category and Two-Category assimilations combined.

To determine whether any of the factors predicted discrimination performance, a separate standard multiple regression was conducted for each composite L1-L2 assimilation type. A high L2 vocabulary size predicted poorer discrimination accuracy for Single-Category assimilations (i.e., a negative correlation), $F(6, 21) = 4.114$, $p = .007$, and accounted for approximately 47% of the variance ($R^2 = .540$, adjusted $R^2 = .409$). A younger age of acquisition was a significant predictor of better discrimination accuracy for Uncategorised-Categorised assimilations, $F(6, 131) = 10.90$, $p < .001$, and accounted for approximately 32% of the variance ($R^2 = .333$, adjusted $R^2 = .303$). Similarly, a younger age of acquisition significantly predicted higher discrimination accuracy scores for Uncategorised-Uncategorised assimilations, $F(6, 138) = 2.560$, $p = .022$, and accounted for approximately 8% of the variance ($R^2 = .100$, adjusted $R^2 = .061$). None of the factors significantly predicted discrimination accuracy for Two-Category assimilations.

Table 2. *Bivariate Pearson correlations between the mean percent discrimination accuracy scores for each composite L1-L2 assimilation type with each of the six factors.*

Composite L1-L2	Mean discrimination accuracy	LOR	AOI	AOA	EFL	Prop. L2 use	L2 VS
SC	71	-.122	**-.407***	.001	.221	-.077	**-.476***
TC	87	-.296	-.142	**-.490****	.164	.352	.231
UC	82	**-.376****	-.146	**-.480****	**.445****	**.202***	**.364****
UU	79	**-.205***	-.123	**-.266****	**.248****	.036	.116

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

4. Discussion

This study aimed to examine new L2 category formation and discrimination accuracy in learners varying in L2 experience. Vowel perception was shown to be highly variable [e.g., 12]. Despite this variability, PAM's predictions of discrimination were upheld such that both Two-Category and Uncategorised-Categorised assimilations were discriminated more accurately than Single-Category assimilations.

By accounting for variability within individual participants who differed on factors related to L2 experience, we have shown that, to some extent, differences in L2 discrimination accuracy may be explained by such factors. A high L2 vocabulary size predicted poor discrimination accuracy. According to PAM-L2, a steadily expanding L2 vocabulary size is beneficial for L2 learners as it forces them to attend to important phonetic details in the L2 that are not employed in the L1, and in turn, help learners distinguish between minimally contrasting L2 words. But, a rapidly expanding L2 vocabulary may be detrimental for L2 learners as it may cause them to fossilise, or settle on a suboptimal common L1-L2 phonological category, thus curtailing further L2 development. At this initial stage of testing, the learners' L2 vocabulary size was high, averaging 9200 words. The current study did not assess the rate of L2 vocabulary acquisition. However, the acquisition of English vocabulary and grammar are normally the key focuses of L2 acquisition in schools and universities in Cairo. It may be tentatively inferred that the L2 vocabulary was acquired rapidly prior to L2 immersion. As this study forms part of a larger longitudinal study, there will be an opportunity to track how changes in L2 vocabulary size affect discrimination accuracy over L2 learning time. Vocabulary size for an average native English speaker is roughly 20,000 words [13], so there remains room for vocabulary expansion.

While discrimination accuracy was affected to some extent by some of the factors, none of the factors reliably predicted discrimination accuracy for Two-Category assimilations. It is unsurprising given that it is L1 attunement that helps the listener distinguish between phones assimilated as Two-Category, which is consistent with PAM's framework.

Individual differences may also play a role in new L2 phonological category formation. The results revealed that none of the factors significantly predicted the likelihood of new L2 phonological category formation. However, as only 21% of cases were of an L2 phone that was uncategorised in the L1 but categorised in the L2, there may not be sufficient statistical power to detect those influences. Consequently, no firm conclusions may be made at this stage of the longitudinal study. The effect of the six factors on category formation will be examined longitudinally as a function of changes in L2 immersion experience. This should in turn be reflected in changes in discrimination performance over L2 learning time.

The next stage of this project will be to examine the developmental changes over a 12-month period of L2 immersion by tracking changes in discrimination performance as a function of perceptual assimilation, and how discrimination performance is affected by the six factors.

5. References

[1] C.T. Best 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, Eds. 2007, John Benjamins: Amsterdam. p. 13-34.

[2] J.E. Flege and I.R.A. MacKay, "Perceiving vowels in a second language," *Stud. Second Lang. Acquis.*, vol. 26, pp. 1-34, 2004.

[3] J.E. Flege, O.S. Bohn and S. Jang, "Effects of experience on non-native speakers' production and perception of English vowels," *J. Phonetics*, vol. 25, no. 4, pp. 437-470, 1997.

[4] R. Bundgaard-Nielsen, C.T. Best and M.D. Tyler, "Vocabulary size is associated with second-language vowel perception performance in adult learners," *Stud. Second Lang. Acquis.*, vol. 33, pp. 433-461, 2011.

[5] J.E. Flege and S. Liu, "The effect of experience on adults' acquisition of a second language," *Stud. Second Lang. Acquis.*, vol. 23, no. 4, pp. 527-552, 2001.

[6] G. Jia, W. Strange, Y. Wu, J. Collado and Q. Guan, "Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure," *J. Acoust. Soc. Am.*, vol. 119, no. 2, pp. 1118-1130, 2006.

[7] W. Baker, P. Trofimovich, M. Mack and J. E. Flege, "The effect of perceived phonetic similarity on non-native sound learning by children and adults," in BUCLD, Boston, MA, vol. 26, no. 1, pp. 36-47, 2002.

[8] J. Cebrian, "Experience and the use of non-native duration in L2 vowel categorization," *J. Phonetics*, vol. 34, no. 3, pp. 372-387, 2006.

[9] M.M. Faris, C.T. Best and M.D. Tyler, "An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized," *JASA*, vol. 139, no. 1, pp. EL1-EL5, 2016.

[10] F. Cox and S. Palethorpe, "Australian English," *J. Int. Phonetic Assoc.*, vol. 37, pp. 341-350, 2007.

[11] P. Nation and D. A. Beglar, "A vocabulary size test," *Lang. Teacher*, vol. 31, pp. 9-13, 2007.

[12] M.D. Tyler, C.T. Best, A. Faber and A.G. Levitt, "Perceptual assimilation and discrimination of non-native vowel contrasts," *Phonetica*, vol. 71, pp. 4-21, 2014.

[13] P. Nation and R. Waring, Vocabulary size, text coverage and word lists, in Vocabulary: Description, Acquisition and Pedagogy, N. Schmitt and M. McCarthy, Eds. 1997, Cambridge University Press: Cambridge, pp. 6-19.