Predicting Discrimination Accuracy of English Vowels by Native-Mandarin Learners

Yuan Chai

University of California San Diego

Acknowledgement

I would like to thank Rebecca Scarborough and Marc Garellek for the guidance in the project and the manuscript revision; Mans Hulden and Barbara Fox for their constructive comments; Dongmei Ma and Hong Jiang for helping recruit subjects. I also thank the two anonymous reviewers for their helpful suggestions.

Abstract

The Perceptual Assimilation Model (PAM) maintains that the discrimination accuracy of two non-native sounds can be predicted by their assimilation patterns from the non-native language to the native language (Best, 1991). However, previous empirical studies have found deviations from PAM's predictions. We argue that this is due to several factors: in previous studies, sounds that were assimilated into multiple categories were defined as uncategorized; the criterion of categorization was arbitrary; and discrimination was predicted by both assimilation and the acoustic similarity between the contrasting sounds. In order to improve discrimination predictions, this study proposes three alternatives: revising PAM by redefining categorization: an L2 sound is considered 'categorized' if it is labeled as the same L1 category significantly above chance, whose level is determined by the number of native categories that differ from the target sound in no more than one articulatory feature. Further, we use overlap scores to measure the degree of overlap in the L1 assimilation categories, and add acoustic similarity to the predictors. Thirty first-year middle-school students from China participated in three experiments: discriminating English contrasts /i/-/ɪ/; /eɪ/-/ɪ/; /ɛ/-/æ/; /ʊ/-/u/; assimilating those English vowels into Mandarin categories; and rating their goodness in the Mandarin categories. The results show that L1 assimilation categories which have assimilation percentages slightly lower than the categorization threshold still influence discrimination, even after the adjustment in what counts as 'categorized'. In contrast, the overlap score predicts discrimination accuracy better than the assimilation pattern. Listeners are sensitive to the spectral differences between contrastive sounds, but not as much as the degree of assimilation overlap.

Keywords: discrimination, assimilation, PAM, overlap score

Predicting Discrimination Accuracy of English Vowels by Native-Mandarin Learners

When listening to a non-native language, people sometimes have the experience that two sounds that are different to native speakers sound the same to them. Such a phenomenon indicates that individuals' perception of non-native speech sounds is different from their perception of native sounds. Some non-native sounds are more discriminable than others. Researchers have strived to find ways to predict the degree of discriminability of non-native sounds, and numerous studies propose that listeners' perception of non-native sounds is largely influenced by the phonological system of their native languages (Kuhl, 1992; Best, 1991; Best, McRoberts, & Goodell, 2001; Flege, 1995). Language users become sensitive to the differences among their native phonetic categories, while also learning to ignore non-native phonetic differences (Kuhl, 1992; Kuhl et al, 2008). As a result, if two non-native sounds are perceived as belonging to different native sounds, they will be well discriminated. In contrast, if two non-native sounds are perceived as two instances of the same native sound, they are likely to be non-discriminable.

How can we predict the extent to which two non-native sounds will be discriminable to a second language learner? One possibility is provided by the Perceptual Assimilation Model (PAM) (Best, 1991; Best et al., 2001, Best & Tyler, 2007), which suggests that the degree of discrimination between two contrastive L2 sounds is predicted by the extent to which, and the way in which, the sounds are assimilated into native phonetic categories. If the contrasting sounds are assimilated into two different categories, it is a "Two Category (TC)" contrast. If the contrasting sounds are assimilated into the same category and have equal distance to the prototype, it is a "Single Category (SC)" contrast; on the other hand, if the contrasts are assimilated into the same category but differ in their distance to the prototype of that category, the pattern is called "Category Goodness (CG)" since the two sounds differ in their category goodness. There are also foreign sounds that cannot be assimilated into a single native category constantly, making them "uncategorized" sounds. If one member of a contrast is categorized while the other is not, the contrast is "Uncategorized-Categorized (UC)"; if both phones are not categorized, the contrast becomes "Uncategorized-Uncategorized (UU)". Sounds that are perceived as non-speech sounds are labeled as "Non-Assimilable (NA)". Among those six types of assimilation, TC contrasts should be discriminated with near-native accuracy because listeners could distinguish them as two different native phonemes. UC contrasts should also be well discriminated since one of the sounds falls into a native category while the uncategorized one is not assimilated into any native category. CG contrasts will be better discriminated than SC contrasts because the difference in goodness gives the listener a clue to distinguish them. The discrimination accuracy of UU contrasts depends on the phonetic difference between the two sounds and how they differ with respect to their nearby native phonemes (Best et al., 2001; Tyler, Best, Faber, & Levitt, 2014). The performance of NA discrimination depends on their articulatory differences (Best, 1995). Thus, the theory predicts a hierarchy of those types of contrast: TC, UC > CG > SC. The position of UU and NA on this continuum depends on the phonetic and/or acoustic similarity of the component sounds, but PAM does not specify how to calculate that relationship.

The predictions by PAM have been tested by numerous empirical studies; however, results often deviated from these predictions to some degree. The objective of the current study is to analyze the causes of those deviations and explore ways of improving the accuracy of PAM's predictions. In addition, this study conducts an empirical study testing the discrimination and assimilation of English vowels by Mandarin native speakers to verify the methods proposed.

The Validity of PAM's Predictions

Among the empirical studies verifying the predictions by PAM, several of them obtained results supporting the model’s predictions. Best, McRoberts, and Goodell (2001) tested the perceptual assimilation and discrimination accuracy of Zulu and Tigrinya consonants by English listeners, finding that the TC contrast (/ɬ/-/ɮ/) (95% correct) was better discriminated than the CG contrast (/k^h/-/k’/) (89.4% correct), which was better discriminated than the SC contrast (/b/-/ɓ/) (65.9% correct). The assimilation pattern of Tigrinya contrast /p’/-/t’/ showed individual variation: discrimination accuracy by listeners who categorized /p’/-/t’/ as TC was near-ceiling (94.09% correct), whereas discrimination accuracy by those who categorized it as SC was lower (64.58% correct). Lengeris (2009) explored how Greek and Japanese native speakers assimilated and discriminated ten English vowels (/iː/-/ɪ/, /ɪ/-/e/, /æ/-/ʌ/, /æ/-/aː/, /æ/-/ɜː/, /ʌ/-/aː/, /ɒ/-/ɔː/, /ʊ/-/uː/ and /ɔː/-/uː/) in two contexts (/bVb/, /bVp/). The results indicated that, for both Greek and Japanese speakers and averaging across two consonantal contexts, the discrimination accuracy of the attested contrasts ranked in TC > UC, CG > SC. Tyler et al. (2014) tested the discrimination and assimilation of six vowel contrasts by American English speakers (Norwegian: /ki/-/ky/, /ki/-/kʉ/; Thai: /bɯ/-/bɤ/; French: /bo/-/bõ/, /dø/-/dœ/). They found that the assimilation type varied significantly across individuals. Therefore, instead of assigning an assimilation type to each contrast, the discrimination scores were grouped and compared by individuals’ assimilation type. TC and UC contrasts generally had higher discrimination scores than CG and SC contrasts, and CG contrasts generally had higher scores than SC contrasts.

However, many PAM studies have found results that run counter to the model’s predictions. UC contrasts are not always well discriminated, with discrimination scores ranging from as poor as SC to as good as TC (Antoniou, Tyler, & Best, 2012; Bohn, Best, Avesani, & Vayra, 2011; Bundgaard-Nielsen, Best, & Tyler, 2011a; Guion, Flege, Akahane-Yamada, & Pruitt, 2000;

Harnsberger, 2001; Reid et al., 2015; Sisinni & Grimaldi, 2009; Tyler et al., 2014). TC contrasts sometimes received low discrimination scores (Best & Tyler, 2007; Bohn et al., 2011; Reid et al., 2015), while SC contrasts sometimes received high discrimination scores (Best, Hallé, Bohn, & Faber, 2003).

We propose that PAM's predictions were not supported in those studies in part because, for UC contrasts, their discrimination accuracy was related to the degree of overlap between the assimilation category of the categorized sound and those of the uncategorized sounds. However, the assimilation categories of the uncategorized sounds are ignored in the current version of PAM. Another reason for unexpected results is that the criterion of whether a non-native sound was categorized or not was arbitrary and varied greatly across studies, making the results of each study less generalizable. Finally, the assimilation pattern itself is not a perfect predictor of discrimination accuracy: listeners also pay attention to acoustic cues when discriminating non-native contrasts. The following section discusses each of those causes in detail and introduces the existing solutions to those limitations.

The Assimilation Categories of Uncategorized Sounds

PAM only allows single-category assimilation: if a given sound is not consistently assimilated into a single category, it is not categorized at all. However, several studies have reported that the assimilation categories of such uncategorized sounds are informative in predicting the discrimination of UC contrasts. When there is a larger degree of overlap in the assimilation categories between the categorized sound and the uncategorized sound, discrimination accuracy declines. For example, Guion et al. (2000) tested how Japanese speakers discriminated and assimilated English consonants into their L1. Contrasts /s/-/θ/ and /ɹ/-/w/ were both treated as UC-type contrasts, but the discrimination score for /s/-/θ/ (A' < 0.5) was higher than that for /ɹ/-/w/

(A’ > 0.75). For the /s/-/θ/ contrast, /θ/ was most frequently labeled as Japanese /s/ (39%) and /ɸ/ (38%), whereas English /s/ was categorized as Japanese /s/ (87%). For the /ɹ/-/w/ contrast, /ɹ/ was most frequently labeled as Japanese /ɯɾ/ (50%) and /ɾ/ (46%) while English /w/ was categorized as /ɰ/ (79%). Since the L1 assimilation categories of /s/-/θ/ overlapped but those of /ɹ/-/w/ did not, the former contrast was less distinguishable than the latter. In another study, Antoniou, Tyler, and Best (2012) tested how Greek monolinguals discriminated and assimilated Ma’di stops. They found that there was large variation in the discrimination scores of UC contrasts, ranging from 59.2% (/b/-/ɓ/) to 80.6% (/b/-/p/). For the /b/-/ɓ/ contrast, Ma’di /ɓ/ was usually categorized as Greek /b/ (97.5%) while Ma’di /b/ was labeled as Greek /b/ only 63.8% of the time. For the /b/-/p/ contrast, Ma’di /p/ was always categorized as Greek /p/ (100%) while Ma’di /b/ was labeled as Greek /p/ 22.5% of the time. The degree of overlap in the assimilation categories of /b/-/ɓ/ was larger than that of /b/-/p/. As a result, /b/-/ɓ/ was less discriminable than /b/-/p/. In yet another example, Sisinni and Grimaldi (2009) found that Salento Italian speakers perceived British English vowel contrasts /æ/-/ɑ/, /æ/-/ʌ/, /uː/-/ʊ/, /ɒ/-/ɜː/, and /ɛ/-/ɜː/ as UC-type contrasts, and the discrimination scores (A’) of those contrasts ranged from 0.69 (/æ/-/ɑ/) to 0.83 (/ɛ/-/ɜː/). Within the /æ/-/ɑ/ contrast, English /æ/ was categorized as /a/ (80%), while English /ɑ/ was similarly labeled as Italian /a/ 74% of the time. In terms of the /ɛ/-/ɜː/ contrast, English /ɛ/ was categorized into Italian /ɛ/ (100%) while English /ɜː/ was labeled as Italian /ɛ/ 43% of the time only. Similar to the study by Best and Tyler (2006), contrasts with a larger degree of overlap in their L1 assimilation categories received a lower discrimination score. Such a relationship between the degree of overlap and the discrimination accuracy have been found in many other studies (Bohn et al., 2011; Bundgaard-Nielsen et al., 2011a; Harnsberger, 2001; Tyler et al., 2014).

Previous studies have proposed two solutions to the inaccurate prediction of UC-type contrasts by PAM. The first solution is to divide UC contrasts into UCs with and without overlap in the L1 assimilation categories (Faris, Best, & Tyler, 2016, 2018). Faris et al. (2016, 2018) have proposed that the discrimination accuracy of contrasts involving uncategorized sounds is affected by the native categories that the uncategorized contrasts were assimilated into above chance. They have further claimed that non-native sounds that are consistently assimilated into a single native category 50% of time should be considered as ‘categorized.’ Moreover, uncategorized sounds can also have their own assimilation category/categories. If a given native category is selected as the assimilation sound of the target non-native sound significantly above chance, it qualifies as a native assimilation category of the target sound. The chance level of the non-native sounds is calculated as 100% divided by the total number of vowels/consonants within the inventory of the native language. A given uncategorized sound can have either one, or multiple, or no native assimilation category/categories. Finally, the discrimination accuracy of sounds involving an uncategorized sound depends on the degree of overlap in the assimilation categories of the target sounds: UC or UU contrasts with no overlap in the assimilation categories should be better discriminated than those with partial overlap in the assimilation categories, followed by those with complete overlap in the assimilation categories. The advantage of this solution is that it accounts for the assimilation categories of uncategorized sounds. The limitation of this solution is that it still adopts arbitrary criterion for categorization, the consequence of which will be discussed in the following section.

The second solution is to use overlap score instead of assimilation pattern to predict discrimination (Flege & MacKay, 2004; Levy, 2009). Overlap score is defined as the smaller percentage of responses for which the components of a contrast were labeled as the same L1 category. For instance, given a French /i/-/y/ contrast where French /i/ is labeled as English /i/ 80%

of the time while French /y/ is labeled as English /i/ 10% of the time, the overlap score between French /i/ and /y/ in terms of English /i/ is 10%. The total overlap score of a contrast is the sum of the overlap scores of all the L1 assimilation categories. The higher the total overlap score is, the lower the discrimination accuracy will be. The advantage of using the overlap score is that it accounts for all the overlap in the L1 assimilation categories of the non-native contrasts, regardless of how little the overlap is or whether the non-native sound is categorized or not. As a result, the degree of overlap in UC-type contrast assimilation categories can be incorporated within PAM. The first limitation of overlap score is that it does not account for the category goodness. The second limitation is that it may overestimate the degree of overlap (Faris et al., 2018). It is usually sufficient to distinguish the degree of overlap among contrasts by looking at the top one or two assimilation categories of the uncategorized sounds, indicating that L1 categories that receive relatively small assimilation percentages in the assimilation test usually do not contribute a lot to the perceptual overlap between two non-native sounds. Consequently, the sum of the overlap score of all their assimilation categories could be larger than the actual degree of overlap the listeners perceive, especially when the members of a contrast are dispersedly assimilated.

Categorization Criterion

Studies testing PAM have used arbitrary and divergent categorization criterion thresholds. Some studies (Mokari & Werner, 2017; Harnsberger, 2001; Lai, 2010) used the criterion that if a given sound was designated with the same label over 90% of time, that sound was categorized. Other studies have set their criterion threshold at 75% (Guion et al., 2000; Sisinni & Grimaldi, 2009), 70% (Antoniou et al., 2012; Best & Tyler, 2007; Bohn et al., 2011; Guion et al., 2000; Strange, Akahane-Yamada, Kubo, Trent, & Nishi, 2001; Tyler et al, 2014), 60% (Lengeris, 2009), or 50% (Bundgaard-Nielsen, Best, & Tyler, 2011b; Faris et al., 2018; Reid et al., 2015). The categorization criterion directly determines the assimilation type, so it is likely that changing the criterion will lead to a shift in assimilation pattern. For instance, Antoniou et al. (2012) used 70% as the criterion and classified the Greek /b/-/p/ contrast as a TC-type contrast by English monolinguals. However, the discrimination of /b/-/p/ was lower (63.3%) than other TC contrasts. The authors maintained that if a more stringent criterion had been used (e.g. 75%), the /b/-/p/ contrast would have become a UC-type contrast and the overlap between Greek /b/ and /p/ in English assimilation category /b/ would have explained the low discrimination accuracy of the Greek /b/-/p/ contrast. In addition, having a cut-off threshold is problematic when the assimilation percentage of some L1 category is at the borderline of the threshold (Harnsberger, 2001): for instance, if Foreign Sound 1 is assimilated into Native Category A 76% of the time while Foreign Sound 2 is assimilated into native category B 74% of the time, it is difficult to conclude that the assimilation of 1 to A is more consistent than the assimilation of 2 to B. However, if the categorization criterion were set to 75%, Foreign Sound 1 would be labeled as "Categorized" while Foreign Sound 2 would be "Uncategorized". Since the goal of PAM is to use assimilation patterns to predict discrimination accuracy, arbitrary and varied categorization criterion thresholds can lead to imprecise determination of the assimilation patterns.

Acoustic Similarity Influences Discrimination

Aside from the issues regarding the assimilation categories of uncategorized sounds and the arbitrariness of the categorization criterion, another reason why PAM's predictions have not borne out in previous work is that assimilation pattern is not a perfect predictor of discrimination. For example, Best et al. (2003) found for Danish speakers that, although the assimilation categories of Norwegian /y/ and /ʉ/ overlapped substantially with the Danish vowel inventory and that their goodness of fit of the overlapping categories did not differ much, Danish listeners could discriminate well between /y/ and /ʉ/ (97-98% correct). Likewise, Flege and MacKay (2004) found that for Italian listeners, while the English /ɪ/-/ɛ/ contrast had a low overlap score (35%), their discrimination of /ɪ/-/ɛ/ was the third lowest among the eight contrasts attested. The authors argued that the reason why a low overlap score did not lead to a high discrimination accuracy was that the formant frequencies and durations of /ɪ/ and /ɛ/ were very similar. The acoustic similarity made discrimination harder for the listeners. Bohn et al. (2011) found that, although Danish listeners assimilated English /ð/ and /v/ with less overlap in their native categories than them assimilating English /w/ and /v/, their average discrimination score of /ð/-/v/ contrast (84%) was significantly lower than that of /w/-/v/ (95%). The authors claimed that this was due to /ð/ and /v/ having very similar acoustic and articulatory properties. In another study involving how Cantonese speakers discriminate and assimilate Thai tones, Reid et al. (2015) found that TC contrasts did not differ significantly from SC contrasts in terms of discrimination accuracy.

Taken together, these results suggest that assimilation patterns may not by themselves be able to predict discrimination accuracy; the acoustic/articulatory similarity of the foreign contrasts may also play a role in determining their discrimination accuracy.

The Current Study

Motivated by the aforementioned limitations of PAM and the availability of possible solutions, the research question of the current study is: how can we improve the accuracy of predicting discrimination based on assimilation? The current study addresses this by testing the discrimination and assimilation of English vowels by Mandarin native speakers. We use the extended PAM (PAM-e), which specifies UC contrasts as UC-Non-overlapping (UC-N), UC-Partial-overlapping (UC-P), and UC-Completely overlapping (UC-C) (Faris et al., 2016, 2018), in addition to the overlap score (Flege & MacKay, 2004; Levy, 2009) as predictors of discrimination accuracy. In addition, this study modifies both the extended PAM and the calculation of the overlap score in order to account for the limitations discussed in the previous sections.

Modifications to PAM-e

This study modified the categorization criterion of PAM-e. Instead of using an arbitrary threshold using a percentage, we examine whether a given non-native sound is assimilated into one or more L1 categories more often than chance and if so, into how many such L1 categories. If one or more than one L1 category is selected more consistently than chance in the assimilation test, that non-native sound is assimilated into a single L1 category or multiple L1 categories. Only if none of the L1 categories is selected above chance is that non-native sound regarded as "uncategorized". The chance level of a given non-native sound is determined by the number of native categories that the non-native sound can possibly be assimilated into. Possible native assimilation categories of a given non-native sound are native sounds that differ from the non-native sound in terms of no more than one articulatory feature (or 'gesture'), such as height, backness, and roundness. For example, for English /i/, Mandarin /iε/ is a possible native assimilation category because /i/ is a high front unrounded vowel while /iε/ is a diphthong consisting of a high front unrounded vowel and a mid front unrounded vowel. The only different phonological feature between /i/ and /iε/ is the height. In contrast, Mandarin /u/ is not a possible native assimilation category because /u/ is a high back rounded vowel, differing from English /i/ in both the backness and the rounding features. Therefore, participants will be provided with native assimilation categories that differ by no more than one feature in the assimilation (identification) test. This study uses the articulatory feature to determine the possible assimilation category because PAM is based on Direct Realist Theory (Fowler, 1996), which suggests that articulatory gestures are the primary cues used to distinguish sounds phonologically (Best, 1991; Best et al.

2001). After the adjustment, this study predicts that the discrimination accuracy will be highest for contrasts with no overlapping assimilation categories, followed by contrasts with partial overlapping assimilation categories, while contrasts with complete overlapping assimilation categories will be most difficult to discriminate. Within contrasts with partial and complete assimilation category overlap, contrasts whose category goodness in the overlapping category differs from each other would be better discriminated than those whose category goodness does not differ.

There are three major differences between the criterion proposed in this study and that in Faris et al. (2016, 2018). Firstly, Faris et al. preserved the arbitrary cut-off threshold for categorized sounds, whereas this study does not. Secondly, the chance level in the previous studies was used to further classify uncategorized sounds, and was used only if there was no native category that received an assimilation percentage above the cut-off threshold (50%). In contrast, the current study uses the chance level to distinguish categorized sounds from uncategorized sounds. Third, the definition of chance level is different. Faris et al. set the chance level as 100% divided by the number of vowels within the phonological inventory of the native language; on the other hand, in the current study we set the chance level to 100% divided by the number of vowel in native language inventories that differ with the target sounds in no more than one articulatory feature.

The current study adopts such a different criterion from Faris and colleagues for two reasons. First, although the previous criterion allowed multiple assimilation patterns for uncategorized sounds, multiple assimilation patterns for categorized sounds were still prohibited. However, it is possible that categorized sounds are assimilated into multiple categories (Escudero & Boersma, 2002), and such information could influence the prediction of the discrimination accuracy. For example, if we suppose that Sound 1 is assimilated into Sound A 51% of the time, and into Sound B 35% of the time. Sound 2 is assimilated into Sound A 35% of the time, and into Sound B 51% of the time. Faris et al.'s method will classify them both as a TC assimilation contrast, because Sound 1 is assimilated into Sound A, and Sound 2 is assimilated into Sound B. The fact that both Sound 1 and 2 are also sometimes assimilated into the other categories is ignored. However, such overlap can probably result in a decrease in discrimination accuracy. In contrast, the categorization criterion employed in the current study allows for the possibility of multiple assimilation patterns for every non-native sound attested.

Second, the criterion adopted in the current study is determined by the number of native categories that a non-native sound can plausibly be assimilated into, which is based on an a priori threshold, namely the number of native categories that differ with non-native categories no more than one articulatory feature. This study did not use Faris et al.'s criterion (100%/the number of vowels in the entire inventory) for two reasons. First, such a criterion is too lenient: the target non-native sound can be categorized into a native category with a very low assimilation percentage, which may not reflect the actual assimilation status of the target sound. In addition, there can be much overlap in the assimilation categories for most contrasts, resulting in a lack of variability in the assimilation pattern, which would result in assimilation patterns poorly predicting discrimination accuracy.

The second reason for not adopting the criterion from previous work is that the criterion adopted in this study allows sounds to have different chance levels for non-native sound categorization, while Faris et al.'s does not. Non-native sounds with more native phonological neighbors are expected to be more confusable, since there are more available options for them to be assimilated into. Thus, non-native sounds with more neighbors in the native phonological space should have a relatively lower chance level in the assimilation test. Likewise, non-native sounds with fewer native phonological neighbors have fewer competitors in the assimilation process. Thus, their chance level should be relatively higher. The criterion adopted in this study allows the chance level of the non-native sounds to vary by their density of neighborhood in the native phonological space.

Modifications to Overlap Score

Because this study provides the listeners with Mandarin vowels that differ from the target English vowel in no more than one articulatory feature as its assimilation options in the assimilation test, the calculation of the total overlap score must be modified. Previous studies provided the listeners with the entire native vowel/consonant inventory in the assimilation test, which may lead to an overestimation of the degree of overlap in the assimilation categories of a given non-native contrast (Faris et al, 2018). But because this study reduces the number of assimilation options in the assimilation test, the overlap in trivial L1 categories will not be added to the total overlap score.

Acoustic Similarity

Because previous studies indicated that assimilation pattern itself may not be sufficient in predicting the discrimination of non-native phones, this study adds two acoustic variables to the predictors of discrimination accuracy: the Euclidean spectral (F1-F2) distance and vowel duration difference between the members of each contrast.

Euclidean spectral distance and duration difference are chosen to represent the acoustic similarity between the members of each non-native contrasts because studies have shown that listeners utilize the spectral and duration properties during the non-native speech sound assimilation and discrimination. For example, the Second Language Linguistic Perception (L2LP) model proposes that the L1-L2 assimilation can be predicted by the acoustic similarity between

L1 and L2 (Escudero, 2005, 2009). L2 sounds are most likely to be assimilated into the L1 category to which they are acoustically most similar. Several empirical studies have supported this proposal. The combination of F1, F2, F3, and duration difference between Canadian French (CF) and Peruvian Spanish (PS) and between Canadian English (CE) and PS largely predicted how PS native speakers assimilated CF and CE vowels into PS vowels (Escudero & Vasiliev, 2011). The F1 and F2 similarity between Dutch and PS vowels generally predicted how PS speakers categorized Dutch into PS (Escudero & Williams, 2011). Studies have also shown that acoustic cues can be used for L2 discrimination. Catalan speakers relied on the duration difference more than the spectral difference between /i/ and /ɪ/ when distinguishing /i/ from /ɪ/ (Cebrian, 2006). Mandarin and Korean speakers relied more on the duration cue than the spectral cue when distinguishing English lax-tense vowel contrasts (/i/-/ɪ/, /ɛ/-/æ/), while German and Spanish speakers behaved reversely (Flege, Bohn, & Jang, 1997). Japanese speakers were more sensitive to the variation in F2 than in F3 while German speakers were more sensitive to F3 than F2 when discrimination English /ɹ/-/l/ contrast (Iverson et al., 2003). Taken together, these studies indicate that spectral and duration cues are informative in predicting both assimilation and prediction. Thus, this study tests how well the acoustic similarity within the target contrasts predicts the discrimination accuracy of the target contrast, and explores whether acoustic similarity or L1-L2 assimilation status is a better predictor of L2 discrimination.

How Mandarin Speakers Perceive English Vowels

Wang (1997) reported that in an identification test of English vowels /i, ɪ, ɛ, æ, eɪ/ by Mandarin native speakers, English vowels /i/ received the highest identification accuracy; /eɪ/ received the lowest; /ɪ, ɛ, æ/ were in between. English /i/ was mostly misidentified as /ɪ/ (14%), /ɪ/ as /ɛ/ (22%), /eɪ/ as /i/ (23%), /ɛ/ as /æ/ (21%), and /æ/ as /ɛ/ (18%), indicating that Mandarin speakers may have difficulty discriminating the English /ɛ/-/æ/ contrast. Mandarin speakers were also inclined to use duration rather than spectral cues to distinguish the English /i/-/ɪ/ contrast, but were not able to use either cue systematically to distinguish English /ɛ/-/æ/ and /u/-/ʊ/ contrasts (Wang, 2006; Wang & Munro, 1999). This indicates that Mandarin speakers may distinguish /ɛ/-/æ/ and /u/-/ʊ/ poorly since they are not sensitive to their acoustic similarity, but may distinguish /i/-/ɪ/ well based on duration. Jia, Strange, Wu, Collado, and Guan (2005) tested the discrimination of six English vowel contrasts by Mandarin monolinguals and their discrimination accuracy ranked as /u/-/ɑ/ (97.9%) > /i/-/eɪ/ (90.2%) > /æ/-/ɑ/ (88.9%) > /i/-/ɪ/ (82.6%) > /ɛ/-/æ/ (76.3%) > /ɑ/-/ʌ/ (71.7%).

In order to create variability in the discrimination result, this study uses English /i/-/ɪ/, /ɪ/-/eɪ/, /ɛ/-/æ/, and /u/-/ʊ/ contrasts. Based on the results of previous studies, we hypothesize that their discrimination accuracy will rank in /ɪ/-/eɪ/ & /i/-/ɪ/ > /ɛ/-/æ/ & /u/-/ʊ/ since there is a distinct duration difference between /i/-/ɪ/ and Mandarin speakers used duration consistently for the /i/-/ɪ/ contrast (Wang, 2006; Wang & Munro, 1999). The duration difference between the monophthong /ɪ/ and the diphthong /eɪ/ should also facilitate the discrimination. And since the speakers did not use acoustic cues consistently for /ɛ/-/æ/ or /u/-/ʊ/ contrasts (Wang, 2006; Wang & Munro, 1999), their discrimination accuracy should be relatively lower than the other two.

The assimilation status of those target vowels into Mandarin varied substantially in previous studies. Sun and Heuven (2007) suggested that English (E) /i/, /ɪ/, and /eɪ/ were all assimilated into Mandarin (M) /eɪ/ (66%, 73%, and 91% respectively); E /ɛ/ and /æ/ were both labeled as M /ɑ/ (33% and 66% respectively) and /ɪə/ (16% for both) most frequently. E /u/ was assimilated into M /u/ (91%) while E /ʊ/ was assimilated into M /oʊ/ (63%). The assimilation pattern obtained by Lai (2010) differed from Sun and Heuven (2007). E /i/ and /ɪ/ were labeled as

M /i/ most frequently (97% and 60% respectively). E /eɪ/ was assimilated into M /ei/ (92%). E /ɛ/ and /æ/ were labeled as M /e/ (82% and 77% respectively). E /u/ and /ʊ/ were both labeled as M /u/ most frequently (97% and 70% respectively). Since there is a lot of individual variation in L1-L2 assimilation and the assimilation options in the assimilation test differ among previous studies and between the literature and the current study, it is hard to predict the assimilation pattern of the target vowels beforehand. However, given the assimilation results from the previous study, we expect various assimilation patterns to emerge in this study.

In summary, the research questions of the current study are: 1) with the modification to PAM-e and the overlap score calculation proposed by this study, are the assimilation patterns from English to Mandarin vowels good predictors of the discrimination of English /i/-/ɪ/, /ɪ/-/eɪ/, /ɛ/-/æ/, and /u/-/ʊ/ contrasts by Mandarin native speakers? 2) is the assimilation pattern, overlap score, or the acoustic similarity between the members of the contrasts a better predictor of the discrimination of target English contrasts by Mandarin native speakers?

The hypotheses are twofold: first, with the modifications proposed by this study, the discrimination of the target English contrasts will be well predicted by the assimilation from the target English vowels to the Mandarin categories. Specifically, the difficulty of contrasts discrimination will be predicted by their degree of overlap in the L1 assimilation: contrasts with non-overlapping assimilation categories will be best discriminated, followed by those with partial overlapping assimilation categories while those with complete overlapping assimilation categories will be discriminated least accurately. In terms of the overlap score, contrasts with higher overlap scores in the L1 assimilation categories will receive poorer discrimination accuracy. The second hypothesis is that both assimilation pattern and overlap score will predict discrimination accuracy more accurately than acoustic similarity.

Methods

In order to test the relationship between discrimination accuracy and assimilation, three experiments were conducted: 1) an English vowel discrimination test to obtain the accuracy hierarchy between contrasts; 2) an identification test to decide how English vowels are assimilated into Mandarin vowels; 3) a goodness rating test to figure out the "fitness" of the English vowels in the Mandarin vowel category that they are assimilated into. All three experiments were completed in one sitting and it took around 30 minutes to finish those experiments. Future studies may arrange breaks among the experiments to avoid fatiguing the participants.

Participants

30 native Mandarin speakers were recruited among first-year middle school students in Qiqihaer, Heilongjiang Province, China. Those Mandarin speakers' exposure to English was mainly from English class in school, which focused on English reading and writing and was taught by non-native English speakers. Thus, the subjects had minimal exposure to native spoken English and could be considered as near-monolinguals of Mandarin. In addition, all the subjects were from the same dialect region of China.

This study also recruited 30 native speakers of American English from introductory linguistics classes at a public university in the Southwest of US. The purpose of involving English speakers in this study is to compare their scores on the English vowel discrimination test to the scores of the Mandarin speakers. For a given vowel contrast, if Mandarin speakers' discrimination scores are significantly lower than the English speakers', Mandarin speakers can be regarded as having difficulties discriminating that vowel contrast. Every participant signed a consent form to participate in the study. This study was approved by the university's Institutional Review Board.

Experiment 1: English Vowel Discrimination

This experiment aimed to test how well Mandarin speakers discriminated English vowel contrasts /i/-/ɪ/; /eɪ/-/ɪ/; /ɛ/-/æ/; and /ʊ/-/u/ using an AXB discrimination task. Both English and Mandarin subjects participated into this experiment.

Stimuli

The target vowel contrasts for the test were /i/-/ɪ/; /eɪ/-/ɪ/; /ɛ/-/æ/; and /ʊ/-/u/. The vowels /i/, /eɪ/, /ɛ/, /æ/, /ʊ/, and /u/ were elicited in a t_d environment. The environment was the same for all the vowels because previous studies suggested that consonant context would affect the discrimination accuracy of foreign sounds (Strange et al., 2001). The discrimination test was presented in AXB mode, following the procedure of the perception test by Højen and Flege (2006). A variation in F0 was created among A, X, and B within each triad. The purpose of creating F0 variation within AXB triad was to increase the acoustic variability of the stimuli, which would increase the task difficulty since subjects had to recognize the phonological properties of the sounds despite the interference of F0. For example, for an AAB triad where the first A was in high F0 and the second A and B were in low F0, if the subject responded that the first two sounds were the same, he/she was paying attention to the phonological similarity between the first two sound rather than the F0 difference between them. Moreover, the acoustic variance can be further increased by having the stimuli produced by different speakers, which could be realized in future studies. The F0 of the target words was either falling or rising. In order to manipulate the pitch of the target words, they were elicited in carrier sentences "I say ____." and "Can I say ____?" The first carrier sentence yielded a target word in falling F0 while the second rising F0. The F0 of the AXB triple was either "neutral" or "incongruent". A "neutral" F0 means that the measure did not indicate whether the adjacent pair was the same word or not. For example, A (rising F0) X=A (falling F0) B (rising F0). "Incongruent" meant F0 was the same between different words. For example, A (rising F0), X=A (falling F0), B (falling F0). The stimuli were produced by a female who is a native speaker of American English. Each target word was presented 10 times with falling F0 and 10 times with rising F0. In order to ensure that speakers were able to ignore the variations in the realizations of the same word, AA/BB were all different tokens of A/B.

The experiment consisted of four blocks. Each block contained 16 different trials for each contrast. Within each block, each type of AXB combination occurred 4 times (AAB, ABB, BBA, BAA). The chance of each token occurring at each position was equal. Half of the trials had neutral F0 and the other half had incongruent F0. The inter-stimulus interval was 500ms. There were 256 (4 contrasts * 4 combinations * 4 repetitions * 4 blocks) trials in this experiment in total.

The F1 and F2 of the English vowels in the stimuli of Experiment 1 were measured using Praat (Boersma, 2001). The F1 and F2 of monophthongs were measured at the vowel midpoint. In diphthongs, the nucleus and off-glide's formants were measured at the midpoint before and after the formant-shift. The results are presented in Table 1.

Table 1. *Means and standard deviations for F1, F2, and duration of English vowels*

Language	Sound	*n*	F1 (Hz)		F2 (Hz)		Duration (s)	
			Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
English	/i/	8	402.625	31.268	2563.000	48.908	0.174	0.028
	/ɪ/	8	542.750	26.174	2065.625	28.046	0.168	0.014
	/e/ in /eɪ/[1]	8	527.750	19.002	2307.375	76.131	0.208	0.015
	/ɪ/ in /eɪ/	8	455.000	36.249	2393.250	61.946		
	/ɛ/	8	738.625	20.085	1882.500	38.597	0.146	0.012
	/æ/	8	863.750	31.089	1602.000	46.031	0.194	0.015
	/ʊ/	8	574.875	30.866	1814.875	47.688	0.162	0.014
	/u/	8	407.750	29.060	1751.875	87.726	0.166	0.013

[1]/eɪ/ is split into /e/ and /ɪ/ when measuring the formant in order to reflect the formant changing within the diphthong.

The Euclidean spectral (F1-F2) distance and vowel duration difference between every pair of contrastive sounds was calculated and those two factors were compared across the four contrasts. The Euclidean distances between the sounds within contrasts were computed by the following formula. Suppose there is a token of /i/ and a token of /ɪ/. The equation for computing the spectral distance between /i/ and /ɪ/ is:

$$Euclidean\ Distance\ (\text{i}\sim\text{ɪ}) = \sqrt{(F1(i) - F1(\text{ɪ}))^2\ + (F2(i) - F2(\text{ɪ}))^2}$$

There were 8 tokens for each vowel in the discrimination test. Thus, each contrast had 64 samples of Euclidean distance. The mean and standard deviation of the spectral distance within each contrast are presented in Table 2.

Table 2. *Means and standard deviations for Euclidean distance between sounds within contrasts*

Contrast	*n*	*Mean*	*SD*
/i/-/ɪ/	64	518.032	54.242
/ɪ/-/eɪ/[1]	64	290.408	69.976
/ɛ/-/æ/	64	310.847	45.822
/u/-/ʊ/	64	198.057	54.922

[1] The F1 and F2 of /eɪ/ is the average of the nucleus /e/ and the offglide /ɪ/

A one-way ANOVA revealed that there are differences in Euclidean distance among the four contrasts, $F(3, 252) = 360.289$, $\eta^2 = .809$, $p < .001$. The post-hoc tests show that the Euclidean distance hierarchy among those four contrasts is: /i/-/ɪ/ > /ɪ/-/eɪ/ = /ɛ/-/æ/ > /u/-/ʊ/ (alpha level adjusted to .008 by Bonferroni adjustment).

The differences in vowel length between the two sounds in each contrast were computed by the following equation using /i/-/ɪ/ contrast as an example:

$$Length\ Diff\ (\text{i}\sim\text{ɪ}) = Length(i) - Length(\text{ɪ})$$

As mentioned previously, each vowel was repeated 8 times, creating 64 samples of vowel length difference. The mean and standard deviation of the vowel length of each sound in each contrast are illustrated in Table 3.

Table 3. *Means and standard deviations for duration difference between sounds within contrasts*

Contrast	*n*	*Mean*	*SD*
/i/-/ɪ/	64	0.026	0.016
/ɪ/-/eɪ/	64	0.040	0.019
/ɛ/-/æ/	64	0.048	0.019
/u/-/ʊ/	64	0.015	0.011

A one-way ANOVA revealed that there are differences in vowel length among four contrasts, $F(3, 252) = 50.468$, $\eta^2 = .368$, $p < .001$. The post-hoc tests show that the vowel length hierarchy among those four contrasts is: /ɛ/-/æ/ > /ɪ/-/eɪ/ > /i/-/ɪ/ > /u/-/ʊ/ (alpha level adjusted to .05/6 = .008 by Bonferroni adjustment). The post-hoc statistical results of vowel length and duration difference are in Table 4.

Table 4. *Post-hoc tests of Euclidean distance and duration difference comparison among contrasts*

Comparison	Euclidean Distance				Duration Difference			
	df	*F*	η^2	*p*	*df*	*F*	η^2	*p*
/i/-/ɪ/ vs. /ɪ/-/eɪ/	1	511.936	.670	<.001	1	23.873	.087	<.001
/i/-/ɪ/ vs. /ɛ/-/æ/	1	424.113	.627	<.001	1	59.429	.191	<.001
/i/-/ɪ/ vs. /u/-/ʊ/	1	1011.622	.801	<.001	1	12.752	.048	<.001
/ɪ/-/eɪ/ vs. /ɛ/-/æ/	1	4.129	.016	.043	1	7.969	.031	.005
/ɪ/-/eɪ/ vs. /u/-/ʊ/	1	84.272	.251	<.001	1	71.538	.221	<.001
/ɛ/-/æ/ vs. /u/-/ʊ/	1	125.687	.333	<.001	1	127.238	.336	<.001

Procedure

Both Mandarin and English speakers participated in this experiment. In this AXB discrimination test, subjects listened to three words in a sequence, and were then asked to decide whether the first two or the last two words were the same. They were asked to press “F” on the keyboard if they considered the first two words the same; “J” if the last two were the same. Mandarin speakers participated in this experiment in a quiet conference room of a local middle school. Due to time and venue limit, subjects participated in groups of ten, working simultaneously on separate laptops. English speakers participated in this experiment in the sound booth in the phonetics lab of University of Colorado Boulder. For both groups, sounds were played through headphones and the instructions were presented on a computer screen using the Psychopy program (Peirce, 2007).

Evaluation. The keyboard responses and the reaction time of participants were recorded. The reaction time was recorded because the studies by Hallé, Chang, and Best (2004) and Hallé and Best (2007) suggested that when people were confident with their responses, they responded faster. Their results show that a higher discrimination score is associated with shorter reaction time. Therefore, this study recorded and analyzed the reaction time as a supplement to the discrimination accuracy results to see whether it will show the same pattern as the previous studies. The discrimination accuracy was calculated as the number of times that the participants picked out the same tokens correctly divided by the total number of trials. In other words, the discrimination accuracy was represented by the percentage of correct answers in the overall test. The keyboard was disabled until the end of the third token in the AXB triple in order to ensure that the participants have listened to all three sounds before making a keyboard response. Thus, the reaction time of each response was measured from the end of the third token in the AXB triple to the moment when a participant pressed a key. However, in AXB test, it is possible for the participants to make a decision before hearing the third sound in the triad while the design of the current experiment disables this possibility. Thus, it is possible that the reaction time measured in this study is longer than the actual time that the participants used to make a decision. However, because the major concern of this study is the difference in reaction time among different contrasts, the lengthening of the reaction time will not influence the result significantly.

Results

The discrimination test examined how English and Mandarin speakers discriminate /i/-/ɪ/, /ɪ/-/eɪ/, /ɛ/-/æ/, and /u/-/ʊ/. Table 5 presents the accuracy and reaction time for the discrimination of each English vowel contrast by English and Mandarin speakers respectively. The discrimination accuracy is the percentage of trials where a subject made correct choice out of all trials. The reaction time is calculated from the end of stimuli to the moment when subjects made a response. The reaction time is log-transformed in order to avoid violations to the assumptions of homogeneity variance and normal distribution of errors.

Table 5. Discrimination Accuracy and Reaction Time by English and Mandarin Groups

L1	Contrast	*n*	Discrimination		Reaction time		Log reaction time	
			Mean (%)	*SD*	*Mean (s)*	*SD*	*Mean*	*SD*
English	/i/-/ɪ/	30	97.812	0.030	0.583	0.156	-0.403	0.112
	/ɪ/-/eɪ/	30	97.969	0.027	0.550	0.163	-0.426	0.125
	/ɛ/-/æ/	30	97.865	0.035	0.611	0.173	-0.380	0.115
	/u/-/ʊ/	30	98.333	0.032	0.570	0.168	-0.402	0.117
Mandarin	/i/-/ɪ/	30	93.750	0.065	1.671	0.548	0.064	0.132
	/ɪ/-/eɪ/	30	71.406	0.138	2.483	0.864	0.233	0.145
	/ɛ/-/æ/	30	75.260	0.156	2.525	1.097	0.218	0.168
	/u/-/ʊ/	30	92.865	0.074	1.566	0.496	0.053	0.123

The results are divided by language and contrast, creating eight groups in total (4 contrasts * 2 languages). The discrimination accuracy and reaction time of those eight groups were submitted to a two-way repeated measure of ANOVA test, examining whether contrast and native language affect discrimination results.

Discrimination Accuracy. ***Effect of Language.*** A two-way between (language: Mandarin vs. English) by within (4 contrasts) ANOVA reveals a significant main effect of language; not surprisingly, English speakers discriminate English vowels more accurately than Mandarin speakers ($F(1, 58) = 71.827$, $\eta^2 = .553$, $p < .001$).

Effect of Contrasts. 12 planned comparisons were conducted to compare the effect of contrast within each language group. The dependent variable of those post-hoc comparisons is the discrimination accuracy score. In order to avoid Type I error, the alpha level was adjusted to .004 (.05/12) in comparisons using a Bonferroni adjustment. There is no significant difference in discrimination between any two contrasts for English participants. For Mandarin participants, /i/-/ɪ/ and /u/-/ʊ/ were better discriminated than /ɪ/-/eɪ/ and /ɛ/-/æ/. There is no significant difference between /i/-/ɪ/ vs. /u/-/ʊ/ or /ɪ/-/eɪ/ vs. /ɛ/-/æ/. The discrimination accuracy for each contrast by each language group are illustrated in Figure 1. The statistics of each planned discrimination accuracy score comparison is presented in Table 6.

Figure 1. Discrimination accuracy by contrast and group

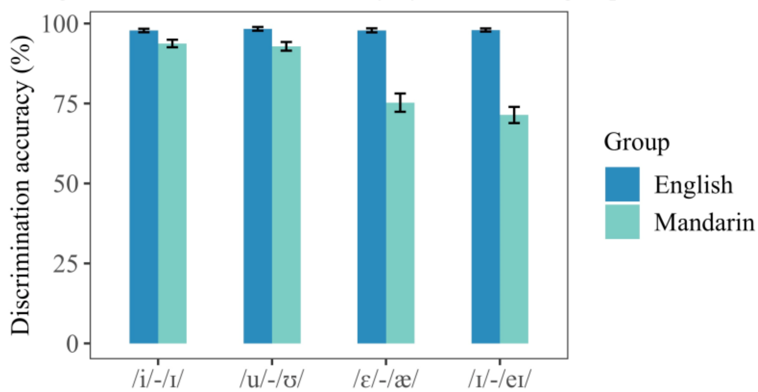


Table 6. *Post-hoc tests of discrimination accuracy score ANOVA*

Language	Comparison	Discrimination Accuracy						
		df	b[1]	CI_2.5	CI_97.5	F	η^2	p
English	/i/-/ɪ/ vs. /ɪ/-/eɪ/	1	-.002	-.033	.030	0.010	<.001	.920
	/i/-/ɪ/ vs. /ɛ/-/æ/	1	-.001	-.032	.031	0.001	<.001	.973
	/i/-/ɪ/ vs. /u/-/ʊ/	1	-.005	-.020	.010	0.476	.008	.493
	/ɪ/-/eɪ/ vs. /ɛ/-/æ/	1	.001	-.037	.040	0.003	<.001	.957
	/ɪ/-/eɪ/ vs. /u/-/ʊ/	1	-.004	-.035	.028	0.053	.001	.819
	/ɛ/-/æ/ vs. /u/-/ʊ/	1	-.005	-.036	.027	0.088	.002	.768
Mandarin	/i/-/ɪ/ vs. /ɪ/-/eɪ/	1	.223	.192	.255	206.037	.780	<.001
	/i/-/ɪ/ vs. /ɛ/-/æ/	1	.185	.154	.216	140.897	.708	<.001
	/i/-/ɪ/ vs. /u/-/ʊ/	1	.009	-.006	.024	1.376	.023	.246
	/ɪ/-/eɪ/ vs. /ɛ/-/æ/	1	-.039	-.077	.000	4.016	.065	.050
	/ɪ/-/eɪ/ vs. /u/-/ʊ/	1	-.215	-.246	-.183	183.223	.760	<.001

/ɛ/-/æ/ vs. /u/-/ʊ/	1	-.176	-.208	-.144	124.211	.682	<.001

[1] The b value denotes the estimated difference in the discrimination accuracy between two contrasts. For example: the estimated difference between /i/-/ɪ/ vs. /ɪ/-/eɪ/ is the discrimination accuracy of /i/-/ɪ/ subtracting the discrimination accuracy of /ɪ/-/eɪ/.

Effect of Acoustic Similarity. In order to examine whether the variations in the Mandarin speakers' discrimination scores among contrasts were affected by the Euclidean distance and duration difference, the Mandarin speakers' discrimination scores were regressed on the Euclidean distances, duration differences, and their interactions in a linear mixed model (Discrimination accuracy percentage ~ Euclidean distance * Duration Difference). All the effects were taken as random at the participant level in order to control the correlation within each participant.

The result shows that for the main effect of duration difference is non-significant. For contrasts with an average Euclidean distance, the duration difference does not have a significant effect on the discrimination score, given b = 0.536, $F(1, 73.04) = 0.084$, $p = .765$. The main effect of Euclidean distance is significant. For contrasts with an average duration difference, the Euclidean distance is positively related to this discrimination score, b = 1.64, $F(1, 58) = 30.173$, $p < .001$. A 10-unit increase in the Euclidean distance predicts a 1.64% increase in the discriminate accuracy. There is a significant interaction between Euclidean distance and duration difference, b = 1.49, $F(1, 58) = 21.197$, $p < .001$. As the duration difference becomes larger, the effect of Euclidean distance becomes stronger. The effect of Euclidean distance become 1.49% more positive per 10-unit increase in Euclidean distance as the duration difference increases by .01 second. For example, when the duration difference between two contrastive sounds is 10ms longer than the mean level, a 10-unit increase in their Euclidean distance predicts 3.13% (1.64% + 1.49%) increase in the discrimination accuracy.

Reaction Time. In general, a higher discrimination accuracy was associated with shorter reaction time, which was in accordance with previous studies (Hallé et al., 2004; Hallé & Best, 2007). The reaction time was log transformed in order to satisfy the normality distribution assumption of linear regression. On average, the log reaction time of English speakers in the discrimination test is shorter than that of Mandarin speakers ($F(1, 58) = 305.061$, $\eta^2 = .84$, $p < .001$). Contrasts with a higher discrimination accuracy were responded more quickly. Within the Mandarin group, /i/-/ɪ/ and /u/-/ʊ/ contrasts have significantly shorter reaction time than /ɪ/-/eɪ/ and /ɛ/-/æ/. There is no significant difference between /i/-/ɪ/ vs. /u/-/ʊ/ or /ɪ/-/eɪ/ vs. /ɛ/-/æ/. The only exception is that English listeners responded to the /ɪ/-/eɪ/ contrast more quickly than the /ɛ/-/æ/ contrast ($F(1, 58) = 9.672$, $p = .003$). The reason might be that /ɪ/-/eɪ/ contrasted a monophthong with a diphthong, creating a more significant the duration difference than other contrasts, which may accelerate the listeners' responses. The mean reaction times are graphed in Figure 2. The F and p statistics of each planned reaction time comparison are presented in Table 7.

Figure 2. Reaction time by contrast and group

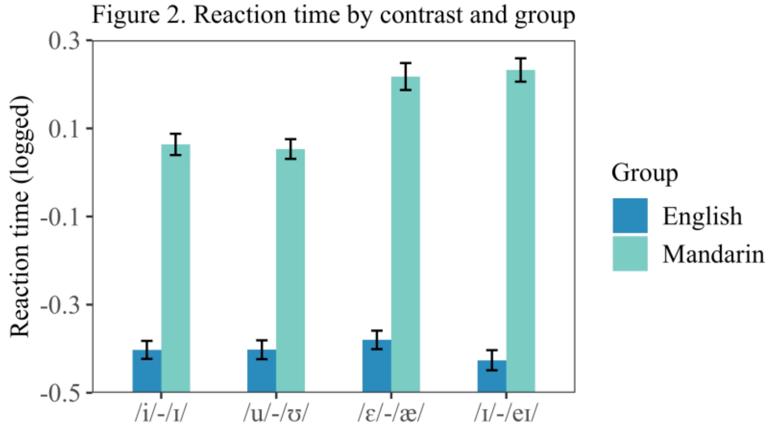


Table 7. Post-hoc tests of reaction time ANOVA

Language	Comparison	Log Reaction Time						
		df	*b*[1]	*CI_2.5*	*CI_97.5*	*F*	η^2	*p*
English	/i/-/ɪ/ vs. /ɪ/-/eɪ/	1	.023	-.011	.058	1.882	.031	.175
	/i/-/ɪ/ vs. /ɛ/-/æ/	1	-.023	-.053	.007	2.271	.038	.137
	/i/-/ɪ/ vs. /u/-/ʊ/	1	.000	-.021	.020	<0.001	<.001	.976
	/ɪ/-/eɪ/ vs. /ɛ/-/æ/	1	-.046	-.076	-.016	9.672	.143	.003
	/ɪ/-/eɪ/ vs. /u/-/ʊ/	1	-.024	-.056	.009	2.149	.036	.148
	/ɛ/-/æ/ vs. /u/-/ʊ/	1	.022	-.007	.052	2.268	.038	.138
Mandarin	/i/-/ɪ/ vs. /ɪ/-/eɪ/	1	-.169	-.203	-.135	97.456	.627	<.001
	/i/-/ɪ/ vs. /ɛ/-/æ/	1	-.154	-.184	-.124	104.673	.643	<.001

/i/-/ɪ/ vs. /u/-/ʊ/	1	.010	-.010	.031	1.016	.017	.317
/ɪ/-/eɪ/ vs. /ɛ/-/æ/	1	.015	-.015	.045	0.996	.017	.323
/ɪ/-/eɪ/ vs. /u/-/ʊ/	1	.179	.147	.212	122.169	.678	<.001
/ɛ/-/æ/ vs. /u/-/ʊ/	1	.165	.135	.194	122.611	.679	<.001

[1] The b value denotes the estimated difference in reaction time between two contrasts. For example: the estimated difference between /i/-/ɪ/ vs. /ɪ/-/eɪ/ is the reaction time of /i/-/ɪ/ subtracting the reaction time of /ɪ/-/eɪ/.

Experiment 2: Assimilation identification

This experiment tested how Mandarin speakers categorized English vowels into Mandarin vowel categories.

Stimuli. The English stimuli were the same set used in Experiment 1. Because Mandarin does not have obstruents in codas, English tVd tokens were clipped before the stop closure, converting them into tV form. The reason why the English stimuli were not elicited in tV form was that English does not allow lax vowel (i.e. ɛ, ɪ, æ) to be in open syllable (Ladefoged & Johnson, 2014). At the same time, only target English syllables with a low F0 were included in this test in order to match tone 4 (falling tone) in Mandarin.

There are disagreements on the vowel inventory of Mandarin Chinese. This study adopted the inventory listed in the book *Modern Mandarin* (Huang & Liao, 1997) which claims there are 7 monophthongal vowels in Mandarin /a, o, ɤ, ɛ, i, u, y/, 9 diphthongs /ai, ei, ɑu, ou, ia, iɛ, ua, uo/, and 4 triphthongs /iɑu, iou, uai, uei/. As mentioned above, possible native assimilation categories of a given non-native sound are defined as native sounds that differ from the non-native sound in no more than one articulatory feature (height, backness, and roundness). Based on that definition, the possible assimilation options for English /i, ɪ, eɪ, ɛ, æ/ were narrowed into four and the options for English /u, ʊ/ into two. The correspondences between English vowels and their possible counterparts in Mandarin are listed in Table 8.

Table 8. The assimilation options for each target English vowel

English vowel	Mandarin vowel

/i/[1]	/i/; /iɛ/; /ei/; /ai/
/ɪ/	/i/; /iɛ/; /ei/; /ai/
/eɪ/	/i/; /iɛ/; /ei/; /ai/
/ɛ/	/i/; /iɛ/; /ei/; /ai/
/æ/	/i/; /iɛ/; /ei/; /ai/
/u/[2]	/u/; /ou/
/ʊ/	/u/; /ou/

[1] Based on the criterion as possible native assimilation category, Mandarin /y/ should also be included as an assimilation option for English /i/.

[2] Mandarin /y/ and /uo/ should also be included as an assimilation option for English /u/.

These correspondences explain why t__d was selected as the particular environment in Experiment 1. First, all tV combinations in tone four except /tei/ are actual Mandarin words. And this is the syllable environment that could elicit the maximum number of real Mandarin words. Using real Mandarin words might better invoke subjects' memory of the corresponding Mandarin pronunciation.

Procedure. The subjects of Experiment 2 are the same thirty Mandarin subjects who participated in Experiment 1. During the experiment, participants listened to the English non-words. Then they were shown several Mandarin words that they could probably categorize the English syllables as (as shown in Table 9), and asked to choose which word they thought they just heard by pressing the number of that option on the keyboard. During the experiment, the Mandarin candidates were presented in both Chinese characters and *pinyin* (the Romanized transcription system of Mandarin). The experiment consisted of eight different tokens for each English vowel. There were 56 trials (7 vowels * 8 tokens) in total. The instructions and trials were presented on a computer screen. The stimuli were played through headphones. The subjects made responses using a keyboard.

Table 9. English non-words and possible Mandarin options

English Non-Word	Mandarin Words

	(Presented in an order of Mandarin character – pinyin transcription – IPA transcription)
/ti/	1. 替 (tì) /ti/ 2. 帖 (tiè) /tiɛ/ 3. tèi /tei/ 4. 太 (tài) /tai/
/tɪ/	1. 替 (tì) /ti/ 2. 帖 (tiè) /tiɛ/ 3. tèi /tei/ 4. 太 (tài) /tai/
/teɪ/	1. 替 (tì) /ti/ 2. 帖 (tiè) /tiɛ/ 3. tèi /tei/ 4. 太 (tài) /tai/
/tɛ/	1. 替 (tì) /ti/ 2. 帖 (tiè) /tiɛ/ 3. tèi /tei/ 4. 太 (tài) /tai/
/tæ/	1. 替 (tì) /ti/ 2. 帖 (tiè) /tiɛ/ 3. tèi /tei/ 4. 太 (tài) /tai/
/tu/	5. 兔 (tù) /tu/ 6. 透 (tòu) /tou/
/tʊ/	5. 兔 (tù) /tu/ 6. 透 (tòu) /tou/

Results. In this experiment, if subjects selected, for example, the Mandarin word [ti] when hearing English [ti], they were assimilating English /i/ (E/i/) into Mandarin /i/ (M/i/). For every target English vowel, the percentage that it was assimilated into each Mandarin vowel category was calculated. For example, 90.8% for the E/i/-M/i/ pair means that out of all the identifications that participants made for E/i/, 90.8% of time they selected M/i/ as the most similar Mandarin vowel to E/i/. The assimilation percentages of target English vowels into Mandarin categories are presented in Table 11 along with the goodness rating which will be introduced in Experiment 3.

Criterion of Categorization. Instead of using an arbitrary threshold when deciding whether one sound is categorized or not, this study adopted a new method, as mentioned above. For a target sound, if only one label received above chance percentage, that sound is one-category-assimilated. If more than one label received above chance percentage, that sound is multiple-category-assimilated. If there is no label having a significant above chance percentage, that sound is uncategorized. For English /i ɪ eɪ ɛ æ/, the chance level is 25% because each of them all have four options to assimilate into. For English /u ʊ/, the chance level is 50% because each of them has two options to assimilate into.

The advantage of this criterion is that the threshold of categorization is less arbitrary, and multiple-category assimilation becomes possible. In order to figure out the categorization of the attested English vowels, each assimilation percentage is compared with the chance level using t-tests. Using this criterion, in this study, English /i/ is categorized into Mandarin /i/, $t = 31.798$, $df = 29$, $p < .001$. English /eɪ/ is categorized into Mandarin /ei/, $t = 9.146$, $df = 29$, $p < .001$. English /ɪ/ is multiple-uncategorized into /iɛ/ and /ei/ significantly above chance (English /ɪ/ to Mandarin /iɛ/: $t = 2.943$, $df = 29$, $p = .006$; English /ɪ/ to Mandarin /ei/: $t = 6.014$, $df = 29$, p < .001). English /ɛ/ and /æ/ are both categorized into Mandarin /ai/ (English /ɛ/ to Mandarin /ai/: $t = 13.453$, $df = 29$, $p < .001$; English /æ/ to Mandarin /ai/: $t = 51.177$, $df = 29$, $p < .001$). English /u/ is categorized into Mandarin /u/, $t = 7.495$, $df = 29$, $p < .001$. English /ʊ/ is categorized into Mandarin /ou/, $t = 61.255$, $df = 29$, $p < .001$.

Experiment 3: Exemplar Goodness Rating

As with Experiment 2, this experiment only involved the Mandarin participant group who have participated in Experiment 1 and 2. Subjects were asked to rate the similarity between a given English word and several Mandarin words. The purpose of this experiment was to reveal the goodness of a given English vowel in a given Mandarin category through rating scores.

Stimuli. The English stimuli in Experiment 3 were the same set used in Experiment 2. The Mandarin stimuli were the same Mandarin words used in Experiment 2. The difference was that the subjects had audio access to the Mandarin stimuli in Experiment 3. Mandarin stimuli were produced by the author, a native Mandarin speaker coming from the same dialect region as the participants. Each Mandarin word was presented 10 times. The F1 and F2 of the Mandarin vowels in Experiment 2 stimuli were measured using Praat (Boersma, 2001). The F1 and F2 of monophthongs were measured at vowel midpoint. In diphthongs, the nucleus and off-glide were measured separately. The results are presented in Table 10. The F1 and F2 frequencies of the Mandarin stimuli were plotted in Figure 3 together with the English stimuli used in all three experiments.

Figure 3. Vowel plot of English and Mandarin stimuli

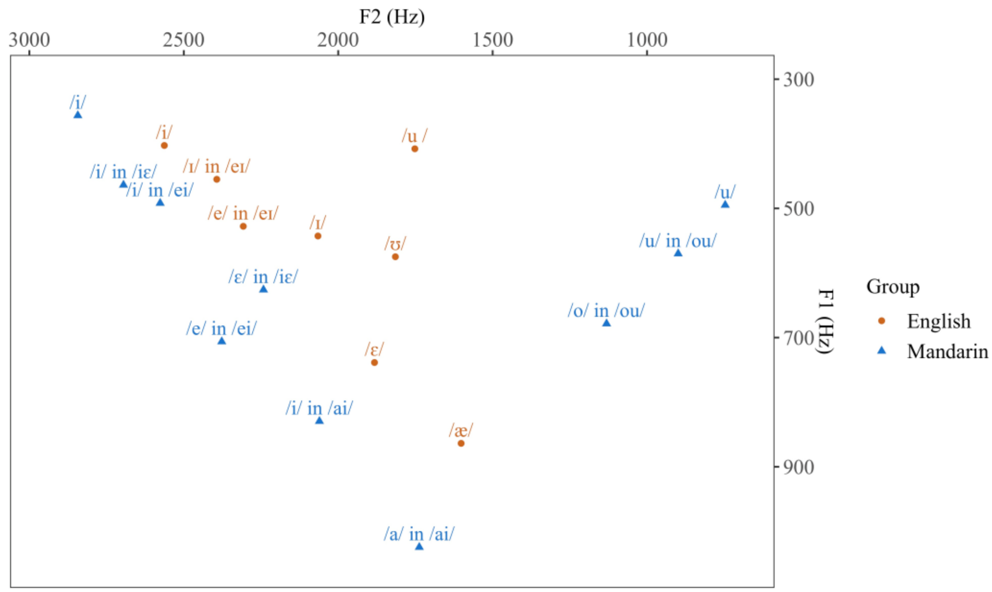


Table 10. Means and standard deviations for F1, F2, and duration of Mandarin vowels

Language	Sound	*n*	F1 (Hz)		F2 (Hz)		Duration (s)	
			Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Mandarin	/i/	8	356.250	21.907	2843.125	68.811	0.147	0.026
	/i/ in /iɛ/[1]	8	463.750	13.446	2695.250	48.902	0.164	0.026
	/ɛ/ in /iɛ/	8	626.000	30.393	2242.000	50.878		
	/e/ in /ei/	8	706.250	33.465	2377.125	62.256	0.124	0.006
	/i/ in /ei/	8	491.875	37.479	2576.625	79.816		
	/a/ in /ai/	8	1024.500	60.321	1737.500	74.682	0.146	0.025
	/i/ in /ai/	8	829.500	78.569	2060.875	90.430		
	/u/	7[2]	494.857	16.077	747.000	45.273	0.096	0.019
	/o/ in /ou/	8	678.500	28.835	1131.125	51.388	0.134	0.020
	/u/ in /ou/	8	570.125	27.885	899.250	46.438		

[1] /iɛ/, /ai/, and /ou/ are split into nucleus and offglide respectively when measuring vowel formants in order to reflect formant changing within the diphthongs.
[2] One token of /u/ was excluded because its information cannot be captured by Praat

Procedure. Subjects listened to one English non-word and one Mandarin word in each trial. Then they were asked to rate how similar those two words were using a scale from 1 (very different) to 9 (very similar). The order of English and Mandarin stimuli was balanced. Each English non-word and Mandarin word was tested by eight different tokens. There were 192 trials (5 English vowels * 4 Mandarin options * 8 tokens + 2 English vowels * 2 Mandarin options * 8 tokens) in this experiment in total. The instructions and questions were presented on a computer screen using the Psychopy program. Subjects listened to the stimuli through headphones and responded by keyboard. Instead of the canonical category goodness rating, this study used the exemplar goodness rating. Those two types of rating serve the same purpose. Category goodness rating measures the prototypicality of a token in a given category. This exemplar goodness rating task asks the listeners to rate the similarity between a given token and the prototype of a given category, which eventually reflects the prototypicality of that token in the category. We claim that the Mandarin exemplars in the stimuli is close to the prototype of their corresponding categories for two reasons. First, the speaker is from the same dialect region as the participants and produced the Mandarin stimuli in careful speech. Second, adults' perception of the prototypical tokens of the native phoneme categories were fairly consistent (i.e. they generally agreed on which token is the prototype of the phoneme) (Davis & Kuhl, 1992; Iverson & Kuhl, 1996; Kuhl, 1991). This reasoning is supported by Iverson and Kuhl (1996). They tested the relationship between exemplar similarity rating and goodness rating, finding that exemplar similarity was related to goodness: listeners rated the tokens more similar to each other when they perceived both of them as good exemplars of the same category. Therefore, the results of the current exemplar goodness rating task should be generally equivalent to the results of a category goodness rating: a higher exemplar goodness rating score indicates that a greater degree of prototypicality in the category. We adopted an exemplar test rather than a category goodness rating task because exemplar test provides concrete references for participants in the comparison task.

Results. The goodness score (in parenthesis) of each attested English vowel into each attested Mandarin category is presented together with the assimilation percentage in Table 11.

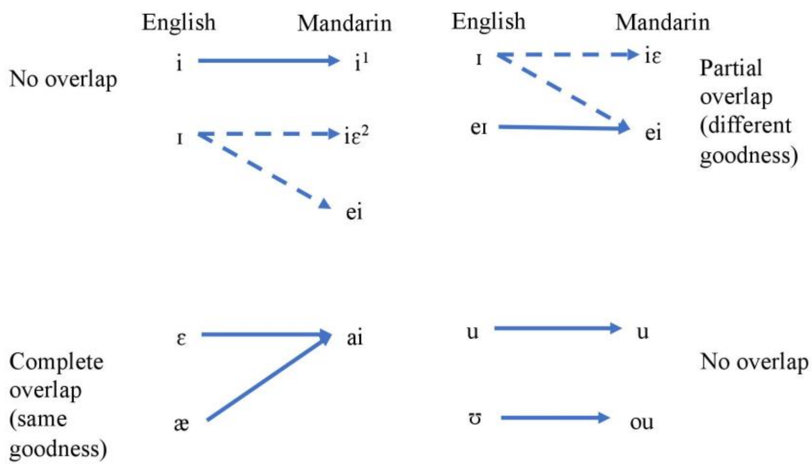
Table 11. Assimilation percentage and goodness (out of 9) of English into Mandarin

Mandarin(M)	M/i/	M/iɛ/	M/ei/	M/ai/	M/u/	M/ou/

English(E)						
E/i/	90.833	8.750	0.417	0.000		
	(7.504)	(4.775)	(3.617)	(3.208)		
E/ɪ/	10.000	35.833	51.667	2.500		
	(4.371)	(4.738)	(5.942)	(4.358)		
E/eɪ/	15.833	22.083	60.833	1.250		
	(4.183)	(4.250)	(7.275)	(4.104)		
E/ɛ/	0.833	4.167	17.500	77.500		
	(3.600)	(4.133)	(5.250)	(6.983)		
E/æ/	0.000	1.250	2.917	95.833		
	(3.354)	(3.396)	(4.358)	(7.338)		
E/ʊ/					20.000	80.000
					(4.942)	(5.779)
E/u/					98.333	1.667
					(7.083)	(4.667)

According to the results of Experiment 2, English /eɪ/ was assimilated into Mandarin /ei/ while English /ɪ/ was partially assimilated into Mandarin /ei/ as well. Thus, the goodness scores of English /ɪ/ and /eɪ/ in Mandarin /ei/ were submitted to a paired t-test. The result shows that the average goodness score of English /eɪ/ is 1.333 points higher than English /ɪ/, $t = 4.709$, $df = 29$, $p < .001$. In addition, English /ɛ/ and /æ/ were assimilated into the same Mandarin category /ai/. The average goodness rating scores of English /ɛ/ and /æ/ in Mandarin /ai/ by each person were submitted to a paired t-test. The result showed that there was no significant difference in goodness score between /ɛ/ and /æ/, $t = 1.783$, $df = 29$, $p = .085$. Combining the results from Experiments 2 and 3, the assimilation patterns of each attested contrasts are illustrated in Figure 4.

Figure 4. The assimilation pattern of contrasts



[1]Solid line indicates that the sound was categorized into a single category.
[2]Dashed line indicates that the sound was categorized into multiple categories.

Overlap Score

With the results of the assimilation test, the overlap score of each English contrast attested was obtained (Table 12). As introduced above, overlap score is defined as the smaller assimilation percentage when the components of a contrast were labeled as the same L1 category. The total overlap score of a contrast is the sum of the overlap score of all the L1 assimilation categories of the contrastive sounds. For example, given that the assimilation percentage from English /i/ to Mandarin /i/ was 90.833% and the assimilation percentage from English /ɪ/ to Mandarin /i/ was 10%, the overlap score of English /i/-/ɪ/ contrast in Mandarin /i/ was 10%. The total overlap score of English contrast /i/-/ɪ/ was 10% (M /i/) + 8.5% (M /iɛ/) + 0.417 (M /ei/) = 19.167%.

Table 12. Overlap Score of the English Contrasts

Contrast	Overlap score
/i/-/ɪ/	19.167%
/ɪ/-/eɪ/	85.000%
/ɛ/-/æ/	81.667%
/u/-/ʊ/	21.000%

General Discussion

The current study aims to improve the accuracy of predicting discrimination by assimilation. The assimilation status of a given contrast is represented both by overlap score and the assimilation pattern (complete overlap; partial overlap; non-overlap; overlap with or without category goodness difference). In addition, we determine whether discrimination patterns are predicted better by assimilation vs. acoustic similarity. In this discussion session, the effectiveness of assimilation pattern, overlap score, and acoustic similarity will be analyzed and compared.

Assimilation Pattern

The assimilation and exemplar goodness tasks showed that the English /i/-/ɪ/ contrast and the /u/-/ʊ/ contrast did not have overlap in their assimilation categories, meaning that their discrimination accuracy should be the highest. This is confirmed by the discrimination results. The /ɪ/-/eɪ/ contrast had partial overlap in the assimilation categories, while the /ɛ/-/æ/ contrast showed complete overlap. In addition, there was a significant category goodness difference within the /ɪ/-/eɪ/ contrast, but no difference within the /ɛ/-/æ/ contrast. We hypothesized that the discrimination of /ɪ/-/eɪ/ should be better than that of /ɛ/-/æ/. However, results showed that the discrimination scores did not differ significantly. This is likely due to our using a threshold for categorization: in this study, a sound is "categorized" if it is assimilated into at least one L1 category significantly above chance. English /eɪ/ was assimilated into Mandarin /iɛ/ 22.083% of the time (with the chance level set to 25%), meaning that Mandarin /iɛ/ was not considered as an assimilation category for English /eɪ/. However, if the Mandarin /iɛ/ had been counted as an assimilation category, there would have been a complete overlap in the assimilation category of the English /ɪ/-/eɪ/ contrast. Consequently, the discrimination accuracy of /ɪ/-/eɪ/ contrast would have been predicted to be the same as the /ɛ/-/æ/ contrast. This reasoning reflects that the borderline problem of PAM

(Harnsberger, 2001) still exists in the current model: when there is a cut-off categorization line and an L1 category receives an assimilation percentage slightly below the border, it is unclear whether it should be counted as an assimilation category or not.

One of the limitations of the current study is the choice of the potential assimilation options. The categorization criterion is decided by whether a category receives an assimilation percentage above chance and the chance level is decided by how many candidates the subjects have to choose from. Thus, the more candidates there are, the lower the chance level will be, and the lower the assimilation threshold will be. In order to avoid over-assimilation, the chance level has to be controlled by eliminating unlikely assimilation options. This study minimizes the assimilation candidates by assuming that the subjects will not assimilate the target non-native sound into native sounds which differ from it in more than one articulatory feature. However, such an assumption is also somewhat arbitrary, and it is likely that the subjects perceive foreign speech sounds as sounds that are not included in the candidates provided in the experiment. For example, in the current study, the mid-back English /ʊ/, which is not always distinctively rounded in American English, could probably be perceived as a mid-back unrounded Mandarin /ɤ/. However, this candidate was not included in the assimilation test. In addition, several valid options were missing in the assimilation task due to the deficiency in the experiment design (Mandarin /y/ for English /i/; Mandarin /y/ and /uo/ for English /u/). Future studies should therefore ideally conduct a pilot assimilation test where the entire L1 vowel inventory is provided, and select the assimilation candidates based on the results of the pilot test.

Although it is a limitation of this study, the absence of one or two possible assimilation categories does not falsify the conclusions. The mismatch between the prediction and the outcome is in the English contrast /ɪ/-/eɪ/. The aforementioned possible assimilation categories were not their assimilation candidates. The addition of those categories will not change the assimilation pattern of /ɪ/-/eɪ/. In addition, no matter which assimilation categories were included, the borderline issue would still persist: L1 categories that might be important to the prediction will be ignored if it receives an assimilation percentage slightly under the chance level. The mismatch between prediction and the results in the current study indicates that it is generally problematic to have a hard cut-off line when determining whether a sound is categorized or not. Thus, it might be more accurate to discard the threshold in general, and instead use the overlap score to represent the assimilation status of the contrasts.

Overlap Score

The overlap scores of the four English contrasts ranked as: /ɪ/-/eɪ/ > /ɛ/-/æ/ > /u/-/ʊ/ > /i/-/ɪ/. Their raw mean discrimination scores were in the exact reverse order: /ɪ/-/eɪ/ < /ɛ/-/æ/ < /u/-/ʊ/ < /i/-/ɪ/. The prediction that the higher the overlap score is, the lower the discrimination accuracy will be is therefore confirmed.

The overlap score overcome two limitations of PAM: overlooking the assimilation categories of the uncategorized sounds, and having an arbitrary and absolute cut-off threshold for categorization. Instead of discarding the L1 categories receiving relatively small percentage, the overlap score accounts for all the categories where the members of a given contrast overlap. Its advantage can be shown by comparing the prediction-by-assimilation pattern with the prediction-by-overlap score. The overlap between English /ɪ/ and /eɪ/ in Mandarin /iɛ/ was overlooked in the assimilation pattern because the assimilation percentage of English /ɪ/ to Mandarin /iɛ/ did not reach the chance level. In contrast, this overlap was captured by the overlap score. In addition, while previous studies (Faris et al., 2018) argued that overlap score is likely to overestimate the degree of overlap, the current study has mitigated this problem by reducing the number of assimilation category options in the identification task. In addition, having a category goodness difference did not make /ɪ/-/eɪ/ contrast more discriminable, indicating that category goodness may not have a strong effect on discrimination. Therefore, the exclusion of category goodness in the overlap score is no longer a limitation.

Acoustic Similarity

This study compared assimilation patterns with vowel length differences and Euclidean distances in terms of their effectiveness in predicting the discrimination accuracy of non-native contrasts. The results indicate that Euclidean spectral distance has a positive relationship with discrimination accuracy, while the effect of vowel duration difference on discrimination is not significant. However, the Euclidean distance hierarchy does not align with the discrimination accuracy hierarchy among contrasts: the Euclidean spectral distance rank was /i/-/ɪ/ > /ɪ/-/eɪ/ = /ɛ/-/æ/ > /u/-/ʊ/, while the discrimination score rank was /i/-/ɪ/ = /u/-/ʊ/ > /ɪ/-/eɪ/ = /ɛ/-/æ/. Thus, this study acknowledges that the increase in Euclidean distance facilitates contrast discrimination, but Euclidean distance does not perfectly predict discrimination score. In contrast, the discrimination accuracy can be accurately predicted by the overlap score. Thus, the assimilation status represented by the overlap score is a better predictor for discrimination accuracy compared with duration and Euclidean distance.

The Significance of Using Assimilation to Predict Discrimination

The results of this study also have implications for L2 learning. Knowing that assimilation pattern can predict discrimination accuracy, the next step could be to explore how to increase discrimination accuracy in L2 learners based on assimilation patterns. According to the Speech Learning Model (Flege 1995), it is possible for L2 speakers to create a separate category for a given L2 sound; it does not have to be assimilated into an L1 category. Flege and colleagues have found that the earlier the learning is (Flege et al., 2006; Flege, MacKay & Meador, 1999; Flege, Munro & MacKay, 1995), and the longer one learns an L2 (Cebrian, 2006; Flege et al., 2006; Flege et al., 1997), the more likely it is that a separate phonetic category would be established for that L2 sound. This study envisions that incorporating the L1-L2 assimilation results into laboratory training will also facilitate the category-establishment. Given the L1 categories that an L2 sound is assimilated into, we can design discrimination or identification task of the L2 sound and its most similar L1 sounds and provide learners with feedback to train them to attend to the difference between L1 and L2. For example, in order to help Mandarin listeners separate English /ɛ/ and /æ/ from Mandarin /ai/ and establish a novel category for each them, an identification task with the tokens of English /ɛ/, /æ/, and Mandarin /ai/ as the stimuli supplemented with immediate feedback will be helpful. Studies have shown that such laboratory training are effective in improving L2 sound discrimination accuracy (Wang & Munro, 1999, 2004). Consequently, speakers should be able to discriminate that L2 sound from L1 sounds and other L2 sounds better and produce them in a more native-like fashion.

Conclusions

This study proposes that assimilation is a useful tool in predicting discrimination. For two given L2 contrastive sounds, the larger the degree of overlap between them is, the more likely it is that they will be poorly discriminated. More specifically, this study suggests some modifications to the methodology measuring the degree of overlap. First, in terms of PAM, this study adjusts the criterion for deciding whether a given sound is categorized or not. Instead of using an arbitrary threshold of the categorization percentage, this study suggests using the chance level as the categorization threshold. One of the advantages of this criterion is that it is not experiment-specific, but can be applied to any study testing PAM's predictions. The other contribution of this study is that it allows multiple-category assimilation for all L2 sounds, rather than just for the "uncategorized" sounds defined in Faris et al. (2016, 2018). One of the reasons why previous studies disagreed on how accurately UC type contrasts can be discriminated is that they misclassified multiple-category assimilated sounds as uncategorized sounds and overlooked their L1 assimilation categories. Using the new criterion proposed in this study, researchers will be able to look into all the possible L1 categories that a given L2 sound is assimilated into, and use the degree of overlap in the L1 assimilation categories to predict the discrimination accuracy of a given L2 contrast. The results confirm the hypothesis that contrasts with no overlap in their assimilation categories (e.g. English /i/-/ɪ/, /u/-/ʊ/ by native Mandarin speakers) would be better discriminated than those with partial (/ɪ/-/eɪ/) or complete overlap (/ɛ/-/æ/).

Despite the improvements to PAM, the prediction of discrimination in this study based on the assimilation pattern is still not accurate enough. Due to the existence of a cut-off threshold for categorization (the chance level), L1 categories that receive a percentage at the border of the threshold might be ignored in the assimilation categories, which may lead to inaccurate predictions of the discrimination (e.g. English /ɪ/-/eɪ/ contrast in this study). However, the overlap score avoids the borderline problem. Because in this study, we reduced the assimilation options in the assimilation task by including only the mostly possible assimilation categories, the original limitation of the overlap score – overestimating the degree of overlap – has been reduced. In summary, the methodological modification made by this study does not solve the borderline limitation in PAM, but it does mitigate the overestimation of overlap score, allowing it to serve as an effective tool in predicting the discrimination of non-native contrasts.

This study also compared assimilation with acoustic cues in terms of their effectiveness in predicting the discrimination accuracy of non-native contrasts. Although in general, a higher

Euclidean distance between the members of the contrasts was associated with a higher discrimination accuracy, Euclidean did not predict the discrimination as well as the overlap score. This suggests that while speakers are sensitive to the spectral differences when discriminating non-native sounds, they may rely more on the assimilation status of the foreign sounds in the native categories.

Finally, this study is also of pedagogical significance. Knowing that the degree of overlap in L1 assimilation categories predicts discrimination difficulty, L2 educators can more easily assess which L2 contrasts will be challenging for discrimination by investigating how L2 learners map L2 sounds into their L1. It would therefore be helpful to design exercises differentiating L2 sounds from their L1 assimilation categories in order to create new phonetic categories for such L2 sounds, turning assimilation-overlapping contrasts into assimilation-non-overlapping contrasts. L2 phonetic learning could consequently be made more effective.

References

Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode?. *Journal of phonetics*, *40*(4), 582-594.

Best, C. T. (1991). The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model. *Haskins Laboratories Status Report on Speech Research*, *107–108*(July-Dec), 1-30.

Best, C. (1995). A direct-realist view of cross-language perception. Speech perception and linguistic experience: issues in cross-language research, ed. by W. Strange, 171-204.

Best, C. T., Hallé, P., Bohn, O. S., & Faber, A. (2003). Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners' native languages. *Proceedings of the 15th international congress of phonetic sciences* (pp. 2889-2892). Barcelona, Spain: Causal Productions.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775-794.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. *Language experience in second language speech learning: In honor of James Emil Flege*, *1334*.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5, 341-345.

Bohn, O. S., Best, C. T., Avesani, C., & Vayra, M. (2011). Perceiving through the lens of native phonetics: Italian and Danish listeners' perception of English consonant contrasts. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 336-339).

Department of Chinese, Translation and Linguistics, City University of Hong Kong, Hong Kong.

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*, *33*(3), 433-461.

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.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*(3), 372-387.

Davis, K., & Kuhl, P. K. (1992). Best exemplars of English velar stops: A first report. In *Proceedings of Second International Conference on Spoken Language Processing* (pp. 495-498). Alberta, Canada.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* (Doctoral dissertation). Netherlands Graduate School of Linguistics.

Escudero, P. (2009). Linguistic perception of “similar” L2 sounds. *Phonology in perception*, *15*, 152-190.

Escudero, P., & Boersma, P. (2002). The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish. In *Proceedings of the 26th annual Boston University conference on language development* (pp. 208-219). Somerville, MA: Cascadilla.

Escudero, P., & Vasiliev, P. (2011). Cross-language acoustic similarity predicts perceptual assimilation of Canadian English and Canadian French vowels. *The Journal of the acoustical society of America*, *130*(5), EL277-EL283.

Escudero, P., & Williams, D. (2011). Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners. *The Journal of the Acoustical Society of America*, *129*(1), EL1-EL7.

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 uncategorised non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics*, *70*, 1-19.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 233-277.

Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics*, *34*(2), 153-175.

Flege, J. E., Bohn, O-S. & Jang, S. (1997). Effects of Experience on Non-native Speakers' Production and Perception of English Vowels. *Journal of Phonetics*, *1997*(25), 437-470.

Flege, J. E., & MacKay, I. R. (2004). Perceiving vowels in a second language. *Studies in second language acquisition*, *26*(1), 1-34.

Flege. J. E., MacKay, I. R. A., Meador, D. (1999). Native Italian Speaker's Perception and Production of English Vowels. *Acoustical Society of America*, *106*(5). 2973-2987.

Flege, J. E., Munro, M. J., MacKay, I. R. A. (1995). Effects of Age of Second-Language Learning on the Production of English Consonants. *Speech Communication, 1995*(16). 1-26.

Fowler, C. A. (1996). Listeners Do Hear Sounds, not Tongues. *The Journal of the Acoustical Society of America*, *99*(3), 1730-1741.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: the case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, *107*(5 Pt 1), 2711-2724.

Hallé, P. A., & Best, C. T. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/clusters. *The Journal of the Acoustical Society of America*, *121*(5), 2899-2914.

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.

Harnsberger, J. D. (2001). On the relationship between identification and discrimination of non-native nasal consonants. *The Journal of the Acoustical Society of America*, *110*(1), 489-503.

Højen, A., Flege, J. E. (2006). Early Learner's Discrimination of Second-Language Vowels. *J. Acoust. Sco. Am., 119*(5). 3072-3084.

Huang, B. & Liao, X. (1997). *Xiandai Hanyu [Modern Mandarin] (2nd Edition) Volume 1* (Vol. 251, p. 256). Beijing: High Education Press.

Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of/r/and/l. *The Journal of the Acoustical Society of America*, *99*(2), 1130-1140.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57.

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.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & psychophysics*, *50*(2), 93-107.

Kuhl, P. K. (1992). Infants' Perception and Representation of Speech: Development of a New Theory. *Proceedings of the International Confer- ence on Spoken Language Processing*, 449-456.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979-1000.

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics*. Nelson Education.

Lai, Y. H. (2010). English vowel discrimination and assimilation by Chinese-speaking learners of English. *Concentric: Studies in Linguistics*, *36*(2), 157-182.

Lengeris, A. (2009). Perceptual assimilation and L2 learning: Evidence from the perception of

Southern British English vowels by native speakers of Greek and Japanese. *Phonetica*, *66*(3), 169-187.

Levy, E. S. (2009). On the assimilation-discrimination relationship in American English adults' French vowel learning. *The Journal of the Acoustical Society of America*, *126*(5), 2670-2682.

Mokari, P. G., & Werner, S. (2017). Perceptual assimilation predicts acquisition of foreign language sounds: The case of Azerbaijani learners' production and perception of Standard Southern British English vowels. *Lingua*, *185*, 81-95.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, *162*(1), 8-13.

Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., & Best, C. T. (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception, & Psychophysics*, *77*(2), 571-591.

Sisinni, B., & Grimaldi, M. (2009). "Second language discrimination vowel contrasts by adults speakers with a five vowel system", *Proceedings of 10th Annual Conference of the International Speech Communication Association* (pp. 1679-1682). Brighton, UK.

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.

Sun, L & Heuven, V. J. (2007). Perceptual assimilation of English vowels by Chinese listeners: Can native-language interference be predicted? *Linguistics in the Netherlands*, *24,* 150–161.

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.

Wang, X. (1997). The Acquisition of English Vowels by Mandarin ESL Learners: A Study of Production and Perception. Unpublished MA Thesis, Simon Fraser University.

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 effect of training on attention to temporal and spectral cues. In *Proceedings of the 14th international congress of phonetic sciences* (Vol. 3, pp. 125-128). Berkeley, CA: University of California.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4), 539-552.