
Perceptual Overlap in Classification of L2 Vowels: Australian English Vowels Perceived by Experienced Mandarin Listeners

Yizhou Wang

School of Languages and Linguistics
University of Melbourne, Australia
wyz2014tw@gmail.com

Brett J. Baker

School of Languages and Linguistics
University of Melbourne, Australia
bjbaker@unimelb.edu.au

Rikke L. Bundgaard-Nielsen

MARCS Institute
University of Western Sydney, Australia
rikkelou@gmail.com

Olga Maxwell

School of Languages and Linguistics
University of Melbourne, Australia
omaxwell@unimelb.edu.au

Abstract

This paper presents a study examining the role of classification overlap in second language (L2) vowel perception. Eighteen experienced L1 Mandarin Chinese (MC) listeners completed a classification task on twelve Australian English (AusE) monophthongal vowels. It was found that the AusE vowels were not equally difficult for the MC listeners, and the vowels also showed various overlapping patterns in the classification space. Correlation tests revealed that for individual L2 vowels, the classification accuracy was negatively correlated with the mean reaction time (RT), $r = -0.817$, $p = 0.001$; the level of classification overlap was negatively correlated with the mean accuracy, $r = -0.860$, $p < 0.001$, while positively correlated with the mean RT, $r = 0.760$, $p = 0.004$. The findings suggest that classification overlap can inform how experienced L2 listeners' first language (L1) phonology influences vowel perception in a non-native language.

1 Introduction

To successfully understand speech, listeners must process phonetic information and match this information to the phonemic categories of the language in question. For non-native or L2 listeners, this task can sometimes be difficult because their perceptual system is shaped by the L1 phonology. Mismatches between the L1 and L2

phonological systems can lead to perceptual difficulty of various linguistic units in the target language, e.g., segmental (Bundgaard-Nielsen et al., 2011a; Højen and Flege, 2006) and suprasegmental structures (Hallé et al., 2004; Tremblay, 2009). One prevalent theory of non-native speech perception, the Perceptual Assimilation Model (PAM, and its extension, PAM-L2) (Best, 1995; Best and Tyler, 2007) attributes the perceptual difficulty to the interference of the L1 phonological system: L2 phones tend to be assimilated into the listener's L1 phonological categories, which can at times cause a loss of phonemic contrast in perception.

PAM and PAM-L2 posit that the discriminability between a pair of non-native (L2) sounds depends on the level of perceptual overlap between the two sounds, e.g., a pair of perceptually overlapping L2 sounds are more difficult to discriminate than those that do not overlap in cross-language perception (Faris et al., 2018). For instance, L1 Japanese listeners assimilate the AusE vowels [i:] and [iə] as their native bimoraic category [i:], while mapping the AusE vowel [i] onto the Japanese monomoraic category [i]; and this perceptual overlap pattern has led them to fail to discriminate between AusE [i:] and [iə], though not between AusE [i:] and [i] (Bundgaard-Nielsen et al., 2011a, 2011b).

Since PAM and PAM-L2 provide precise predictions in terms of pairwise discriminability, previous research adopting the framework often

uses discrimination tasks (e.g., AXB tasks) to examine perceptual performance by L2 learners. Besides, an assimilation task is used in association to offer predictions of potential sound pairs that show perceptual overlap (Bundgaard-Nielsen et al., 2011a; Faris et al., 2018; Tyler et al., 2014). For a discussion on the methodology, see (Strange and Shafer, 2008). However, discrimination tasks can only examine one pair of L2 sounds at a time, and the assimilation task only offers the native phonemic inventory as the set of classification candidates, which cannot effectively represent the non-native categories that are already learned by *experienced* L2 listeners.

Here, we present a study examining the perception of Australian English (AusE) vowels by experienced L2 listeners whose native language is MC. We use a classification task which allows all target L2 categories to be offered as candidates, and the listener's response is not limited by the range of phonemic categories in their native phonology. In addition, whereas the AXB paradigm only examines one pair of L2 sounds at a time, the classification task used here requires forced phonological processing of the sound stimulus among a set of phonemic categories and thus is able to measure the perceptual similarity between multiple phonemic pairs. These results can potentially reveal scenarios of multiple category assimilation.

To estimate the level of perceptual overlap for a specific pair of L2 sounds, we calculate the *phonological overlap score* following the method reported in Faris et al. (2018). The phonological overlap score of a specific pair of L2 sounds takes into account all classification overlaps that are above chance; and researchers have demonstrated that the score can successfully quantify the perceptual similarity between two sounds (Faris et al., 2016, 2018), especially when the assimilation pattern is not effectively differentiated in the classic framework of PAM (Best, 1995).

The aim of the present study is to directly examine how the perceptual overlap level of L2 vowels links to the overall perceptual difficulty, as indicated by accuracy and latency measures. In particular, the present study aims to answer three research questions:

1. *How are AusE vowels perceived by experienced MC listeners as reflected in the classification results?*
2. *Which AusE vowels show perceptual overlaps when perceived by experienced MC listeners?*
3. *Does the classification performance, accuracy and RT, of AusE vowels correlate with the corresponding perceptual overlap score?*

We hypothesise that, under the influence of native phonological interference, L2 vowel perception can be affected by the level of perceptual overlap between the target L2 sound and other similar categories (possibly more than one). In particular, we hypothesise that the perceptual accuracy should be negatively correlated with the perceptual overlap level; and that the RT measure should be positively correlated with the overlap level. This prediction is also consistent with the Automatic Selective Perception (ASP) model (Strange, 2011; Strange and Shafer, 2008) which posits that difficult L2 sounds tend to take longer to process than easy categories.

2 Methods

2.1 Participants

Eighteen L1 MC listeners participated in the study. The participants were first-year international students at an Australian university ($M_{\text{age}} = 23.6$, range: 21–27). All spoke English as a second language, and their mean length of English education was 16.2 yr ($SD = 3.6$, range: 10–24). None reported a history of speech or hearing disorders. All participants were experienced L2 learners with advanced target language proficiency (B2-C1 level), and they were able to complete the perception task directly in the target language, i.e., English.

2.2 Stimuli

All twelve AusE monophthongal vowels [i: ɪ e e: æ ɐ ɜ: ɔ o: ʊ ʊ: ɜ:] (as in *beat, bit, bet, paired, bat, but, Bart, bird, bot, bought, boot, and who'd*) were included in the study (excluding only schwa). All the vowel stimuli were generated in the phonetic software Praat (Boersma, 2001), and the values for F1, F2, and duration were the same as those

reported in (Cox, 2006) for AusE male speakers. The pitch contour dropped linearly from 120 Hz to 100 Hz for all vowels, and the intensity was also set identically at a level of 70 dB SPL for all the stimuli. These manipulations ensured that listeners only attended to the first two formants and the characteristic duration value when making classification decisions.

2.3 Procedure and analysis

All participants completed a vowel classification task in the phonetic software Praat (Boersma, 2001) on a laptop computer. Classification responses were made by clicking boxes with keyword labels. Stimuli were delivered via acoustic headphones, and both the responses and the RTs were automatically recorded. Each participant completed 576 trials (12 vowels, 48 repetitions), and the stimuli were presented in a pseudo-randomised order. An additional 60 trials were given in a practice block before the experiment for familiarisation. To address research question (1), classification results were aggregated to generate a 12×12 confusion matrix in which each cell represents a unique “stimulus-response” combination. Any cell with a response percentage over 70% is considered as a “categorised” case, which indicates a consistent perceptual pattern (Bundgaard-Nielsen et al., 2011a, 2011b; Tyler et al., 2014). Classification is deemed as correct if the response category matches the stimulus phoneme. To address the research question (2), we calculated the between-category overlap score following the method introduced by Faris et al. (2018), which is formulated as below:

$$\text{Overlap}(AB) = \sum_1^n \min\{P_n(A), P_n(B)\}$$

where A and B represent two L2 categories, n represent the number of above chance response categories, $P_n(A)$ and $P_n(B)$ represent the percentage of choice of a specific vowel as the corresponding above chance category, and the function $\min\{\}$ selects the smaller value of the two percentage scores. In the present study, since there are twelve alternatives, the chance level is $1/12 \approx 8\%$.

To address research question (3), the between-category perceptual overlap scores were further analysed on the basis of every single AusE vowel.

For instance, some vowels can be overlapping with multiple categories, while in optimal performance, no perceptual overlap should be observed if every vowel is correctly classified in all trials. For each vowel, a sum score of perceptual overlap can be added up from all the pairwise overlaps where the target vowel is involved. To confirm the hypotheses, a negative correlation is expected between the classification accuracy and the sum score of perceptual overlap, and a positive correlation is expected between the perceptual overlap and RT.

3 Results

3.1 Classification results

The overall results of the classification task are summarised in Table 1. Cells on the diagonal line represent the average categorisation accuracy of each AusE vowel. The AusE vowel [ɪ] reached the highest accuracy of 90%, with an average RT of 1.60 sec, while [æ] was associated with the lowest accuracy of 41%, and the average RT was 2.98 sec. When the 70% criterion was applied (Bundgaard-Nielsen et al., 2011a, 2011b), seven AusE vowels were considered as categorised: [u:], [ɜ:], [ɛ:], [i:], [ʊ], [ɔ], and [ɪ]; while the other five vowels fail to reach the criterion: [æ], [e:], [e], [o:], and [ɐ]. A general tendency was observed that the vowels with higher classification accuracy values have shorter RT measures. To confirm this tendency, a Pearson’s correlation test revealed that the two measures were negatively correlated, $r = -0.817$ (95%CI: -0.459 ~ -0.947), $p = 0.001$.

3.2 Pairwise perceptual overlap

In the present study, we tested the classification of twelve AusE vowels, which potentially derive $C_{(12, 2)} = 66$ pairwise contrasts. However, not all of the vowel pairs were perceived as similar or overlapping by the MC listeners. Following previous research (Faris et al., 2018), we calculated the phonological overlap scores for all the 66 potential contrasts and find fifteen pairs that have shown above chance ($> 8\%$) overlap patterns, see Table 2. The AusE vowel pair [æ]-[e] received the highest overlap score of 44, followed by the [æ]-[e:] pair with an overlap score of 34, [o:]-[ʊ] with a

Table 1. Classification matrix of AusE vowels, percentage (mean RT, in seconds)

Stim.	Response Category											
	bad [æ]	paired [e:]	bed [e]	bought [o:]	but [ʊ]	who'd [u:]	bird [ɜ:]	Bart [ɛ:]	beat [i:]	hood [ʊ]	bot [ɔ]	bit [ɪ]
[æ]	41 (2.98)	8 (3.67)	27 (3.36)		5 (3.46)		17 (3.16)	1 (4.54)				
[e:]	4 (4.67)	55 (2.96)	10 (4.06)			7 (4.02)	16 (3.29)		4 (2.62)	1 (3.44)		2 (2.88)
[e]	9 (3.86)	15 (3.53)	56 (3.15)			4 (3.19)	5 (3.26)					10 (2.14)
[o:]				66 (2.27)		8 (3.18)				15 (3.34)	11 (3.03)	
[ʊ]	7 (3.73)	1 (4.26)	3 (4.05)		67 (2.48)		7 (3.39)	14 (2.93)				
[u:]		4 (6.07)	4 (4.71)			71 (2.68)	8 (3.29)			11 (3.46)		
[ɜ:]	2 (5.54)	8 (4.31)	5 (5.27)		1 (3.82)	10 (2.96)	72 (2.94)	1 (4.87)				
[ɛ:]	2 (5.81)	2 (3.98)			7 (4.70)		8 (4.13)	81 (2.26)				
[i:]		3 (3.40)				5 (2.71)			81 (1.86)			9 (2.35)
[ʊ]				1 (2.27)		7 (3.57)				81 (2.40)	11 (2.26)	
[ɔ]				10 (3.51)	3 (2.27)		2 (3.76)			2 (3.63)	82 (1.90)	
[ɪ]			1 (2.99)			1 (4.34)			4 (2.76)	4 (1.95)		90 (1.60)

Note. Cells with a percentage under 1% were removed; identification percentages over 70% are in boldface.

score of 25, [e:]-[e] with a score of 25, and [o:]-[ɔ] with a score of 20. Other overlapping pairs had a score under 20.

Table 2. Pairwise perceptual overlap scores

No.	AusE Vowel	Overlap
1	[æ]-[e]	44
2	[æ]-[e:]	34
3	[o:]-[ʊ]	25
4	[e:]-[e]	25
5	[o:]-[ɔ]	20
6	[æ]-[ɜ:]	17
7	[e:]-[ɜ:]	16
8	[ʊ]-[ɛ:]	14
9	[o:]-[u:]	11
10	[u:]-[ʊ]	11
11	[ʊ]-[ɔ]	11
12	[u:]-[ɜ:]	10
13	[e]-[ɪ]	10
14	[e]-[i:]	9
15	[i:]-[ɪ]	9

According to PAM and PAM-L2 (Best, 1995; Best and Tyler, 2007; Faris et al., 2018), the perceptual similarity between L2 vowels can predict their mutual discriminability. The present study has found that perceptual overlap of vowels seems to be gradient, e.g., both the mid vowels [e] and [ɜ:] were perceived as overlapping with [æ] and [e:], but to different degrees: [e]-[æ] (44) > [ɜ:]-[æ] (17), and [e]-[e:] (25) > [ɜ:]-[e:] (16). According to PAM, high levels of perceptual overlap lead to the assimilation types that are difficult to discriminate, including single-category (SC type), category-goodness (CG type), and some uncategorised-uncategorised (UU type, with classification overlaps) assimilations (Best, 1995; Best and Tyler, 2007; Faris et al., 2018).

3.3 Classification accuracy and overlap

The previous section has reported that all twelve AusE vowels are perceived as overlapping with at least one other category, and the level of perceptual overlap also differs from pair to pair. In addition, some vowels share overlaps with more than one other vowel category, e.g., the low front

vowel [æ] was perceived as overlapping with [e] (44), [e:] (34), and [ɜ:] (17), which yields a sum score of 95. This calculation probes into how much the overlapping area is taken up in the full classification space, and also estimate the likelihood of multiple category assimilation. Following this method, the sum scores for the other eleven AusE vowels were also calculated, and the values are summarised in Table 3, together with their mean accuracy and RT measures.

Table 3. *Overlap sum score and classification accuracy*

AusE Vowel	Overlap (Sum)	Accuracy (%)	RT (Sec)
[æ]	95	41	2.98
[e]	88	56	3.15
[e:]	74	55	2.96
[o:]	56	66	2.27
[ɒ]	47	81	2.40
[ɜ:]	43	72	2.94
[u:]	32	71	2.68
[ɔ]	31	82	1.90
[i:]	19	81	1.86
[ɐ]	14	67	2.48
[ɛ:]	14	81	2.26
[ɪ]	10	90	1.60

The analysis showed that three front vowels [æ], [e], and [e:] receive the highest sums of 95, 88, and 74, respectively. Note that these three vowels also showed the lowest accuracy measures of 41%, 56%, and 55%, respectively. These responses were also made with the highest level of decision uncertainty, as indicated by the RT measure of 2.98, 3.15, and 2.96 sec. The short high front vowel [ɪ] had the lowest total score of perceptual overlap (10), and it was also the category that showed the highest accuracy (90%). To confirm this tendency and answer research question (iii), a Pearson's correlation test was carried out, and it confirmed that the sum score of perceptual overlap was negatively correlated with the mean accuracy for the twelve AusE vowels, $r = -0.860$, 95% CI: $-0.565 \sim -0.960$, $p < 0.001$. For the RT measure, it was found to be positively correlated with the sum overlap score, $r = 0.760$, 95% CI: $0.330 \sim 0.929$, $p = 0.004$. These results confirmed our hypotheses that both the

classification accuracy and speed are closely associated with the perceptual overlap level.

4 Discussions

In the present study, we have examined the perception of the whole monophthong system of AusE by experienced L2 listeners whose native language is MC. In answering research question (i), we analysed the relative difficulty of vowel perception using both accuracy and latency measures. The findings suggest that not all the AusE vowels are easy to perceive even after more than ten years of English study. The negative correlation between classification accuracy and RT found in our data is consistent with the Automatic Selective Perception (ASP) model (Strange, 2011; Strange and Shafer, 2008), which predicts that successfully learned sound categories are perceived in a more automatic manner by non-native (L2) listeners. In answering research question (ii), we have diagnosed fifteen AusE vowel pairs that may be vulnerable to category merging and discrimination problems. In particular, our findings of the perceptual similarity patterns are consistent with reports of pairwise discrimination studies based on other English dialects, e.g., Canadian English (Wang, 2006; Wang and Munro, 1999, 2004), and American English (Jia et al., 2006; Lai, 2010).

In answering research question (iii), we extend the notion of perceptual overlap from pairwise discrimination towards an application in a multiple-choice classification task. The overlap score method used in the present study was introduced by a perceptual assimilation study which aimed to differentiate the subtypes of uncategorised assimilations (Faris et al., 2018). While the classic framework of PAM and PAM-L2 makes predictions based on qualitative differences in terms of assimilation mapping, e.g., two-category assimilation type means two L2 sounds are categorised as two distinct native categories, and excellent discrimination is expected (Best, 1995; Best and Tyler, 2007). The overlap score method (Faris et al., 2018) enables us to describe the perceptual similarity in quantitative terms.

The importance of analysing misclassification patterns is often neglected in previous studies, but for L2 phoneme classification the error patterns are

not purely random: certain vowels are more likely to be confused than others, e.g., MC listeners can confuse between AusE [æ] and [e], but less often between [æ] and [ɐ], see Table 1, 2.

Here we offer a schematic explanation of perception overlap in a minimal scenario, where A, B, and C are three non-native (L2) phonemes, and X and Y are two native (L1) phonemes: if both A and B are both perfectly mapped to category X (i.e., single-category assimilation), and C is mapped to category Y, then an unbiased listener should be unable to distinguish A from B, and should thus randomly categorise the sounds using the two labels, which will cause the classification results to be highly overlapping, see Figure 1. On the other hand, category C will show non-overlapping classification patterns with either A or B, because C forms a two-category assimilation pair with them. To conclude, classification overlaps reflect the level of perceptual similarity between L2 phonemes, and the patterns also reveal how these non-native sounds are assimilated by the native phonological system. As for experienced L2 listeners, the perceptual overlap between categories is more likely to be partial rather than a complete merging, because new L2 categories can be developed through extensive phonetic exposure and learning.

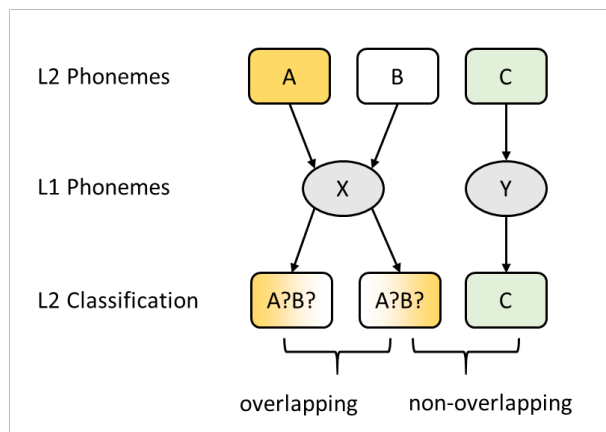


Figure 1. The relation between perceptual assimilation and L2 classification.

The present study, however, does have some limitations. Since the vowel stimuli used in the present classification task are presented in isolation, we have not examined whether different phonological contexts can influence the L2 listeners' classification performance, while a

growing body of research has reported that vowel perception is sensitive to the phonetic-phonological context in which the target vowel is situated (Levy, 2009a, 2009b; Strange et al., 2001). Additionally, since the vowel stimuli were synthesised signals with invariant formant values, the dynamic nuances of monophthongal vowels are ignored in the present study, while some studies have found that authentically produced vowel dynamics such as the inherent spectral change are important cues in vowel recognition (Jin and Liu, 2013; Morrison, 2009). Finally, since our study uses the classification paradigm, which is not often used by perceptual assimilation studies, further experiment comparison research is needed to inform the correspondence between the patterns revealed by a classification task and those from more conventional paradigms, e.g., assimilation and discrimination tasks (Bundgaard-Nielsen et al., 2011a; Faris et al., 2018; Tyler et al., 2014). In particular, how does classification overlap score (as shown in Table 2) corresponds to mutual discriminability remains an urgent topic for future studies.

5 Conclusion

In the present study, we have used the perceptual overlap score to analyse confusion patterns in L2 vowel perception, and we have also compared the overlap scores with the accuracy and RT measures and found robust correlation properties between them. These findings suggest that the misclassification patterns in a confusion matrix can reveal the perceived similarity between sound categories. For L2 listeners, in particular, this information should be analysed together with the potential interference from the listeners' native phonology, e.g., perceptual assimilation. The method used in the present study has the potential to be incorporated with other experimental paradigms (e.g., assimilation and discrimination tasks) in the cases where L2 listeners have relatively high proficiency in the target language and therefore have access to the emerging L2 categories.

References

- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Press.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn (Ed.), *Language experience in second language speech perception* (pp. 13–34). John Benjamins.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9–10), 341–345.
- Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011a). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, 22, 433–461. <https://doi.org/10.1017/S0272263111000040>
- Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011b). Vocabulary size matters: The assimilation of second language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, 32(1), 51–67. <https://doi.org/10.1017/S0142716410000287>
- Cox, F. (2006). The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics*, 26(2), 147–179.
- Faris, M. M., Best, C. T., & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America*, 139(1), EL1–EL5. <https://doi.org/10.1121/1.4939608>
- Faris, M. M., Best, C. T., & Tyler, M. D. (2018). Discrimination of uncategorized non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics*, 70, 1–19. <https://doi.org/10.1016/j.wocn.2018.05.003>
- Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32(3), 395–421. [https://doi.org/10.1016/S0095-4470\(03\)00016-0](https://doi.org/10.1016/S0095-4470(03)00016-0)
- Højen, A., & Flege, J. E. (2006). Early learners' discrimination of second-language vowels. *The Journal of the Acoustical Society of America*, 109, 3072–3084.
- Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America*, 119(2), 1118–1130.
- Jin, S.-H., & Liu, C. (2013). The vowel inherent spectral change of English vowels spoken by native and non-native speakers. *The Journal of the Acoustical Society of America*, 133(5), EL363–EL369. <https://doi.org/10.1121/1.4798620>
- Lai, Y. (2010). English vowel discrimination and assimilation by Chinese-speaking learners of English. *Concentric: Studies in Linguistics*, 36(2), 157–182.
- Levy, E. S. (2009a). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America*, 125(2), 1138–1152. <https://doi.org/10.1121/1.3050256>
- Levy, E. S. (2009b). On the assimilation-discrimination relationship in American English adults' French vowel learning. *The Journal of the Acoustical Society of America*, 126(5), 2670–2682. <https://doi.org/10.1121/1.3224715>
- Morrison, G. S. (2009). L1-Spanish Speakers' Acquisition of the English /i-i/ contrast II: Perception of Vowel Inherent Spectral Change. *Language and Speech*, 54(4), 437–462.
- Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39(4), 456–466. <https://doi.org/10.1016/j.wocn.2010.09.001>
- Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *The Journal of the Acoustical Society of America*, 109(4), 1691–1704.
- Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners. In J. G. Hansen-Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153–192). Benjamins.
- Tremblay, A. (2009). Phonetic variability and the variable perception of L2 word stress by french canadian listeners. *International Journal of Bilingualism*, 13(1), 35–62. <https://doi.org/10.1177/1367006909103528>
- Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, 71(1), 4–21. <https://doi.org/10.1159/000356237>
- Wang, X. (2006). Mandarin listeners' perception of english vowels: Problems and strategies.

-
- Canadian Acoustics*, 34(4), 15–26.
- Wang, X., & Munro, M. J. (1999). The perception of English tense-lax vowel pairs by native Mandarin speakers: The effects of training on attention to temporal and spectral cues. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (Vol. 3, pp. 125–128). American Institute of Physics.
- Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, 32(4), 539–552.