JANUARY 11 2016

An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized

Mona M. Faris; Catherine T. Best; Michael D. Tyler

J. Acoust. Soc. Am. 139, EL1–EL5 (2016)
https://doi.org/10.1121/1.4939608

An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized

Mona M. Faris[a)]
The MARCS Institute, Bankstown Campus, Western Sydney University, Locked Bag 1797, Penrith, New South Wales 2751, Australia
m.faris@westernsydney.edu.au

Catherine T. Best
School of Humanities and Communication Arts, Bankstown Campus, Western Sydney University, Locked Bag 1797, Penrith, New South Wales 2751, Australia
c.best@westernsydney.edu.au

Michael D. Tyler
School of Social Sciences and Psychology, Bankstown Campus, Western Sydney University, Locked Bag 1797, Penrith, New South Wales 2751, Australia
m.tyler@westernsydney.edu.au

Abstract: This study examined three ways that perception of non-native phones may be uncategorized relative to native (*L*1) categories: focalized (predominantly similar to a single *L*1 category), clustered (similar to > 2 *L*1 categories), and dispersed (not similar to any *L*1 categories). In an online study, Egyptian Arabic speakers residing in Egypt categorized and rated all Australian English vowels. Evidence was found to support focalized, clustered, and dispersed uncategorized assimilations. Second-language (*L*2) category formation for uncategorized assimilations is predicted to depend upon the degree of perceptual overlap between the sets of *L*1 categories listeners use in assimilating each phone within an *L*2 contrast.

[DOS]
Date Received: June 10, 2015 **Date Accepted:** December 24, 2015

1. Introduction

There is ample evidence to suggest that native language (*L*1) attunement influences non-native speech perception in adults (e.g., Best *et al.*, 2001; Tyler *et al.*, 2014). One way to shed light on how experience with the *L*1 shapes speech perception is by testing listeners' perceptual assimilation of non-native phones (see Escudero and Williams, 2011; Gilichinskaya and Strange, 2010). Modeling the effects of *L*1 attunement on cross-language speech perception also has the applied benefit of accounting for the initial state of a second language (*L*2) learner and allowing predictions to be made about subsequent *L*2 development. According to the Perceptual Assimilation Model (PAM: Best, 1995), when a non-native phone is perceived as speech-like, but fails to resemble any particular *L*1 category, it is deemed *uncategorized*, and is likely to be acquired as a new category by *L*2 learners (PAM-*L*2: Best and Tyler, 2007). We postulate that uncategorized assimilations can be further differentiated, and present evidence here for three different ways in which non-native phones may be perceptually assimilated as uncategorized.

Non-native speech perception involves the detection of fine-grained phonetic information and identification of the abstract phonological-categorical functions of those phonetic details. According to PAM, listeners perceive non-native speech relative to the phonological and phonetic similarities and differences of the phonemes of their *L*1. A non-native phone may be *categorized* to an *L*1 category, and conceptually, with respect to its specific phonetic details, it may be perceived as an identical, an acceptable, or a deviant exemplar of that *L*1 category. Alternatively, a non-native phone may be perceived as speech-like, but without closely resembling any particular native category (i.e., *uncategorized*), and so falls in an untuned region of the listener's phonological system. The final possibility is that a non-native phone is heard as non-speech, thus falling

[a)]Author to whom correspondence should be addressed.

outside the listener's native phonological space (i.e., *non-assimilable*). Studies conducted within a PAM framework have predominantly examined phones that were assimilated as categorized and non-assimilable, but uncategorized assimilations remain relatively understudied. Therefore, it is important to examine how uncategorized phones map onto the *L*1 phonological system in order to provide a comprehensive assessment of non-native speech perception, and the potential for *L*2 perceptual learning.

While PAM was developed to account for the perceptual assimilation of all speech segments (i.e., consonants and vowels), previous research has focused primarily on non-native consonants. In a recent study on the perceptual assimilation of non-native vowels, there was a high degree of within- and between-subject variability (Tyler *et al.*, 2014). Unlike the perceptual assimilation of consonants, many of the vowels were uncategorized, possibly due to the less well-defined category boundaries for vowels than consonants. For this reason, vowels are ideal for examining the differing degrees to which non-native phones may be uncategorized.

In studies testing PAM predictions, uncategorized phones are operationally defined as those that are not consistently assigned to a single *L*1 category above a pre-defined threshold (e.g., 50%). However, observations of perceptual assimilation data patterns suggest three possible ways in which a phone might be uncategorized: (1) *focalized responses*: the non-native phone is perceived as primarily similar to a single *L*1 category but responses are below the categorization threshold; (2) *clustered* responses: the uncategorized non-native phone is perceived as similar to a small set of *L*1 categories; and (3) *dispersed* responses: the listeners select a range of different *L*1 categories, which may reflect random responding when none of the available *L*1 categories are very similar to the stimulus phone.

This study evaluated the theoretical notion of uncategorized phones forming focalized, clustered, and dispersed responses. Listeners with a small *L*1 vowel inventory size are more likely to assimilate non-native vowels as uncategorized than those with a larger vowel inventory (Escudero and Williams, 2011). Therefore, Egyptian Arabic (EA) is ideal because it has 10 diphthongs and monophthongs (Lehn and Slager, 1959) as compared to the 19 of Australian English (AusE) (Cox and Palethorpe, 2007). To increase the possibility of observing uncategorized vowels, participants categorized and rated all AusE vowels in relation to their full EA vowel inventory. To ensure that listeners have had minimal exposure to the stimulus language, an innovative online method was used to collect cross-language speech perception data from participants residing in Egypt.

2. Method

2.1 Participants

Twelve native adult speakers of EA (10 females, $M_{\mathrm{age}} = 38.67$ yrs, age range: 20–67 yrs) were recruited from universities in Egypt and through snowball sampling whereby each participant was asked to forward the link to the online experiment to some of their personal contacts. All were native-born speakers of EA, with no hearing, vision, or language impairments. They were not remunerated.

2.2 Stimuli and apparatus

The online experiment was programmed using Python and hosted using Google App Engine. The experiment was presented to participants using Firefox, Google Chrome, or Safari browsers. Participants were shown a grid containing /CVC/ or /CV/ keywords in Arabic orthography representing the entire Arabic short /a, i, u/ and long /aː, iː, uː, eː, oː/ vowels, and the diphthongs /aw, aj/ (transcribed as vowel-glide sequences: see Thelwall and Sa'Adeddin, 1990). To allow the EA listeners to be able to make choices reflecting fine-grained phonetic distinctions, we also included Arabic keywords containing allophonic variants of vowels [æ, æː, ɑ, ɑː, ɛː, ɛ̽ː, e, o, ɪ, ɪː, ʊ, ʊː, ə], and /ʔ/ (Lehn and Slager, 1959; Selim and Anbar, 1987). The vowels in each of the 24 keywords were highlighted in red, and a 7-point rating scale was also displayed.

Two native AusE female speakers (34 and 44 yrs old) from the South-Western Sydney region were recorded in a sound-attenuated booth. The 12 AusE monophthongs /ɐ, e, ɪ, ɔ, ʊ, æ, ɐː, eː, iː, oː, ʉː, ɜː/, six diphthongs /ɑe, æɔ, æɪ, əʉ, oɪ, ɪə/, and /ə/ (Cox and Palethorpe, 2007) were produced in /ˈhVbə/ non-words. The stimuli were recorded using a Lenovo T520 laptop, running on Windows XP, at a 44.1 kHz sampling rate using a Shure SM10A headset microphone connected to an Edirol UA-25EX external USB sound card, with all equipment used in the present study provided by the MARCS Institute.

The recordings were high-pass filtered at 70 Hz to attenuate any low-frequency noise and correct for the direct current component. The onset and offset of each token were ramped by 10 and 20 ms, respectively. Any audible clicks were excised from the tokens. Four tokens of each nonsense word were selected from each speaker based on similar intonation and speaking rate, resulting in 152 tokens (19 vowels × 4 repetitions × 2 speakers). All tokens from a given speaker were amplified by a constant amount to normalize the vowel intensity across speakers while maintaining natural intensity differences across vowel categories.

2.3 Procedure

Written Arabic text instructed participants to sit in a quiet area and use headphones to listen to the auditory stimuli. Completion of the online experiment took 30–40 min. In an initial familiarization phase, participants were guided through three step-by-step examples using the Arabic vowels /i, aː, aj/ in /ˈhVbə/ non-words, spoken by a native female Arabic speaker. Participants were instructed to attend to the first vowel in the non-words. In each example, participants clicked to play an auditory token. They were then asked to refer to the categorization grid and find an Arabic keyword containing the most similar vowel to the one they heard. Participants were instructed that the same auditory token would then be presented a second time after which they needed to indicate how well the target vowel in the non-word matched the vowel in the chosen Arabic keyword on a scale from 1 (very strange) to 7 (perfect).

The trial structure for the category assimilation task with goodness-of-fit ratings was identical to the familiarization phase, except that no feedback was provided and AusE tokens were used. The AusE vowels were pseudorandomized such that there were no more than three consecutive tokens from the same speaker, and that tokens from the same vowel category were not presented consecutively, regardless of the speaker. All participants completed the same pseudorandom order. There were 152 trials, with the last 10 trials additionally presented at the beginning to serve as warm-up trials; these were not included in the final analyses.

3. Results

There were no systematic differences in the way in which the AusE vowels were assimilated to the EA core phonemic versus allophonic categories, so the allophonic vowel categories were collapsed into the appropriate main phonemic category: [æ, ɑ, ɛː, ə] were collapsed into /a/, [ɪ] was collapsed into /i/, [ʊ] was collapsed into /u/, [e] was collapsed into /eː/, [o] was collapsed into /oː/, [æː, ɑː, ɛ̽ː] were collapsed into /aː/, [ɪː] was collapsed into /iː/, [ʊː] was collapsed into /uː/, and /ʔ/ was excluded from the analysis as it was rarely chosen. As in Bundgaard-Nielsen *et al.* (2011), a non-native vowel was considered *categorized* if it was consistently assimilated to a particular *L*1 vowel category label more than 50% of the time, otherwise, it was deemed *uncategorized.* As can be seen in Table 1, only 3 of the 19 AusE vowels were consistently assimilated to an EA vowel category: AusE /ɐː/, /e/, and /ɔ/, were assimilated to EA /aː/ (87%), /i/ (60%), and /u/ (65%), respectively.

To differentiate among focalized, clustered, and dispersed uncategorized responses, *t*-tests were first conducted comparing the mean percent categorization of an AusE vowel with each EA response option against a chance score of 10%, a value that takes into account the 10 possible EA phonological categories (see So and Best, 2014). For the average percent categorization of a given AusE vowel to an EA response option, a significant *p*-value ($p < 0.05$) indicates that a specific EA label was selected significantly more often than chance. Focalized responses were defined as those where participants selected only one EA response label above chance for a given AusE vowel. Clustered responses were identified as those where more than one EA response label was selected above chance, and dispersed responses were those where no EA response label was selected more often than chance.

As can be seen in Table 1, the AusE vowels /ʊ, iː, ʉː, ɪə, ɑe, æɔ/ were assimilated as focalized, as EA /u, iː, uː, iː, eː, aː/, respectively. AusE /ɐ, æ, ɪ, oː, æɪ, əʉ, ə/ were clustered assimilations, with 2–3 EA response labels selected above chance per AusE vowel. Interestingly, responses for some clustered assimilations were split between short and long EA vowels (e.g., /ɪ/ as EA /i/ and /iː/). The AusE vowels /eː, ɜː, oɪ/ were dispersed assimilations, although EA label selection appears to be much more variable for /ɜː, oɪ/ than for /eː/.

4. Discussion

The predicted differentiation among focalized, clustered, and dispersed uncategorized assimilation types was clearly supported. These findings contribute to a better understanding of how listeners make use of both gradient phonetic details and abstract phonological categories in speech perception. Focalized and clustered assimilations suggest

Table 1. Mean percent categorization and goodness ratings of AusE vowels by EA speakers, with EA allophonic categories collapsed across appropriate main categories.

Assimilation type	AusE vowel	EA category label a	i	u	aː	eː	iː	oː	uː	aw	aj
Categorized[b]	e	8	**60 (4.93)**[a]		3	12	12	1	1	1	2
	ɔ	2		**65 (5.06)**	1	1		16	10	2	3
	ɐː	7		2	**87 (5.41)**	2		1			
Focalized	ʊ	1		***45 (4.93)***[c]	1			22	27	3	
	iː	9	16	1	6	14	***49 (5.64)***	1		1	3
	ʉː	1	7	22	2	5	6	19	***35 (3.95)***		1
	ɪə	5	14		7	12	***36 (4.86)***			2	24
	ɑe	8			17	***36 (5.12)***	7	2	1	2	26
	æɔ	12	1	4	***43 (4.81)***	4		13	6	16	1
Clustered	ɐ	***45 (5.25)***	2	1	***40 (4.88)***	6	3			1	1
	æ	***30 (5.26)***	11	1	***46 (5.01)***	4	5		2	1	
	ɪ	5	***44 (5.24)***		5	8	***36 (4.66)***				1
	oː	1	1	***29 (4.82)***	1		1	***39 (5.49)***	***21 (4.83)***	7	
	æɪ	7	4		***22 (4.62)***	***25 (5.40)***	19		1	1	20
	əʉ	3	1	***34 (4.93)***	1	2	4	21	***27 (3.91)***	6	
	ə	18	***29 (5.29)***	***31 (4.22)***	1	3	6	2	8		1
Dispersed[d]	eː	11	23		20	17	23				6
	ɜː	7	12	23	7	11	16	6	9	6	2
	oɪ	8	3	13	5	9	14	7	14	11	16

[a]Numbers represent the percentage of each AusE vowel assimilated to an EA label, averaged across participants. The goodness-of-fit ratings are on a scale of 1 (sounds poor) to 7 (sounds perfect), also averaged across participants.

[b]Numbers in bold indicate the mean percent categorization scores that have reached the 50% assimilation criterion, with the averaged goodness rating presented within parentheses.

[c]For focalized and clustered responses, the mean percent categorization scores that are presented in bold italics represent the EA response label/s that was selected significantly more often than would be predicted by chance ($p < 0.05$).

[d]For dispersed responses, none of the response choices were significantly greater than chance.

that, to a certain degree, listeners detect some phonetic information in non-native phones that is phonologically meaningful in the *L*1. However, with dispersed responses, listeners are sensitive only to the phonetic-gradient level of detail.

In the same way that PAM differentiates between the various ways that non-native segments may be categorized to an *L*1 phoneme, we have shown that uncategorized phones also differ in the way they map onto the *L*1 phonological system. Although this study has focused on perceptual assimilation of individual phones, it is possible to make different discrimination predictions for the three uncategorized assimilation types. According to PAM, discrimination accuracy for pairs of uncategorized phones (i.e., *uncategorized-uncategorized* assimilation type; UU) will range from poor to moderate, depending on whether the two phones are perceived as similar to the same set of *L*1 categories and the similarity of the two phones to one another. By classifying uncategorized phones as focalized, clustered, or dispersed, more precise discrimination predictions may be made for UU contrasts. Assuming that a pair of uncategorized contrasting phones are each assimilated to a different *L*1 category or sets of *L*1 categories, phones assimilated as focalized (i.e., focalized-focalized) are predicted to be relatively easy to discriminate, followed by focalized-clustered, clustered-clustered, focalized-dispersed, clustered- dispersed, with dispersed-dispersed predicted to be the most difficult to discriminate.

Interestingly, certain pairs of AusE vowels were perceived as similar to the same set of EA vowel categories, such as the clustered AusE /ɐ/ and /æ/ vowels, where responses were both split between EA /a/ and /aː/. Recall that discrimination accuracy for contrasts assimilated as UU depends on the degree to which the contrasting phones are assimilated to the same set of *L*1 categories. Indeed, the degree of overlap between contrasting phones may affect discrimination performance (Tyler *et al.*, 2014). As no single response is above chance in dispersed assimilations, it is only in focalized and clustered assimilation types that meaningful overlap could be encountered. UU focalized/clustered contrasts should be more accurately discriminated if both are assimilated to a *different* set of L1 categories (i.e., non-overlapping) such as AusE /iː/-/ʉː/ (focalized assimilation to EA /iː/ vs /uː/), than if they are assimilated to the *same* set of *L*1 categories (i.e., completely overlapping) such as AusE /iː/-/ɪə/ (focalized assimilation of both to EA /iː/). Those that overlap with some of the same *L*1 categories (i.e., partially overlapping) such as AusE /ɐ/-/æɪ/ (clustered assimilation to EA /a, **a̲ː**/ vs to /eː, **a̲ː**/, respectively), should be discriminated more accurately than contrasts that completely overlap, but less well than non-overlapping contrasts. Non-overlapping phones may be more discriminable because listeners are sensitive to the phonetic similarity of each non-native phone to different *L*1 phonological categories.

Our perceptual assimilation results also inform PAM-*L*2 predictions. Given that listeners fail to detect clear higher-order *L*1 phonological category invariants for dispersed assimilations, a new *L*2 phonological category is likely to be formed because there will be no systematic interference from previous *L*1 attunement. This is similar to the concept of a "new" *L*2 phone in the Speech Learning Model (Flege, 1995). For focalized and clustered assimilations, however, acquisition of a new *L*2 category is likely to depend on the degree of phonetic overlap with contrasting *L*2 phones. If the *L*2 focalized or clustered phone does not overlap with any other *L*2 category, then a new *L*2 category is likely to be acquired.

Differentiating between categorized and focalized phones depends upon the predefined categorization threshold. Using a 50% assimilation criterion, AusE /iː/ was deemed uncategorized-focalized at 49%. The individual data reveals that 8 out of the 12 participants consistently categorized AusE /iː/ to EA /iː/. Due to the large individual differences, Tyler *et al.* (2014) split individuals' data by assimilation type rather than contrast type. However, it was not possible to employ this approach here due to an insufficient sample size.

Future studies are required to evaluate the hierarchy of discrimination predictions, and the effect of perceptual overlap on discrimination performance and *L*2 category formation in learners. Additionally, the development of alternate methods is required to replace the use of an arbitrary cut-off criterion, particularly for vowel perception.

Acknowledgments

We thank Lei Jing for his technical assistance in helping to develop the online study and all of those who helped with recruitment, in particular, Ahmed Moustafa and Rawhia Ali. We also thank Mark Antoniou for his helpful suggestions on the final draft.

References and links

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Baltimore, MD), pp. 171–204.

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," J. Acoust. Soc. Am. **109**, 775–794.

Best, C. T., and Tyler, M. D. (**2007**). "Non-native and second-language speech perception: Commonalities and complementarities," in *Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production*, edited by M. Munro and O.-S. Bohn (John Benjamins, Amsterdam), pp. 13–24.

Bundgaard-Nielsen, R., Best, C. T., and Tyler, M. D. (**2011**). "Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories," J. Appl. Psychol. **32**, 51–67.

Cox, F., and Palethorpe, S. (**2007**). "Australian English," J. Int. Phonetic Assoc. **37**, 341–350.

Escudero, P., and Williams, D. (**2011**). "Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners," J. Acoust. Soc. Am. **129**, EL1–EL7.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Baltimore, MD), pp. 233–276.

Gilichinskaya, Y. D., and Strange, W. (**2010**). "Perceptual assimilation of American English vowels by inexperienced Russian listeners," J. Acoust. Soc. Am. **128**, EL80–EL85.

Lehn, W., and Slager, W. R. (**1959**). "A contrastive study of Egyptian Arabic and American English: The segmental phonemes," Lang. Learn. **9**, 25–33.

Selim, H., and Anbar, T. (**1987**). "A phonetic transcription system of Arabic text," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (Dallas, TX).

So, C. K., and Best, C. T. (**2014**). "Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories," Stud. Second Lang. Acquis. **36**, 195–221.

Thelwall, R., and Sa'Adeddin, M. A. (**1990**). "Arabic," J. Int. Phonetic Assoc. **20**, 37–39.

Tyler, M. D., Best, C. T., Faber, A., and Levitt, A. G. (**2014**). "Perceptual assimilation and discrimination of non-native vowel contrasts," Phonetica **71**, 4–21.