



Kwon, Joy & Starr, Glenn. 2023. How L1-Mandarin L2-English learners perceive English front vowels: A phonological account. *Glossa: a journal of general linguistics* 8(1). pp. 1–36. DOI: <https://doi.org/10.16995/glossa.9282>



Open Library of Humanities

How L1-Mandarin L2-English learners perceive English front vowels: A phonological account

Joy Kwon, University of South Florida, US, joy.b.kwon@gmail.com

Glenn Starr, University of Wisconsin-Milwaukee, US, starrg@uwm.edu

Second language acquisition involves readjusting features from one's L1 onto counterparts in the L2. Learners often face difficulty during this process due to the presence of an already firmly rooted L1 grammar. Furthermore, a learner's L1 serves to constrain sensitivity to non-native contrasts during the acquisition process. If a learner's L2 grammar lacks the phonological feature that can differentiate a non-native contrast, then that learner may experience persistent difficulties in representing the L2 sounds as a result. Mandarin learners of English as a second language have to contend with a substantially expanded L2 vowel inventory in the early stages of acquisition, grappling with the addition of pronounced features less prevalent in their L1. In an attempt to account for front vowel acquisition difficulties and possible routes to progress for L1-Mandarin L2-English using a direct transfer approach, this work follows the Toronto School of contrastive phonology which holds that phonological representation is determined primarily through the ordering of contrastive features. We present data from recent phonetic research that catalogues Mandarin learners' progress in incorporating English front vowels while, at the same time, examining the underlying phonological processes. This serves as the basis for a preliminary model of contrastive hierarchy in language acquisition using elements of a feature geometry paradigm. The model provides a theoretical roadmap showing that, as Mandarin learners progress and gradually incorporate English front vowels into their L2 repository, the learner's L2 hierarchy evolves through successive stages as contrasts are perceived and categorized.

Glossa: a journal of general linguistics is a peer-reviewed open access journal published by the Open Library of Humanities. © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

OPEN ACCESS



1 Introduction

One of the core tenets of second language acquisition (SLA) research involves the idea of interlanguage grammar. When a second language (L2) learner embarks upon the path to learning a language, the system that emerges is not only the result of combining elements from their native tongue onto the newly created grammar of the L2. Nor does it only involve identifying differences between the first language (L1) and the L2 and attempting to correct errors that arise during L2 development. Though L1 transfer and error correction undoubtedly play a role, these two phenomena do not account for the entirety of the language learning process. Rather, the interlanguage generated during this undertaking consists of a unique system all its own with its own systemic rules and developmental markers.

Lardiere's (2009) Feature Reassembly Hypothesis frames this idea well. The theory states that successfully acquiring the formal features of a second language—whether they are phonological, pragmatic, syntactic, or otherwise—often involves a substantial period of time and investment wherein the learner must either readjust features from their own first language onto formal counterparts in the second language or contend with novel, unfamiliar features not encountered previously. These formal linguistic features are packaged onto lexical items of every language and are purported to come from a universal repository that all children have access to. Language learners, however, often face considerable difficulty in their attempts to regain access to that once readily available archive. In language learning scenarios later in life, the already firmly rooted features and representations in one's own native language can serve to make the acquisition process much more difficult.

In the fields of phonetics and phonology, research into L2 learners' perception and learnability has steadily increased. Many researchers have investigated the L1 influence on L2 speech perception and acquisition by examining the roles of L1-L2 similarity, phonotactic rules, and mental lexicon (Major 1987; Flege 1995; Laeufer 1996; Flege et al. 1997; Sebastián-Gallés & Soto-Faraco 1999; Sebastián-Gallés et al. 2006; Best & Tyler 2007; Johnson 2012; Sebastián-Gallés & Díaz 2012; Chang 2015). One of the common findings among the previous literature is the heavy influence of L1 background. Thus, it is paramount to understand the influence of the L1 grammar on the acquisition of a new or unfamiliar L2 inventory. Phonological features are often organized language-specifically and consist of distinct phoneme inventories and representations.

White's (1988) Transfer Hypothesis holds that aspects of the native language transfer initially and, given appropriate input, access to universal grammar (UG), and time, the L2 grammar correctly adjusts, and acquisition can be successful. Brown (1998) notes, however, that if a learner's L2 grammar lacks the phonological feature that can aid in the differentiation of a non-native contrast, then that learner may face considerable difficulty in acquiring accurate L2 representations as a result. This difficulty may persist indefinitely. Thus, detection of novel

or unfamiliar contrasts is another aspect to consider in language development. This can pose considerable difficulties to the learner. Brown's work raises several intriguing questions that are valuable to researchers in the L2 acquisition of phonology. It is worthwhile to consider, for example, how we can accurately describe acquisitional difficulties if the phonological properties of the L2 are purported to be provided by UG. Moreover, what aspects of a learner's L1 grammar contribute to successful L2 acquisition? Do phonological components in the L1 inventory play a role in the construction of an emerging interlanguage grammar and is this process one of construction, constraint, or restriction?

Many have proposed models that account for the L2 acquisition of speech sounds. For example, the speech learning model (SLM; Flege 1987; 1995; Flege & Bohn 2021) is a well-known learning-based theory of acquisition hypothesizing that acquisition of intelligible speech is an adaptive and life-long process. The core argument of this model is that the processes which resulted in the successful acquisition of the L1 sound system in early childhood remain active over the course of one's life and go on to have an influence on the acquisition and learning of L2 sound systems.

Flege's (1995) initial SLM categorized L2 sounds into three groups—new, similar, and identical—but Flege and Bohn's (2021) revised SLM-r mentions only two—new and similar. Furthermore, these groupings are categorized based on comparison with the existing L1 feature set. New L2 phones are defined as sounds that do not exist in the L1 system. Thus, they result in the creation of new L2 phonetic categories. Similar L2 phones are those that differ from the source sounds but may share certain features. In this way, a composite L1-L2 phonetic category emerges. As for identical sounds, Flege did not provide an explicit definition in the initial model of the SLM and left this category out of Flege and Bohn's SLM-r, but based on the revised new and similar definitions, we can infer that identical sounds likely refer to phones which systematically match to sounds in the L1 inventory. In short, the SLM and the updated SLM-r hypothesize that difficulties in L2 perception and production are caused mainly by discernment barriers between native and non-native sounds. It should be noted as well that, according to Flege, developmental accuracy in both perception and production of L2 sounds are often related and co-evolve. However, the degree to which the co-development process overlaps necessarily varies in relation to individual differences in proficiency. One must also strongly consider how different the target sounds are in relation to the existing L1 inventory.

One striking feature of the SLM is that it leaves bare the role of phonology in L2 acquisition. One of Flege's main hypotheses in the SLM, for example, is that L1 and L2 sounds are related perceptually to one another at an allophonic level rather than at a more abstract phonemic level. As such, phonology has little impact in L2 learning since sounds are purported to be learned

at a surface phonetic level.¹ Another position is that new phonetic categories in the L2 are possible if learners can discern phonetic differences between L1 and L2 sounds and the bigger this difference is, the better. Yet unexplained in this proposition is the degree to which learners access their phonological grammars in doing so. Is it the case that they are solely relying on phonetic information from their L1? And how dissimilar do sounds have to be in order to be discerned successfully? Adding to this, Flege argues that the creation of new categories will be slowed if the L1 and L2 sounds are closely linked perceptually. We are in agreement with this position, but argue that Flege's intuitive approach regarding perception and production of L2 sounds would be complemented by phonological analyses based on Modified Contrastive Specifications (Dresher et al. 1994) (see §2 for more details).

Along with the SLM, another commonly cited model for the acquisition of L2 sounds is Best and colleague's Perceptual Assimilation Model (PAM; Best 1994; Best et al. 1988; Best & Tyler 2007). Best's direct realist view of cross-language speech perception puts forth the claim that L2 learners will perceptually assimilate L2 phonemes to phonemes already present in their L1 and that the L1 serves to constrain one's sensitivity to non-native contrasts during acquisition. As a result, the perception patterns of a language learner are influenced most heavily by established L1 inventories and contrasts. PAM's fundamental tenet is that listeners discriminate foreign sounds by comparing the pattern of contrast assimilation. It assumes that naïve listeners (not language learners) are likely to assimilate the nonnative sound segments to the most similar L1 phoneme in terms of articulation. The PAM suggests three different categories for nonnative sounds: an example of an L1 phoneme, either good or bad (*categorized*), no match to L1 phonemes (*uncategorized*), and in rare cases, non-speech sounds (*non-assimilated*).

Best and Tyler (2007) later clarified PAM and proposed the PAM-L2 in an attempt to accommodate L2 learners' perceptual learning at both phonetic and phonological levels. In it, possible cases of speech segment perception were offered. To illustrate, one such case states that when an L2 segment has a matching L1 segment, these sounds will be considered identical. Another instance refers to L2 sounds which have comparable counterparts in the L1 inventory. In this case, learners may regard one sound as either less or more similar to an existing L1 counterpart because of nuanced non-matching properties but will still likely associate it within the analogous L1 category. The last case refers to instances wherein L2 sounds cannot be matched with any of those in the L1 inventory and since there is no possibility of L1-L2 assimilation on a phonological level, the sounds are deemed as uncategorized.

¹ Flege's (1995; 2021) SLM does state, however, that simultaneous bilinguals are able to maintain clear contrasts between their L1 and L2 sound systems throughout their lives. In order for this to be achieved, it is necessary for both languages to be continually accessed in an ongoing manner. Thus, Flege argues that continuous exposure and upkeep throughout life allows for disparate L1 and L2 sound categories to exist in a shared phonological space.

Both the SLM/SLM-r and PAM/PAM-L2 models provide criteria for how L2 sounds are perceived, and the SLM/SLM-r maintains that production of L2 sound models along with advances in perception. Similar L2 sounds have equivalent or similar counterparts in the L1 phonological grammar whereas new sounds are L2 sounds which are not present in the L1 inventory. These models offer a way to distinguish new/similar sounds, yet a major problem is that the suggested criteria neglect the continuous aspect of the new-similar system. From a learner's perspective, one may encounter completely new sounds, sort-of-new sounds, similar sounds, and sounds that are identical; however, the SLM and PAM both present a system wherein an L2 sound either maps to an L1 category (identical; perceptually assimilated) or when it does not (new; uncategorized). The grey area between these two instances is described as similar, per the SLM, and categorized, per the PAM.

In addition, the SLM/SLM-r and PAM do not capture the possibility of various levels of representation to illustrate the similarity, e.g., phonological, morpho-phonological, and phonological-phonetic levels. To illustrate, the SLM's main stance is that phonetic surface properties are more significant when it comes to perceiving and producing L2 sounds. However, the SLM's hypotheses do not clearly delineate the boundary of phonetics and phonology. Although the SLM/SLM-r and PAM have provided substantial groundwork and elaborated on new/uncategorized and similar/categorized sounds, the descriptions leave out which underlying features may impede progress. For example, it is difficult to gauge which properties specifically affect one's perception of similarity.

Thereby, one of the goals of this paper is to provide a complimentary stance of previous phonetic research (e.g., SLM, SLM-r, PAM, PAM-L2) from a phonological perspective that seeks to provide a better explanation for obstacles in L2 sound learning. To this end, this article re-examines and re-analyzes data from a selection of vowel perception studies involving L1 Mandarin Chinese (Mandarin) learners of L2 English. What follows is an explanation of the phonological architecture involved in this undertaking. Subsequent sections lay out a model for the discernment of distinctive features as well as providing a hierarchy of contrasts that help in understanding the scope and form of English and Mandarin phonology. How phonetic elements are represented in a feature geometry and set up in contrastive dimensions sets the stage for the work that must be done by L1-Mandarin L2-English learners.

Another goal is to explore two phonological approaches that predict vowel confusion patterns and account for L2 sound learnability, i.e., the featurally underspecified lexicon (FUL) model (Larihi & Reetz 2002) and the phonological-phonetic representation (Kwon 2021). We attempt to demonstrate how these phonological accounts will contribute to investigating perceptual similarity, providing phonological evidence of L2 learning, and displaying developmental stages of interlanguage grammar.

2 Phonological underpinnings for L2 perception and similarity

To understand L2 perception and subsequent difficulties learners may face in a more comprehensive manner, we argue that phonological representations can provide insightful explanations. Thus, it is imperative to look at the phonological structure and how L1 phonological features influence L2 perception. In this section, we first examine the phonological structure of both Mandarin and English.

2.1 Abstract vowel space

First, let us consider the conceptual vowel spaces of Mandarin and English. Mandarin has five monophthongal vowel phonemes, though this classification differs somewhat from researcher to researcher. For the purposes of the present work, the vowel phonemes used are those that have been presented in Duanmu (2007). They are listed as [i, y, u, ə, a], though [ɤ] is sometimes used in place of [ə] in some interpretations. The vowels [i, y, u] are categorized as high vowels and Duanmu states that [i] and [u] are ordinary high vowels that resemble the ones found in English. The remaining two are a mid [ə] and a low [a] vowel. The Mandarin vowel characterization scheme lends itself well to a triangular vowel space structure with the vowels [i, a, u].

Most studies in Chinese vowel phonology ignore length as a characteristic. Duanmu (2007) lays out the reasoning behind this characterization extensively in his work and concludes that all vowels in Mandarin, due to the language's weighted syllable structure, can be essentially classified as long and/or tensed. This will play an important role in the discussion of the L2 acquisition of English vowels.

Moving on to English, there are at least 11 monophthongs in its inventory, listed as [i, ɪ, e, ɛ, æ, u, ʊ, o, ɔ, ɑ, ʌ]. Perception and production of these vowels relies heavily on tense/lax distinctions, a contrast absent in Mandarin. The presence of tense/lax characteristics necessitates different features. According to Lai (2010), tenseness is marked by laryngeal height and when the tense vowels are produced, the root of the tongue is drawn forward (advanced tongue root) and the larynx is in a lowered position. For lax vowels, the tongue root is not advanced (retracted tongue root), and the larynx not lowered. These gestures, then, are not realized in Mandarin and may evoke difficulty in acquisition.

At first glance, the Mandarin vowel system differs from the English vowel space as Mandarin has fewer vowels and smaller space (**Figure 1**). Because of this difference in the conceptual space, we can predict that Mandarin learners will struggle with several English vowels because Mandarin only includes one unrounded high vowel, /i/, and one unrounded low vowel, /a/. In order to perceive English front vowels, learners will have to learn how to distinguish different sounds: lax and tense, mid from low.

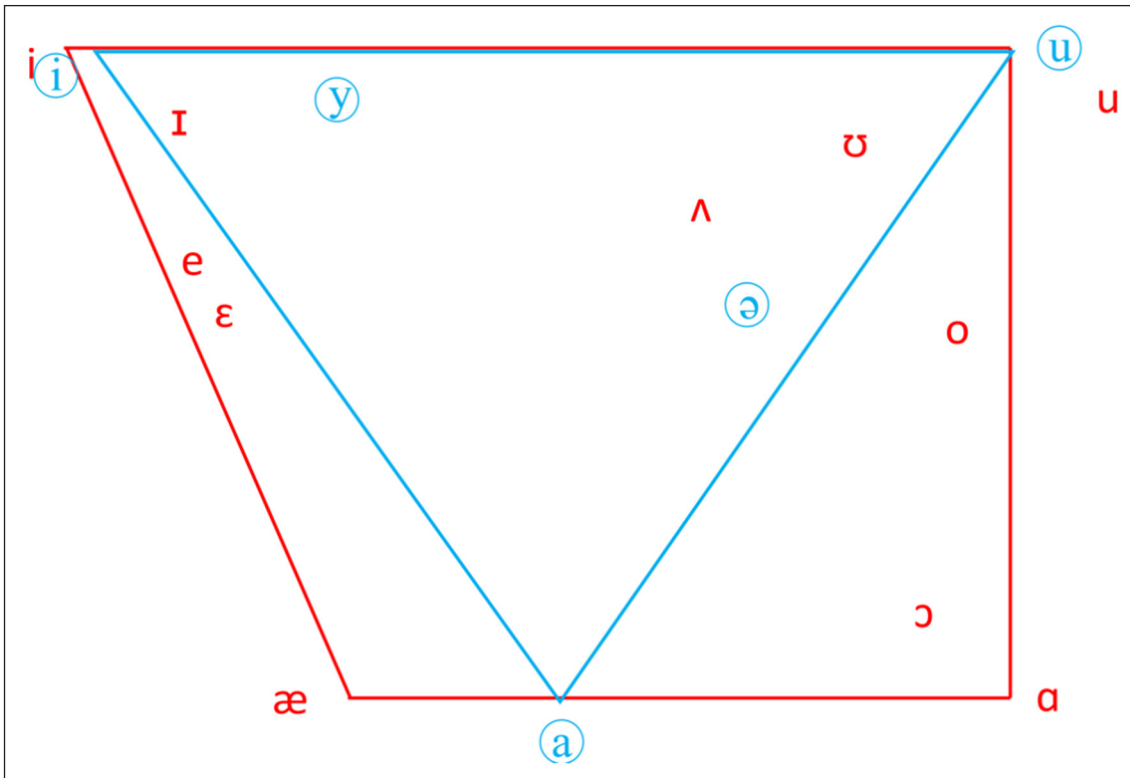


Figure 1: Conceptual vowel space of English and Mandarin phonemes (English vowels in red; Mandarin vowels in blue and circled).

2.2 Phonological-phonetic representation model

This study assumes that L1 phonology plays a pivotal role in L2 perception. In order to understand the underlying mechanism of perception, it is crucial to recognize how L1 phonology influences L2 perception. As remarked upon above, acquisition of L2 sounds can be constrained considerably due to the learner already possessing robust linguistic resources established in the L1 grammar (Archibald 2022a; 2022b; Brown 2000). Nevertheless, our approach to sound learning is akin to that referenced in Flege (1995), Flege and Bohn (2021), and more broadly in White (1988), namely that the mechanisms responsible for L1 acquisition can be applied to L2 acquisition as well because this system stays intact and evolves throughout the lifespan. With regard to L2 sound acquisition, the key distinction between our initial endeavor here and that of Flege and Bohn's SLM/SLM-r is that our approach provides a strictly phonological account for how such acquisition proceeds.

The general theory of phonological representation in this paper is founded upon Modified Contrastive Specification (MCS) developed from the Toronto School of Phonology (Dresher et al. 1994; Zhang 1996; Dresher & van der Hulst 1998, to name a few). The main tenet of the MCS

is that phonological contrasts form a hierarchical tree of ranked features. The MCS also assumes that the contrastive hierarchy is the universal resource that a language learner has access to, but not the features or the feature ordering (Dresher 2018). The feature ordering—which is language specific—is derived from the Successive Division Algorithm (SDA; Dresher 2008; 2009) and the selection and ranking of the features are determined by phonological activities of a given language (Dresher 2003; 2018). In addition, only phonologically active features are contrastive as different feature rankings can trigger different phonological activities (see Dresher 2018 and Archibald 2022a for further discussions).

Based on the MCS, Purnell and Raimy (2015) incorporated Avery and Idsardi's (2001) model of distinctive features (AI model), an articulator-based feature system, into the contrastive hierarchy. The AI model can be distinguished from other feature geometries as it introduces two sets of features, namely *dimensions* and *gestures*. The dimension is a superordinate unit that comprises a set of antagonistic phonetic gestures. The realization of each dimension is processed by L1 phonological rules named *completions*. For example, if one of the contrasts is the Tongue Height dimension in a language, it can be realized as either [high] or [low] (an antagonistic pair) at the surface level depending on the completion rule. That is, if the completion rule is Tongue Height > [high], a segment marked with Tongue Height will be completed with a [high] phonetic feature at the surface representation. Dimensions are the contrastive phonological unit that demonstrate an underlying representation while the gestures illustrate a surface level of representation. In addition to the completion rules, Avery and Idsardi introduced enhancement rules which portray redundant features (gestures, in the AI model) at the surface level. Enhanced gestures are distinguished from completed gestures as they are not expressed in the contrastive hierarchy (i.e., inert in the phonological representation) (Avery & Idsardi 2001; Stevens et al. 1986). For instance, long vowels in English are redundantly tense (Halle & Mohanan 1985: p. 73); following the AI model, tenseness is specified with one of the Tongue Root dimension's gestures, advanced tongue root (ATR). Thus, at the phonetic level of representation, English tense vowels, e.g., /i/ and /e/, will have an [ATR] gesture as an enhanced—not contrastive—feature. In short, Purnell and Raimy's model incorporating AI feature geometry and contrastive hierarchy is capable of capturing both phonological and phonetic representations and their interface. In this paper, the Purnell and Raimy contrastive hierarchy model is referred to as the phonological-phonetic representation (PR; Kwon 2021).

Table 1 summarizes the phonological feature orders of Mandarin and English vowels. The PR for Mandarin vowels is taken from Raimy (manuscript) and **Figure 2** illustrates the feature order hierarchy in the Mandarin phonological grammar. Based on Mandarin vowels' phonological rules, Raimy suggests that the feature order of Mandarin vowels is Tongue Height (TH; completed with [high]), Tongue Thrust (TT; completed with [front]), Tongue Root (TR; completed with [RTR (retracted tongue root)]), and Labial (Lab; completed with [round]). **Figure 3** demonstrates the

English long vowel hierarchy adopted from Purnell and Raimy (2015). The feature order for this group of vowels is TR (completed with [RTR]), TT (completed with [front]), and TH (completed with [high]), and Lab (completed with [round]). The feature order of short vowels is TT and TH (see Purnell & Raimy 2015; Purnell et al. 2019 for further discussion on rank order motivation and determination).

Language	Length ²	Feature hierarchy
Mandarin	long	Root > Tongue Height > Tongue Thrust > Tongue Root > Labial
English	long	Root > Tongue Root > Tongue Thrust > Tongue Height > Labial
	short	Root (> Tongue Root) > Tongue Thrust > Tongue Height (> Labial)

Table 1: Phonological feature order of Mandarin and English vowels.

2.3 Quantifying perceptual similarity

To avoid vague categorization of (dis)similarity, this section suggests a measure that can express perceptual similarity on a 0 to 1 continuous scale. Lahiri & Reetz (2002) proposed a scoring system, the featurally underspecified lexicon (FUL) model, that calculates similarity scores by comparing the number of (mis)matching features. Originally, the FUL model was proposed to illustrate L1 speech perception accounting for within- and across-speaker variations. Following the recognition process of Lahiri and Reetz, Natvig (2017) extended the FUL model to attest how it can explain variations of Norwegian vowel integration in English loanwords. The current study adopts the FUL model in an attempt to examine L2 perception and to further support the development of the FUL model within a different but related domain.

(1) FUL score formula:

$$Score = \frac{(No. of matching features)^2}{(No. of features from signal) \times (No. of features in the lexicon)}$$

The FUL model score formula presents a theoretical degree of similarity presuming that the learners perceive and unpack relevant features of L2 sounds accurately. The model operates on an analysis of features and calculates the similarity score as in (1) by comparing matching features from the signal and the lexicon ranging from 0 to 1. In terms of L2 perception, the signal will be the L2 input and the lexicon will be the L1 grammar. In other words, the scoring formula

² Length is separated from the feature hierarchy column as it is not one of the dimensions of Avery and Idsardi's (2001) feature geometry. Purnell and Raimy (2015) distinguish English vowels as long and short but the feature hierarchy for both long and short vowels is identical.

can be used to demarcate similarities on a quantitative scale, from 0 (dissimilar) to 1 (similar). That is, the FUL scores can generate hypotheses regarding easier or more challenging sounds for L2 learners and evaluate learners' perceptual difficulty of the target sounds (i.e., the higher the FUL score is, the easier the sound is to perceive).

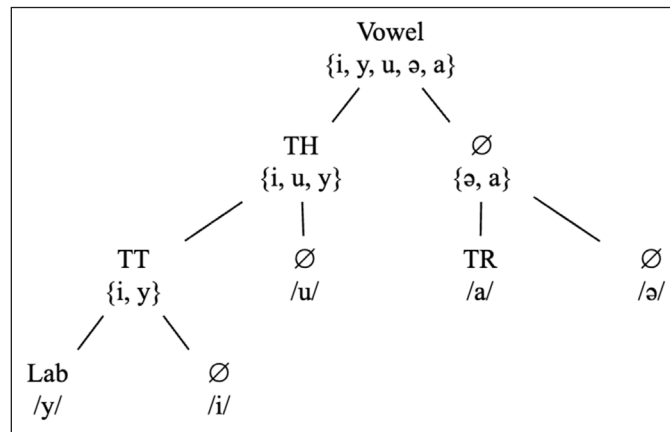


Figure 2: PR of Mandarin vowels.³

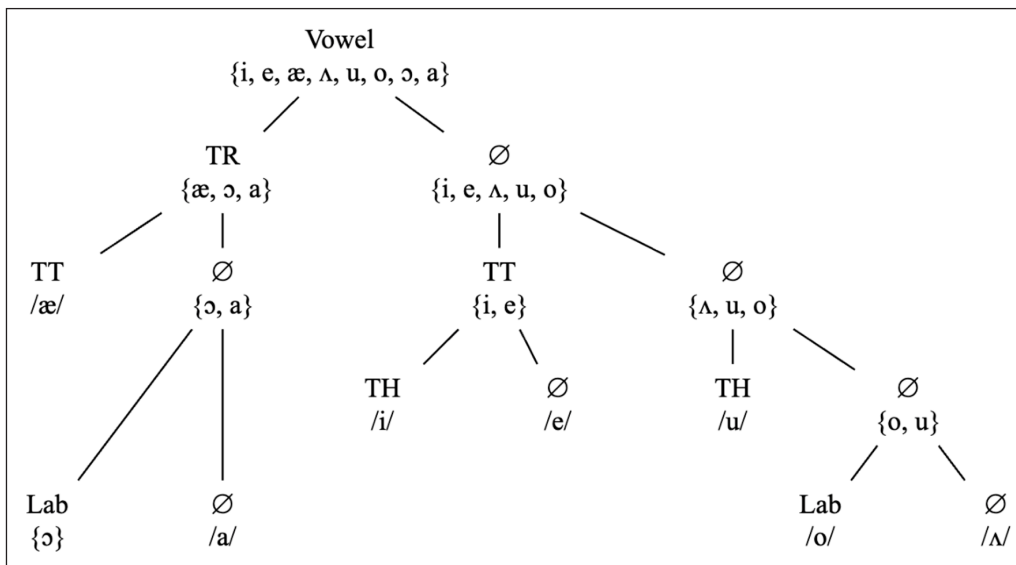


Figure 3: PR of English long vowels.

³ For brevity, Figure 2 only illustrates relevant dimensions that contrast phonemes. That is, TT and Lab are not depicted in the non-TH branch.

For example, **Table 2** demonstrates the scoring system of English [i] FUL scores of Mandarin learners. Theoretically, five possible scores can be rendered because there are five vowels in Mandarin: {i-i}, {i-y}, {i-u}, {i-ə}, and {i-a}. The first column exhibits features of each Mandarin vowel—the marked features of each Mandarin vowel and the second column displays the features of English [i]. Besides the dimensional features, [i] presents an enhancement feature [advanced tongue root (ATR)]. To compute the similarity score between Mandarin vowels and English [i], we compare the matching features of Mandarin vowels and English [i] as an example. Each combination will compare how many matching features they have in common which becomes the numerator of the FUL score formula. Thus, the pair with the highest score suggests the most likely English vowel that L1-Mandarin speakers will associate with perceptually.

	L1 features	L2 features of [i]	FUL Score
/i/	Root[vowel], TH[high], TT[front], Length[long]	Dimensions: Root[vowel], TH[high], TT[front], Length[long], Enhancement: [ATR]	$4^2/(5*4) = 0.80$
/y/	Root[vowel], TH[high], TT[front], Labial[round], Length[long]		$4^2/(5*5) = 0.64$
/u/	Root[vowel], TH[high], length[long]		$3^2/(5*3) = 0.60$
/ə/	Root[vowel], Length[long]		$2^2/(5*2) = 0.40$
/a/	Root[vowel], TR[RTR], Length[long]		$2^2/(5*3) = 0.27$

Table 2: Mandarin learners FUL scores for English [i].

The FUL predicts Mandarin /i/ to be the most perceptually similar sound to English [i] as it has the highest FUL score (0.80). The FUL score between Mandarin /i/ and English [i] is rendered by the feature analysis of both sounds. Mandarin /i/ is comprised of four features, i.e., Root, TH, TT, and Length; English [i] consists of four phonological features, i.e., Root, TH, TT, and Length. It is also enhanced with [ATR] because long English vowels are redundantly tense. Since the completion rules are identical (e.g., for both Mandarin and English, TH is completed with [high], TT is completed with [front]), there are four matching features between Mandarin /i/ and English [i]: Root, TH, TT, and Length. The number of matching features will be squared where the denominators are the number of features of each L1 and L2, four and five, respectively.

In contrast, Mandarin /a/ has the lowest FULS (0.27) which implies the lowest similarity between it and English [i]. The process of calculating this FULS is identical to the case of {i-i}. Mandarin /a/ is composed of three features, i.e., Root, TR, and Length. When compared with the features of English [i], there are two matching features, i.e., Root and Length. Thus, the

numerator is two squared and the denominators are five and three because there are five features for English [i] and three features for Mandarin [a].

Utilizing the FUL system, we can quantify the degree of perceptual similarities and predict which L1 phoneme(s) will be the most perceptually similar candidates for the target L2 sound. **Table 3** presents a comprehensive chart that shows the FULS for English and Mandarin. The English (L2) column is the signal sound, and the Chinese (L1) column is the lexicon resource. Since there are 11 English phonemic vowels and 5 Mandarin phonemic vowels, there are 55 combinations in total. Higher FULS suggest higher similarity between the compared sound segments in English and Chinese vowels. The following subsections summarize the results of the FULS analysis.

Eng Mandarin	i	ɪ	e	ɛ	æ
i	0.80	0.56	0.56	0.33	0.56
y	0.64	0.45	0.45	0.27	0.45
u	0.60	0.33	0.33	0.11	0.33
ə	0.40	0.13	0.50	0.17	0.50
a	0.27	0.08	0.33	0.11	0.75

Table 3: Perceptual similarity of English front vowels and Mandarin vowels using FUL scores (the highest score of each column is bold-faced).

The FUL model predicts what Mandarin vowel will associate with what English vowel when inputted. For example, when Mandarin learners listen to a stimulus including /i/, Mandarin /i/ will be the first candidate (0.8) that matches with the input and the next candidate vowel will be /y/ (0.64). The least similar vowel is /a/. In the same way, each possible combination is computed in the FUL equation and generates FUL scores ranging from 0 to 1. The descending order of the FUL scores implies the rank of the most to least similar vowels.

Table 3 shows the FUL scores of possible combinations between English front vowels and all Mandarin vowels. The highest score of each column indicates the most likely Mandarin vowel evoked by the L2 input. For example, the most similar Mandarin vowel elicited by English /i/ is Mandarin /i/ (0.8). The first candidate for other English vowels {ɪ, e, ɛ} is also Mandarin /i/, which is not surprising as the only unrounded front vowel is /i/ in the Mandarin inventory. However, the absolute value of a FUL score should also be considered. For instance, the highest FUL score in the /ɛ/ column is 0.33 (matched with Mandarin /i/). The lower score implies that learners will have difficulty learning English /ɛ/ in the early stages of acquisition because scores

closer to 0 indicate dissimilarity between the two sounds. This is in line with Flege's (1995) SLM and Flege and Bohn's (2021) SLM-r in that markedly different L2 sounds are initially much more difficult to assimilate than similar ones, though in the long run they are easier to acquire due to their distinctive properties.

The rank order of the highest FUL score (of each column) entails the order of perception-and/or acquisition-related difficulty. When ranked from high to low, the order becomes $i > \text{æ} > e = \text{ɪ} > \text{ɛ}$. Based on this information, we predict that English $\{i, \text{æ}\}$ are less difficult while $\{\text{ɪ}, e, \text{ɛ}\}$ are more challenging to learn.

2.4 Strong phonology transfer hypothesis

In her work, Kwon (2021) advocates the strong phonology transfer hypothesis, arguing that it can contribute to analyzing phonetic confusion patterns and delineating developmental stages of learners' interlanguage grammar. In the spirit of Brown (1998; 2000), Kwon argues that L2 learners (specifically, sequential L2 learners who have already developed their L1 grammar) will apply their phonological feature order directly to the L2 input. Thus, the strong phonology transfer hypothesis can explain learners' perceptual confusion and provide a phonological account of obstacles in L2 sound learning. For example, employing the PR, Kwon explains how L1-Korean L2-English beginner learners perceive English vowels. She claims that the hypothesis can explain vowel confusion and predict interlanguage development with more accurate representations.

This paper aims to expand the strong phonology transfer hypothesis to L1-Mandarin L2-English learners. Thereby, we demonstrate how results from previous literature (see §3) can be accounted for from a phonological viewpoint. The dimension order of Mandarin grammar is $\text{TH} > \text{TT} > \text{TR} > \text{Lab}$ (Table 1). Novice L1-Mandarin L2-English learners will utilize this representation when perceiving L2 sounds (English front vowels, in this study). This stage of L2 perception is characterized by what we refer to as the Phase 1 interlanguage grammar (IG). Because English front vowels are not processed with the target grammar (i.e., native speakers' phonological feature order), not all English vowels are correctly identified by the interlanguage grammar since each node should be occupied by a single vowel. When a node consists of more than one vowel, this causes perceptual confusion. As learners' proficiency improves, they will learn how to begin initiating contrasts between vowels that were unavailable in the Phase 1 IG. In short, L2 learners will rely on their existing L1 grammar and try to process L2 sounds filtered through their L1 grammar (i.e., strong phonology transfer hypothesis). Naturally, confusing pairs occur since the L2 sounds are not processed by the target L2 grammar. The next section discusses this issue further.

3 CasestudiesinL1-MandarinChineseL2-Englishvoweldevelopment

This section reviews perception data of English vowels by low proficiency learners whose experience is restricted to English as a foreign language (EFL) environments (Jia et al. 2006; Ho 2009; Lai 2010; Hu et al. 2019). This will be followed by examination of vowel perception with more advanced proficiencies in English or of those who have had more experience and exposure to English while living in North America (Jia et al. 2006; Xiahou 2012; Yu 2012). We selected these studies because they all test vowel perception capabilities in L1-Mandarin L2-English learners and provide insight into how front vowels are acquired over time. We were careful to select studies that measured L2 learner performance at various stages in the acquisition process. To achieve this, we report on research that specifically targets learners who are at different points in the L2 learning process regarding proficiency level. We also consider the degree of immersion in the L2, environmental setting,⁴ and individual differences as factors in our evaluation. This allowed us to gather evidence of learner progression across a range of tasks, measures, and samples and map out the process by which English front vowels are acquired in the L2.⁵ The findings reported in these studies reveal systematic difficulties in learners' ability to perceive certain front vowels in English in the early stages of acquisition. Data from these studies also reveal significant challenges among L2 learners in establishing contrasts among front vowel pairs. As proficiency improves and as learners receive more exposure to the L2, however, the findings show noticeably similar paths of development. By re-examining the data from previous work in this way, we intend to propose a phonological account that elaborates on why L1-Mandarin L2-English speakers struggle to learn certain English vowels and explains how these L2 speakers distinguish English front vowels phonologically. **Table 4** provides a summary of the research we review. The selected research in L2 perception and acquisition will be revisited in Section 4 where we apply a phonological perspective to L2 learner front vowel development.

⁴ We'd like to thank a reviewer for pointing out that, at some points in our analysis, we chose to compare learners by learning environment, noting differences in vowel perception accuracy based on whether learners were situated in either EFL or ESL learning contexts. Some authors in our review also chose to compare participating learner groups based on length of residence (LOR) in an English-speaking country and noted differences in vowel perception ability in their results based on this metric. It must be noted, however, that the degree of "immersion" in a foreign country may not be a reliable indicator of how much L2 exposure the learner receives. Some learners in EFL contexts, for example, may consume large amounts of English language media or converse with others in their L2 more often than those residing in countries where English is the dominant language. Thus, we ask readers to consider these caveats.

⁵ Due to the nature of our inquiry and because of the limited number of empirical studies that examine the specific parameters we have just outlined, some of the data we have included in our search may not be reliable due to methodological concerns and/or incomplete statistical analyses conducted by the original authors. We comment on these concerns in the relevant studies and revisit this issue in the discussion. Though we acknowledge this as a potential drawback of our analysis, the data we report are consistent with other findings in the peer-reviewed literature.

Study		Target vowels	Task	Task Voice
Ho (2009)	EFL	{i, ɪ, e, ɛ, æ}	Identification task	Native speaker of American English – male (1)
Lai (2010)	EFL	{i, ɪ, e, ɛ, æ, u, ʊ, ʌ, o, ɔ, ɑ}	AX discrimination task	Native speakers of American English – male (2)
Hu et al. (2019)	EFL	{æ, ɛ, e, i, ɪ, ɜ, ʌ, u, ʊ, o, ɔ, ɑ}	Identification and perception tasks	Native speaker of American English – female (1)
Jia et al. (2006)	EFL/ESL	{i, ɪ, e, ɛ, æ, u, ʌ, ɑ}	AXB discrimination task	Native speaker of American English – female (1)
Yu (2009)	ESL	{i, ɪ, e, ɛ, æ, u, ʊ, o, ɔ, ʌ}	Identification task	Native speakers of Canadian English (15)
Xiahou (2012)	ESL	{i, ɪ, e, ɛ, æ}	AXB discrimination task	Native speakers of American English – one male and one female (2)

Table 4: Summary of previous literature.

3.1 Ho (2009)

Ho (2009) recruited 40 EFL learners in Taiwan with a mean age of approximately 18 years who were divided into higher and lower proficiency groups. All of them had Mandarin as their L1 but were also fluent in (Taiwanese) Hokkien, a prominent dialect of Chinese in Taiwan and member of the Minnan language family. They had no exposure to other second languages. Their English proficiency level was determined by the GEPT (General English Proficiency Test) which is “a standardized English proficiency test administered by the Language Training and Testing Center in Taiwan” (Ho 2009: 60–61). The GEPT provides five levels of proficiency: Elementary, Intermediate, Higher Intermediate, Advanced, and Superior. The higher proficiency group (HEFL) exceeded the GEPT Intermediate proficiency in reading, listening, speaking, and writing. Some students were Higher Intermediate in reading and listening, but none were Advanced or Superior in the four subjects. The lower proficiency group (LEFL) were regarded as Elementary based on a commercialized version of the GEPT.

Ho (2009) examined perception and production of English front vowels {i, ɪ, e, ɛ, æ} in [b_t] (i.e., *beat, bit, bait, bet, bat*) and [b_d] (i.e., *bead, bid, bade, bed, bad*) contexts (Table 5). During the perception task, the learners were presented with five choices and had to identify which word was aurally presented for each trial (Ho 2009: 64). The target vowels were presented ten times (2 recordings and 5 repetitions) in random order. For the purpose of this study, only the perception task results are reported.

Target vowel	LEFL	Target vowel	HEFL
æ	42.8	e	75.8
i	39.8	æ	69.5
e	37.0	i	68.5
ɛ	34.5	ɛ	62.5
ɪ	32.5	ɪ	62.5

Table 5: Mandarin learners' vowel identification task accuracy (%) results (adapted from Ho 2009: 84–85).

The HEFL correctly identified English front vowels approximately 67.8% of the time and the mean correct identifications for {i, ɪ, e, ɛ, æ} were 68.5%, 62.5%, 75.8%, 62.5%, and 69.5%, respectively. Ho stated that the perception of {e, æ, i} was significantly better than the perception of {ɪ, ɛ}. The LEFL group showed markedly lower scores across the board, correctly identifying the vowels at 37.3%. The mean accuracies for {i, ɪ, e, ɛ, æ} were 39.8%, 32.5%, 37%, 34.5%, and 42.8%, respectively.⁶ Ho reported that no significant differences in perception across vowels were found.

The performance differences between the HEFL and the LEFL groups were statistically significant—the HEFL group performed more accurately than the LEFL group in the vowel discrimination perception task. One common behavior between both groups was misidentification of /ɪ/ as /i/: LEFL was inconsistent in mapping errors, but the HEFL group also displayed difficulties distinguishing /ɪ/ as /i/ which are closer together in the vowel space.

In addition, Ho (2009) provided mean confusion (%) matrices of the two groups (Tables 6 and 7) which are critical for investigating how L1 interacts with L2 input.⁷ Not surprisingly, the LEFL group performance indicates that they were confused more often with other English vowels compared to the HEFL. A bidirectional confusion between /i/ and /ɪ/ was found in the HEFL group; /i/ was identified as /ɪ/ 31.5% of the time and /ɪ/ was misidentified as /i/ 31.75 % of

⁶ A reviewer wonders why, since the Mandarin vowel inventory has an obvious height contrast, participant error rates for English /ɪ/ and /ɛ/ were so prevalent. Ho (2009) remarks that participants may have confused /ɪ/ and /ɛ/ because of their close phonetic distance to nearby vowels in English. That is, the pairs /i/ and /ɪ/ as well as /ɪ/ and /ɛ/ are near to each other in the English vowel space and, as such, are likely to be consistently misidentified, particularly by lower-level learners (see Tables 6 and 7).

⁷ Confusion matrices and vowel accuracy measures have been included in this summary and in work reported later in this section as preliminary evidence of the difficulties learners encounter with respect to the acquisition of front vowels in English. Some (but not all) authors have used these reporting tools in a descriptive sense without conducting further statistical analyses on the results. As such, they should not be interpreted as definitive. We include mention of them here because the evidence provided merits further investigation, which is one of the purposes of this work.

the instances. Regarding mid vowels, /e/ was confused with /ɛ/ (20%) while /ɛ/ was confused with /e/ (13%). Lastly, /æ/ and /ɛ/ also showed a bidirectional confusion pattern: /æ/ was often confused with /ɛ/ (22.75%) and /ɛ/ fell into /æ/ 21.25% of the time.

Target Response	i	ɪ	e	ɛ	æ
i	39.75	32.25	7.5	6.25	3.5
ɪ	50.75	32.5	3.25	0.75	1.25
e	6.5	16.25	37.0	24.0	25.75
ɛ	3.0	14.0	26.5	34.5	26.75
æ		5.0	25.75	34.5	42.75

Table 6: LEFL group's mean confusion matrices in perception of English front vowels (adapted from Ho 2009: 91).

Target Response	i	ɪ	e	ɛ	æ
i	68.5	31.75	0.5	1.0	0
ɪ	31.5	62.5	0.5	2.25	0
e	0	1.75	75.75	13.0	7.75
ɛ	0	3.75	20.0	62.5	22.75
æ	0	0.25	3.25	21.25	69.5

Table 7: Mean confusion matrices of HEFL front vowel perception (adapted from Ho 2009: 91).

In terms of LEFL, /i/ was mainly misidentified as /ɪ/ (50.75%) (Ho 2009: 91). /ɪ/ diverged to /i/ (32.25%) and /e/ (16.25%) which suggests beginner learners' perceptual difficulty. The mid vowels {e, ɛ} were often confused with their neighboring vowels and /æ/ was also mapped to mid vowels.

3.2 Lai (2010)

In a study on vowel discrimination, Lai (2010) tested high and low proficiency Taiwanese learners in delineating English tense-lax contrasts. They recruited 90 Mandarin speakers ranging in age from 19 to 22 years old who learned English as a foreign language at school and had no experience of living in English-speaking countries. Based on significant differences in mean TOEIC scores (full score: 900) as well as undergraduate major designation, the participants

were grouped into two levels, low and high. The high proficiency group (HEFL) consisted of 45 English majors who recorded a mean TOEIC test score of 530 (SD = 36). The low group (LEFL) was comprised entirely of non-English majors and had mean scores of 352 (SD = 33). Lai reported that the difference between the two groups was statistically significant ($t = 23.8^{***}$, $p < .000$).

Lai (2010) observed eleven English vowels in the context of [h_t] and tested seven minimal pairs: {i-i}, {e-ε}, {æ-ε}, {æ-e}, {o-ɔ}, {u-ʊ}, and {a-ʌ}. They performed an AX discrimination task in which participants had to discern whether the two recorded stimuli were the same or not and identify the target vowel. The vowel pairings were presented in three different forms which consisted of A-A/B-B, A-B, and B-A trial types. Participants then had to decide if the two sounds were the same or different. If they had identified the sounds as being the same (i.e., A-A/B-B trial types), then they were told to identify which word from an accompanying multiple-choice menu matched the sound they heard. If perceived as different (i.e., A-B and B-A trial types), they were instructed to identify and circle both the first and second words that they considered to have the matching segments. If their decision on both discernment and identification were correct, the authors gave one point. The task had three trials for each target contrast in total, thus the highest possible score was three. Lai's one-way ANOVA comparing the effect of proficiency on mean scores in the discrimination task is reported in **Table 8**.

Target	Group	Mean (SD)	F
{æ-ε}	LEFL (N = 45)	2.33 (0.60)	26.57***
	HEFL (N = 45)	2.87 (0.34)	
{æ-e}	LEFL (N = 45)	2.27 (0.94)	4.10**
	HEFL (N = 45)	2.60 (0.58)	
{e-ε}	LEFL (N = 45)	1.98 (0.94)	9.91**
	HEFL (N = 45)	2.49 (0.54)	
{i-i}	LEFL (N = 45)	1.71 (0.87)	1.25
	HEFL (N = 45)	1.87 (0.34)	

Table 8: Mandarin learners' mean scores in vowel discrimination task (highest possible score = 3) (adapted from Lai 2010: 167). Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

A perceptual saliency hierarchy was {æ-ε} > {æ-e} > {e-ε} > {i-i} for both proficiency groups. In other words, {æ-ε} was the easiest pair to discriminate while the {i-i} pair was the hardest. Overall, the higher proficiency learners obtained higher mean scores and the score differences were statistically significant except for {i-i}.

3.3 Hu et al. (2019)

Using both vowel identification and vowel perception tasks, Hu et al. (2019) set out to examine what effects participants' native Mandarin vowel inventory had on the identification of L2 English vowels during the acquisition process. In this study, they focused on /i/ and /ɪ/ since this contrast has proven to be a persistent area of difficulty for Mandarin learners of English (Lai 2010). English /i/ and Mandarin /i/ are argued to be identical counterparts with each other (Duanmu 2007), so realizing a contrast between English /i/ and /ɪ/ was hypothesized to be particularly difficult due to their phonetic proximity.

The study tested 46 Mandarin learners of English (aged 22 to 26) who had started learning English (their only second language) between five and thirteen years of age. The students were all undergraduate or graduate students in the People's Republic of China. Proficiency was measured through the Oxford Placement Test, but this did not correlate with any of the results reported later. The vowel identification task tested participants' perceptions of twelve English monophthongs /æ, ε, e, i, ɪ, ɜ, ʌ, ʊ, u, o, ɔ, ɑ/. For the initial test, each of the English vowels were recorded in monosyllabic /hVd/ contexts by a young female native speaker of American English. The identification test had participants listen to a vowel in isolation and then click on a text box on a computer screen which consisted of different words representing all the twelve vowel sounds in the study. They were given ten seconds to respond. The words were all situated in monosyllabic /hVd/ contexts. Participants also took part in English and Mandarin vowel categorization tasks. In the English version, they would listen to the twelve English monophthongs and in the Mandarin version they would hear five of their native vowels and try to identify which vowel category each belonged to. After each categorization attempt, they were asked to rate that vowel for how representative it was for the category they had just selected using a five-point scale (one being least representative and five being most).

Results showed poor identification of several vowels in the English inventory, though we only report on English front vowels. English /ɪ/ was the least correctly identified vowel of all with an accuracy rate of just 13.9%. Moreover, English front vowels /æ/ and /ε/ as well as /ɪ/ and /i/ were often confused for one another. The vowel /ɪ/ was confused for /i/ 35% of the time, the highest misidentification rate of all front vowels. It was correctly identified in only 14% of cases. The /i/ vowel itself had the second highest accuracy rate, being correctly identified at a rate of 58%. The second most accurate was /e/ at 45%. /æ/ saw an accuracy rate of just 27% and was confused as /ε/ 20% of the time and /e/ in another 20% of cases. /ε/ was identified correctly at 24% but was confused as /æ/ 24% of the time and /ɪ/ in 20% of cases (see **Table 9**).

However, a curious finding regarding individual differences emerged. Perception results were closely related with how well learners could distinguish and identify L2 vowels in the categorization task. It was found that Chinese learners' vowel identification and perception accuracy of twelve L2 English monophthongs, but most notably /i/ and /ɪ/, were related to the

how representative (and thus how distant) the learners perceived the English vowel counterparts to be. The researchers reported statistically significant correlations between participants' ability to accurately perceive English /i/ and /ɪ/ and then identify them correctly as representatives of their respective vowel designations. This was true as well as for English /i/ and Mandarin /i/, but not for English /ɪ/ and Chinese /i/. The results in **Table 9** attest to the latter finding.

Target Response	i	ɪ	e	ɛ	æ
i	58	35	2	2	
ɪ	8	14	23	2	7
e	4	11	45	7	20
ɛ	2	7	13	43	20
æ	1	5	5	24	27

Table 9: Confusion matrices (%) of front vowels in Hu et al. (2019) (adapted from Hu et al. 2019: 4539).

Regarding the target pair under investigation, if English /i/ and Mandarin /i/ were judged to be strongly representative of their respective categories, participants were more likely to correctly identify not only English /i/ but all other English vowels as well. It was found that L2 vowel contrasts were more easily established if an L2 English vowel pair that was near each other in the vowel space could be differentiated in the ears of the listener. If not, listeners registered both poor identification and perception rates, particularly with the /i/ and /ɪ/ pairing. The ability to perceive perceptual contrasts in the categorization task was related to accuracy in the vowel identification task. The authors suggested these individual differences may be due to each participant's particular L2 learning experience, but no data for this was collected.

3.4 Yu (2012)

Yu (2012) investigated the production and perception of 10 English vowels by 15 Mandarin speakers living in Canada for approximately two years. Yu collected information regarding length of residence, age of arrival, and years spent learning English to examine learner progression. All participants were students at a Canadian university with a mean length of residence in country of 3.5 years and with a mean age of 25.8 years. On average, the students had 17 years of experience learning English and their mean age of arrival in Canada was 22.

The study consisted of two experiments; however, for the purposes of the present study, only the perception task in Experiment 2 is summarized. In addition, the original study investigated

ten vowels, {i, ɪ, e, ε, æ, u, ʊ, o, ɔ, ʌ}; here, we report results concerning front vowels since the target vowels are the five front vowels.

In the perception experiment, learners identified vowels that had been recorded by 15 native Canadian English speakers in a previous experiment. The recorded vowels were situated within [bVt] words and nested in carrier sentences such as “Now I say the word beat”. Collectively, the vowels {i, ɪ, æ, ε} were correctly identified more than 70% of the time by the learners. With regard to individual vowels, the vowel /i/ was misidentified just 15% percent of the time while {æ, ɪ, ε, e} had higher mean error rates at, 18%, 23%, 26%, and 37%, respectively (see **Table 10**). Using simple correlational analysis, Yu found that length of residence and years spent studying English influenced the error rate. That is, the longer the participants lived in the English-speaking country, the higher the perceptual accuracy and the longer they had studied English, the lower the error rate.⁸

Target vowel	Accuracy rate (%)
i	85
æ	82
ɪ	77
ε	74
e	63

Table 10: Mandarin learners’ accuracy rate with English front vowels.

Based on the results, **Table 11** presents the learners’ confusion matrices. The /i/ vowel was mischaracterized as /ɪ/ in 12% of instances. The /æ/ vowel was misidentified as /ε/ 12% percent of the time. The /ɪ/ vowel was confused for /ε/ in 10% percent of instances and as /e/ another 8% of the time/. The /ε/ vowel was confused for /æ/ in 15% of instances. Finally, the /e/ vowel was misidentified as /i/ 20% of the time and as /ε/ at 15%. The vowel accuracy results for the ESL learners described here are descriptive in nature since no statistical analyses were carried out, but a general comparison between the vowel confusion matrices here and in Hu et al. (2019) and Ho (2009), for example, nevertheless reveal evidence of significant progress in perceptual accuracy.

⁸ It should be noted that Yu’s (2012) use of correlational analysis here is a descriptive measure, and so the results should not be interpreted as statistically significant. Nevertheless, we include mention of them because they are consistent with other findings that we report in the peer-reviewed literature.

Target Response	i	æ	ɪ	ɛ	e
i	85	2	3	0	20
æ	0	82	2	15	1
ɪ	12	2	77	6	0
ɛ	1	12	10	74	15
e	2	2	8	5	63

Table 11: Mandarin learners' mean confusion matrices (%) for English front vowels (adapted from Yu 2012: 60). Note: Bold-faced numbers indicate the correct identification.

3.5 Xiahou (2012)

This study reported experiments that analyzed perception of English phonemes {i, ɪ, e, ɛ, æ}. Xiahou (2012) recruited 30 Mandarin speakers who had been residing in the U.S. for less than five years. The mean length of residence was 1.6 years. Participants were all international students studying at university with a mean age of 22.7 years. They had begun studying English at a mean age of 10 and all had no immersion experience prior to the age of 18 (Xiahou 2012: 16).

In the perception portion of the study, learners completed a categorical AXB perceptual discrimination task which involved listening to three stimuli in sequence. All three stimuli were made up of /bVt/ syllables which had the above-mentioned target phonemes situated within them. The order of presentation had listeners first listen to stimulus A, then the target stimulus X, and finally stimulus B. Participants were then asked to identify whether stimulus X was the same as either stimulus A or B.

All possible vowel contrasts were then created with the following 10 pairings: [beat-bit], [beat-bait], [beat-bet], [beat-bat], [bit-bait], [bit-bet], [bit-bat], [bait-bet], [bait-bat], and [bet-bat]. An example trial provided by Xiahou shows how the [bit-bet] A-B contrast was presented to listeners: *bit – bit – bet; bit – bet – bet; bet – bit – bit; bet – bet – bit*. The participants simply had to decide whether the second word was the same as the first or the third. As shown in **Table 12**, it was found that the {i-æ} and {i-ɪ} vowel pairs were easiest for participants to discern as evidenced by 99% accuracy rates. Similarly, high accuracy rates were recorded with the {ɪ-æ}, {e-æ}, {ɪ-ɛ}, {i-ɛ}, {e-ɛ}, and {e-i} contrasts. The {e-i} contrast saw a 90% accuracy rating. The lowest accuracy was recorded with the vowel pairing {ɛ-æ} with an 82% accuracy rating. Discrimination accuracy scores were also strongly related to length of residence and L2 use in daily life. These results show that, at later stages of progression, L2 learners begin to develop improved discriminatory skills with pairings {i-æ}, {i-ɪ}, {ɪ-æ}, {e-æ}, {ɪ-ɛ}, {i-ɛ}, {e-ɛ}, and {e-i} but the {e-ɪ} and {ɛ-æ} still present some lingering difficulty. Planned pairwise contrasts revealed that scores for vowel pairings with the highest accuracy (e.g., {i-æ} and {i-ɪ}) did

not different statistically with one another. The difference in accuracy rate between {i-ɪ} and {ɪ-æ} was statistically significant, as were differences between {e-i} and {e-ɪ} as well as {e-ɪ} and {e-æ}.

Target vowel pair	Accuracy rate (%)
{i-æ}	99
{i-ɪ}	99
{ɪ-æ}	98
{e-æ}	97
{ɪ-ɛ}	97
{i-ɛ}	96
{e-ɛ}	95
{e-i}	94
{e-ɪ}	90
{ɛ-æ}	82

Table 12: Mandarin learners' AXB perceptual discrimination task accuracy rate % (adapted from Xiahou 2012: 21).

3.6 Jia et al. (2006)

Jia et al. (2006) recruited three groups of speakers, namely, monolinguals,⁹ recent arrivals, and past arrivals. The monolingual groups were 91 Mandarin speakers living in China. The monolinguals received formal English education from school, and none had extensive training or private lessons from native English speakers. The other two groups were recruited from the L1-Chinese L2-English speakers living in New York City. Based on their length of residence (LOR), the participants were categorized as recent arrivals (LOR less than 2 years) or past arrivals (LOR 3–5 years) (**Table 13**). According to Jia et al., the participants' exposure to native English sounds was minimal prior to their arrival in the U.S.

Jia et al. (2006) investigated six contrasts: {i-ɪ}, {i-e}, {ɛ-æ}, {æ-ɑ}, {ɑ-ʌ}, {u-ɑ}. In their perception task, the target vowels were presented in a [dV-pə] disyllabic structure so that the stimuli conformed to both English and Mandarin phonotactic constraints. The task was a

⁹ Although we borrow Jia et al.'s (2006) use of the term "monolingual" here and in later sections of the paper, the Mandarin speakers in this and other studies may have had some knowledge of local Chinese dialects. Thus, it should be noted from this point that "monolingual" is used by Jia et al. (2006) and in our analysis to describe participants who have only just begun studying English as a second language in a formal classroom environment and, thus, have only had limited exposure to the language. We'd like to thank a reviewer for bringing this to our attention.

categorical AXB discrimination task where the participants pressed “1” if the middle vowel sounded like the first stimulus or “3” if the middle vowel sounded like the third one. Each of the vowels was presented at inter stimulus intervals of 500ms. The perception accuracy was calculated by the number of correct responses out of the total trials.

Group	LOR	Mean Age	<i>n</i>
Monolinguals	N/A	N/A (Range: 7–20)	91
Recent arrivals	1.3 (SD = 0.7)	20.5 (SD = 8.7)	77
Past arrivals	3.7 (SD = 0.8)	24.4 (SD = 8.0)	54

Table 13: General characteristics of participants in Jia et al. (2006).

For the purposes of this study, we report results of the front vowel pairs, i.e., {i-ɪ}, {i-e}, {ɛ-æ}, {æ-ɑ} (see Jia et al 2006: 1124, for the comprehensive result table). According to Jia et al. (2006), the overall accuracy rates of monolinguals, recent arrivals, and past arrivals were 84.6%, 94.5%, and 96.2%, respectively. The authors reported that the monolingual group had significantly lower accuracy than the other two groups (**Table 14**). Also, a one-way ANOVA test revealed that there was a significant effect of group for all vowel pairs. Across the three groups, the {i-e} pair was the easiest contrast and {ɛ-æ} was the most difficult to discern. The confusion patterns of the perception task were not reported in the original study.

	Monolinguals	Recent arrivals (~1yr)	Past arrivals (> 4 yrs)
[i]-[ɪ]	82.6	97.4	97.8
[i]-[e]	90.2	99.5	99.7
[ɛ]-[æ]	76.3	89.4	91.8
[æ]-[ɑ]	88.9	96.1	96.6

Table 14: Mandarin learners’ perception accuracy of English front vowel pairs (adapted from Jia et al 2006: 1124).

3.7 Interim summary

The findings from these previous studies offer interesting results; however, the mechanism behind these patterns is not fully expounded, and thus, requires further investigation. We discuss possible explanations for the confusion patterns in Sections 4 and 5. Our position is that the ability to both identify and discriminate among vowels closely relates to how far along the learners are

at realizing perceptual contrasts in the L2. We argue that a phonological representation can provide concrete accounts regarding the confusion patterns and can help depict the developments of learners' L2 grammar. In our review of these phonetic studies, we have documented the persistent difficulties that Mandarin learners of English face regarding certain vowels and vowel pairs in L2 English. These difficulties are consistent and are reported across both identification and discrimination tasks. **Table 15** summarizes prominent difficulties in front vowel acquisition faced by low and intermediate learners. The lax vowels /ɪ/, /ɛ/, and /æ/ consistently present hurdles to both low and intermediate learners. As proficiency improves, certain contrasts begin to take shape, and this will be important for our later predictions regarding front vowel perceptual development. With time comes expansion of the L2 learner vowel space. Learners begin to create contrasts among vowel pairs and correctly recognize and perceive lax English vowels that were initially mapped to L1 equivalents based on Mandarin's triangular vowel space. For example, these studies have shown that English lax /ɪ/ is often confused for /i/, which exists both in Mandarin and English.

Study	Proficiency level	Persistent difficulties	Emerging contrasts
Ho (2009)	EFL – low	/ɪ/, /ɛ/	
	EFL – intermediate		/i/, /e/, /æ/
Lai (2010)	EFL – low	{i-ɪ}	
	EFL – intermediate		{æ-ɛ}, {æ-e}, {e-ɛ}
Hu et al. (2019)	EFL – low to intermediate	{i-ɪ}	{i-ɪ}, {e-ɛ}, {e-æ}
Yu (2009)	ESL – intermediate	/e/, /ɛ/	/i/, /æ/, /ɪ/
Xiahou (2012) ¹⁰	ESL – intermediate	{ɛ-æ}	{i-ɪ}, {i-æ}, {ɪ-æ}, {e-æ}, {ɪ-ɛ}, {i-ɛ}, {e-ɛ}, {e-ɪ}, {e-ɪ}
Jia et al. (2006)	EFL – low	{ɛ-æ}	
	ESL – low to intermediate		{i-ɪ}, {i-e}

Table 15: Summary of English vowel acquisition studies.

¹⁰ Though both Yu (2012) and Xiahou (2012) only included participants with intermediate proficiency levels (thus making comparison with lower levels difficult), we still record their results in the emerging contrasts column. We did this because the learners' data here show that some of the initial difficulties experienced by less proficient learners in other studies are beginning to fade along with increasing proficiency.

4 Applying phonological-phonetic representation to phonetic data

This section demonstrates how our analysis captures the pattern of emerging contrasts and persistent difficulties. We also use L1-Mandarin L2-English speakers' interlanguage grammar (IG) to elucidate how perceptual difficulties can be predicted based on PRs. The hierarchy of contrastive features is the universal repository that learners have access to, and we assume that language learners' L1 phonological hierarchy heavily influences the structure of their interlanguage grammar. The contrastive hierarchies of the L1-Mandarin L2-English IG are displayed in **Figures 4 and 5** and represent early stages (low to intermediate levels) of L2 acquisition.

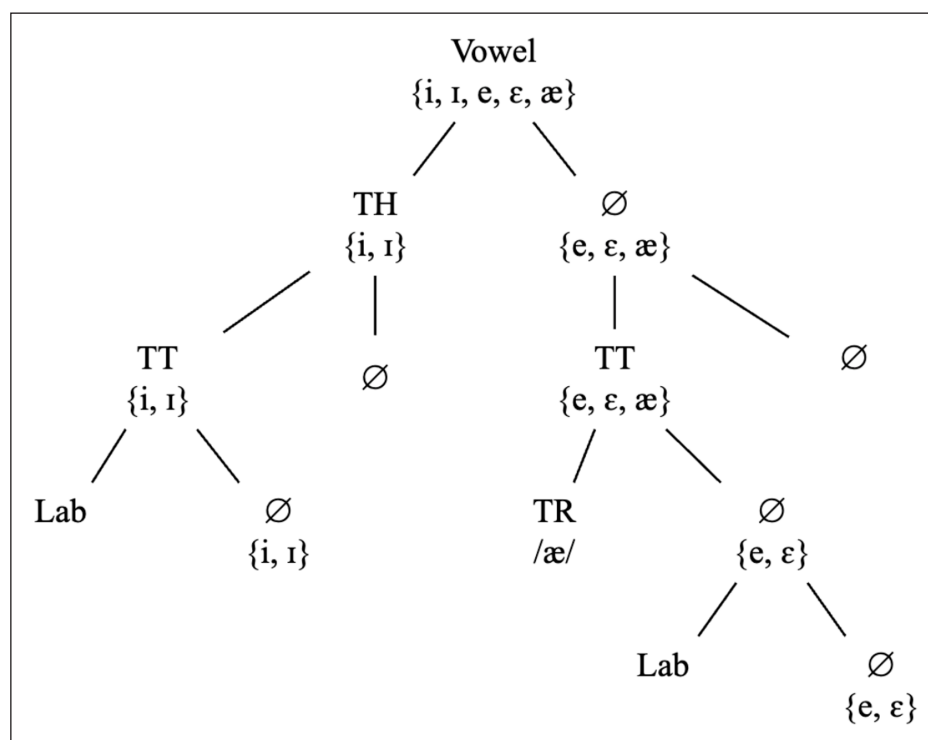


Figure 4: Phase 1 IG. Terminal nodes which contain more than one vowel segment suggest possible vowel confusion.

We claim that Mandarin learners construct a L2 phonological grammar which superficially resembles the target grammar but deviates from it slightly. As learners become more familiar with novel English vowel features (e.g., [ATR]) that are not present in the native inventory, Mandarin learners will begin to establish the necessary contrasts for better perception. To test our hypothesis, we have regrouped the participants of the previous studies we have summarized in Section 2 (see **Table 16**). Phase 1 represents the monolingual learners from Jia et al. (2006), Ho (2009), Hu et al. (2019), and Lai (2010). These monolingual learners are participants who

are learning English as a foreign language. This implies that their language input and output opportunities are largely limited to formal classroom settings as they do not live in an environment where English is a dominant language. Phase 2 represents the intermediate ESL learners from Xiaohou (2012), Jie et al. (2006), and Yu (2012) who have arguably had more exposure to English based on LOR in English-speaking countries.

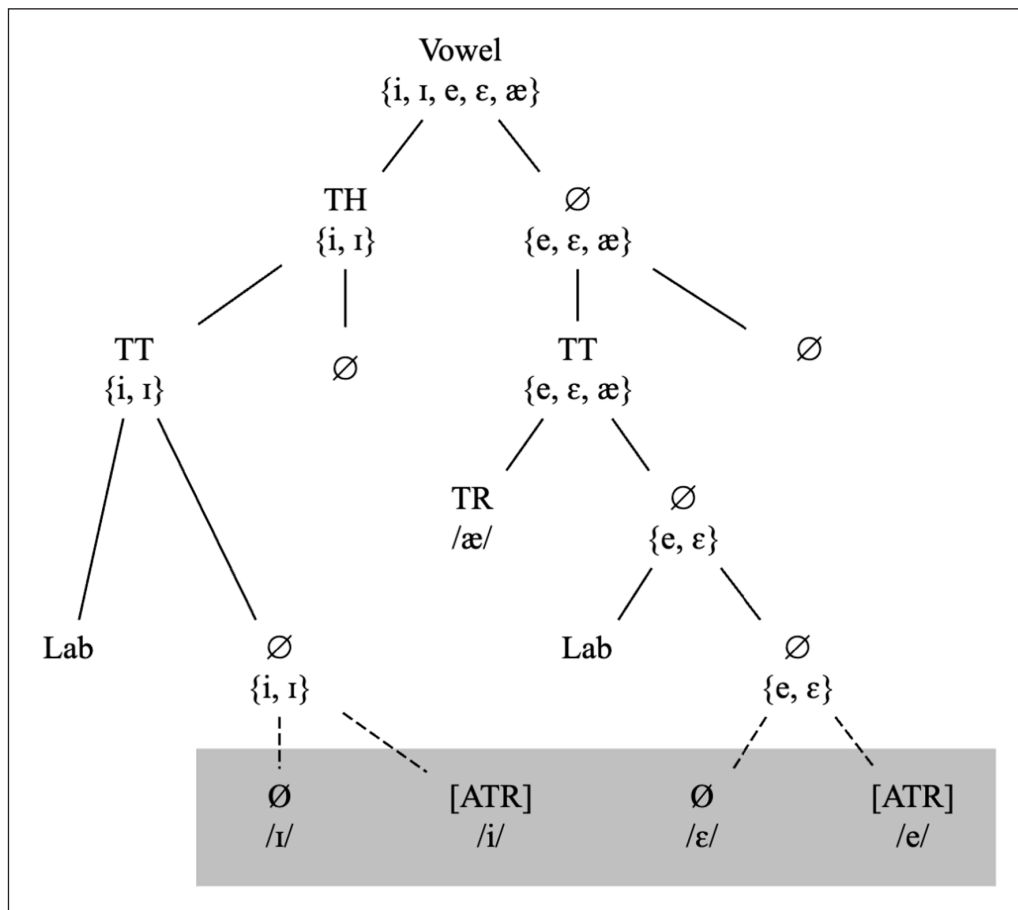


Figure 5: Phase 2 IG. The dashed lines indicate emerging contrasts by assembling a feature in the L1 grammar that can distinguish the two sounds.

In the beginning, L1-Mandarin L2-English speakers will transfer their L1 feature hierarchy directly to the L2 input. The feature hierarchy of Mandarin is Root (Vowel), Tongue Height (TH), Tongue Thrust (TT), Tongue Root (TR), and Labial (Lab; see **Table 1**). Since the Mandarin feature order is not identical to the English feature ranking, not all vowels are parsed (i.e., terminal nodes having more than one segment). Phase 1 learners were all monolingual speakers of Mandarin, received a limited amount of English education (EFL environment), and had no

prior experience living in English-speaking countries. Thus, the Phase 1 group can attest to beginners' sound perception. In general, beginner learners had a hard time discerning {i, ɪ, ε}. For example, Jia et al. (2006), Hu et al. (2019), and Lai (2010) revealed that the {i-ɪ} pair was problematic through discrimination tasks; Ho found that /ɪ/ and /ε/ had low accuracy rates. As for the non-high vowels (the sister node of TH), learners will apply TT as it is the second dimension of their native language. However, since the non-high vowels are all front vowels, it does not further distinguish {e, ε, æ}. The next resolution is applying the next ranked dimension, which is TR. TR discerns /æ/. The last dimension, Lab, is not a relevant feature to contrast {e, ε}; hence, we argue that learners will have difficulty in distinguishing these two vowels. Perception data from previous studies support our argument (Ho 2009; Lai 2010).

Phase	Studies	LOR (years)
Phase 1	Jia et al. (2006): monolingual/EFL learners Ho (2009): monolingual/EFL learners Hu et al. (2019): monolingual/EFL learners Lai (2010): monolingual/EFL learners	none
Phase 2	Xiahou (2012): ESL learners	1.6 (SD = 1.1)
	Jia et al. (2006): recent arrivals & past arrivals ¹¹	1.3 (SD = 0.7) 3.7 (SD = 0.8)
	Yu (2012): ESL learners	3.5 (SD = 1.6)

Table 16. Re-grouping of Phase 1 and Phase 2 learners.

Then the next question should be: why do learners struggle with these sounds? The Phase 1 IG (**Figure 4**) predicts that learners will have difficulty in discerning {i-ɪ} and {e-ε} pairs because they are not further distinguished in the IG (note that each terminal node should have one vowel segment). Thus, the results from Ho (2009), Hu et al. (2019), Lai (2010), and Jia et al. (2006) are not surprising. The confusion patterns and lower accuracy rates are due to the learners' Phase 1 IG that cannot distinguish all vowels. That is, beginner learners apply their L1 feature order to the L2 input because they strongly rely on their L1 grammar. Specifically, the TH dimension will first distinguish {i, ɪ} from {e, ε, æ} because it is the highest dimension in the Mandarin grammar. Then, there are two vowels marked with TH which cannot be further distinguished because the next available dimensions are TT and Lab, which mark front and rounded vowels, respectively. As we know, {i, ɪ} are high front unrounded vowels, so none of the dimensions are able to distinguish /i/ from /ɪ/. This is where the first overt perceptual confusion occurs.

¹¹ We have grouped recent and past arrival groups together from Jia et al.'s (2006) study because the discrimination task results did not show a significant difference.

As for the non-high vowels, i.e., {e, ε, æ}, the TR dimension is employed and separates /æ/ from {e, ε}. It is the only vowel that establishes a felicitous terminal node in the contrastive hierarchy. Beginner learners seem to perceive the [æ] sound relatively earlier than other sounds as evinced in Ho (2009) and Lai (2010). In contrast, the Phase 1 IG predicts that the other two vowels, {e, ε}, are strong candidates for confusion because they are further parsed. The possible dimensions that Mandarin learners have include TR, but it does not distinguish the {e-ε} pair because it marks low vowels. The findings from Ho, Hu et al., Lai, and Jia et al. all report on the learning difficulty of these vowels.

With regard to Phase 2 learners, the Phase 2 IG suggests that learners will employ the [ATR] gesture which is an enhanced feature of English tense vowels. The [ATR] gesture is the reverse of [RTR] and both gestures are commanded by their superordinate unit TR. Thus, [ATR] is not an entirely new feature in Mandarin as the phonologically relevant dimension (i.e., TR) is retrievable from the L1 grammar. We claim that the learners will utilize [ATR] by inserting it at the bottom of the IG before they acquire the appropriate length contrast in English. The premise is that reinterpreting the L1 completion rule is easier and faster (i.e., more efficient) than learning a new contrast (length, in this context). By the [ATR] insertion, Mandarin learners are able to discern both {i-ɪ} and {e-ε} pairs. This argument is supported by the results from Xiahou (2012), Yu (2012), and Jia et al. (2006) whose participants are considered as belonging to the Phase 2 category (intermediate ESL learners). Data from these three studies strongly suggest that Mandarin learners begin to perceptually differentiate /i/ from /ɪ/. The discrimination task results from Xiahou and Jia et al. show that learners are able to tell the difference between /i/ and /ɪ/. According to Xiahou, the {i-ɪ} pair was accurately distinguished 99% of the time; Jia et al. reported that participants had higher than 97% accuracy rate for discriminating the {i-ɪ} pair. The {e-ε} pair was correctly discriminated in 95% of the cases (Xiahou 2012). The identification task results from Yu also reveal that /i/ and /ɪ/ are relatively easier to perceive accurately than the mid front vowels /e/ and /ε/ (see Section 2 for full results). Thus, it can be concluded that Phase 2 learners have established the emerging contrast by adding an [ATR] gesture at the bottom of their IG. However, it should be noted that it is beyond the scope of the current study to analyze how the learning transition occurs (i.e., from **Figure 4** to **Figure 5**). We leave this for future studies.

5 A phonological window into L2 perception

We posit that the L1-Mandarin L2-English learners will utilize the [ATR] gesture to discern confusing pairs like /i/ and /ɪ/. In other words, the improved discrimination can be attributed to the insertion of [ATR] in their developing IG. Why [ATR] then? At the surface level, e.g., when the phonological units are realized to pronounceable phonetic gestures, the [ATR] gesture is an enhanced feature for English /i/ and /e/ (Halle & Mohanan 1985). **Table 17** is mostly identical

to **Figure 3** (the PR of English vowels) except that the dimensions are filled in with their gestures and arrayed into columns. Since enhanced gestures ([ATR], in the case of English front vowels) are redundant features and phonologically inert, they are not encoded in the phonological representation.

	i	ɪ	e	ɛ	æ
Length	long	short	long	short	long
Gesture	[vowel] [front] [high]	[vowel] [front] [high]	[vowel] [front]	[vowel] [front]	[vowel] [RTR] [front]
Enhancement	[ATR]		[ATR]		

Table 17: English front vowels phonetic specification.

The Phase 1 IG (**Figure 4**) shows that {i-ɪ} and {e-ɛ} will cause perceptual difficulty because they cannot be separated unlike /æ/. The L1-Mandarin L2-English learners have to learn the length contrast in order to perceive them like the target English grammar. However, the length contrast is a novel contrast for Mandarin speakers; thus, for Phase 2 L1-Mandarin L2-English learners, utilizing resources they have in their L1 grammar as a temporary solution will be more efficient to help distinguish the confusing pairs. Here, we argue that learners will discern the [ATR] enhancement because Mandarin has the dimension that governs the [ATR] gesture (i.e., TR). The retrieval of [ATR] will lead Mandarin learners to reinterpret the TR dimension as [ATR]. Eventually, they will establish a distinction between /i/ and /ɪ/ by employing [ATR] to mark /i/; /ɪ/ remains underspecified (\emptyset) (**Figure 5**). Similarly, they will discern /e/ from /ɛ/ by marking [ATR] to /e/. The Phase 2 IG is not identical with the target English grammar; however, in terms of perceptual competence, [ATR] insertion distinguishes the confusing pairs and the data from the selected previous studies supports this position.

Besides /i/, even beginner learners had less difficulty discerning /æ/, confirming our prediction that English {i, æ} will be less challenging to discriminate compared to the other front vowels {e, ɪ, ɛ}. English /æ/ had a high FUL score when matched with Mandarin /a/ (score = 0.6), which means L1-Mandarin learners are likely to perceive /æ/ as a similar sound to their /a/ vowel. The PR model supports the idea that /æ/ will be easier to perceive because it is the sole vowel that is distinguished by applying their L1 grammar to the L2 input. In addition, better perception of /æ/ can be attributed to the fact that [a] in Mandarin has several allophonic variants that have characteristics of English counterparts (Duanmu 2007: 38). Xu (1980) explains that one of the variants, [æ], occurs in closed syllables before [n] and after [j], like in the word 'salt' [jæn]. As the learner's proficiency develops, accuracy with these two vowels also increases.

This study makes an important contribution in demystifying why certain English front vowels are easier or more difficult to discern for L1-Mandarin L2-English learners by utilizing FUL scores and the IG based on the PR model. The proposal also provokes a meaningful discussion regarding communicative competence versus the ideal L2 target grammar. The Phase 2 IG does not resemble the English contrastive hierarchy perfectly, but it may be sufficient to achieve L2 sound perception and production for advanced beginners and intermediate learners (Darcy et al. 2012); however, since the interlanguage grammar deviates from the target L1 grammar, the learners' vowel production may vary from the native speakers' production. Also, the learners may not accurately distinguish a pair of English vowels (e.g., /i/ and /ɪ/) but they may distinguish the vowels based on the lexical context. We speculate that learners will use the surface phonological grammar before they acquire the length difference in English. To confirm whether they can attain a native-like phonological grammar, more research should be done with advanced L1-Mandarin L2-English learners.

Another strength of the current theoretical framework is that the PR analysis is not limited to Mandarin and English, but it can be applied to various languages. The PR is based on MCS which posits that the contrastive feature hierarchy is universal (not the features or the feature order) (Dresher 2018); hence, contrastive hierarchy models like PR can be utilized to various L2 or L3 acquisition contexts (e.g., Natvig 2017; Kwon 2021; Archibald 2022a; 2022b). This study has demonstrated how PR can be applied to Mandarin and English vowels and accounts for Mandarin learners' perceptual confusions. With more applications to various languages, the PR model will further complement the existing line of phonetic-based research of perception and production (e.g., SLM and PAM) and deepen our understanding of L2 sound acquisition.

The current article is limited in that the role of length has not been fully delineated. Although some studies (e.g., Flege et al. 1997; Liu et al. 2014; Mi et al. 2016) have concluded that learners' dependency on duration cues is not substantial, length should not be disregarded since it is part of the L1 grammar. It is possible that length or mora has a different effect or weight on perception (or at the surface representation) as it is a structural feature, but we leave this matter to future research. Lastly, there is a need for an increase in the number of studies investigating L1-Mandarin L2-English vowel perception and, in particular, acquisition of these features over time using comparative measures of proficiency, length of residency, and years studying English. The studies we used in our analysis used a diverse array of tasks, measures, and samples. Though this selection procedure may be considered as a net positive owing to the effectiveness of the model in relation to the trends reported in the included research, the number of studies available to us at the time of writing was limited. As a result, we included some studies whose findings could have benefited from more rigorous and comprehensive statistical analysis, particularly with respect to the use of descriptive statistics in the vowel accuracy measures and confusion

matrices. Therefore, we must acknowledge that some of the data reported in section 3, though consistent with other findings, may not be reliable because of the methods and/or analysis.

6 Conclusion

With respect to the previous line of L2 perception research, one of the limitations included the lack of phonological accounts for either good or bad L2 perception or concrete criteria for perceptual (dis)similarity. That is, not many studies delved into explaining what triggers sound discrimination or confusion. For instance, previous studies adopting SLM/SLM-r and PAM/PAM-L2 were not able to fully elaborate on perceptual challenges that L2 learners face. In order to fill the gap and suggest a complementary viewpoint from a phonological perspective, we argue that the PR framework can aid our understanding of L2 perception and learnability and provide better accounts for perceptual challenges. The current paper re-examined data from six previous studies that investigated L1-Mandarin L2-English learners' perception of English front vowels. We introduced how PR can be utilized to demonstrate developmental stages of IG for L1-Chinese L2-English learners.

By re-examining selected previous studies with the PR framework, two conclusions can be drawn. First, the PR accounts for the possible confusing pairs based on the idea of the strong phonology transfer hypothesis (i.e., learners applying their L1 grammar to L2 input). For L1-Mandarin L2-English learners, /i/ and /æ/ are easier to perceive compared to the other front vowels, even in Phase 1, because they are perceptually similar to their existing L1 vowel segments. Second, the PR can capture the developmental stages of learners' IG (beginner and intermediate, in this study). It can also suggest how learners are processing the confusing sounds or learning new segments. In order to bolster the PR framework on L2 acquisition, more studies on advanced learners and different languages are encouraged to confirm the feature (re)assembly process exemplified by the PR.

Acknowledgements

We'd like to take this opportunity to thank Professor Eric Raimy of UW-Madison for helping to get this project off the ground and for encouraging us to develop it into a manuscript. We'd also like to thank Associate Professor Hanyong Park of UW-Milwaukee for his helpful insight during the writing and revision of this paper. We also owe a sincere debt of gratitude to the three anonymous reviewers for their patience and very useful suggestions during the revision process. All remaining errors are our own.

Competing interests

The authors have no competing interests to declare.

References

- Archibald, John. 2022a. Phonological parsing via an integrated I-language: The emergence of property-by-property transfer effects in L3 phonology. *Linguistic Approaches to Bilingualism*. DOI: <https://doi.org/10.1075/lab.21017.arc>
- Archibald, John. 2022b. Segmental and prosodic evidence for property-by-property transfer in L3 English in Northern Africa. *Languages* 7(1). 28. DOI: <https://doi.org/10.3390/languages7010028>
- Avery, Peter & Idsardi, William J. 2001. Laryngeal dimensions, completion and enhancement. In Hall, Tracy (ed.), *Distinctive feature theory*, 41–70. Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110886672.41>
- Best, Catherine T. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In Goodman, Judith & Nusbaum, Howard (eds.), *The development of speech perception: The transition from speech sounds to spoken words*, 167–224. MIT Press.
- Best, Catherine T. & McRoberts, Gerald W. & Sithole, Nomathemba M. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14(3). 345–360. DOI: <https://doi.org/10.1037/0096-1523.14.3.345>
- Best, Catherine T. & Tyler, Michael D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn, Ocke-Schwen & Munro, Murray (eds.), *Language experience in second language speech learning*, 13–34. John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/llt.17.07bes>
- Brown, Cynthia A. 1998. The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14(2). 136–193. DOI: <https://doi.org/10.1191/026765898669508401>
- Brown, Cynthia. 2000. The interrelation between speech perception and phonological acquisition from infant to adult. In Archibald, John (ed.), *Second language acquisition and linguistic theory*, 4–63. Blackwell.
- Chang, Charles B. 2015. Determining cross-linguistic phonological similarity between segments. In Raimy, Eric & Cairns, Charles (eds.), *The segment in phonetics and phonology*, 199–217. John Wiley & Sons. DOI: <https://doi.org/10.1002/9781118555491.ch9>

- Darcy, Isabelle & Dekydtspotter, Laurent & Sprouse, Rex A. & Glover, Justin & Kaden, Christiane & McGuire, Michael & Scott, John H. 2012. Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English-L2 French acquisition. *Second Language Research* 28(1). 5–40. DOI: <https://doi.org/10.1177/0267658311423455>
- Dresher, B. Elan. 2003. Contrast and asymmetries in inventories. In di Sciullo, Anna-Maria (ed.), *Asymmetry in Grammar, volume 2: Morphology, phonology, acquisition*, 239–57. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/la.58.10dre>
- Dresher, B. Elan. 2008. The contrastive hierarchy in phonology. *Toronto Working Papers in Linguistics* 20. 47–62. DOI: <https://doi.org/10.1017/CBO9780511642005>
- Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511642005>
- Dresher, B. Elan. 2018. Contrastive hierarchy theory and the nature of features. In Bennett, Wm. G., Hrats, Lindsay & Storoshenko, Dennis R. (eds.), *Proceedings of the 35th west coast conference on formal linguistics*, 18–29. Cascadilla Proceedings Project.
- Dresher, B. Elan & Piggott, Glyne & Rice, Keren. 1994. Contrast in phonology: Overview. *Toronto Working Papers in Linguistics* 13(1). iii–xvii.
- Dresher, B. Elan & van der Hulst, Harry. 1998. Head-dependent asymmetries in phonology: Complexity and visibility. *Phonology*, 15(3). 317–352. DOI: <https://doi.org/10.1017/S0952675799003644>
- Duanmu, San. 2007. *The phonology of Standard Chinese*. Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199267590.003.0005>
- Flege, James E. 1987. A critical period for learning to pronounce foreign languages? *Applied Linguistics* 8(2). 162–177. DOI: <https://doi.org/10.1093/applin/8.2.162>
- Flege, James E. 1995. Second language speech learning theory, findings, and problems. In Strange, Winifred (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*, 233–277. York Press.
- Flege, James E. & Bohn, Ocke-Schwen. 2021. The revised Speech Learning Model (SLM-r). In Wayland, Ratre (ed.), *Second language speech learning: Theoretical and empirical progress*, 3–83. Cambridge University Press. DOI: <https://doi.org/10.1017/9781108886901.002>
- Flege, James E. & Bohn, Ocke-Schwen & Jang, Sunyoung. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25(4). 437–470. DOI: <https://doi.org/10.1006/jpho.1997.0052>
- Halle, Morris, & Mohanan, Karuvannur Puthanveetil. 1985. Segmental Phonology of Modern English. *Linguistic Inquiry* 16(1). 57–116.
- Ho, Yen-kuang. 2009. *The perception and production of American English front vowels by EFL learners in Taiwan: The influence of first language and proficiency levels*. University of Kansas dissertation.
- Hu, Wei & Tao, Sha & Li, Mingshuang & Liu, Chang. 2019. Distinctiveness and assimilation in vowel perception in a second language. *Journal of Speech, Language, and Hearing Research* 62(12). 4534–4543. DOI: https://doi.org/10.1044/2019_JSLHR-H-19-0074

- Jia, Gisela & Strange, Winifred & Wu, Yanhong & Collado, Julissa & Guan, Qi. 2006. Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America* 119(2). 1118–1130. DOI: <https://doi.org/10.1121/1.2151806>
- Johnson, Keith. 2012. *Acoustic and auditor phonetics*. John Wiley & Sons.
- Kwon, Joy. 2021. *Defining perceptual similarity with phonological levels of representation: Feature (mis)match in Korean and English*. University of Wisconsin-Madison dissertation.
- Laeufer, Christiane. 1996. Towards a typology of bilingual phonological systems. In James, Allan & Leather, Jonathan (eds.), *Second-language speech: Structure and process*, 325–342. De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110882933.325>
- Lahiri, Aditi & Reetz, Henning. 2002. Underspecified recognition. In Gussenhoven, Carlos & Warner, Natasha (eds.), *Laboratory phonology 7*, 637–676. De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110197105.2.637>
- Lai, Yi-hsiu. 2010. English vowel discrimination and assimilation by Chinese-speaking learners of English. *Concentric: Studies in Linguistics* 36(2). 157–182.
- Lardiere, D. 2009. Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research* 25(2). 173–227. DOI: <https://doi.org/10.1177/0267658308100283>
- Liu, Chang & Jin, Su-Hyun & Chen, Chia-Tsen. 2014. Durations of American English vowels by native and non-native speakers: Acoustic analyses and perceptual effects. *Language and Speech* 57(2). 238–253. DOI: <https://doi.org/10.1177/0023830913507692>
- Major, Roy. 1987. Phonological similarity markedness, and rate of L2 acquisition. *Studies in Second Language Acquisition* 9(1). 63–82. DOI: <https://doi.org/10.1017/S0272263100006513>
- Mi, Lin & Tao, Sha & Wang, Wenjing & Dong, Qi & Guan, Jingjing & Liu, Chang. 2016. English vowel identification and vowel formant discrimination by native Mandarin Chinese- and native English-speaking listeners: The effect of vowel duration dependence. *Hearing Research* 333. 58–65. DOI: <https://doi.org/10.1016/j.heares.2015.12.024>
- Natvig, David. 2017. A model of underspecified recognition for phonological integration: English loan vowels in American Norwegian. *Journal of Language Contact* 10(1). 22–55. DOI: <https://doi.org/10.1163/19552629-01001003>
- Purnell, Thomas C. & Raimy, Eric. (2015). Distinctive features, levels of representation, and historical phonology. In Honeybone, Patrick & Salmons, Joseph (eds.), *The Oxford handbook of historical phonology*, 522–544. Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199232819.013.002>
- Purnell, Thomas C. & Raimy, Eric & Salmons, Joseph. 2019. Old English vowels: Diachrony, privativity, and phonological representations. *Language* 95(4). e447–e473. DOI: <https://doi.org/10.1353/lan.2019.0083>
- Raimy, Eric. *Contrastive specification for Mandarin vowels*. Manuscript.

- Sebastián-Gallés, Núria & Díaz, Begoña. 2012. First and second language speech perception: Graded learning. *Language Learning* 62. 131–147. DOI: <https://doi.org/10.1111/j.1467-9922.2012.00709.x>
- Sebastian-Gallés, Núria & Rodríguez-Fornells, Antoni & de Diego-Balaguer, Ruth, & Díaz, Begoña. 2006. First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience* 18(8). 1277–1291. DOI: <https://doi.org/10.1162/jocn.2006.18.8.1277>
- Sebastián-Gallés, Núria & Soto-Faraco, Salvador. 1999. Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition* 72(2). 111–123. DOI: [https://doi.org/10.1016/S0010-0277\(99\)00024-4](https://doi.org/10.1016/S0010-0277(99)00024-4)
- Stevens, Kenneth & Keyser, Samuel & Kawasaki, Haruko. 1986. Toward a phonetic and phonological theory of redundant features. In Perkell, Joseph & Klatt, Dennis (eds.), *Variance and invariability in speech processes*, 426–449. Hillsdale: Erlbaum.
- White, Lydia. 1988. Universal grammar and language transfer. In Pankhurst, James (ed.), *Learnability and second languages: A book of readings*, 36–60. De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110874150-003>
- Xiahou, Xiafan. 2012. *Cross-language transfer in acquisition of English front vowels: Mandarin to English*. Murray State University dissertation.
- Xu, Shirong. 1980. *Putonghua Yuyin Zhishi* [Phonology of Standard Chinese]. Beijing: Wenzhi Gaige Chubanshe.
- Yu, Zhaoru. 2012. *The production and the perception of English vowels by Mandarin speakers*. University of Victoria dissertation.
- Zhang, Xi. 1996. *Vowel systems of the Manchu-Tungus languages of China*. University of Toronto dissertation.

