

Acoustic Cue-Weighting in L1 and L2 English Vowel Perception

Ziyue Zheng

MSc in Applied Linguistics and Second Language Acquisition,
2023

DECLARATION BY THE CANDIDATE AS AUTHOR OF THE DISSERTATION

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.
2. I allow the Department to deposit on my behalf a copy of this dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].
3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals (unless permission has been obtained from the individuals) and that such material should be removed before this dissertation is deposited in ORA.
4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [https://ora.ox.ac.uk/deposit_agreements] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.
5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]:	Zifue Zheng
Date:	2023.8.8

Table of Contents

Abstract 1
Acknowledgments 2
List of Tables 3
List of Figures 4
Chapter 1 Introduction 5
Chapter 2 Literature review 8
2.1 Factors influencing L2 cue weighting 8
2.1.1 The cross-linguistic influence on L2 cue weighting 8
2.1.2 Influence of perceptual cue salience on L2 cue weighting 10
2.1.3 The role of instruction on L2 cue weighting 12
2.2 Mandarin L2 English learners' perceptual cue weighting of /i:/-/ɪ/ 13
2.2.1 Mandarin vowel inventory 13
2.2.2 Native Mandarin speakers' perceptual cue weighting of English /i:/-/ɪ/ 14
2.3 Phonetic training studies to shift cue-weighting patterns 15
2.4 Summary and research gaps 16
2.5 The current study 17
Chapter 3 Experiment 1 19
3.1 Methodology 19
3.2 Results 25
3.2.1 Cue weighting patterns of native English speakers and L2 English learners 25
3.2.2 Individual differences in the cue weighting patterns 29
3.2.3 Potential factors affecting L2 English learners' cue weighting patterns 30
3.2.4 Summary of results of Experiment 1 33
3.3 Discussion 33
3.3.1 Different response patterns of the two stimuli pairs 33
3.3.2 The role of previous instruction 35
Chapter 4 Experiment 2 36
4.1 Methodology 36
4.2 Results 39
4.2.1 Data trimming 39
4.2.2 Results of the pretest 39
4.2.3 Treatment's role on shifting the cue weighting patterns of L2 English learners 42

4.2.4 Summary of Experiment 2 45
4.3 Discussion 45
Chapter 5 General Discussion 46
5.1 RQ1: the L1 and L2 cue weighting patterns at the group level 46
5.2 RQ 2: individual differences 47
5.3 RQ 3: the influence of instruction 49
5.4 Pedagogical implications 51
5.5 Limitations 51
5.6 Conclusion 52
References 53
Appendix 1: CUREC Approval Letter 58
Appendix 2: Participant Information Sheet for Experiment 1 (for native English speakers) 59
Appendix 3: Participant Information Sheet for Experiment 1 (for L2 English learners) 61
Appendix 4: Participant Information Sheet for Experiment 2 63
Appendix 5: Recruitment Poster for Experiment 1(for L2 English learners) 65
Appendix 6: Recruitment Poster for Experiment 2 66
Appendix 7: Consent Form (English version) 67
Appendix 8: Consent Form (Chinese version) 68
Appendix 9: Language Background Questionnaire for Native English Speakers in Experiment 1 69
Appendix 10: Language Background Questionnaire for L2 English Speakers in Experiment 1 (English version) 70
Appendix 11: Language Background Questionnaire for L2 English Speakers in Experiment 1 (Chinese version) 71
Appendix 12: Treatment Materials for the Experimental Group in Experiment 2 (Chinese version) 72
Appendix 13: Treatment Materials for the Experimental Group in Experiment 2 (English version) 75
Appendix 14: Examples of R Code 78
Appendix 15: Individual weighting scores and transformed scores 79

Abstract

Listeners identify phonemes of a language via acoustic cues. For instance, the English /i:/-/ɪ/ vowels are distinguished by the vowel length and spectral range between the first (F1) and second (F2) formant frequency of the vowels. The research in this dissertation examined perceptual weighting of two acoustic cues, namely vowel duration and spectra, in the perception of English /i:/-/ɪ/ vowel contrasts by native speakers of British English and Chinese L2 learners of English. First, it asked how native speakers and L2 English learners weighted the cues at the group level. Second, it investigates the individual differences for the native and L2 group and what had caused such variability in the L2 speakers. Third, this research examined instruction's role on cue weighting of English /i:/-/ɪ/. In Experiment 1, a total of 29 native speakers of British English and 29 Chinese L2 learners of English identified the vowels in synthesized 'bead-bid' and 'steet-stit' continua which differed in 7 spectral and temporal steps. After the forced-choice task, each participants completed a background questionnaire. Experiment 2 adopted a pre- and post-test design, where a total of 48 L1-Mandarin L2 English learners (23 participants in the control group and 25 in the experimental group) completed similar word identification task as in Experiment 1. The experimental group completed a treatment session where they were explicitly instructed in the differences between English /i:/-/ɪ/, while the control group heard the same minimal pairs of the contrast without the instruction. Participants in Experiment 2 also completed the background questionnaire. The results showed that, at the group level, both the native speakers and L2 learns relied predominantly on vowel duration. Substantial variability of cue weighting behaviors was found at the individual level for both L1 and L2 English speakers, but neither self-perceived English proficiency nor experience of living in English-speaking countries were found to have significant influence on cue weighting at the group level. Regarding the role of instruction, Experiment 1 found previous instruction as a significant predictor of the L2 cue weighting pattern. Experiment 2 found no evidence of such influence of instruction from the correlational perspective, but it did find that, from the causal perspective, intervention of explicitly instructing the participants about the differences between English /i:/-/ɪ/ had led the experimental group to down-weigh the vowel duration.

Acknowledgments

For this dissertation, I am extremely grateful to Professor Elizabeth Wonnacott, my supervisor, for her very useful advice and guidance throughout this research, and, of course, her tireless efforts to review and provide feedback on this work. Thanks are also due to my boyfriend, who is now my fiancé, for helping me to recruit participants. I would also like to express my gratitude to all the participants, who were actively engaged in my experiments. Special thanks to my friend Rachel, without whose company the last year at Oxford would be a lot more challenging.

List of Tables

Table 1 Self-perceived English proficiency of native Mandarin speakers in Experiment 1 ... 20
Table 2 Third languages spoken by native Mandarin speakers in Experiment 1 20
Table 3 Third languages spoken by native English speakers in Experiment 1 20
Table 4 Acoustic information of the 'bead', 'bid', 'steet' and 'stit' natural speech stimuli 21
Table 5 Model output of the 29 L2 English speakers and the 29 native English speakers for the 'bead-bid' and 'steet-stit' stimuli 28
Table 6 Model output of the 29 native English speakers and 29 L2 learners' overall cue weighting patterns for the 'bead-bid' stimuli 28
Table 7 Model output of the 29 native English speakers and 29 L2 learners' overall cue weighting patterns for the 'steet-stit' stimuli 28
Table 8 Summary of individual cue weighting behaviors of the 29 native English speakers and 29 L2 learners. 30
Table 9 Results of the Mann-Whitney test comparing the 29 native speakers and 29 L2 learners in Experiment 1. 30
Table 10 Output of model examining previous instruction's influence on the cue weighting patterns for the 23 L2 English speakers 31
Table 11 Statistical output of the mixed-effects logistic regression model to examine the cue weighting pattern for the 13 L2 English speakers who had received proper instruction 32
Table 12 Statistical output of the mixed-effects logistic regression model to examine the cue weighting pattern for the 10 L2 English speakers who had received improper instruction 32
Table 13 Self-perceived English proficiency of the control group 36
Table 14 Third languages spoken by the control group 36
Table 15 Self-perceived English proficiency of the experimental group 37
Table 16 Third languages spoken by the experimental group 37
Table 17 Model output for the 44 L2 English learners' overall cue weighting pattern in the pretest of Experiment 2. 40
Table 18 Results of the Mann-Whitney test comparing the transformed scores of native speakers in Experiment 1 (N=29) and L2 learners in both Experiments. 41
Table 19 Output of model examining the influence of Instruction and Major on the cue weighting patterns of the 41 L2 English speakers in Experiment 2. 42
Table 20 Statistical output of the mixed-effects logistic regression model to account for the test-group interaction 44

List of Figures

Figure 1 IPA chart of Mandarin vowels 13
Figure 2 Arrangement of Mandarin /i/ and the English /i:/ and /ɪ/ regarding the tongue height and position. 14
Figure 3 Distribution of vowels in the stimuli used for the word recognition task. 21
Figure 4 Procedure of each trial for the word recognition task. 23
Figure 5 Cue weighting patterns of the 29 native speakers and 29 L2 learners for the 'bead-bid' pair and 'steet-stit' pair. 26
Figure 6 Distribution of transformed scores of native speakers (N=29) and L2 learners (N=29) for the 'bead-bid' and 'steet-stit' stimuli. 30
Figure 7 Design of Experiment 2. 38
Figure 8 Heat graph of the 44 L2 English learners' response pattern in Experiment 2. 40
Figure 9 Distribution of transformed scores of native speakers in Experiment 1 (N=29) and L2 learners in both Experiments (N=73) for the 'bead-bid' stimuli. 41
Figure 10 The control group and experimental group's beta coefficients of Spec and Dur in pretest and post-test. 43
Figure 11 Distribution of gain scores of the control group (N=20) and experimental group (N=24). 44

Chapter 1 Introduction

The different sounds of a language are distinguished by acoustic properties, which are combined to distinguish phonological contrasts. For example, English voiced and voiceless consonants differ in two acoustic dimensions—the voice onset time (VOT) and the fundamental frequency (F0) of the following vowel (Schertz & Clare, 2020). In comparison with the voiced consonant (such as /d/ as in 'dear'), the voiceless consonant (such as /t/ in 'tear') tends to have longer VOT and the following vowel is more likely to have higher F0 (Lehet & Holt, 2017). In the case of the lax-tense English vowel contrasts, the tense vowel /ɪ/ (as in 'ship') has a shorter duration and narrower spectrum between the frequencies of the first (F1) and second (F2) formant, while its lax counterpart /i:/ (as in 'sheep') has longer vowel length and larger spectrum (Hillenbrand, Clark, & Wheeler, 1995).

Listeners rely on these acoustic cues to distinguish phonological contrasts. But the multiple dimensions do not necessarily contribute equally to the acoustic differences within a contrast (Holt & Lotto, 2006; Nittrouer, 2004). Therefore, listeners assign different weight to different acoustic dimensions when perceiving the sound pairs. This process is known as the 'perceptual cue weighting'. Adult native listeners' perceptual cue weighting patterns usually reflect the acoustic regularities of the speech production environment. For instance, the /r/ and /l/ sounds produced by native American English speakers differ in terms of the F2 and the F3 of the onset, with the latter being a more distinct acoustic feature than the former (Idemaru & Holt, 2013; Lotto, Sato, & Diehl, 2004). Accordingly, native American English listeners have been found to rely more on the F3 than then F2 when categorizing the two sounds (e.g., Ingvalson, McClelland, & Holt, 2011; Iverson et al., 2003).

However, given that for many languages most speakers are second language learners, an important question is how perceptual cue weighting differs for non-native speakers. Therefore, researchers have also investigated the perceptual cue weighting patterns of non-native speakers of a language regarding various speech contrasts at both segmental (e.g., Cerviño-Povedano & Mora, 2015; Kivistö-De Souza, Carlet, Julkowska, & Rato, 2017) and suprasegmental (e.g., Baek, 2022; Kim & Tremblay, 2021) levels. Results suggest that cue weighting behaviors of L2 listeners tend to differ from native speakers. For instance, Lipski, Escudero and Benders (2012) investigated Spanish listeners' cue weighting of Dutch vowels. They found that participants relied heavily on duration to distinguish Dutch vowels, while Dutch listeners relied primarily on spectral cue. Such discrepancy between L1 and L2 listeners suggests the perceptual cue weighting pattern might be largely influenced by listeners' language experience. The focus of the current dissertation is also L2 perceptual cue weighting, and it looks at the context of the English /i:/-/ɪ/ contrast comparing native English speakers and Chinese learners of English in particular.

L1 cue weighting of English /i:/-/ɪ/

The English vowels /i:/ and /ɪ/ differ in vowel length and the spectrum between the first (F1) and second (F2) vowel formant frequency. The vowel spectrum serves as the primary difference between the contrast, since native English speakers produce the vowel spectrum of the two vowels contrastively (Cohen, 1971). In comparison, vowel duration in nature is not a contrastive feature in English. But English does make a distinction between lax and tense vowels, which may influence the vowel length with the former being slightly shorter than the latter (Crystal & House, 1988). In the case of the English /i:/-/ɪ/, /i:/ has larger vowel spectrum and is the tense vowel with longer duration, while /ɪ/ has smaller spectrum with shorter vowel length (Hillenbrand et al., 1995). There is a close correspondence between such distributional properties of the English vowel inventory and L1 English speakers' perception behaviors of /i:/ and /ɪ/ (Lehet & Holt, 2017; Schertz & Clare, 2020). There have been several speech perception studies which found native English speakers to rely strongly and almost exclusively on vowel spectrum (e.g., Bohn, 1995; Hillenbrand, Clark, & Houde, 2000). However, it is notable that the majority of these studies have been done on American English rather than other English varieties. Few studies have been done on other dialects, though one by Escudero and Boersma (2004) focusing on British English found more mixed findings.

L2 cue weighting of English /i:/-/ɪ/

In terms of L2 cue weighting in English, the differentiation of the /i:/-/ɪ/ contrast is of great importance for L2 learners. This is because this contrastive vowel pair is highly functional in English, as the /i:/-/ɪ/ contrast distinguishes many high-frequency English word pairs (Kivistö-De Souza et al, 2017). As is listed in Higgins (2013), 466 minimal pairs of English words are distinguished by the /i:/-/ɪ/ contrast.

However, extensive studies have found that it might be difficult for L2 English learners to accurately categorize this vowel pair (e.g., Nishi & Kewley-Port, 2007; Ruan & Saito, 2023). Ruan and Satio (2023) investigated Chinese English learners' categorization of English /i:/-/ɪ/ and found a mean accuracy of 70.2% for the experimental group and 73.9% for the control group; In Nishi and Kewley-Port (2007), Japanese participants identified English /i:/ and /ɪ/ among other English vowels with a nine-alternative forced-choice task, and their accuracy was less than 75% and 50% respectively. One possible reason for such difficulty could be because native speakers of English discriminate the /i:/-/ɪ/ based not only on vowel length but more on the vowel spectrum, whereas these L2 learners with various L1 backgrounds such as Japanese (e.g., Udovchenko, 2016; Yazawa, Whang, Kondo, & Escudero, 2020), Spanish (e.g., Bohn, 1995; Escudero & Boersma, 2004) and Mandarin (e.g., Bohn, 1995; Wang & Munro, 1999) have been found to rely heavily and almost exclusively on the vowel duration. Therefore, L2 English learners might experience some

might pose some difficulties in correctly distinguishing this English vowel pair.

For some of these languages, such as Japanese, participants' heavy reliance on the temporal cue might be caused by L1 influence (Grenon & Sheppard, 2019; Yazawa et al., 2020), since Japanese distinguishes the /ii/-/i/ vowels which contrast only in duration rather than the spectral dimension. In comparison, although Chinese English learners have also been found to relied predominantly on vowel duration, what has caused such cue weighting behaviors is not clear. This is because Mandarin does not have the contrastive pair similar to the /iː/-/ɪ/ contrast in English. Instead, it has a vowel spectrally similar to the English /iː/ sound. Thus, results and implications from existing studies on L1 speakers of other languages such as Japanese cannot be generalized to Mandarin speakers. Therefore, it is important to investigate Chinese English learners' cue weighting of the /iː/-/ɪ/ English contrast and, more importantly, the rationales behind their cue weighting patterns. It is also worth looking into whether their cue weighting patterns could be shifted to approximate cue weighting behaviors of the L1 English speakers. In this way, this study may hopefully contribute to helping Mandarin-L1 EFL learners discriminate English vowels more accurately.

Structure of dissertation

The introduction is followed by Chapter 2 Literature Review, where the influencing factors of L2 perceptual cue weighting are discussed. Studies examining Chinese English learners' cue weighting of English /iː/-/ɪ/ are also discussed before moving on to phonetic training studies which aimed to shift L2 cue weighting patterns. Research objectives and an overview of the research paradigm of the current study are presented in the end of Chapter 2. Chapter 3 Experiment 1 and Chapter 4 Experiment 2 present the two experiments conducted to address the research questions. For each chapter, methodology and results of the experiment are presented first, followed by a discussion of findings of this experiment. In Chapter 5, there is a general discussion which aims to link the two experiments and revisit the research objectives. Finally, the pedagogical implications and limitations of this research are discussed before the conclusion.

Chapter 2 Literature review

This chapter begins by discussing three factors which may have caused the different cue weighting patterns of L2 learners from L1 speakers, namely the cross-linguistic influence, the perceptual salience of certain cues, and the role of instruction. More specifically, the literature on Chinese English learners' cue weighting of English /i:/-/ɪ/ is reviewed. Then, this chapter discusses previous attempts to shift listeners' perceptual cue weighting patterns, followed by a summary and identification of research gaps. Finally, outline of the current research as well as the research questions are presented.

2.1 Factors influencing L2 cue weighting

2.1.1 The cross-linguistic influence on L2 cue weighting

The L2 cue weighting pattern could be influenced by L2 speakers' experience with phonetic distributions in their L1 (Grenon & Sheppard, 2019; Zhang, 2012). Speakers may pay more attention to the cues carrying more robust phonetic information in their native language. When the L1 and L2 phonetic systems differ systematically, L2 learners' cue weighting pattern may deviate from L1 speakers. Evidence supporting the cross-linguistic influence has shown that L2 learners' cue weighting patterns in the target language vary according to their L1 backgrounds (e.g., Flege, Bohn, & Jang, 1997; Udovchenko, 2016). For instance, Escudero, Benders and Lipski (2009) investigated how Dutch, German and Spanish listeners perceived the Dutch /a:/-/a/ contrast. Results showed that native Dutch and L1-German L2-Dutch listeners tended to be spectral listeners, while the L1-Spanish group relied more on the duration cue. The perceptual behavior of the German listeners can be explained by L1 transfer, since German and Dutch share similar spectral and duration values for the /a:/-/a/ contrast. As a result, these two groups of participants showed similar cue-weighting patterns.

Models and theories have been proposed to systematically account for how L1 phonetic structure influences L2 speech perception patterns, among which the Speech Learning Model (SLM) (Flege, 1995) focuses on discriminating phonemic contrasts rather than just perception of individual non-native sounds. Therefore, the SLM could account for the relationship between L1 and L2 in terms of L2 cue weighting (Harnsberger, 2001). To begin with, the SLM assumes that L2 learners, when judging an L2 sound, tend to relate it to the closest phone in their L1. Therefore, the existing L1 sound categories may shape the way L2 listeners perceive the L2 sounds (Chan, 2012). This model specifies two possible ways of how the L1 sound categories influences L2 sound categorization, namely assimilation and dissimilation. Assimilation of L1 and L2 sound categories happens when the perceived phonetic difference between the target L2 sound and its closest L1 sound is not great that listeners fail to detect it. As a result, they assimilate the L2 sound and this L1 sound into the

same category. Under such circumstances, listeners one single phonetic category is adopted to process the L1 and L2 sounds which are perceptually similar. Therefore, the cue weighting patterns for the L1 and L2 sounds might be similar. However, dissimilation of the L1 and L2 sounds may take place as long as listeners can detect some phonetic dissimilarities between the L1 and L2 sounds. Then, listeners will deliberately 'deflect' away from the L1 category when perceiving L2 sound in order to retain distinct contrasts between the L1 and L2 sound categories. This will lead to the formation of a new phonetic category for the L2 sound, which strives to steer away from the L1 category.

Studies in perceptual cue weighting and speech categorization have yielded results which fit into the SLM framework. For instance, Yazawa et al. (2020)[1] investigated the cue weighting patterns of Japanese-L1 English-L2 speakers in terms of how they categorized the English /i:/-/ɪ/ contrast and the Japanese /ii/-/ɪ/ contrast. They found that when perceiving the Japanese contrasts, participants used duration as the primary cue and vowel spectrum as the secondary. But the patterns were mixed regarding English vowel classification: some participants still relied on duration as the primary cue, while others demonstrated strong reliance on the spectral cue. This discrepancy can be interpreted under the SLM framework. On the one hand, those who consistently used duration might be listeners who failed to notice the phonetic differences between the English and Japanese vowel contrasts. As a result, their perception of each sound in the English /i:/-/ɪ/ pair is assimilated to a different L1 category (English /i:/ to Japanese /ii/ and English /ɪ/ to Japanese /ɪ/). On the other hand, those who shifted to the spectral cues when perceiving English sounds might have formed new phonetic categories because they have noticed the phonetic difference. Another study which provides some evidence for the SLM frame was Grenon and Sheppard (2019). They found that native Japanese speakers distinguished English vowels /i:/-/ɪ/ largely based on duration. After an eight-week phonetic training session, some participants formed new vowel categories which weighted vowel spectrum as the primary cue. But the newly formed category was still different from that of native speakers' in that the boundary of categorization moved further away from the category center of the English /ɪ/ in comparison to the boundary held by native English speakers. The authors interpreted this to be a result of the dissimilation mechanism: in order to maintain two distinct systems, participants who noticed the phonetic difference between the Japanese and English vowel contrasts after training strove to create a new category for L2 vowels as far away from the L1 correspondence as possible.

[1] The research paradigm adopted in Yazawa et al. (2020) is the most widely used paradigm in perceptual cue weighting research. They created 49 artificially manipulated acoustic stimuli to form an acoustic continuum between /i:/ and /ɪ/ along the two acoustic dimensions respectively, namely vowel duration and spectrum. There were seven acoustic steps along each dimension, and the two dimensions co-varied. Therefore, a total of 49 (7*7) stimuli were generated, with one stimulus having the the most /i:/-like properties in terms of vowel spectrum and duration, one stimulus having the most and /ɪ/-like properties, and the other 47 stimuli falling in the middle of these two stimuli. These stimuli were presented to participants, and they had to identify which of the two sounds they thought they had heard using a binary decision task.

As we can see, the SLM systematically emphasizes the importance of phonetic similarities and differences between L1 and L2 in L2 cue weighting and phonetic categorization (Chan, 2012). But it is worth noting that evidence supporting the cross-linguistic influence on cue weighting relies on the precondition that there are either similarities or differences between L1 and L2 in terms of weight assigned to the target cues. For the Japanese-L1 speakers investigated in Yazawa et al. (2020) and Grenon and Sheppard (2019), their behavior of assigning greater weight to duration when differentiating the English /i:/-/ɪ/ could be due to duration's dominance in differentiating the Japanese /ii/-/i/ contrast. But cross-linguistic influence might not be able to account for L2 listeners' behaviors when the cues differentiating the target L2 contrast are essentially absent in L1. For example, Previous studies have shown that vowel duration serves as the predominant cue for Spanish (Escudero, 2001; Escudero & Boersma, 2004) and Mandarin (Bohn, 1995; Wang & Munro, 1999) English learners. Unlike Japanese English learners, these Spanish and Mandarin participants' strong reliance on duration cannot be interpreted by the cross-linguistic influence of L1 features on L2 speech perception. This is because neither vowel duration nor spectrum carries enough phonetic information in Mandarin and Spanish. First, vowels in both Spanish and Mandarin are not distinguished by duration, which suggests duration does not carry enough phonetic information in Mandarin and Spanish. Second, vowel spectrum should not carry much phonetic information for native speakers of the two languages either, because both Spanish and Mandarin only have one vowel category in the high-front area of vowel space—a vowel spectrally similar to the English /i:/. As a result, they do not use spectral cues to distinguish vowel contrasts in that acoustic vowel space. In this sense, neither of the two cues—vowel spectrum and duration—distinguishing English /i:/-/ɪ/ contrasts carries enough phonetic information for L2 English speakers of Mandarin and Spanish. Yet results of these studies above have indicated that both groups only relied heavily on duration but not vowel spectrum when distinguishing the two English vowels. Thus, cross-linguistic influence cannot be the only factor that determines L2 cue weighting, despite its significant role in speech perception. There must be other factors at play in the cue weighting process.

2.1.2 Influence of perceptual cue salience on L2 cue weighting

Perceptual salience of cues is another factor which may compensate for the limited explanatory power of cross-linguistic influence on L2 cue weighting, since perceptual salience of cues may influence L2 cue weighting behaviors regardless of listeners' L1 background (Kajouj, Kager, Acquisition, & Engelse, 2019). Holt and Lotto (2006) and Pisoni (1977) suggested that some acoustic dimensions, such as vowel length and VOT of consonants, are more robustly encoded by the human auditory system, therefore more discriminable compared with other acoustic dimensions. Cues with higher discriminability

are more acoustically informative. Therefore, listeners are likely to assign more perceptual weight to this cue when discriminating L2 phonetic contrasts.

In terms of L2 cue weighting of English vowels, for example, two acoustic dimensions—vowel length and vowel spectrum—distinguish the English vowel contrast. Since the temporal cue has been suggested to be a generally salient cue in comparison with the spectral cue (Bohn, 1995; Kajouj, Kager, Acquisition, & Engelse, 2019; Kivistö-De Souza et al, 2017), vowel duration in this case is more universally accessible to L2 listeners. This may explain why Spanish and Mandarin participants rely heavily on vowel duration rather than spectrum even though neither of the two cues is informative in participants' L1. For instance, Bohn (1995) found that the dominant cue in vowel categorization tasks for both Mandarin and Spanish participants was vowel duration. Bohn attributed the dominance of duration solely to the L1-independent perceptual salience of temporal cues. Based on this, Bohn put forward the 'Desensitization Hypothesis' to explain such phenomenon. According to this hypothesis, Spanish and Mandarin only have one vowel category in the high-front area of vowel space, so they do not use spectral cues to distinguish vowel contrasts in that acoustic vowel space. Thus, L1 speakers of Spanish and Mandarin have been 'desensitized' to the spectral cue in that vowel space. In this sense, whenever the previous linguistic experience desensitizes the L2 learners to the spectral cue, the duration cue will be applied as primary cue to distinguish the L2 vowel contrasts, because it is more perceptually salient and easier to access.

However, Bohn (1995) was too absolute when attributing the results to salience of vowel duration while ignoring other possibilities which may have also contributed to the results. This is because one crucial characteristic of the participants was ignored in Bohn's (1995) explanation: the selection criterion of participants was 'minimum exposure to authentic English' (p. 288). In other words, their knowledge of the English /i:/-/ɪ/ came mostly from their non-native teachers rather than from the authentic English environment. As a result, the English they heard from their teachers might differ in the weightings of the cues compared to input received from native speakers, which could be the cause of their cue weighting patterns. Furthermore, it was also likely that the explicit instruction from the teachers might have a major influence on how these English learners perceived the English sounds. In other words, these participants' reliance on duration might be due to the improper instruction from their teachers specifying that the difference between English /i:/-/ɪ/ only lies in duration. Thus, although acoustic cue salience might be a plausible explanation of participants' reliance of duration, it might not be the only factor contributing to Spanish and Mandarin speakers' heavy reliance on vowel length. We cannot rule out the possibility that such pattern was caused by input from their English teachers or improper instruction students had received from their teachers. Therefore, the next section focuses on the role of explicit instruction on L2 perceptual cue weighting.

2.1.3 The role of instruction on L2 cue weighting

Regarding the influence of previous instruction on cue weighting, relevant studies focused on instruction of English vowels and found that teachers indeed tended to emphasize the vowel duration when teaching students the difference between English vowel contrasts (Schertz, Lotto, & Warner, 2015; Wang & Munro, 1999; Wang & Munro, 2004). For example, almost all Mandarin participants in Wang and Munro's (1999) research on English vowel categorization reported that their English teachers had characterized the difference between English /i:/-/ɪ/ as difference in length; Flege et al. (1997) also reported that Korean English learners had received improper instruction in schools that English /i:/-/ɪ/ contrast should be distinguished by length only; Considering this, the results of Bohn (1995) may be at least in part due to the inappropriate instruction that the Spanish and Mandarin speakers had previously received.

However, one limitation of studies discussing the role of previous instruction is that instruction's role has remained an assumption, but whether and to what extent it actually plays a role in shaping L2 cue weighting patterns have never been addressed properly. For instance, Flege et al. (1997) found Chinese and Korean students rely heavily on vowel duration and interpreted such pattern to be possibly due to the inappropriate instruction participants received. But this remained an assumption since Flege et al. did not investigate the instruction participants actually received and relied entirely on anecdotal knowledge of how EFL learners were instructed in countries such as China and Korea. On the other hand, Wang and Munro (1999) paid more attention to the role of instruction by asking L1-Mandarin participants to report how they had been taught about the difference between the English /i:/-/ɪ/. They found that—at the group level—participants relied predominantly on duration, while all of them reported to have been instructed that the only difference between /i:/ and /ɪ/ lies in duration. But, again, this does not necessarily mean incorrect instruction is correlated with preference on vowel duration. This is because all the participants reported incorrect instruction, and there was no sample of participants who had not had appropriate instruction. Thus, we do not know what the latter's cue weighting behaviors would be. If listeners who had been properly instructed relied on duration as heavily as those who had received improper instruction, it may indicate that instruction does not influence the L2 cue weighting patterns. As we can see, previous studies may have indeed confirmed the existence of the problem that EFL learners have received improper instruction on English /i:/-/ɪ/. But essentially whether and how instruction affects L2 cue weighting has not been thoroughly explored.

2.2 Mandarin L2 English learners' perceptual cue weighting of /i:/-/ɪ/

2.2.1 Mandarin vowel inventory

It is generally agreed that Mandarin has a five-vowel phonemic system (Hao, 2018). The five contrastive phonemes in Mandarin are accepted as /i, u, y, e, a/. The five Chinese vowels can be categorized as high (/y, i, u/), mid (/e/) and low (/a/) vowels in terms of height of the tongue, and they can also be classified into front rounded vowel (/y/), unrounded vowel (/i, e, a/) and back rounded vowel (/u/) by the lip gesture and tongue position (Cheng, 1966). Figure 1 shows the IPA chart of the five vowels (Lee & Zee, 2003).

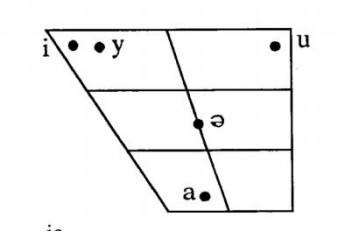


Figure 1 IPA chart of Mandarin vowels

The Mandarin vowels do not use duration contrastively to signal segmental contrasts, and Mandarin does not distinguish lax and tense vowels like English does. Therefore, duration does not function as an acoustic cue in Mandarin vowel perception. But it is worth noting that duration serves as a cue for Mandarin tonal contrasts, with the tone 3 consistently longer than tone 2 (Blicher, Fiehl, & Cohen, 1990). Be that as it may, it has been suggested that such L1 experience of tonal duration may have only marginal, if any, influence on Mandarin speakers' categorization of English vowel contrasts (Bohn, 1995), since the duration difference in Mandarin is confined to the suprasegmental level and is deemed as a reliable cue to Mandarin tonal contrasts only when listeners are differentiating tone 2 and 3.

Directly relevant here is that Mandarin only has one vowel category (/i/) in the high-front acoustic vowel space, where English has two vowels (/i:/ and /ɪ/). Figure 2 illustrates the arrangement of Mandarin /i/ as well as the English /i:/ and /ɪ/ regarding the tongue height and position. As we can see, properties of the Mandarin /i/ approximate the English /i:/ compared with the English /ɪ/. Since English has two contrasting vowels in this high-front vowel space, which are contrastive in their vowel spectra, native English speakers tend to be sensitized to the spectral difference when identifying the vowel contrast. In comparison, Mandarin has only one vowel category in this acoustic area. Therefore, native Mandarin speakers might be theoretically desensitized to spectral differences between vowels in the high-front area of the acoustic vowel space (Bohn, 1995). Overall, neither vowel spectrum nor duration carries sufficient acoustic information in the high-front vowel space of Mandarin. Although duration might be a cue in tone differentiation, it is marginal at best in terms of vowel perception.

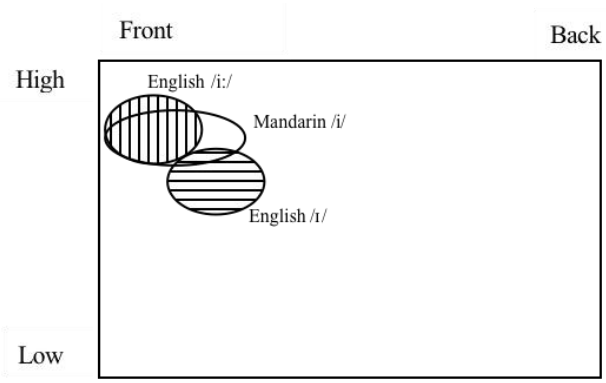


Figure 2 Arrangement of Mandarin /i/ and the English /i:/ and /ɪ/ regarding the tongue height and position.

2.2.2 Native Mandarin speakers' perceptual cue weighting of English /i:/-/ɪ/

Results of the few empirical studies so far investigating Chinese English learners' categorization of English /i:/-/ɪ/ indicate that duration serves as the primary cue for Chinese English learners (Bohn, 1995; Wang & Munro, 1999). For example, Bohn (1995) examined how Mandarin-L1 English learners, who had minimum exposure to authentic English, weighted the two cues in English /i:/-/ɪ/ vowel perception tasks. The Mandarin participants relied almost exclusively on duration, and Bohn attributed this pattern to vowel length's perceptual salience. In Wang and Munro (1999), the Chinese participants relied on both vowel duration and spectrum when completing the vowel identification task, but overall they gave much more weight to duration. Moreover, Yang, Shi, Liu and Zhao (2016) found that compared with the the American participants, the Chinese participants in general weighted duration more heavily in comparison with spectral quality when identifying English /i:/-/ɪ/.

Although all of these studies on Chinese L2 English learners suggest the predominance of vowel duration over spectrum, a limitation is that these few studies focused more on analyzing data of all participants as a whole to yield an overall pattern for participants at group level. However, individual differences within the language group might be masked by the group effect (Schertz et al., 2015). Bohn (1995) compared Mandarin speakers' cue weighting patterns of English vowels against other language groups by looking at changes of the overall percentages of 'beat-bit' response along the spectral and duration steps. Doing so generated an overall pattern representing the entire group, while individual cue weighting behaviors were not mentioned in Bohn's research. Likewise, Wang and Munro (1999) also calculated the percentages of all participants' responses on the multiple-choice questions regarding the six spectral steps respectively to form an overall tendency to account for participants' cue weighting behavior at group level. But through visual examination of percentages of each participant's response along the spectral and temporal steps, the authors were aware of the individual differences: four of the 14 native Mandarin speakers investigated relied exclusively on the spectral cues; eight demonstrated heavy reliance on duration; and two showed equal reliance on the two acoustic cues. However, the authors

only pointed out the existence of such individual differences but did not examine or discuss what had caused such individual differences. Therefore, considering the few studies on L1-Mandarin speakers' weighting of the /i:/-/ɪ/ contrast have not properly addressed the individual differences which might have been masked by the group effect, it is precipitant to conclude that the Mandarin EFL learners rely more on duration in perceiving English /i:/-/ɪ/. Therefore, further research on Mandarin speakers should be conducted, which should investigate whether there are individual differences in their cue weighting patterns as well as what might have caused the individual differences.

2.3 Phonetic training studies to shift cue-weighting patterns

There have been attempts to shift listeners' perceptual cue weighting patterns in an implicit way (e.g., Grenon & Sheppard, 2019; Holt & Lotto, 2006; Lehet & Holt, 2017; Wang & Munro, 1999). For instance, Wang and Munro (1999) adopted a pretest-training-posttest paradigm to examine how L1-Mandarin English learners perceived the English /i:/-/ɪ/ contrast and /u:/-/ʊ/ contrast and whether the cue weighting patterns could be shifted through brief training. The training stimuli for each continuum involved long and short tokens which respectively varied along the spectral dimension in six acoustic steps. Therefore, there were in total 24 stimuli in each training block randomly presented to participants, and each trial was followed by a forced-choice question as well as immediate feedback. The feedback aimed to help participants realize that they should identify the stimuli based on vowel quality regardless of the length. Before the training session, participants weighted vowel duration as the primary cue, while their perceptual performance on both /i:/-/ɪ/ and /u:/-/ʊ/ was comparable to the native English speakers after the training session, with spectrum becoming the primary cue. However, Wang and Munro (1999) did not specify whether there was a correct answer for stimuli with intermediate spectral steps. This is an essential drawback for studies like Wang and Munro (1999) which attempted to change cue weighting behaviors by providing feedback to the acoustic continuum. Because although we can label stimuli with extreme spectral values as either /i:/or /ɪ/, it is inappropriate to simply label intermediate values into the two categories. Therefore, it might be unjustifiable to tell participants their choice of the intermediate stimuli was right or wrong.

Moreover, Holt and Lotto (2006) demonstrated that the cue weighting pattern could be changed even without such feedback. They tried to shift the cue weighting patterns of a novel artificial sound contrast through short-term exposure alone. Before moving on to label the test stimuli, participants in their experiment completed a passive listening section where acoustic property of the training stimuli varied along the acoustic dimension of the weaker cue, while the stronger acoustic cue was held constant at a numerically intermediate value. Such short-term exposure reliably enhanced the weaker cue. However, it is notable that

participants did not have any previous exposure to this novel sound contrast outside the laboratory. For the real sound, in comparison, participants might have previously experienced or learned it in real life. Such accumulated knowledge and input might make it more difficult to shift cue weighting of real sound contrasts to a different direction. Therefore, the effect of such treatment based on the novel sound might not be generalized to real phonetic contrasts which language learners had already experienced in their L1 or L2 linguistic environments.

The rationale behind these training procedures is to shift listeners' attention towards the variation along the less-weighted acoustic dimension (Wang & Munro, 2004). In this way, participants may be aware of the acoustic dimension which they have attached less weight, thus relying more on this cue when differentiating the sound contrast. Therefore, it is logical to explore if cue weighting patterns could be altered through explicit instruction which raises learners' metalinguistic awareness directly.

However, one limitation of the training studies is that only few studies have examined whether cue weighting patterns of English /i:/-/ɪ/ could be shifted via explicit instruction. Grenon and Sheppard (2019) spent 8 weeks training Japanese participants to give more weight to vowel spectrum when distinguishing English /i:/-/ɪ/. Participants in Grenon and Sheppard's study were exposed to the /i:/-/ɪ/ stimuli which varied in spectral dimension, and the trainees were at the same time explicitly instructed about the differences between the vowels. The results demonstrated increased weighting of the spectral cue. However, since this study involved both explicit instruction and implicit exposure, it is impossible to tell whether explicit instruction was useful in shifting participants' cue weighting behaviors. However, other studies have explored whether instruction could influence the accuracy of speech categorization. For instance, Ruan and Saito (2023) investigated whether meaning-oriented instruction could lead to increased accuracy of native Chinese speakers in identifying the English /i:/-/ɪ/. In this study, the experimental group received 10 minutes' explicit instruction on the differences between the two vowels in terms of length and tongue position. Then they were involved in a debate task, where the content focused on argumentative writing and the instructor provided recast on the vowels at the same time. The experimental group significantly enhanced their English /i:/-/ɪ/ identification accuracy in the post-test. But it is not clear whether the accuracy was improved by changing the cue weighting patterns. Therefore, it is important to further explore how explicit instruction affects L2 speakers' cue weighting patterns.

2.4 Summary and research gaps

To sum up, L2 cue weighting patterns differ systematically from L1 speakers according to L2 learners' native language experience. Models such as SLM have been proposed to account for such cross-linguistic influence. However, cross-linguistic influence cannot be

the sole explanation for L2 speakers' deviance from the L1 group, since there are some phenomenon it fails to explain. For instance, L1 influence cannot explain why Mandarin and Spanish EFL learners relied predominantly on duration when perceiving English /i:/-/ɪ/, since neither vowel duration nor spectrum carries enough acoustic information in their L1 regarding identification of this contrast. Therefore, other factors may also be at play, such as universal salience of certain cues and previous instruction participants had received from their English teachers. However, a gap in studies discussing the role of previous instruction is that these studies did not look explicitly into whether and how instruction had affected L2 cue weighting.

This literature review section has also reviewed the existing literature on native English speakers and Chinese English learners' cue weighting patterns of the English /i:/-/ɪ/ contrast. Evidence demonstrated that native speakers of American English relied almost exclusively on the spectral cue, whereas the only one study examining native speakers of British English showed they relied on both cues at the group level and their cue weighting patterns demonstrated great individual differences. On the other hand, the Chinese English learners tended to rely on vowel duration at the group level. But the limitation is that it is not clear whether there were any individual differences among these L1-Mandarin English learners as well as what had caused the individual variability if there was any.

Besides examining the L2 cue weighting patterns, researchers have also investigated whether such patterns could be sifted through phonetics training of contrast categorization so that L2 speakers can be more native-like in cue weighting. The research gap is that trainings were mostly carried out through implicit exposure, while it is not clear whether L2 cue weighting behaviors can be shifted via explicit instruction.

2.5 The current study

Research objectives

This study aims to answer the following research questions:

1) How do native speakers of British English and L1-Mandarin L2 English learners weight the duration and spectral cues when differentiating the English /i:/-/ɪ/ vowel contrast?

2) Do Mandarin speakers demonstrate individual differences in terms of cue weighting of the English /i:/-/ɪ/? A secondary question under this objective is to investigate the potential predictors of such individual variability if there is any.

3) To what extent can cue weighting be shifted through explicit instruction?

Research paradigm

A core goal of the current study is to look at Chinese speakers' cue weightings when identifying English /i:/-/ɪ/ and to further elucidate the factors that affect L2 cue weighting. To that end, Experiment 1 used a paradigm similar to Yazawa et al. (2020) discussed above,

where participants heard acoustic stimuli which varied along five temporal steps and seven spectral steps and had to identify which of the two sounds they thought they had heard using a binary decision task. Their responses were analyzed to see the extent to which they had relied on the temporal and spectral cues respectively. Critically, data were collected regarding participants' previous language experience and in particular about the previous instruction they had received. This allowed us to include instruction as a predictor in the mixed-effect logistic regression analyses and examine whether previous instruction had an impact on cue weighting. In this way, the first gap regarding the influence of previous instruction was addressed.

For comparison purposes, Experiment 1 also tested native speakers of British English. Note that the majority of previous studies have been done with speakers of American English, and this is only the second study with British English speakers. The expectation following Escudero and Boersma (2004) is that the differences between British native speakers and Mandarin Chinese native speakers will be smaller than in the previous studies with American native speakers, since most of the native speakers of southern British English tended to rely on both vowel duration and spectrum when identifying English /i:/-/ɪ/. But this expectation needs to be tested.

As Experiment 1 took a correlational approach to looking at the role of instruction. Experiment 2 further probed if the relationship was causal and took an intervention approach, where the cue weightings at pre- and post-test were examined using the same method as in Experiment 1. But an intervention was given to the experimental group between the pre- and post-test to provide them with explicit instruction on the two cues distinguishing English /i:/ and /ɪ/. Change in weightings of cues of the experimental group was compared with a control group who received a matched intervention without the explicit instruction. This allowed us to investigate whether and how L2 listeners' reliance on the cues could be shifted when they received appropriate instruction of the importance of the acoustic dimensions, thus addressing the gap regarding explicit instruction's role in shifting the L2 cue weighting patterns.

To deal with the gap regarding individual differences, this study aims to focus on not only group patterns but also individual differences. Individual differences were quantified through mixed-effect logistic regression, which generated the cue weighting pattern for each participant with the beta coefficients.

Chapter 3 Experiment 1

3.1 Methodology

Experiment 1 set out to examine the cue weighting patterns of L1 English speakers and L2 English learners regarding the /i:/-/ɪ/ English contrast, and it also sought to identify the potential predictors of L2 speakers' cue weighting patterns. Acoustic stimuli from the 'bead-bid' and 'steet-stit' pair were randomly presented to participants, who were asked to label each stimulus they heard as 'bead', 'bid', 'steet' or 'stit' through a forced-choice task.

Participants

29 L1-Mandarin L2-English speakers (17 females and 12 males; mean age=23) and 29 native English speakers (15 female and 14 male; mean age=38.3) participated. All of the L2 English learners were university students from China. The recruitment criteria for L2 learners in this study were: 1) Participants should not be native speakers of any language other than Mandarin; 2) They should be L2 speakers of English at intermediate level and above; 3) They were unable to use any foreign language other than English at 'high proficiency level', which meant that they gave a rating of 6 or above on a 1 to 7 point Likert scale measuring self-reported proficiency across all four skills of English-speaking, writing, listening and reading. The second recruitment criterion was to guarantee that participants had the knowledge that the letter combinations 'ea' (as in 'bead') and 'ee' (as in 'steet'), which would appear in the experiment, corresponded to the /i:/ sound, while 'i' (as in 'bid' and 'stit') corresponded to the /ɪ/ sound; the third criterion was designed to reduce the effect of any third language on the cue weighting of English /i:/-/ɪ/.

Table 1 summarizes the self-perceived English proficiency of the 29 L2 participants. 16 of them reported to speak a third language, and three reported to speak two foreign languages other than English. Table 2 summarizes these foreign languages spoken by the participants. Furthermore, 11 of the 29 participants reported to have over 6-month experience of studying in English-speaking countries (min=7 months; max=108 months; mean=22.1 months). In terms of the instruction the L2 participants had received on the English /i:/-/ɪ/, 6 of them reported they had never been instructed about the differences between the two vowels; of the 23 participants who reported to have received the instruction from their teachers, 13 received proper instruction, while 10 received improper instruction.[2]

[2] In the language background questionnaire (see Appendix 10), L2 participants were asked whether they have been explicitly instructed about the differences between /i:/ and /ɪ/. Those who selected "Yes" wrote down the exact instruction they had received. Responses were coded into two categories: 'proper instruction' and 'improper instruction'. A response was coded as 'proper instruction' if it correctly pointed out the difference regarding the mouth shape or position of the tongue; otherwise, the response would be labeled as 'improper instruction'. For instance, responses indicating that the difference only lies in duration were improper instruction.

Table 1 Self-perceived English proficiency of native Mandarin speakers in Experiment 1

		Maximum	Minimum	Mean
English proficiency of native Mandarin speakers (N=29)	Speaking	6	2	4.55
	Listening	6	2	4.86
	Reading	7	3	5.48
	Writing	6	2	4.76

Table 2 Third languages spoken by native Mandarin speakers in Experiment 1

	French	Spanish	Japanese	German	Korean
Number of participants speaking the third language (N=19)	6	6	6	2	2

The 29 native speakers of British English were recruited from Prolific, an online participant recruitment platform (https://www.prolific.co/) and were paid at a rate of GBP 15 per hour. The criteria for native speaker recruitment were: 1) Participants should not be native speakers of any language other than English; 2) Participants had not previously learned Mandarin as a foreign language; 3) Participants were unable to use any language other than English at 'high proficiency level'. 'High proficiency' was defined by the same criterion as that in the above L2 speaker part. 11 of the 29 native speakers reported they spoke one to three foreign languages. Table 3 summarizes the foreign language spoken by the participants.

Table 3 Third languages spoken by native English speakers in Experiment 1

	Spanish	French	German	Polish	Irish
Number of participants speaking the foreign language (N=11)	7	5	1	1	1

Stimuli

Stimuli for the word recognition task consisted of two sets of monosyllabic word pair continua: 'bead-bid', the real-word pair and 'steet-stit', the pseudo-word pair. The pseudo-word pair was included considering the 'word frequency effect' (Dufour, Brunellière, & Frauenfelder, 2013), whereby high frequency words are recognized and responded to faster and with more ease than low frequency words in word recognition tasks (Fox, 1984). This may in turn bias the results of cue weighting studies which require listeners to make decisions between real-word pairs, since listeners may be inclined to choose the high-frequency token over its low-frequency counterpart.

The Original four natural recordings

The four original recordings used to create the stimuli were from Giannakopoulou, Brown, Clayards and Wonnacott (2017), and the recordings were natural 'bead', 'bid', 'steet' and 'stit' speech exemplars recorded by a male native speaker of English. The spectral and duration information of these four samples is listed in Table 4.

Table 4 Acoustic information of the 'bead', 'bid', 'steet' and 'stit' natural speech stimuli

	Bead	Bid	Steet	Stit
F1 (Hz)	325	495	346	517
F2 (Hz)	2510	2057	2468	1888
Duration (ms)	287	127	161	98

Distribution of vowels in the target stimuli

Vowels of the target stimuli were 35 (7×5) synthetic vowels manipulated from the natural recordings, which varied in vowel duration and spectrum. The spectral dimension varied in seven steps and the temporal dimension varied in five steps. Values for the spectrum and duration of these vowels are shown in Figure 3. The horizontal and vertical axis show how duration and spectrum of the vowels vary respectively. To approximate the naturally produced speech samples, F1 and F2 of the most /i:/-like vowel (Token 5 in Figure 3) were set at 340 Hz and 2600 Hz respectively, while F1 and F2 of the most /i:/-like vowel (Token 31) were set at 488 Hz and 2114 Hz. The duration of the former was 260ms, and duration of the latter was set at 100ms. In order to form a continuum from /i:/ to /ɪ/, five linearly equal steps were created, with duration decreasing from 260ms to 100ms. At the same time, F1 and F2 co-varied in seven equal mel steps[3], where F1 increased from 340Hz to 430 Hz in steps of 25 mels and F2 decreased from 2600 Hz to 2114 Hz in steps of 30 mels.

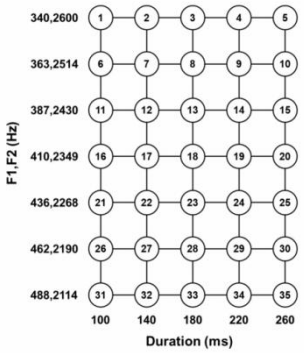


Figure 3 Distribution of vowels in the stimuli used for the word recognition task.

Manipulation of stimuli

To begin with, the /i:/ vowel was cut from the natural 'bead' sample. F1 and F2 of the vowel were manipulated using the Klatt synthesizer (Klatt & Klatt, 1990) in Praat (Boersma & Weenink, 2023). A filter was first extracted from the original vowel sample. Then, values of F1 and F2 were edited through the FormantGrid function based on this filter according to values of the seven steps, creating seven filters with varying spectral information in F1 and

[3] In speech perception research, the value of power spectrum is usually converted from linear frequency scale (Hz) to the mel-scale (Kathania, Shahnawazuddin, Ahmad, & Adiga, 2018). This is because humans perceive sound frequency in a non-linear way. Linearly equal F1-F2 steps might not be perceptually equal to listeners. Therefore, F1 and F2 of the stimuli varied in equal mel steps rather than linear steps. The formula converting the linear frequency into mel-frequency is as follows:

$$m = 2595 \lg\left(1 + \frac{f}{700}\right)$$

F2. The F3, F4 and F5 value of all seven vowels were maintained at a constant level of 3100Hz, 4100Hz and 5000Hz respectively across all seven filters. The original sound was then pre-emphasized and resampled into 10000Hz. Based on the resampled sound, the 'source' was created to be combined with the seven manipulated filters, yielding 7 manipulated vowels with differing F1 and F2 values: F1 (340Hz)/F2 (2600Hz), F1 (363Hz)/F2 (2514Hz), F1 (387Hz)/F2 (2430Hz), F1 (410Hz)/F2 (2349Hz), F1 (436Hz)/F2 (2268Hz), F1 (462Hz)/F2 (2190Hz), F1 (488Hz)/F2 (2114Hz). After spectral manipulation, the duration of the seven vowels was manipulated via the duration points in Praat's 'To Manipulation' function. This yielded five stimuli with varying duration for each of these seven vowels: 100ms, 140ms, 180ms, 220ms, 260ms. Therefore, a total of 35 synthetic vowels presented in Figure 3 were generated. Finally, vowels in the original 'bead' and 'steet' sample were cut and replaced with the 35 manipulated vowels. In this way, 35 'bead-bid' word stimuli and 35 'steet-stit' pseudo-word stimuli were created.

Procedure

The experiment was conducted through the 'Gorilla' (https://gorilla.sc/) experiment builder platform (Anwyl-Irvine, Massonnié, Flitton, Kirkham & Evershed, 2019). Participants accessed the experiment online using their own computers. Task instructions appearing on the screen were given to the L2 learners received in Chinese, their native language. Before the experiment began, participants were asked to wear headphones. In order to proceed to the experiment, they completed a sound check to ensure that they could hear the sound clearly and their browser could play the audio files automatically. They then read some basic information about the research and provided informed consent.

Participants who passed the audio test then completed the main experimental task: a forced-choice word recognition task. In this task, participants identified what they heard by clicking one of the two boxes presented on their screen. Each box contained an orthographic option. The trial started with an 800ms presentation of a fixation (+) placed at the center of the screen. After that, there was a 100ms pause, during which participants saw a blank screen. Then, the auditory stimulus was played, and the two options were presented at the center of the screen at the same time the audio started. Participants were expected to respond after the audio had finished by clicking one of these two options, and the next trial started 100ms later. Figure 4 demonstrates the procedure of each trial. There were five blocks of the word labeling task. Each block contained the 35 'bead-bid' stimuli and the 35 'steet-stit' stimuli. For each participant, presentations of the 70 trials within each block were randomized. Each trial was played only once. Participants took a short break after completing each block.

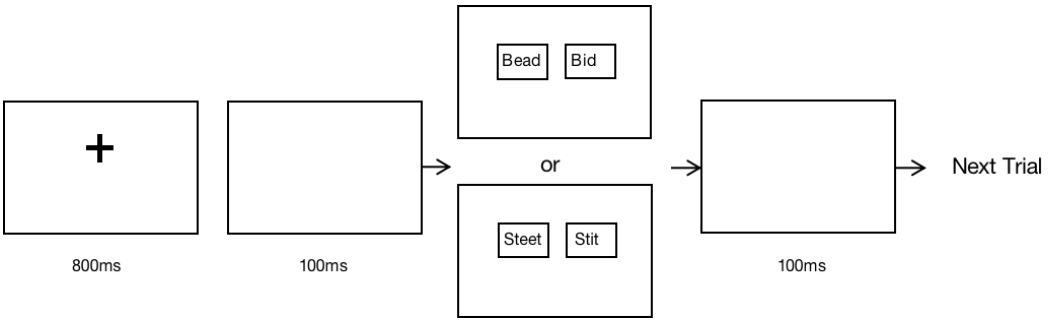


Figure 4 Procedure of each trial for the word recognition task.

After the main experimental task, a language experience questionnaire was presented to each participant. The questionnaire for native speakers (see Appendix 9) was divided into three parts: the first part collected the demographic information including age and gender; the second part asked participants to list all the languages they spoke and rate those languages in terms of self-perceived proficiency in speaking, listening, writing and reading with a 7-point Likert scale, where 1 corresponded to 'very poor' and 7 indicated 'native like'; the third question asked what types of English dialects they spoke. For the L2 group, the questionnaire was presented in Mandarin. Besides demographic information and participants' languages, this questionnaire (see Appendix 10) contained another two sections: one asked participants to list the English-speaking countries or regions where they had previously stayed for over 6 months, the age they arrived at these places and how long they stayed; the last section investigated previous instruction they had received from their teachers regarding the differences between the English /i:/ and /ɪ/.

Analysis

Mixed-effects logistic regression analysis was used to quantify the extent to which a group of listeners relied on the spectral cue and duration cue when perceiving /i:/ versus /ɪ/. To introduce the mixed-effects logistic regression model, this section proceeds from logistic regression analysis and zooms in on the mixed-effect model with a focus on explaining the fixed and random effects.

Logistic Regression Analysis

Logistic regression analysis is a powerful tool to quantify perceptual cue weighting data obtained through forced-choice research design (Shertz & Clare, 2020), since the dependent variable is binary. This means that the regression model is predicting the extent to which the listeners perceive the /i:/ versus /ɪ/ contrast. The equation for logistic regression is:

$$ln(\frac{\hat{Y}}{1-\hat{Y}}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_i X_i$$

As we can see, logistic regression model operates in a logistic space: $\hat{Y}$ represents the probability of one binary outcome versus the other, but the dependent variable is not

expressed by $\hat{Y}$; rather, $\hat{Y}$ is transformed into the logit scale, such that the dependent variable represents the 'logged odds'. X_i represents each independent variable, while β_i is the beta coefficient which represents the slope of for that variable. Therefore, the larger β_i means steeper slope, indicating that the corresponding independent variable (X_i) is more likely to contribute to the observed outcome. β_0 is the intercept independent of changes in independent variables. The regression model for this study can be expressed as:

$$ln(\frac{\hat{Y}}{1-\hat{Y}}) = \beta_0 + \beta_{spec} \times spectral\ step + \beta_{dur} \times duration\ step$$

The two independent variables in this model were the acoustic information of the spectral and duration cues. Therefore, the reliance of the two cues can be quantified by the corresponding beta values. The primary cue should have a larger absolute beta value than the other, and the weighting of the spectral cue versus the temporal cue can be quantified as follows (Yazawa et al., 2020):

$$Cue\ weighting = \frac{|\beta_{spec}|}{|\beta_{spec}| + |\beta_{dur}|}$$

However, using logistic regression to account for participants' cue weighting behaviors at the group level might be problematic. This is because logistic regression requires independence of observations (Stoltzfus, 2011). But in this study, the observations were not truly independent of each other, as observations can be grouped according to different participants. To make things worse, perceptual cue weighting is well-known for its individual variability (e.g., Kong & Kang, 2022; Schertz et al., 2015). Therefore, responses grouped within each participant are increasingly likely to be homogeneous. Considering this, mixed-effects logistic regression analysis was adopted instead of logistic regression to accommodate such features.

Mixed-Effects Logistic Regression Model

The mixed-effects model is applicable to situations where data are clustered or there are both fixed and random effects (UCLA: Statistical Consulting Group, 2021b). It could be understood as an extension of logistic regression model to include random effects aside from the fixed effects (UCLA: Statistical Consulting Group, 2021a). In broad terms, fixed effects are variables that are expected to have an effect on the dependent variable. In the case of this study, two major fixed effect were the spectrum and duration of the stimuli, since we were fundamentally interested in how variation of these two dimensions would influence the response patterns. In comparison, random effects are usually grouping factors for which we are trying to control. In this research, such grouping factor was the participant. Because individual differences of participants might affect the overall pattern at the group level, introducing this factor as the random effect allowed us to explain some part

of the pattern which could not be explained by the fixed effects. In this way, the influence of individual difference on the model was controlled. The equation for mixed-effects logistic regression model can be expressed as:

$$y = X\beta + a + \varepsilon$$

In this equation, y represents the outcome variable; *X* stands for the matrix of the predictors; β is the regression coefficients of the fixed effects; a represents the matrix of the random components; and ε represents the residuals, the part of *y* which cannot be explained by the model. The basic model for this study can be expressed as:

$$y = \beta_{spec} \times spectral\ step + \beta_{dur} \times duration\ step + a_{participant} + \varepsilon$$

In this study, mixed-effects logistic regression analysis was conducted using the *glmmer* function in the lme4 package in R (R Core Team, 2023). Group difference was compared based on the β values of fixed effects of the mixed-effects models; individual difference was investigated based on the β coefficients for each participant, which were extracted from the models.

3.2 Results

3.2.1 Cue weighting patterns of native English speakers and L2 English learners

The four graphs in Figure 5 visualize the cue weighting patterns of each group (the native speaker group and L2 learner group) for each word pair stimuli (‘bead-bid’ and ‘steet-stit’). The vertical axis of the heat map presents the 7 spectral steps and the horizontal axis shows the 5 duration steps. Therefore, each cell represents one stimulus, and the 35 (5*7) cells in the heat graph correspond to the 35 stimuli. The color of each cell was determined by the percentage of ‘bead’ or ‘steet’ responses, which was calculated by dividing the number of the ‘bead’ or ‘steet’ responses to the total observations of ‘bead-bid’ or ‘steet-stit’. The darker color indicates a higher percentage of the ‘bead’ responses, whereas lighter color means a dominance of the ‘bid’ responses. Therefore, reliance on the spectral cue is visualized as darker color along the upper edge of the rectangle compared with the lower edge; reliance on duration is indicated by the darker color along the right edge in comparison with the left edge. As we can see from graph (a) and (b), in terms of the ‘bead-bid’ stimuli, both native English speakers and L2 learners seemed to rely on both cues. The two groups’ overall cue weighting patterns for the ‘steet-stit’ stimuli are illustrated in graph (c) and (d). Like the ‘bead-bid’ stimuli, the two groups again showed similar patterns: they relied on both acoustic dimensions to distinguish the pseudo-word pair ‘steet-stit’. However, the cells were overall darker than those of the ‘bead-bid’ stimuli shown in graph (a) and (b). This indicates a biased perceptual pattern where participants were more inclined to choose ‘steet’ over ‘stit’.

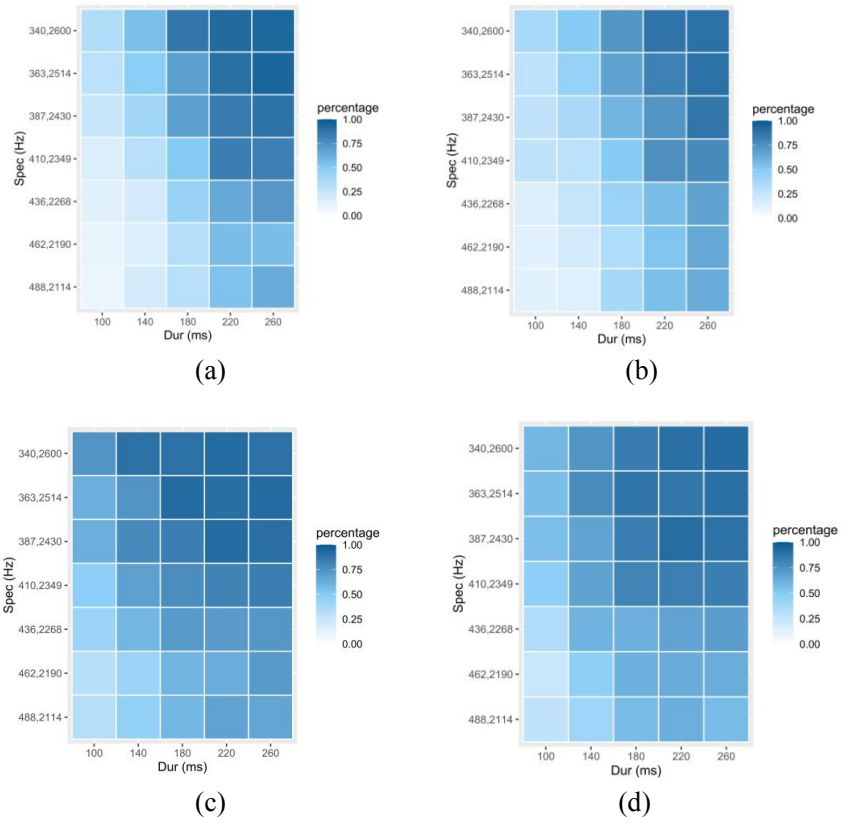


Figure 5 Cue weighting patterns of the 29 native speakers and 29 L2 learners for the 'bead-bid' pair and 'steet-stit' pair. (a) native English speakers' responses of the "bead-bid" stimuli. (b) L2 English speakers' responses of the "bead-bid" stimuli. (c) native English speakers' responses of the "steet-stit" stimuli. (d) L2 English speakers' responses of the "steet-stit" stimuli.

To further compare the native speaker group and the L2 learner group, a mixed-effects logistic regression model was generated combining all the data from both groups of participants for both types of word stimuli. Formula of the model is as follows:

`Response~Spec*Word*Group+Dur*Word*Group+(Spec*Word+Dur*Word|ParticipantID)`.

The response variable was labeled as `Response` in the statistical model. The 'bead' and 'steet' responses corresponded with the vowel /i:/, while 'bid' and 'stit' corresponded with /ɪ/. In the model, /i:/ was encoded as `0.5`, and /ɪ/ was encoded as `-0.5`. The predictors for the fixed effects included `Spec`, `Dur`, `Word` and `Group`. The variable `Spec` stands for the seven spectral steps of the acoustic stimuli. It was regarded as a continuous variable and was encoded as seven integers centered around 0 (`3, 2, 1, 0, -1, -2, -3`). Likewise, the variable *`Dur`* stands for the five temporal steps of the stimuli, which were encoded as `2, 1, 0, -1, -2`. Introducing `Spec` and `Dur` into the model enabled us to quantify how the response patterns changed along with the two acoustic dimensions. Furthermore, the variable of stimulus types was labeled as `Word` in the regression model: the real-word pair 'bead-bid' was encoded as `0.5`, while the pseudo-word pair 'steet-stit' was encoded as `-0.5`. Another variable, the language groups participants belonged to, was labeled as `Group`. This variable has two values: the native English speaker group was encoded as `-0.5` and the L2 English learner group was encoded as `0.5`.

Aside from these fixed-effects predictors, the current study also involved 'participant' as a random effect, which was labeled as `ParticipantID`. 'Participant' was treated as a random effect because observations were grouped according to different participants. Some participants might inherently favor the spectral cue, while others might give more weight to the temporal cue. Moreover, each participant's cue weighting behaviors might differ in terms of the 'bead-bid' and 'steet-stit' pairs. To control for such 'noise', the random slope in the model was set as `Spec*Word+Dur*Word`.

Table 5 presents the output of the statistical model. The positive estimate `Spec` (estimate=0.50731) means that as vowel spectrum increased, these listeners at the group level were more likely to identify the stimulus as 'bead' or 'steet'; the positive estimate of `Dur` (estimate=0.93386) indicates participants tended to label stimuli with longer vowel duration as 'bead' or 'steet'. The effects of both the predictors `Spec` and `Dur` are statistically significant. Therefore, the two groups of participants overall relied on both vowel spectrum and duration when identifying the stimuli. Moreover, the fact that estimate of `Dur` is larger than that of `Spec` means the response pattern was more likely to change with the variation of duration compared with spectrum. This indicates that duration functioned as a more dominant cue for the two groups of participants overall. `Spec:Group` (estimate=0.0292, p=0.858) and `Group:Dur` (estimate=-0.02716, p=0.882) respectively represent `Group`'s interaction with `Spec` and `Dur`. The coefficients were close to 0 and the p values were not significant, meaning there was no significant interaction between `Group` and `Spec` or between `Group` and `Dur`. This indicates there is no evidence that the two groups differed in terms of reliance on both vowel spectrum and duration.

It is worth mentioning that the model had shown two interesting patterns. First, the insignificant interaction between `Word` and `Group` (estimate=-0.06045, p=0.86) as well as the negative and significant influence of `Word` (estimate=-1.32598, p=2.23e-14) indicates that participants in both groups were more likely to choose /i:/ over /ɪ/ when hearing 'steet-stit' than when hearing 'bead-bid'. Second, the positive and significant interaction between `Word` and `Dur` (estimate=0.40396, p=1.12e-07) suggests that participants tended to rely less on duration for the categorization of the 'steet-stit' stimuli in comparison with the 'bead-bid' stimuli.

Table 5 Model output of the 29 L2 English speakers and the 29 native English speakers for the 'bead-bid' and 'steet-stit' stimuli

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	0.80672	0.12668	6.368	1.91e-10***
Spec	0.50731	0.08178	6.204	5.52e-10***
Word	-1.32598	0.17363	-7.637	2.23e-14***
Group	-0.06550	0.25236	-0.260	0.795
Dur	0.93386	0.09165	10.189	<2e-16***
Spec:Word	-0.01625	0.04176	-0.389	0.697
Spec:Group	0.02920	0.16327	0.179	0.858
Word:Group	-0.06045	0.34340	-0.176	0.860
Word:Dur	0.40396	0.07613	5.306	1.12e-07***
Group:Dur	-0.02716	0.18249	-0.149	0.882
Spec:Word:Group	-0.12152	0.08025	-1.514	0.130
Word:Group:Dur	-0.21430	0.14786	-1.449	0.147

The interaction between `Word` and `Dur` motivated us to conduct another analysis where models for 'bead-bid' and 'steet-stit' were run separately. To this end, the same model was built for each word, with the same predictors as before, except without `word` and the interactions with `word`. Therefore, two regression models were generated, and the formula of the models was `Response~Spec*Group+Dur*Group+(Spec+Dur|ParticipantID)`. Output of the regression models is presented in Table 6 and Table 7. As we can see, the two models had similar estimates for `Spec` (0.49332 for 'bead-bid' and 0.50913 for 'steet-stit'). But the model for the pseudo-word pair (estimate=1.4165) had larger estimate for `Dur` than model for the real-word pair (estimate=0.72293), which echoes the output of the overall model above. However, for both models, there is still no significant interaction between `Group` and `Spec` or `Group` and `Dur`, suggesting the two groups of participants did not differ significantly in terms of cue weighting patterns for each of the word stimuli pair.

Table 6 Model output of the 29 native English speakers and 29 L2 learners' overall cue weighting patterns for the 'bead-bid' stimuli

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	0.13963	0.12685	1.101	0.271
Group	-0.09925	0.25330	-0.392	0.695
Spec	0.49332	0.08152	6.051	1.44e-09***
Dur	1.14165	0.10613	10.757	<2e-16***
Group:Spec	-0.04021	0.16270	-0.247	0.805
Group:Dur	-0.17202	0.21131	-0.814	0.416

Table 7 Model output of the 29 native English speakers and 29 L2 learners' overall cue weighting patterns for the 'steet-stit' stimuli

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	1.46414	0.17461	8.385	<2e-16***
Group	-0.02322	0.34731	-0.067	0.947
Spec	0.50913	0.08620	5.906	3.50e-09***
Dur	0.72293	0.09014	8.020	1.06e-15***
Group:Spec	0.07137	0.17192	0.415	0.678
Group:Dur	0.08783	0.17915	0.490	0.624

3.2.2 Individual differences in the cue weighting patterns

For each participant, regression coefficients of `Spec` and `Dur` for both stimulus pairs were extracted from the model objects through the `Coef` function in R. Appendix 15 presents by-participant results of the coefficients and cue weighting for L2 speakers and native English speakers. Table 8 summarizes the individual cue weighting behaviors of the native and L2 groups for the 'bead-bid' and 'steet-stit' stimuli. Here the cue weighting values below 0.1 were categorized as 'relying exclusively on duration'; values above 0.9 were labeled as 'relying exclusively on spectrum'; values between 0.1 and 0.9 were categorized as 'relying on both cues. It seems that more native speakers tended to rely on both cues compared with the L2 learners: for the 'bead-bid' stimuli, 21 of the 29 native speakers relied on both cues, whereas only 15 L2 learners relied on both; for the 'steet-stit' stimuli, 25 native speakers relied on both spectral and temporal cues, while only 16 L2 participants used both cues.

The extent to which participants relied on both cues could also be captured by transforming the original cue weighting score for each participant so that the score was kept the same if it was smaller than 0.5 and was transformed as '1-weighting_score' if it was greater than 0.5. In this way, the greater the transformed score, the higher amount of reliance on both cues. The transformed score for each participant is listed in Appendix 15. Figure 6 visualizes the score distributions of the two groups of participants for the two types of stimuli. The lower edge of the box refers to the 25th percentile of the scores, also known as the 'lower quartile' (referred to as 'Q1'); the upper edge of the box refers to the 75th percentile of the scores, also known as the 'upper quartile' (Q3); the line inside the box refers to the median of the scores (Q2). As can be seen, for both 'bead-bid' and 'steet-stit', the native group has larger Q1, Q2 and Q3 than the L2 group. which seems to suggest that native speakers in general relied more on both cues in comparison with L2 learners. Moreover, the scores for the L2 group concentrate around lower values approximating 0, indicating a large portion of the L2 participants relied primarily on one single cue.

A Mann-Whitney U tests were conducted to further compare the scores of native speakers and L2 learners. Results are summarized in Table 9. For the 'bead-bid' stimuli, the two groups differed marginally insignificantly from each other in terms of the score distribution ($z=1.936$, $p=0.053$). For the 'steet-stit' stimuli, however, the two groups observed significant difference ($z=2.309$, $p=0.021$). Therefore, considering the summary statistics and results of the Mann-Whitney tests, it appears that the native participants tended to use both cues in comparison with the L2 English learners.

Table 8 Summary of individual cue weighting behaviors of the 29 native English speakers and 29 L2 learners.

		Rely exclusively on temporal cue	Rely exclusively on spectral cue	Rely on both cues
bead-bid	Native speakers (N=29)	8	0	21
	L2 learners (N=29)	11	3	15
steet-stit	Native speakers (N=29)	3	1	25
	L2 learners (N=29)	9	4	16

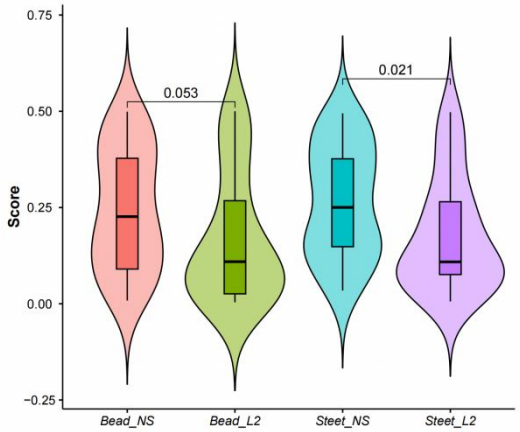


Figure 6 Distribution of transformed scores of native speakers (N=29) and L2 learners (N=29) for the 'bead-bid' and 'steet-stit' stimuli.

Table 9 Results of the Mann-Whitney test comparing the 29 native speakers and 29 L2 learners in Experiment 1. The value outside the parentheses represents the median score (Q2); the first value within the parentheses is the lower quartile (Q1); the second value in the parentheses is the upper quartile (Q3).

	Native speakers (N=29)	L2 learners (N=29)	z	p
Bead-Bid	0.2264 (0.0886, 0.3934)	0.1093 (0.0251, 0.2932)	1.936	0.053
Steet-Stit	0.2503 (0.1472, 0.4047)	0.1089 (0.0685, 0.2654)	2.309	0.021

3.2.3 Potential factors affecting L2 English learners' cue weighting patterns

Given the individual differences found in L2 English learners' cue weighting, follow-up analyses were conducted to investigate the potential factors which might have influenced L2 cue weighting. This was analyzed by rerunning the model just for non-native speakers and with additional predictors of previous instruction, L2 proficiency, and the experience of studying in English-speaking countries. These variables were extracted from the questionnaire responses for each participant. These potential factors are introduced as follows.

Participant variables of L2 learners

1) The variable `Instruction`

The instruction participants had previously received was labeled as `Instruction`. This variable has three values: participants who had never received any instruction (encoded as `0`); participants who had received proper instruction (encoded as `0.5`); and those who had received improper instruction (encoded as `-0.5`).

2) The variable Experience

The variable Experience refers to L2 learners' experience of staying in English-speaking countries. This variable has two values: participants who had such experience (encoded as 0.5) and participants who did not have such experience (encoded as -0.5).

3) The variable LProf

LProf stands for L2 learners' self-perceived listening proficiency. Participants rated their listening proficiency with a 7-point Likert scale ranging from 1 to 7. Therefore, the seven steps of LProf were encoded as 3, 2, 1, 0, -1, -2, -3.

4) The variable OProf

The variable OProf refers to the overall self-reported proficiency of L2 participants. The overall proficiency was calculated by adding the scores of reading, listening, speaking and writing proficiency. The scores of the four aspects were all obtained from the 7-point Likert scale, and the overall proficiency was regarded as a continuous variable ranging from 4 to 28. The 25 steps from 4 to 28 were therefore encoded as 25 integers centered around 0.

Instruction's influence on L2 learners' cue weighting behaviors

First, we were interested in how L2 speakers who had previously had proper (N=13) or improper (N=10) previous instruction differ in the cue weighting patterns. To this end, an overall model involving observations of the 23 participants who had previously been instructed on the English /i:/-/ɪ/ was generated, with Instruction being a fixed effect. The formula for the statistical model was Response~Spec*Word*Instruction+ Dur*Word*Instruction+(Spec*Word+Dur*Word|ParticipantID), and the output of the model is presented in Table 10. The model observes significant interaction between Spec and Instruction (estimate=0.88742, p=0.003154), which suggests participants with proper previous instruction relied more on the spectral cue compared with those with improper instruction. For Instruction:Dur, the estimate of -0.05656 (p=0.868115) indicates there was no statistically significant difference between the two groups of participants regarding their reliance on duration.

Table 10 Output of model examining previous instruction's influence on the cue weighting patterns for the 23 L2 English speakers

	Estimate	Std.Error	zvalue	Pr(>|z|)
(Intercept)	0.81193	0.21734	3.736	0.000187***
Spec	0.46291	0.15039	3.078	0.002083**
Word	-1.53580	0.31002	-4.954	7.28e-07***
Instruction	0.01325	0.43267	0.031	0.975561
Dur	1.00859	0.17078	5.906	3.51e-09***
Spec:Word	0.01177	0.07343	0.160	0.872640
Spec:Instruction	0.88742	0.30059	2.952	0.003154**
Word:Instruction	0.40017	0.61342	0.652	0.514172
Word:Dur	0.28430	0.08105	3.508	0.000452***
Instruction:Dur	-0.05656	0.34062	-0.166	0.868115
Spec:Word:Instruction	0.01839	0.14310	0.128	0.897763
Word:Instruction:Dur	-0.29157	0.14333	-2.034	0.041922*

Then, two separate models were built to respectively account for the cue weighting patterns of the 13 participants with proper instruction and of the 10 participants with improper previous instruction. Data for the 'bead-bid' and 'steet-stit' stimuli were analyzed in combination, with `Word` being a fixed effect to identify whether patterns differed according to the stimuli type. The formula for the two models was as follows:

`Response~Spec*Word+Dur*Word+(Spec*Word+Dur*Word|ParticipantID)`.

Model output for the 13 participants with proper instruction is presented in Table 11, and output for the 10 participants with proper instruction is shown in Table 12. As we can see, the 13 L2 learners who had received proper instruction had similar estimates regarding `Spec` (estimate=0.92498) and `Dur` (estimate=0.98419), which means they tended to weight duration and spectrum equally. Moreover, the coefficients of `Word:Dur` and `Spec:Word` were insignificantly different from 0, meaning that the stimulus type (whether 'bead-bid' or 'steet-stit') did not significantly influence the degrees of reliance on spectrum and duration. For the 10 L2 learners in comparison, there was no evidence that `Spec` was a significant predictor (estimate=0.01231, p=0.83792). The fact that the estimate of `Spec:Word` did not differ significantly from 0 (estimate=`0.02102`, p=0.83251) suggests there was no evidence that participants' reliance on the spectral cue was influenced by the stimuli type.

Table 11 Statistical output of the mixed-effects logistic regression model to examine the cue weighting pattern for the 13 L2 English speakers who had received proper instruction

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	0.83554	0.26780	3.120	0.00181**
Spec	0.92498	0.27024	3.423	0.00062***
Word	-1.35565	0.45207	-2.999	0.00271**
Dur	0.98419	0.27750	3.547	0.00039***
Spec:Word	-0.02538	0.10272	-0.247	0.80485
Word:Dur	0.10833	0.11647	0.930	0.35232

Table 12 Statistical output of the mixed-effects logistic regression model to examine the cue weighting pattern for the 10 L2 English speakers who had received improper instruction

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	0.79829	0.37487	2.130	0.03321*
Spec	0.01231	0.06020	0.205	0.83792
Word	-1.72130	0.38961	-4.418	9.96e-06***
Dur	1.04597	0.15203	6.880	5.99e-12***
Spec:Word	0.02102	0.09938	0.211	0.83251
Word:Dur	0.42289	0.13426	3.150	0.00163**

Other potential factors affecting L2 English learners cue weighting patterns

Besides `Instruction`, another three potential factors (`Experience`, `LProf and OProf`) were examined to see whether they might have affected L2 learners' cue weighting patterns. Effects of these factors were investigated by looking at each factor's interaction with the spectral dimension (`Spec`) and temporal dimension (`Dur`). A

regression model was generated for each of the three predictors based on observations of both the 'bead-bid' and the 'steet-stit' stimuli. The model structure for each of the three factors was formulated as follows:

```
Response~Spec*Word*Factor+Dur*Word*Factor+(Spec*Word+Dur*Word
| ParticipantID)
(Factor=Experience;LProf;OProf)
```

A total of three models were generated, and all three models yielded non-significant results in terms of the factor's interaction with `Spec` or `Dur` (e.g., `Experience:Spec`). Nor was there any further significant interaction with `Word` (e.g., `Experience:Spec:Word`). Therefore, there was no evidence that any of the three factors had an impact on participants' reliance of either the spectral or temporal dimension regarding both 'bead-bid' and 'steet-stit'.

3.2.4 Summary of results of Experiment 1

Native speakers and L2 learners of English did not differ significantly in terms of their cue weighting behaviors of both the 'bead-bid' and 'steet-stit' stimuli. Both groups relied on duration and spectral cues when differentiating the contrasts, with duration being the primary cue. However, participants' cue weighting patterns observed substantial individual variability. In terms of the potential factors affecting the L2 group's cue weighting patterns, there was no evidence that `Experience`, `LProf` and `Oprof` had any significant influence. However, their cue weighting behaviors were predicted by `Instruction`. L2 participants with proper previous instruction tended to rely more on vowel spectrum than those who had been improperly instructed, and the latter relied almost exclusively on the temporal cue.

3.3 Discussion

3.3.1 Different response patterns of the two stimuli pairs

One point needs to be discussed for the results of Experiment 1: for both native English speakers and L2 English learners, response patterns differed between the 'bead-bid' and 'steet-stit' continua in two aspects. First, the 'steet-stit' continuum saw a biased distribution of the binomial responses, where participants were in general more inclined to label the sounds as 'steet' over 'stit'. In comparison, the 'bead-bid' stimuli demonstrated a relatively balanced response pattern. Second, participants were less reliant on the temporal cue when categorizing the 'steet-stit' pair in comparison with 'bead-bid'. One explanation for the biased response patterns could be that 'steet' in the pseudo-word pair sounded similar to some high-frequency words such as 'street' or 'steep', so that participants were more inclined to choose 'steet' over 'stit'. But if this was the case, the variability of response percentages should have decreased equally along both the spectral and the temporal

dimension. Thus, this explanation cannot explain why participants gave less weight to duration but not spectrum when hearing ‘steet-stit’, in comparison with ‘bead-bid’.

Therefore, another explanation was proposed to account for the two discrepancies in the response patterns between the two types of stimuli—the differences could be due to the range of duration steps adopted for the two pairs of stimuli. The temporal steps of both pairs were manipulated to range from 100ms to 260ms with an intention of variable control. However, it has been found that duration of English vowels was influenced by the consonant environment. For instance, House and Fairbanks (1953) measured the duration of ten native American English speakers’ production of six English vowels (/ɪ/, /ʌ/, /ɒ/, /ə/, /æ/, /ʊ/) before various consonants and found that the average vowel duration was longer when vowels were followed by voiced consonants in comparison with voiceless consonants. Hillenbrand et al. (1995) measured the duration of 12 American English vowels in the /hVd/ syllables. They found that /i:/ had an average duration of 243mm for the male speakers and 306mm for female speakers, whereas /ɪ/ had an average duration of 192mm for males and 237mm for females. In comparison, Black (1949)’s measurement of the duration of 11 English vowels in the /tVp/ environment showed that the average duration for /i:/ was 159mm and that for /ɪ/ was 135mm. Hillenbrand et al. (2000) attributed the longer durations in Hillenbrand et al (1995) than in Black (1949) primarily to the final voiced stop.

In the case of this experiment, the real-word stimuli had a voiced final stop /d/, while the pseudo-word pair had a voiceless stop /t/. Therefore, it is possible that the duration range from 100ms to 260ms was closer to the /i:/ end for the ‘steet-stit’ pair, while it was intermediate for the ‘bead-bid’ pair. Since both groups of participants relied predominantly on duration as is shown in the response patterns for ‘bead-bid’, it was reasonable for them to categorize the pseudo-word continuum as ‘steet’ if the duration range of this continuum overall approximated ‘steet’. This may explain the biased response distribution of the pseudo-word pair towards ‘steet’. Furthermore, such biased response towards ‘steet’ caused by the unbalanced range of duration steps might lead to a decreased variability of response percentages along the temporal dimension, which could explain why the regression model indicated that participants had relied less on duration when distinguishing the pseudo-word pair in contrast with the real-word pair. A replication study manipulating the ‘steet-stit’ stimuli to have overall shorter duration is needed to verify this assumption. This account might be confirmed if shortening the vowel length of the ‘steet-stit’ stimuli leads to less biased response patterns and increased reliance on duration. Moreover, future cue weighting studies on English vowels should take into account the consonant environments to make sure the acoustic characteristics of stimuli will not lead to biased response patterns.

3.3.2 The role of previous instruction

Another point that needs to be discussed before moving on to Experiment 2 is instruction's influence on L2 English speakers' cue weighting patterns. The statistical model of the 23 L2 speakers who had previously received either proper or improper instruction showed that previous instruction served as a significant predictor of participants' response patterns. Those who had received proper instruction tended to rely more on vowel spectrum in comparison with those who had been improperly instructed. One possibility is that the previous instruction indeed affected participants' usage of the cues. But we could not rule out that there might be a third variable at play, such as the major of participants. L2 English speakers majoring in English language are essentially more likely to receive proper instruction on English phonology. Therefore, it is possible that the majority of the 13 participants with proper instruction were those with an English major background. However, English majors tend to receive more intensive training in English. As a result, they might have higher English proficiency and higher amount of exposure to English. In this sense, the influence of instruction in Experiment 1 might essentially be the result of other mediator variables linked to major. But such assumption regarding major's impact and its interaction with instruction have not been thoroughly investigated in previous L2 cue weighting research. The only known study is Yang et al. (2016), which examined Mandarin-L1 English speakers' cue weighting behaviors of English /i:/-/ɪ/. The Chinese participants' perception distances from American native speakers were calculated through distance analysis, and a dependent t-test found that English majors ($M = 1.51$, $SD = 1.01$) did not differ significantly from non-English majors ($M = 1.73$, $SD = 0.87$), $t(81) = -1.07$, $p = 0.287$). However, Yang et al. (2016) at best indicates that major in general did not influence cue weighting, whereas interaction of major and instruction was not examined.

Based on what have been discussed, an Experiment 2 was conducted to further investigate the role of previous instruction on L2 cue weighting. Three improvements were made based on Experiment 1. First, Experiment 2 included only the 'bead-bid' stimuli but not the 'steet-stit' stimuli, since the responses were biased regarding the latter. Second, because it was not clear if the influence of instruction observed in Experiment 1 was causal, Experiment 2 took an intervention approach to examine whether direct instruction could manipulate participants' use of the cues. To do so, the instruction given to participants in Experiment 2 made it clear that there are two cues differentiating English /i:/-/ɪ/. But the instruction specifically encouraged participants to use spectrum, the weaker cue, rather than duration. Third, to investigate major's interaction with instruction, Experiment 2 asked participants to report their majors and included the interaction between major and instruction as a fixed effect of the statistical model.

Chapter 4 Experiment 2

4.1 Methodology

The aim of Experiment 2 was to find out to what extent explicit instruction could shift L2 English speakers' cue weighting patterns. To this end, this experiment consisted of a pretest, a treatment section immediately followed by a post-test.

Participants

A total of 48 Chinese English learners were recruited. All of them were university students from China. The recruitment criteria in this experiment were the same as those in Experiment 1. The participants were randomly divided into the experimental group and control group. As in Experiment 1, a language experience questionnaire (see Appendix 10) collected information on participants' language experience. But the questionnaire in Experiment 2 additionally asked what major participants was studying.

The control group involved 23 participants (12 females and 11 males; mean age=22.8). 12 of them studied English language as a major in university. The self-reported English proficiency of the control group is summarized in Table 13. Moreover, Table 14 summarizes the third languages spoken by the participants. Furthermore, 7 participants had stayed in English-speaking countries or regions for over 6 months (min=8 months; max=24 months; mean=13 months). 3 of the 23 participants had never received instruction on the differences between /i:/ and /ɪ/; 9 reported to have received proper instruction from the teachers; and 11 had received improper previous instruction.

Table 13 Self-perceived English proficiency of the control group

		Maximum	Minimum	Mean
English proficiency of native Mandarin speakers (N=23)	Speaking	7	2	4.22
	Listening	6	3	4.61
	Reading	7	4	5.22
	Writing	7	3	4.39

Table 14 Third languages spoken by the control group

	French	Japanese	Spanish	Korean
Number of participants speaking the third language (N=7)	3	3	1	1

There were 25 participants in the experimental group (17 females and 8 males; mean age=22.7). 16 of them were English major students. Table 15 shows the English proficiency reported by the experimental group. 16 of the 25 participants spoke another foreign language other than English, and one participant reported to speak two. These third languages have been listed in Table 16. Furthermore, 7 participants reported to have over 6 months' experience of staying in English-speaking countries (min=8 months; max=12 months; mean=10.3 months). In terms of previous instruction, 2 participants had never

received instruction regarding the differences between the English /i:/-/ɪ/; 13 participants had been properly instructed; and 11 had previously received improper instruction.

Table 15 Self-perceived English proficiency of the experimental group

		Maximum	Minimum	Mean
English proficiency of native Mandarin speakers (N=25)	Speaking	7	3	4.8
	Listening	7	4	5.16
	Reading	7	4	5.8
	Writing	6	3	5.04

Table 16 Third languages spoken by the experimental group

	French	Japanese	Spanish	German
Number of participants speaking the third language (N=17)	8	6	3	1

Stimuli

Stimuli in the pretest and post-test.

Stimuli for the word recognition task of Experiment 2 consisted of only the real-word continua 'bead-bid', whereas the pseudo-word continua 'steet-stit' was not presented. The 35 real-word stimuli were the same as those in Experiment 1.

Treatment

For the experimental group, participants were presented with detailed and explicit instruction materials (see Appendix 13) on the differences between the two English vowels /i:/ and /ɪ/. Before studying the instruction materials, participants were told that they would learn something about the differences between the English vowels /i:/ and /ɪ/, after which they would complete a word recognition task. Then, they were presented with the learning materials and were asked to learn the materials carefully. In the first section of the instruction materials, participants learned that the vowels differ in two aspects—vowel duration and the different shapes of the mouth when producing these sounds. The second part of the instruction zoomed in on exactly how the mouth shapes of the two vowels differed, and participants were told that the degree of mouth opening of /ɪ/ falls between /i:/ and /e/ and for this reason the property of /ɪ/ also falls between /i:/ and /e/. This was followed by a 2-minute video produced by BBC Learning English (https://www.bbc.co.uk/learningenglish). In the first part of this video, a female native speaker of standard southern British English produced minimal pairs of /i:/ and /ɪ/, asking the audience to try to hear the difference; in the second part of the video, the same speaker produced minimal pairs of /ɪ/ and /e/. After watching the video, participants then learned that duration is not so reliable since it is subjected to individual difference of the speakers. The whole instruction took about 10 to 15 minutes to complete.

For the control group, the treatment was to watch the same video as that in the instruction materials for the experimental group without receiving any form of instruction. Therefore, treatment of control group took around 2 minutes to complete.

Procedure

Experiment 2 adopted a pre- and post-test design and contained two sessions: In Session 1, participants completed the pretest, followed by the language background questionnaire; in Session 2, they completed a treatment session followed immediately by the post-test. The forced-choice word recognition task used at pretest and the post test was identical for the two conditions. The task was identical to the main experimental task in Experiment 1 except that each of the five experimental blocks only contained the 35 'bead-bid' stimuli, not the 35 'steet-stit' stimuli.

After giving informed consent, participants completed the pretest and the language background questionnaire on day 1. The language background questionnaire was almost the same as that in Experiment 1, except for two minor changes. First, the questionnaire in Experiment 2 additionally asked what major participants were studying. Second, the last question concerning previous instruction on English /i:/ and /ɪ/ was not presented in the pretest. Instead, this question was presented at the beginning of Session 2 in order to prevent participants from looking it up during the interval between the two sessions. Session 1 took around 15 minutes to complete.

Participants were invited back to Session 2 no less than one day but no more than five days after the first session. In Session 2, they first answered the aforementioned question regarding instruction before completing the treatment. The post-test began immediately after the treatment session. For the control group, Session 2 took about 15 minutes to complete; for the experimental group, Session 2 took about 25-30 minutes. Figure 7 summarizes the design of Experiment 2.

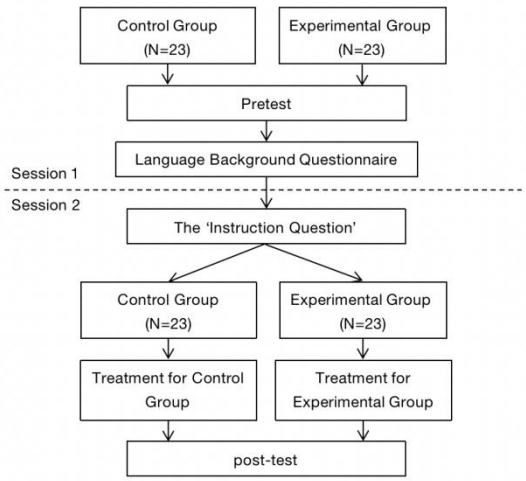


Figure 7 Design of Experiment 2.

4.2 Results

The analysis has two parts. The first part looks at participants' performance in the pretest prior to the intervention. Data were analyzed across the control group and experimental group, as there was no reason the two groups should differ prior to the intervention. The analyses conducted were similar to those for Experiment 1. Doing so examines whether the patterns were consistent with Experiment 1. The second part of the analysis looks at data from the pre- and post-test to examine the effect of intervention.

4.2.1 Data trimming

Originally, 48 participants were recruited (25 participants for the experimental group and 23 for the control group). But preliminary visual examination of the heat graphs generated for each participant showed that four participants (three from the control group and one from the experimental group) showed reverse cue weighting patterns. The 'reverse pattern' means choosing 'bead' when the spectral and duration steps were low and choosing 'bid' when the two steps were high. The possible reason may be that these four participants confused the labels of 'bead' and 'bid' (Yazawa et al., 2020). Data of the four participants were excluded from further analysis. After the data trimming, there were 20 participants in the control group and 24 participants in the experimental group.

4.2.2 Results of the pretest

Cue weighting patterns of L2 learners in the pretest

To investigate the overall cue weighting behavior of L2 learners in Experiment 2, observations of the 44 participants in the pretest were analyzed. The response pattern for the 'bead-bid' stimuli is visualized in Figure 8. A mixed-effect logistic regression analysis was conducted on the pretest data of the 44 participants. Formula of the model is `Response ~ Spec + Dur + (Spec + Dur | ParticipantID)`. Output of the model is presented in Table 17. The model shows that both `Spec` (estimate=0.44509, p=7.58e-06) and `Dur` (estimate=1.32372, p<2e-16) were statistically significant predictors of the response patterns. But participants assigned more weight to the temporal cue than the spectral cue, as estimate of the former is larger than the latter. This is similar to the L2 cue weighting pattern for the 'bead-bid' stimuli in Experiment 1, where the coefficients for `Spec` and `Dur` were respectively 0.5054 and 1.0554.

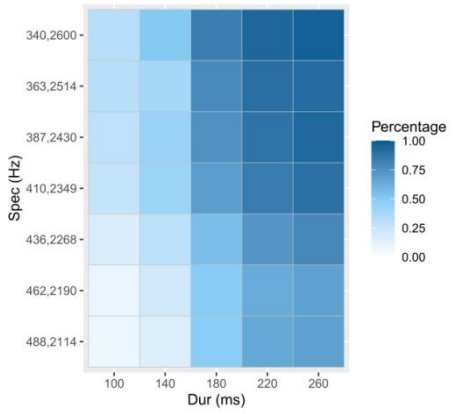


Figure 8 Heat graph of the 44 L2 English learners' response pattern in Experiment 2.

Table 17 Model output for the 44 L2 English learners' overall cue weighting pattern in the pretest of Experiment 2.

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	0.70644	0.12224	5.779	7.52e-09***
Spec	0.44509	0.09942	4.477	7.58e-06***
Dur	1.32372	0.15161	8.731	<2e-16***

Individual differences in the pretest

β_{spec} and β_{dur} of each of the 44 participants were also extracted from the above model to account for each individual's cue weighting behavior in the pretest. The weighting scores and transformed scores were calculated in the same way as in Experiment 1. Appendix 15 presents the coefficients, original weighting scores and transformed scores for each participant. Based on the original weighting scores, cue weighting behaviors of the 44 participants were also categorized into 'relying exclusively on duration', 'relying exclusively on spectrum' and 'relying on both cues' following the same criteria as Experiment 1. Only 19 of the 44 participants relied on both cues, 21 relied exclusively on duration and 4 relied exclusively on vowel spectrum.

Further analyses were conducted to compare the transformed scores for the 'bead-bid' stimuli of the 29 native English speakers in Experiment 1 and a combination of the 29 L2 learners in Experiment 1 and 44 L2 participants in Experiment 2. Data of the two experiments could be combined because at this point it was prior to intervention. Figure 9 visualizes the distributions of the transformed scores for the native speakers (N=29) and L2 learners (N=73). Similar to Experiment 1, the native group had larger Q1, Q2 and Q3 than the L2 group, and the scores of the L2 group concentrated around lower values. A Mann-Whitney test was conducted to compare the transformed scores of the 29 native speakers and 73 L2 learners. Table 18 presents the results, which show that the two groups differed significantly from each other (z=2.756, p=0.006). Therefore, after combining L2 data of Experiment 1 and the pretest of Experiment 2, the summary statistics and the Mann-Whitney test still indicate that the native English speakers tended to use both cues compared with L2 English learners.

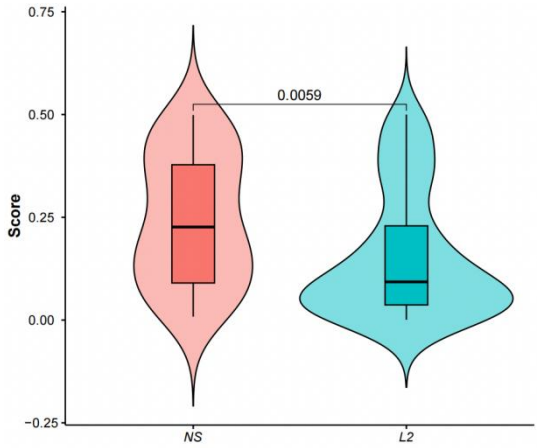


Figure 9 Distribution of transformed scores of native speakers in Experiment 1 (N=29) and L2 learners in both Experiments (N=73) for the 'bead-bid' stimuli.

Table 18 Results of the Mann-Whitney test comparing the transformed scores of native speakers in Experiment 1 (N=29) and L2 learners in both Experiments.

	Native speakers (N=29)	L2 learners (N=73)	z	p
Bead-Bid	0.2264 (0.0886, 0.3934)	0.0929 (0.0341, 0.2430)	2.756	0.006

Potential factors affecting L2 weighting patterns in the pretest

Participant variables in Experiment 2 included the same four variables in Experiment 1: `Instruction`, `Experience`, `LProf`, `OProf`. Besides these, `Major` was introduced into Experiment 2, which stands for the major participants studied at university. This variable has two values: English major (encoded as `0.5`) and non-English majors (encoded as `-0.5`).

Influence of `Instruction` and `Major`

To examine the role of `Instruction`, `Major` and their interaction, a model was generated based on data of the 41 participants who had previously received either proper or improper instruction regarding English /i:/-/ɪ/. The three participants who had never received any instruction were excluded from further analysis. `Instruction` and `Major` were included in the model as fixed effects. Formula of the model was `Response~Spec*Major*Instruction+Dur*Major*Instruction+(Spec+Dur|ParticipantID)`. Output of the model is shown in Table 19. As we can see, `Major` and `Spec` did not have significant interaction (estimate=0.112989, p=0.58889). Likewise, there was no significant interaction between `Major` and `Dur` either (estimate=0.16656, p= 0.962133). This suggests there is no evidence that `Major` served as a significant predictor of participants' reliance on either duration or vowel spectrum. In terms of previous instruction's role, there was no significant interaction between `Instruction` and `Spec` (estimate=0.022033, p=0.916055) or between `Instruction` and `Dur` (estimate=0.046705, p=0.894025). This means there is no evidence that participants' reliance on the two cues was affected by `Instruction`. However, `Spec`, `Major` and `Instruction` observed mildly positive interaction (estimate=0.441905). This suggests

the effect was in the direction that English-major participants who had been properly instructed about the English /i:/-/ɪ/ tended to rely more on vowel spectrum in comparison with other participants. But this effect was not statistically significant (p=0.290355).

Table 19 Output of model examining the influence of `Instruction` and `Major` on the cue weighting patterns of the 41 L2 English speakers in Experiment 2.

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	0.688631	0.144391	4.769	1.85e-0***6
Spec	0.369478	0.104739	3.528	0.000419***
Major	0.089736	0.288103	0.311	0.755442
Instruction	-0.019291	0.287863	-0.067	0.946571
Dur	1.357905	0.176347	7.700	1.36e-14***
Spec:Major	0.112989	0.209066	0.540	0.588890
Spec:Instruction	0.022033	0.209033	0.105	0.916055
Major:Instruction	0.264606	0.575640	0.460	0.645752
Major:Dur	0.016656	0.350811	0.047	0.962133
Instruction:Dur	0.046705	0.350606	0.133	0.894025
Spec:Major:Instruction	0.441905	0.417939	1.057	0.290355
Major:Instruction:Dur	-0.001505	0.700871	-0.002	0.998287

Other potential factors affecting L2 English learners cue weighting patterns

Like Experiment 1, Experiment 2 also investigated the three potential factors' (`Experience, LProf and OProf`) influence on L2 cue weighting. Three separate regression models were generated respectively for `Experience, LProf` and `OProf` to examine these factors' influence. The formula for each of the three models was as follows:

```
Response~Spec*Factor+Dur*Factor+(Spec+Dur|ParticipantID)
(Factor=Experience; LProf; Oprof)
```

All three models yielded non-significant results regarding `Experience`, `LProf` and `OProf`'s interaction with either `Spec` or `Dur`, which means there is no evidence that any of the three factors served as a significant predictor of participants' cue weighting of the 'bead-bid' continuum.

4.2.3 Treatment's role on shifting the cue weighting patterns of L2 English learners

Four mixed-effect logistic regression models were generated to demonstrate the cue weighting patterns of 1) the control group in the pretest; 2) the control group in the post-test; 3) the experimental group in the pretest; 4) the experimental group in the post-test. The formula of the four models was `Response~Spec+Dur+(Spec+Dur|ParticipantID)`. In this way, a β_{spec} and a β_{dur} was generated for each of the four regression models. These eight estimates are demonstrated by Figure 10, which shows that, after the treatment, both the control group and the experimental group saw increase in the reliance of the spectral cue. The reliance of duration decreased for the experimental group, while that of the control group basically remained unchanged.

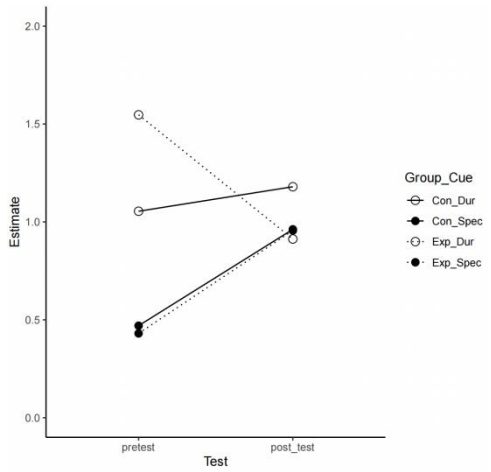


Figure 10 The control group and experimental group's beta coefficients of `Spec` and `Dur` in pretest and post-test. The solid lines demonstrate the control group, whereas the dotted lines represent the experimental group. The open symbols show coefficients of `Dur`, and the black symbols show coefficients of `Spec`.

To further examine the interaction between test sessions and group, a regression model was generated based on all the pre- and post-test data, where the fixed effects were `Spec`, `Dur`, `Group` and `Test`. The formula for this model was `Response~Spec*Group*Test +Dur*Group*Test+(Spec*Test+Dur*Test|ParticipantID)`. The variable `Test` refers to the test condition of the observations, which had two values: the pretest condition (encoded as `-0.5`) and the post-test condition (encoded as `0.5`). The variable labeled as `Group` refers to the control group (encoded as `-0.5`) and experimental group (encoded as `0.5`). In terms of the random slope in this model, because participants might behave differently in pretest and post-test, the random slope was set as `Spec*Test+Dur*Test`.

Table 20 summarizes the results of the model. Significant two-way interaction was observed for both `Spec:Test` (estimate=0.5069, p=9.28e-06) and `Test:Dur` (estimate=-0.247131, p=0.01665). The positive estimate for `Spec:Test` indicates that overall the two groups increased in the amount they used the spectral cue from pre- to post-test. The negative estimate of `Test:Dur` suggests the two groups overall used duration cue less after intervention. On the other hand, through the three way interactions of `Spec:Group:Test` and `Group:Test:Dur`, we can know whether and how the amount of change differed between the two groups. Because estimate of `Spec:Group:Test` did not differ significantly from 0 (estimate=0.064854, p=0.77159), there was no evidence that the two groups differed in the amount of increase in their reliance of the spectral cue. In comparison, estimate of `Group:BA:Dur` was -0.599705 (p=0.00295), indicating the experimental group had significantly larger amount of decrease in their reliance of duration than the control group.

Table 20 Statistical output of the mixed-effects logistic regression model to account for the test-group interaction

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	0.662235	0.110037	6.018	1.76e-09***
Spec	0.693374	0.124810	5.555	2.77e-08***
Group	0.492268	0.219281	2.245	0.02477*
Test	-0.008006	0.143289	-0.056	0.95544
Dur	1.168791	0.136530	8.561	<2e-16***
Spec:Group	0.003751	0.248563	0.015	0.98796
Spec:Test	0.506900	0.114340	4.433	9.28e-06***
Group:Test	0.001239	0.283966	0.004	0.99652
Group:Dur	0.086670	0.272436	0.318	0.75039
Test:Dur	-0.247131	0.103216	-2.394	0.01665*
Spec:Group:Test	0.064854	0.223407	0.290	0.77159
Group:Test:Dur	-0.599705	0.201742	-2.973	0.00295**

Furthermore, whether the intervention had caused participants to rely more on both cues was also examined by comparing the 'gain scores' of the experimental and control group. The gain score for each participant was calculated by subtracting the transformed score in the pretest from the transformed score in the post-test. Figure 11 presents the gain scores distribution for the two groups. As we can see, gain scores of both groups were small and distributed around 0, indicating that for both groups the majority of the participants did not saw large gains regarding the transformed score. An independent t-test was performed and found no evidence that the experimental group (M=0.0104, SD=0.228) and control group (M=-0.0039, SD=0.17) differed significantly, t (42)=-0.232, p=0.818, Cohen d=0.204. Overall, the gain scores distribution, small means and t-test result together suggest there was no evidence that the intervention used in either group had promoted listeners to rely more on both cues.

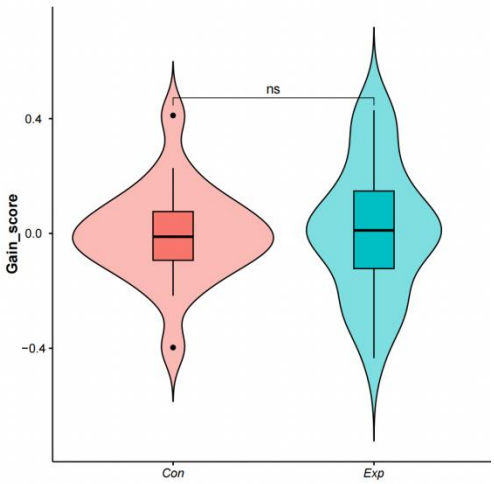


Figure 11 Distribution of gain scores of the control group (N=20) and experimental group (N=24).

4.2.4 Summary of Experiment 2

In the pretest, participants have shown similar cue weighting patterns with the L2 group in Experiment 1, since they used duration as a primary cue. `Experience`, `LProf` and `OProf` did not have significant effect on cue weighting. However, there was no evidence that `Instruction` and `Major` alone influenced cue weighting. As for the role of treatment, both control group and experimental group increased in their reliance on the spectral cue, but only the experimental group observed a significant decrease in their reliance on duration.

4.3 Discussion

Experiment 2 focused on the cue weighting behaviors of L2 speakers and how the patterns could be shifted through explicit instruction. Consistent with Experiment 1, pretest of Experiment 2 showed that Chinese L2 English learners at the group level relied primarily on vowel duration when distinguishing the English /i:/-/ɪ/, and participants in both experiments demonstrated substantial variability at the individual level. In terms of the potential variables influencing the cue weighting patterns, both Experiment 1 and 2 found no evidence that participants' experience of living in English-speaking countries, self-reported listening proficiency or overall proficiency had an impact on L2 cue weighting behaviors. More importantly, unlike Experiment 1 where participants with proper previous instruction relied more on vowel spectrum than participants with improper instruction, in Experiment 2 the two groups did not differ significantly in terms of their reliance on either temporal or spectral cue.

However, from the existing evidence, it is not clear why Experiment 1 and 2 differed regarding previous instruction's influence. One possible reason might be that major and previous instruction must work in tandem to promote reliance on the spectral cue. Although in Experiment 2 major or previous instruction alone was not found to have significant influence on L2 cue weighting, there was positive interaction of `Instruction`, `Major` and `Spec`, which indicates there was a tendency that learners who had an English major background and at the same time had received proper instruction of /i:/-/ɪ/ tended to give more weight to the spectral cue than other participants. Although the effect was numerically in this direction, such direction of effect was after all statistically non-significant. Therefore, the proposed interpretation is largely tentative. However, we cannot rule out such possibility with a larger sample. Therefore, further investigation must be carried out based on larger sample size to examine previous instruction's influence on weighting of the two cues and the interaction with major. It is also worth noting that despite Experiment 2 found no evidence of previous instruction's influence on L2 cue weighting in the pretest, results of the post-test confirmed that explicit instruction indeed had a causal influence on L2 participants' cue weighting, which will be further discussed in the General Discussion section.

Chapter 5 General Discussion

This study seeks to investigate three research questions. First, it compared the cue weighting patterns of English /i:/-/ɪ/ of L1 and L2 English speakers at the group level. Second, it investigated the individual differences for the two groups and what had caused such variability in L2 speakers. Third, this research examined explicit instruction's role in L2 cue weighting. In general, the two experiments found that both native speakers of British English and Chinese English learners relied on duration as the primary cue at the group level. However, the two groups differed in that native speakers used both cues more. Furthermore, both groups observed substantial individual variability in the individual cue weighting patterns. Regarding the potential factors influencing L2 cue weighting, L2 proficiency and experience of staying in English-speaking countries did not serve as significant predictors. In comparison, the role of previous instruction was a significant predictor in Experiment 1 but not in Experiment 2. However, intervention in Experiment 2 suggests explicit instruction had a causal effect of shifting the L2 cue weighting behaviors. Further discussion of these results related to the three research questions is presented below.

5.1 RQ1: the L1 and L2 cue weighting patterns at the group level

Regarding L2 cue weighting, both Experiment 1 and 2 showed that duration was the primary cue for Mandarin-L1 L2 English speakers at the group level, which is consistent with previous /i:/-/ɪ/ cue weighting studies on Chinese English speakers (e.g., Bohn, 1995; Yang et al., 2016; Wang & Munro, 1999). In terms of L1 cue weighting, native speakers of British English investigated in Experiment 1 relied on both cues, with duration being the dominant cue. This differs from previous research on native speakers of American English, where these native speakers relied predominantly and even exclusively on the spectral cue (e.g., Bohn, 1995; Hillenbrand et al., 2000).

Such discrepancy from previous literature might be due to the different varieties of English (American English versus British English) investigated. Findings of this research echo Escudero and Boersma's (2004) study on native speakers of British English, as both studies showed that native speakers of British English tended to use a combination of the two cues. Escudero and Boersma examined the individual cue weighting patterns of 21 native speakers of southern British English and found that only 5 of the subjects relied exclusively on the spectral cue, whereas 14 subjects relied on both vowel spectrum and duration. In this study, for the 'bead-bid' stimuli, 21 of the 29 native speakers in Experiment 1 relied on both cues. Thus, studies indicate native British English speakers were more reliant on both cues when differentiating English /i:/-/ɪ/, which is substantially different from American English speakers, who tended to rely exclusively on vowel spectrum. However, this study is only the second one investigating how native speakers of British English categorize the /i:/-/ɪ/ contrast. Future studies should compare British English and

American English to confirm such discrepancy and to examine the possible reasons behind this phenomenon.

In terms of the comparison between the L1 and L2 speakers, the regression model in Experiment 1 found that the two groups did not differ significantly at the group level. From the distributions of the transformed score calculated for each individual, however, we could see that the L1 and L2 groups differed in the extent of relying on both cues, since the 29 native speakers had significantly larger transformed scores than the 73 L2 learners. This means the native speakers used both cues more in comparison to the L2 listeners. This difference also seems to fit with Escudero and Boersma's (2004), where only 4 of the 30 Spanish listeners while 14 of the 21 native listeners investigated relied on both duration and spectrum when categorizing the English /i:/-/ɪ/.

5.2 RQ 2: individual differences

Although there were distinct patterns at the group level, the study also found substantial individual differences among both native English speakers and L2 speakers. Therefore, it was important to investigate the potential factors which might have influenced cue weighting. This research focused on factors affecting the L2 cue weighting in particular. This section discusses the influence of experience, listening proficiency and overall proficiency, while the role of major and instruction are discussed in Section 5.3.

In terms of influence of L2 proficiency, both Experiment 1 and 2 found no evidence that self-perceived listening proficiency and overall proficiency influenced L2 weighting of vowel spectrum and duration. This supports the findings of Yang et al. (2016) and Yazawa et al. (2020). Yang et al. (2016) compared native Mandarin speakers' cue weighting of English /i:/-/ɪ/ and found no significant difference between the higher proficiency group and lower proficiency group. Likewise, in Yazawa et al. (2020), proficiency was not a significant predictor of Japanese English L2 learners' cue weighting patterns of English /i:/-/ɪ/. However, L2 proficiency was found to influence Korean English speakers' cue weighting patterns of English consonants (Kong & Kang, 2023; Kong & Yoon, 2013) and Dutch speakers' cue weighting of English lexical stress (Tremblay et al., 2021).

One possible reason to account for such mixed finding lies in the measurement of L2 proficiency. The three aforementioned studies which found significance influence of proficiency adopted standardized proficiency measurement tools: Tremblay et al. (2021) adopted the Lexical Test for Advanced Learners of English (LexTALE); Kong & Kang (2023) used participants' latest TOEIC scores (The Test of English for International Communication); Kong & Yoon (2013) classified participants' English proficiency into 'high' and 'low' according to high-school English level standardization. In comparison, this study and the two studies where L2 proficiency did not play a role lacked standardized proficiency measurement: this study adopted a self-report approach to English proficiency;

Yang et al. (2016) did not specifically measure English proficiency of participants but regarded the English major group to have higher English proficiency than the non-English major group; Yazawa et al. (2020) did not specify how proficiency was measured. As a result, proficiency level obtained in these three studies had not necessarily reflected participants' actual English proficiency, which might have masked the influence of L2 proficiency on L2 cue weighting. However, the three studies which adopted standardized measurement focused on acoustic aspects other than English vowel perception. Therefore, future studies should adopt standardized measurement tools to examine whether L2 proficiency has an influence on L2 cue weighting of English vowels in particular.

Regarding the influence of L2 experience, both experiments found no evidence that this was a predictor of L2 participants' cue weighting behaviors. This was consistent with Yazawa et al. (2020), which did not find any evidence that Japanese participants' experience of living in the USA had any impact on cue weighting of English /i:/-/ɪ/. However, Escudero and Boersma (2004) found that experience of living in the target language speaking regions did shape L2 speakers' preferences of acoustic cues. They investigated native Spanish speakers' cue weighting of English /i:/-/ɪ/ and found that participants who had spent more time in Scotland tended to be spectral listeners, while participants who had spent more time in England tended to be duration listeners. These mixed findings of L2 experience's influence might be due to the length of experience investigated in these studies. In general, participants in this research and Yazawa et al. (2020) had shorter experience of living in the English-speaking regions than participants in Escudero and Boersma (2004). For this research, the 11 participants in Experiment 1 who had the experience of living in English-speaking countries had a mean length of 22.1 months and the 14 participants in Experiment 2 had a mean length of 11.7 months. In Yazawa et al. (2020), only four participants had spent more than a year in English-speaking countries. In comparison, in Escudero and Boersma (2004), 14 participants had lived in Scotland for an average of 34 months and 12 participants had lived in England for 43 months on average. Therefore, it is possible that in this study and Yazawa et al. (2020) the time participants spent in English-speaking countries was not long enough. As a result, the L2 input was not enough to shape their cue weighting behaviors and differentiate them from those who had never lived in English-speaking countries. To further investigate the influence of L2 experience on perceptual cue weighting, future studies could focus on how cue weighting patterns vary with the length of living in the regions where the target L2 is spoken.

Furthermore, another point needs to be discussed regarding the role of L2 experience. While discussing L2 experience's influence on L2 cue weighting, one should take into account the dialects of this second language rather than view this language as an integral whole. In the case of English /i:/-/ɪ/, for example, native speakers of British English and American English demonstrated different cue weighting behaviors as is discussed earlier,

suggesting that the nature of the English input environment might have affected L1 cue weighting of English /i:/-/ɪ/. If this is the case for native speakers, the cue weighting behaviors of L2 English speakers might also be affected by the types of English input. This is supported by Escudero and Boersma (2004)'s finding that Spanish-L1 speakers who had spent more time in Scotland observed a heavier weighting of the spectral cue, whereas those who had spent more time in England tended to rely more heavily on duration. Therefore, another potential factor influencing L2 cue weighting is the nature of L2 input, in the case of this study, participants' exposure to different English dialects. For example, those who had been heavily exposed to American English may essentially rely more on the spectral cue, whereas those who had received more input of British may weight duration more heavily. However, this study set out to treat English as a whole and did not take into consideration the nature of English input participants had been exposed to. Future studies should investigate what types of input L2 speakers had previously received and the corresponding percentages of exposure to look at its correlation with cue weighting behaviors.

5.3 RQ 3: the influence of instruction

This study explored instruction's role on L2 cue weighting from both the correlational and causal perspective. From the correlational perspective, instruction appears to play a role in participants' reliance of the spectral cue rather than duration. Experiment 1 found that the difference between those who have been properly and improperly instructed lied in reliance on the spectral cue, while both groups relied heavily on the temporal cue whether they had received proper previous instruction. Therefore, it is worth discussing why the difference lies in vowel spectrum but not in duration. The first possible reason takes into account the perceptual salience of duration. Because duration is salient to the human perceptual system (Kajouj et al., 2019), participants might pay much attention to duration no matter or not they had been instructed about the importance of the spectral cue.

The second reason might be that vowel duration was always emphasized in participants' process of English learning no matter the importance of the spectral cue was emphasized or not. This could be inferred from L2 speakers' responses to the questionnaire asking about how they had been instructed about the differences between English /i:/-/ɪ/. Of the 23 participants in Experiment 1, who had previously received instruction, 22 participants reported the instruction they had received included vowel length as a difference. Likewise, 40 of the 44 participants in Experiment 2 had received previous instruction on the importance of duration. As we can see, the temporal cue had always been emphasized, which may explain why L2 participants assigned great weight to duration even if they had been properly instructed about the role of the spectral cue.

Through an intervention approach, Experiment 2 has also found causal influence of explicit instruction on Chinese English speakers' L2 cue weighting behaviors. Regarding the

changes in the reliance of the temporal cue, results of Experiment 2 showed that the experimental group had significantly larger amount of decrease than the control group. Therefore, explicitly telling participants that duration was not necessarily a reliable cue to distinguish the /i:/-/ɪ/ contrast was indeed effective enough to result in less heavy weighting of duration.

But in terms of the spectral cue, both the control group and experimental group saw an increase in the reliance of the spectral cue after the treatment, but the degrees of increase for the two groups did not differ significantly. This suggests it was not clear whether explicit instruction alone was effective enough in increasing reliance of vowel spectrum. To begin with, the reason why the control group observed increase in the weighting of spectral cue may be due to intensive exposure to the variability of the spectral dimension during the pre- and post-test. As the stimuli in the pretest and post-test essentially contained intensive variability along the spectral dimension, the change of the spectral dimension of the stimuli might be treated by the listeners as 'a potential source of information' (Holt & Lotto, 2006, p. 3068). As a result, their attention might be drawn toward vowel quality, which would help them realize that English /i:/-/ɪ/ differ in spectral quality. Such role of exposure is supported by Holt and Lotto (2006). In this study, participants' cue weighting patterns of an artificial sound contrast was shifted through passive exposure to the training stimuli, where the stronger acoustic cue was held constant at an intermediate level and the weaker cue had varied acoustic properties. Results show that such short-term exposure reliably enhanced the weaker cue.

However, this does not allow us to dissociate the role of passive exposure and explicit instruction. On the one hand, it is indeed possible that instruction was not effective at all in causing participants to rely more on the spectral cue and that all what was working was the intensive exposure to the spectral dimension. This may explain why the control group did not differ significantly from the experimental group regarding the degree of increase in the reliance of the spectral dimension. On the other hand, this cannot rule out the possibility that instruction was effective in enhancing the weighting of vowel spectrum. In this case, it might be that exposure alone had already provided enough information to raise participants' awareness of the spectral cue. But this does not mean the treatment of explicit instruction alone would not be effective had participants only received explicit instruction but not intensive exposure. Therefore, a challenge for future cue weighting studies should be tearing apart the influence of passive exposure and explicit instruction to see whether instruction is effective in increasing the weighting of the spectral cue.

5.4 Pedagogical implications

The current study has practical pedagogical implications for increasing Chinese L2 English learners' /i:/-/ɪ/ contrast identification accuracy via a cue weighting approach. Findings in this research imply that teachers could change English learners' cue weighting behaviors by providing explicit instruction on the acoustic differences between the /i:/-/ɪ/ pair and, importantly, which acoustic cue is more reliable. Because the results also suggest that input of the target language matters in cue weighting behaviors, explicit instruction should at the same time be accompanied by increased exposure to English in general and the variability of the dimension which teachers aim to enhance.

However, there is no unified answer to which cue should be enhanced, since this research indicates that different dialects of English may assign different weights to the temporal and spectral cues. Besides, in some situations students might need to understand as many dialects as possible. In this sense, teachers may need to generally raise students' awareness of both cues and that both cues could be useful in distinguishing the English /i:/-/ɪ/ contrast.

It is also worth mentioning that although teachers might actually decide that the English /i:/-/ɪ/ was not the most important contrast to focus on given the differences across students' native languages, this research may have broader implications in that explicit instruction can shift perceptual cue weighting for contrasts, at least in the short-term.

5.5 Limitations

Limitations of the present research have been pointed out throughout the dissertation, including the biased responses for the 'steet-stit' stimuli in Experiment 1 and self-reported measurement of L2 proficiency. Additionally, it is not clear whether the effectiveness of the treatment in Experiment 2 could be generalized to L2 English learners who speak an L1 other than Mandarin. Mandarin is a unique case in terms of cue weighting of English /i:/-/ɪ/ in that it only has one vowel category in the high-front area of vowel space (Bohn, 1995). As a result, the spectral does not carry enough information in this vowel space. In comparison, languages such as Japanese has /ii/-/i/ contrasting only in the temporal domain rather than the spectral dimension. For this reason, instruction emphasizing the importance of the spectral cue in English /i:/-/ɪ/ might not be as effective for Japanese English learners because such instruction conflicts with their L1 knowledge. Therefore, duplication studies should be conducted to further examine if the influence of instruction treatment can be generalized to English learners from other L1 backgrounds.

The second limitation resides in the experimental design. Experiment 2 of the research adopted a pre- and post-test design. The treatment was immediately followed by the post-test, while participants' delayed performance was not examined. Thus, it is not clear whether the influence of treatment could last. To address this question, future research

should take a long-term perspective by adding delayed post-tests to the experiment.

The third limitation lies in the nature of the intervention in Experiment 2. Although the instruction materials for the experimental group made participants aware of the existence of both cues, the materials also specifically told participants that duration was not always reliable. Therefore, the nature of the instructions might have led participants to downplay duration. The gain scores of participants in Experiment 2 further confirmed that such way of instruction did not lead the experimental group to rely more on both cues. But given that the comparison of the transformed scores of the native speakers and L2 learners showed that the key difference between these native English speakers and L2 learners in this research was that the former tended to use both cues, future research should investigate if explicit instruction could direct L2 participants' attention to both cues.

5.6 Conclusion

This dissertation examined how native speakers of British English and Chinese L2 English speakers weighted the spectral and temporal cues when categorizing English /iː/-/ɪ/ vowel contrast. For both groups of listeners, duration was the dominant cue at the group level. But the native speakers of British English were found to differ from native speakers of American English examined in previous studies, who relied almost exclusively on vowel spectrum. The L1 and L2 group differed in that the former tended to use both cues more.

This research also found substantial individual differences for both native and L2 English listeners, and what might have caused the variability of the L2 group was also investigated. For the L2 group, self-perceived L2 proficiency was not found to influence cue weighting patterns, but further studies are needed to examine L2 proficiency's influence adopting more standardized proficiency measurement tools. Furthermore, there was no evidence that L2 experience influenced cue weighting either. The reason might be that the L2 participants recruited in this research had relatively short experience of living in English-speaking countries.

This study approached instruction's role on L2 cue weighting from both correlational and causal perspective. From the correlational perspective, Experiment 1 and 2 showed discrepant results. In Experiment 1, previous instruction was a significant predictor of L2 cue weighting behaviors, whereas in Experiment 2 it was not. The reason proposed was that instruction must work in tandem with major to influence cue weighting behaviors. From the causal perspective, Experiment 2 took an intervention approach and found that explicit instruction reduced participants' reliance on duration. In contrast, reliance of the spectral cue increased in both experimental and control group, with no evidence that the amount of increase was different.

References

Anwyl-Irvine, A.L., Massonié, J., Flitton, A., Kirkham, N.Z., Evershed, J.K. (2019). Gorilla in our midst: an online behavioural experiment builder. *Behavior Research Methods. 52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Baek, H. (2022). Prosodic cue weighting in the processing of ambiguous sentences by native English listeners and Korean listeners of English. *Journal of the Acoustical Society of America*, *151*(1), 158-167. https://doi.org/10.1121/10.0009171

BBC Learning English. (2014, Nov 04). The sounds of English. BBC Learning English. https://www.bbc.co.uk/learningenglish/english/features/pronunciation/longvowel1

BBC Learning English. (2014, Nov 04). The sounds of English. BBC Learning English. https://www.bbc.co.uk/learningenglish/english/features/pronunciation/shortvowel1

BBC Learning English. (2014, Nov 04). The sounds of English. BBC Learning English. https://www.bbc.co.uk/learningenglish/english/features/pronunciation/shortvowel6

Black, J. W. (1949). Natural frequency, duration, and intensity of vowels in reading. *Journal of Experimental Speech and Hearing Disorder, 14* (1), 216–221.

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics*, *18*(1), 37-49. https://doi.org/10.1016/S0095-4470(19)30357-2

Boersma, P. & Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.14, retrieved 4 August 2023 from http://www.praat.org/

Bohn, O. (1995). Cross language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279-395). Baltimore, MD: York Press.

Cerviño-Povedano, E., & Mora, J. C. (2015). Spanish EFL Learners' Categorization of /i:-I/ and Phonological Short-Term Memory. *Procedia, Social and Behavioral Sciences*, *173*, 18-23. https://doi.org/10.1016/j.sbspro.2015.02.024

Chan, A. Y. W. (2012). Cantonese English as a Second Language Learners' Perceived Relations Between "Similar" L1 and L2 Speech Sounds: A Test of the Speech Learning Model. *The Modern Language Journal*, *96*(1), 1-19. https://doi.org/10.1111/j.1540-4781.2012.01291.x

Cheng, R. L. (1966). Mandarin phonological structure. *Journal of Linguistics*, *2*(2), 135-158. https://doi.org/10.1017/S0022226700001444

Cohen, A. (1971). *The Phonemes of English: A Phonemic Study of the Vowels and Consonants of Standard English*. Springer Dordrecht.

Crystal, T. H., & House, A. S. (1988). The duration of American-English vowels: an overview. *Journal of Phonetics*, *16*(3), 263-284. https://doi.org/10.1016/S0095-

4470(19)30500-5

Dufour, S., Brunellière, A., & Frauenfelder, U. H. (2013). Tracking the Time Course of Word-Frequency Effects in Auditory Word Recognition With Event-Related Potentials. *Cognitive Science, 37*(3), 489-507.

Escudero, P. (2001). The role of the input in the development of L1 and L2 sound contrasts: Language-specific cue weighting for vowels, In A. Dominguez, & A. Johansen (Eds.), *Proceedings of the 25th annual Boston University Conference on Language Development* (pp. 250–261). Somerville, MA: Cascadilla Press.

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, *37*(4), 452-465. https://doi.org/10.1016/j.wocn.2009.07.006

Escudero, P., & Boersma, P. (2004). BRIDGING THE GAP BETWEEN L2 SPEECH PERCEPTION RESEARCH AND PHONOLOGICAL THEORY. *Studies in Second Language Acquisition*, *26*(4), 551-585. https://doi.org/10.1017/S0272263104040021

Flege, J. (1995). Second-language speech learning: theory, findings, and problems, In W. Strange (Ed.), *Speech perception and linguistic experience: theoretical and methodological issues* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437-470. https://doi.org/10.1006/jpho.1997.0052

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(4), 526-540. https://doi.org/10.1037/0096-1523.10.4.526

Grenon, I., Kubota, M., & Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. *Journal of Phonetics*, *72*, 17-34. https://doi.org/10.1016/j.wocn.2018.10.005

Giannakopoulou, A, Brown, H, Clayards, M, & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners.

Hao, Y.-C. (2018). Second Language Perception of Mandarin Vowels and Tones. *Language and Speech*, *61*(1), 135-152. https://doi.org/10.1177/0023830917717759

Harnsberger, J. D. (2001). On the relationship between identification and discrimination of non-native nasal consonants. *Journal of the Acoustical Society of America*, *110*(1), 489-503. https://doi.org/10.1121/1.1371758

Higgins, J. (2013). Minimal pairs for English RP. Retrieved June 24, 2023. Retrieved from http://myweb.tiscali.co.uk/wordscape/wordlist/minpairs.html

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of

American english vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099-3111. https://doi.org/10.1121/1.411872

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America*, *108*(6), 3013-3022. https://doi.org/10.1121/1.1323463

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, *119*(5), 3059-3071. https://doi.org/10.1121/1.2188377

House, A. S., & Fairbanks, G. (1953). The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels. *The Journal of the Acoustical Society of America*, *25*(1), 105-113. https://doi.org/10.1121/1.1906982

Idemaru, K., & Holt, L. L. (2013). The developmental trajectory of children's perception and production of English /r/-/l. *Journal of the Acoustical Society of America*, *133*(6), 4232-4246. https://doi.org/10.1121/1.4802905

Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of Phonetics*, *39*(4), 571-584. https://doi.org/10.1016/j.wocn.2011.03.003

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. i., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57. https://doi.org/10.1016/S0010-0277(02)00198-1

Kajouj, F., Kager, R. W. J., Acquisition, I. L. S., & Engelse, t. (2019). Effects of bilingualism on cue weighting: How do bilingual children perceive the Dutch [ɑ]-[a:] contrast? *International Journal Of Bilingualism: Cross-disciplinary Cross-linguistic Studies of Language Behavior*, *23*(2), 509-524. https://doi.org/10.1177/1367006917745696

Kang, E., & Han, Z. (2015). The Efficacy of Written Corrective Feedback in Improving L2 Written Accuracy: A Meta-Analysis. *The Modern Language Journal*, *99*(1), 1-18. https://doi.org/10.1111/modl.12189

Kathania, H. K., Shahnawazuddin, S., Ahmad, W., & Adiga, N. (2019). Role of Linear, Mel and Inverse-Mel Filterbanks in Automatic Recognition of Speech from High-Pitched Speakers. *Circuits, Systems, and Signal Processing, 38*(10), 4667-4682.

Kim, H., & Tremblay, A. (2021). Korean listeners' processing of suprasegmental lexical contrasts in Korean and English: A cue-based transfer approach. *Journal of Phonetics*, *87*, 101059. https://doi.org/10.1016/j.wocn.2021.101059

Kivistö-De Souza, H., Carlet, A., Julkowska, I. A., & Rato, A. (2017). Vowel inventory size matters: Assessing cue-weighting in L2 vowel perception. *Ilha Desterro*, *70*(3), 33-46. https://doi.org/10.5007/2175-8026.2017v70n3p33

Klatt, D. H, & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America, 87*(2), 820-857.

Kong, E. J., & Kang, S. (2023). Individual Differences in Categorical Judgment of L2 Stops: A Link to Proficiency and Acoustic Cue-Weighting. *Lang Speech*, *66*(2), 354-380. https://doi.org/10.1177/00238309221108647

Kong, E. J., & Yoon, I. H. (2013). L2 Proficiency Effect on the Acoustic Cue-Weighting Pattern by Korean L2 Learners of English: Production and Perception of English Stops. *Phonetics and Speech Sciences*, *5*(4), 81-90. https://doi.org/10.13064/KSSS.2013.5.4.081

Lee, W.-S., & Zee, E. (2003). Illustrations of the IPA: Standard Chinese (Beijing). *Journal of the International Phonetic Association*, *33*(1), 109.

Lehet, M., & Holt, L. L. (2017). Dimension-Based Statistical Learning Affects Both Speech Perception and Production. *Cognitive Science*, *41*(S4), 885-912. https://doi.org/10.1111/cogs.12413

Lipski, S. C., Escudero, P., & Benders, T. (2012). Language experience modulates weighting of acoustic cues for vowel perception: An event-related potential study. *Psychophysiol*, *49*(5), 638-650. https://doi.org/10.1111/j.1469-8986.2011.01347.x

Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/. In J. Slifka, S. Manuel, & M. Matthies (Eds.), Presented at *From sound to sense: 50+ Years of Discoveries in Speech Communication*, (pp. C181-C186).

Nishi, K., & Kewley-Port, D. (2007). Training Japanese Listeners to Perceive American English Vowels: Influence of Training Sets. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1496-1509. https://doi.org/10.1044/1092-4388(2007/103)

Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *Journal of the Acoustical Society of America*, *115*(4), 1777-1790. https://doi.org/10.1121/1.1651192

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, *61*(5), 1352-1361. https://doi.org/10.1121/1.381409

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ruan, Y., & Saito, K. (2023). Less precise auditory processing limits instructed L2 speech learning: Communicative focus on phonetic form revisited. *System*, *114*, 103020. https://doi.org/10.1016/j.system.2023.103020

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of*

Phonetics, *52*(Sep), 183-204. https://doi.org/10.1016/j.wocn.2015.07.003

Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews - Cognitive Science*, *11*(2), e1521-n/a. https://doi.org/10.1002/wcs.1521

Stoltzfus, J. C. (2011). Logistic Regression: A Brief Primer. *Academic Emergency Medicine*, *18*(10), 1099-1104. https://doi.org/10.1111/j.1553-2712.2011.01185.x

Tremblay, A., Broersma, M., Zeng, Y., Kim, H., Lee, J., & Shin, S. (2021). Dutch listeners' perception of English lexical stress. A cue-weighting approach. *The Journal of the Acoustical Society of America*, *149*(6), 3703-3714. https://doi.org/10.1121/10.0005086

UCLA: Statistical Consulting Group. (2021). Introduction to generalized linear mixed models. Retrieved June 24, 2023, from University of California, Los Angeles, Introduction to SAS. Retrieved from https://stats.oarc.ucla.edu/other/mult-pkg/introduction-to-generalized-linear-mixed-models/

UCLA: Statistical Consulting Group. (2021). Mixed effects logistic regression: R data analysis examples. Retrieved June 24, 2023, from University of California, Los Angeles, Introduction to SAS. Retrieved from https://stats.oarc.ucla.edu/r/dae/mixed-effects-logistic-regression/

Udovchenko, I. (2016). *L1 influence on acoustic cue weighting in perception and production of English high front vowel contrasts*. Mater of Arts, California State University.

Wang, X., Munro, M.J. (1999). The perception of English tense-lax vowel pairs by native Mandarin speakers: the effect of training on attention to temporal and spectral cues. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 125–128).

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4), 539-552. https://doi.org/10.1016/j.system.2004.09.011

Yang, X., Shi, F., Liu, X., & Zhao, Y. (2016). Learning styles and perceptual patterns for English /i/ and /I/ among Chinese college students. *Applied Psycholinguistics*, *37*(3), 673-701. https://doi.org/10.1017/S014271641500020X

Yazawa, K., Whang, J., Kondo, M., & Escudero, P. (2020). Language-dependent cue weighting: An investigation of perception modes in L2 learning. *Second Language Research*, *36*(4), 557-581. https://doi.org/10.1177/0267658319832645

Zhang, X. (2012). *A comparison of cue-weighting in the perception of prosodic phrase boundaries in English and Chinese*. Doctor of Philosophy, The University of Michigan.

Appendix 1: CUREC Approval Letter

Subject: Curec approval
Date:
From:
To:
CC:
Attachments:

Dear

Research title:
Acoustic Cue-Weighting in L1 and L2 English Vowel Perception

Research ethics reference:
EDUC_C1A_23_027

The above application has been considered on behalf of the Education Departmental Research Ethics Committee (DREC) in accordance with the University's procedures for ethical approval of all research involving human participants.

I am pleased to inform you that, based on the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted. Please keep in mind that you will be required to update us on any amendments to your study should you need to change your research methods and may need to complete a data protection impact assessment if you use online platforms to conduct and record interviews. There is an application for this, and it is kept separate from the ethics committee. Please see the link below for more information.
https://researchsupport.admin.ox.ac.uk/policy/data/checklist

Please continue to follow all current guidance issued by CUREC during the pandemic, notably COVID-19: CUREC guidance on research involving human participants, https://researchsupport.admin.ox.ac.uk/governance/ethics/coronavirus

If needed, please follow the guidance on online data collection and research methods issued by the University,
(1) https://researchsupport.admin.ox.ac.uk/covid-19/data#collapse2299911
(2) https://infosec.web.ox.ac.uk/article/guidelines-for-using-zoom

If relevant, please also check the CUREC website for their best practice research guides, https://researchsupport.admin.ox.ac.uk/governance/ethics/resources/bpg

Yours sincerely,

Appendix 2: Participant Information Sheet for Experiment 1 (for native English speakers)

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Supervisor XXX

Acoustic Cue-Weighting in L1 and L2 English Vowel Perception

Information for Participants

Dear Participants,
You are being invited to take part in a research study. Before you decide to participate, it is important to understand why the research is being conducted and what your participation entails. Please take time to read the following information carefully. Please ask if there are any aspects of the project that are unclear or if you would like more information. Take time to decide whether or not you would like to take part in this research.

Why have I been chosen?

For this study, we are seeking for both monolingual native speakers of English and advanced English speakers whose first language is Mandarin. You have been identified as someone who fall into one of these categories. The hope is that by comparing speech perception patterns of native and non-native speakers of a language, valuable pedagogical implications can be drawn.

What tasks will I complete in this research?

This study takes around 30 to 40 minutes to complete. You will be asked to do tasks such as listening to English words and choosing the correct spelling. After this, you will complete a language background questionnaire.

Do I have to take part? What are the risks and benefits of taking part?

It is your decision to take part in this study. You can decide to stop participating at any time and if you do your data will be withdrawn from the study. There are no known risks to taking part. Your participation will benefit those trying to explore how English speech sounds are perceived and processed by native speakers of English in comparison with English learners in general, and with Mandarin-speaking learners in particular.

What happens to the data collected?

Data is collected via Gorilla (https://gorilla.sc/). Gorilla is an online platform for conducting psychological research online and all data is held on secure servers within the European Economic Area. As such, Gorilla is the data controller with respect to your personal data and will determine how their personal data is used. Please see their privacy notice here: https://gorilla.sc/privacy. Gorilla will share only fully anonymised data with the University of Oxford, for the purposes of research. Research data will be stored securely by Oxford University. It will be kept for a minimum of 3 years after publication of our research findings. Responsible members of the University of Oxford and funders may be given access to data for monitoring and/or audit of the study to ensure we are complying with guidelines, or as otherwise required by law. We may share completely anonymised electronic data with other researchers, for example on the Open Science Framework (OSF) website (https://osf.io/). Storing data on the OSF website enables data to be shared between funding bodies, researchers and institutions with ease, whilst

offering the opportunity of a wider evaluation and scrutiny by the scientific community, thus allowing a greater and more accurate replication and validation of research results. This will contain no personal data and there will be no way to identify you from the data stored.

What will happen to the results of this research?

The results of this research will form the basis of an Oxford master's dissertation. Some results maybe published in relevant academic journals and presented at conferences. Summaries of the results may also be shared with other interested parties such as language teachers. If you wish to obtain a copy of the published, please contact the researcher.

Who is organizing the research?

The research organized as an independent master's dissertation project in conjunction with the Department of Education, University of Oxford.

Contact for Further Information or Follow-up

Should you have any further questions about this research, please feel free to contact the researcher ([contact details]) or her supervisor ([contact details]). Your queries will normally be answered within 10 working days. If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Social Science & Humanities Interdivisional Research Ethics Committee at the university of Oxford (Email: ethics@socsci.ox.ac.uk; Address: Research Service, University of Oxford, Wellington Square, Oxford OX1 2JD) who will seek to resolve the matter as soon as possible.

Appendix 3: Participant Information Sheet for Experiment 1 (for L2 English learners)

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Supervisor XXX

Acoustic Cue-Weighting in L1 and L2 English Vowel Perception

Information for Participants

亲爱的受试,
您被邀请参加一项研究。在您决定参与之前，重要的是要了解进行这项研究的目的，以及参与条件。请花时间仔细阅读以下信息。若有任何不清楚的地方， 或是需要更多信息，请询问研究者。与此同时，您可以决定是否愿意参加这项研究。

为什么选我?
对于这项研究，我们正在寻找英语母语者和普通话为母语的中高级英语二语习得者。 您已被确定为属于这些类别之一的人。我们希望通过比较二者语音感知模式，从而得出有价值的教学意义。

本实验具体要完成哪些任务?
完成这项研究大约需要 30 到 40 分钟。 您要完成的任务包括听英语单词后选择正确的拼写。 在此之后，您将完成一份语言背景调查问卷。

我必须要参加吗？参加的好处和潜在风险有哪些?
是否参加这项研究完全是您自己的决定。 您可以随时决定停止参与，如果您这样做，那么您的数据将从研究中撤回。 目前参与本研究没有已知的风险。 您的参与将有益于对英语母语者与英语二语者语音感知（尤其是中文母语者）区别的研究这一领域作出贡献。

研究数据去向?
数据将通过 Gorilla (https://gorilla.sc/) 收集。 Gorilla 是一个用于进行心理研究的线上平台。所有数据都储存在欧洲的安全服务器上。 因此，Gorilla 是您个人数据的实际掌握者，并决定个人数据的使用规范。 点击此处查看他们的隐私声明：https://gorilla.sc/privacy。 出于研究目的，Gorilla 只会与牛津大学共享完全匿名的数据。 研究数据将安全地储存在牛津大学。 它将在我们的研究结果发表后至少保存 3 年。 牛津大学的负责成员和资助者可能有权访问数据以监控和/或审计研究，以确保我们遵守指导方针或相关法律。 我们可能会与其他研究人员共享完全匿名的电子数据，例如在开放科学框架 (OSF) 网站 (https://osf.io/) 上。 在 OSF 网站上存储数据使数据能够在资助机构、研究人员和机构之间轻松共享，同时为科学界提供更广泛的评估和审查的机会，从而允许研究者更广泛准确地复制和验证研究结果。 但这些数据不包含任何个人数据，也无法从存储的数据中识别您的身份。

本研究结果将用来干什么?
这项研究的结果将成为牛津硕士论文的基础。 一些结果可能会发表在相关的学术期刊上并在会议上发表。 结果摘要也可以与其他相关方（例如语言教师）共享。 如果您未来希

望获得已发布的副本，请联系研究者。

研究的组织者是谁?
该研究作为一项独立的硕士论文项目，由牛津大学教育系合作组织。

联系研究者以获得更多信息
如果您有任何其他疑问，请联系研究者 ([contact details]) 或她的导师 ([contact details])。您的问题一般会在 10 个工作日内收到答复。如果您仍然感到不满意或希望提出正式投诉，请联系牛津大学社会科学与人文跨学科研究道德伦理委员会主席（电子邮件: ethics@socsci.ox.ac.uk; 地址: Research Service, University of Oxford, Wellington Square, Oxford OX1 2JD)，您的反馈将会被尽快解决答复。

Appendix 4: Participant Information Sheet for Experiment 2

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Supervisor XXX

Acoustic Cue-Weighting in L1 and L2 English Vowel Perception

Information for Participants

Dear Participants,
You are being invited to take part in a research study. Before you decide to participate, it is important to understand why the research is being conducted and what your participation entails. Please take time to read the following information carefully. Please ask if there are any aspects of the project that are unclear or if you would like more information. Take time to decide whether or not you would like to take part in this research.
亲爱的受试,
您被邀请参加一项研究。在您决定参与之前，重要的是要了解进行这项研究的目的，以及参与条件。请花时间仔细阅读以下信息。若有任何不清楚的地方， 或是需要更多信息，请询问研究者。与此同时，您可以决定是否愿意参加这项研究。

Why have I been chosen? 为什么选我?
For this study, we are seeking for both monolingual native speakers of English and advanced English speakers whose first language is Mandarin. You have been identified as someone who fall into one of these categories. The hope is that by comparing speech perception patterns of native and non-native speakers of a language, valuable pedagogical implications can be drawn.
对于这项研究，我们正在寻找英语母语者和普通话为母语的中高级英语二语习得者。 您已被确定为属于这些类别之一的人。我们希望通过比较二者语音感知模式，从而得出有价值的教学意义。

What tasks will I complete in this research? 本实验具体要完成哪些任务?
This study consists of two stages. You can do the first part today and you will need to return in about 2 to 5 days to complete the second stage. Each stage takes around 20 minutes to complete.
研究包括两个阶段。您需要在完成第一阶段的 2 到 5 天完成第二阶段。每阶段耗时约 20 分钟。

Do I have to take part? What are the risks and benefits of taking part? 我必须要参加吗? 参加的好处和潜在风险有哪些?
It is your decision to take part in this study. You can decide to stop participating at any time and if you do your data will be withdrawn from the study. There are no known risks to taking part. Your participation will benefit those trying to explore how English speech sounds are perceived and processed by native speakers of English in comparison with English learners in general, and with Mandarin-speaking learners in particular.
是否参加这项研究完全是您自己的决定。 您可以随时决定停止参与，如果您这样做，那么您的数据将从研究中撤回。 目前参与本研究没有已知的风险。 您的参与将有益于对英语母语者与英语二语者语音感知（尤其是中文母语者）区别的研究这一领域作出贡献。

What happens to the data collected? 研究数据去向?

Data is collected via Gorilla (https://gorilla.sc/). Gorilla is an online platform for conducting psychological research online and all data is held on secure servers within the European Economic Area. As such, Gorilla is the data controller with respect to your personal data and will determine how their personal data is used. Please see their privacy notice here: https://gorilla.sc/privacy. Gorilla will share only fully anonymised data with the University of Oxford, for the purposes of research. Research data will be stored securely by Oxford University. It will be kept for a minimum of 3 years after publication of our research findings. Responsible members of the University of Oxford and funders may be given access to data for monitoring and/or audit of the study to ensure we are complying with guidelines, or as otherwise required by law. We may share completely anonymised electronic data with other researchers, for example on the Open Science Framework (OSF) website (https://osf.io/). Storing data on the OSF website enables data to be shared between funding bodies, researchers and institutions with ease, whilst offering the opportunity of a wider evaluation and scrutiny by the scientific community, thus allowing a greater and more accurate replication and validation of research results. This will contain no personal data and there will be no way to identify you from the data stored.
数据将通过 Gorilla (https://gorilla.sc/) 收集。 Gorilla 是一个用于进行心理研究的线上平台。所有数据都储存在欧洲的安全服务器上。 因此，Gorilla 是您个人数据的实际掌握者，并决定个人数据的使用规范。 点击此处查看他们的隐私声明：https://gorilla.sc/privacy。 出于研究目的，Gorilla 只会与牛津大学共享完全匿名的数据。 研究数据将安全地储存在牛津大学。 它将在我们的研究结果发表后至少保存 3 年。 牛津大学的负责成员和资助者可能有权访问数据以监控和/或审计研究，以确保我们遵守指导方针或相关法律。 我们可能会与其他研究人员共享完全匿名的电子数据，例如在开放科学框架 (OSF) 网站 (https://osf.io/) 上。 在 OSF 网站上存储数据使数据能够在资助机构、研究人员和机构之间轻松共享，同时为科学界提供更广泛的评估和审查的机会，从而允许研究者更广泛准确地复制和验证研究结果。 但这些数据不包含任何个人数据，也无法从存储的数据中识别您的身份。

What will happen to the results of this research? 本研究结果将用来干什么?
The results of this research will form the basis of an Oxford master's dissertation. Some results maybe published in relevant academic journals and presented at conferences. Summaries of the results may also be shared with other interested parties such as language teachers. If you wish to obtain a copy of the published, please contact the researcher.
这项研究的结果将成为牛津硕士论文的基础。 一些结果可能会发表在相关的学术期刊上并在会议上发表。 结果摘要也可以与其他相关方（例如语言教师）共享。 如果您未来希望获得已发布的副本，请联系研究者。

Who is organizing the research? 研究的组织者是谁?
The research organized as an independent master's dissertation project in conjunction with the Department of Education, University of Oxford.
该研究作为一项独立的硕士论文项目，由牛津大学教育系合作组织。

Contact for Further Information or Follow-up 联系研究者以获得更多信息
Should you have any further questions about this research, please feel free to contact the researcher ([contact details]) or her supervisor ([contact details]). Your queries will normally be answered within 10 working days. If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Social Science & Humanities Interdivisional Research Ethics Committee at the university of Oxford (Email: ethics@socsci.ox.ac.uk; Address: Research Service, University of Oxford, Wellington Square, Oxford OX1 2JD) who will seek to resolve the matter as soon as possible.
如果您有任何其他疑问，请联系研究者 ([contact details]) 或她的导师 ([contact details])。您的问题一般会在 10 个工作日内收到答复。如果您仍然感到不满意或希望提出正式投诉，请联系牛津大学社会科学与人文跨学科研究道德伦理委员会主席 (电子邮件: ethics@socsci.ox.ac.uk; 地址: Research Service, University of Oxford, Wellington Square, Oxford OX1 2JD)，您的反馈将会被尽快解决答复。

Appendix 5: Recruitment Poster for Experiment 1(for L2 English learners)

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

导师： Professor XXX



本实验已通过牛津大学研究伦理委员会审核，审批编号 EDUC_C1A_23_027

语音感知实验 志愿者招募

本实验旨在对比英语母语者和非母语者对英语元音的感知。希望借此对二语课堂语音教学提出建议。

实验志愿者应当满足以下条件:

1. **母语为中文**
2. **英语水平中等及以上**
3. **年满18周岁**
4. **无任何语言障碍**

本实验采取线上的方式进行，参与时间非常灵活。实验大概需要20-30分钟来完成。如果您有意参与实验，想确认是否符合条件，或想了解跟多信息，请通过以下方式联系牛津大学的xxx：

电子邮箱: [contact details]

微信号: [contact details]

您可以自行决定是否参加实验，并随时可以选择退出实验。完成实验的志愿者将参与抽奖：一等奖1名，200RMB；二等奖2名，每人100RMB；三等奖4名，每人50RMB（30人左右参与抽奖）。

Appendix 6: Recruitment Poster for Experiment 2

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

导师： Professor XXX

本实验已通过牛津大学研究伦理委员会审核，审批编号 EDUC_C1A_23_027

语音感知实验 志愿者招募

本实验旨在对比英语母语者和非母语者对英语元音的感知。希望借此对二语课堂语音教学提出建议。

实验志愿者应当满足以下条件:

5. **母语为中文**
6. **英语水平中等及以上**
7. **年满18周岁**
8. **无任何语言障碍**

本实验采取线上的方式进行，参与时间非常灵活。实验由两部分组成，您将听到一组读音相近的单词，并选择您所听到的单词。part1大概需要10-15分钟，part2大概需要20分钟。part2会在part1完成后24h发布，如果您有意参与实验，想确认是否符合条件，或想了解跟多信息，请通过以下方式联系牛津大学的xxx：

电子邮箱: [contact details]

微信号: [contact details]

您可以自行决定是否参加实验，并随时可以选择退出实验。完成实验的志愿者每人将获得20RMB保底酬劳，并参与抽奖：一等奖1名，100RMB；二等奖2名，每人50RMB；三等奖4名，每人25RMB（50人左右参与抽奖）

Appendix 7: Consent Form (English version)

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Supervisor Professor XXX



Research Consent Form

Project Title: Acoustic Cue-Weighting in L1 and L2 English Vowel Perception

Researcher: XXX

Declaration of Consent:

☐ I confirm that I have read the information available in this information sheet. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.
☐ I understand that my participant is voluntary and I am are free to withdraw at any time without giving any reason during the experiment. I understand as well that the data can't be deleted when the experiment is over as the data are anonymised and it's not possible to trace back the identity of the participant in the data.
☐ I understand that relevant sections of the data collected during the study may be looked at by individuals from University of Oxford and from regulatory authorities, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.
☐ I give permission for my child's anonymised (not personally identifiable) data to be stored on the Open Science Framework (OSF) website where researchers worldwide have access to my child's anonymised data.
☐I agree to take part in the studies.

Appendix 8: Consent Form (Chinese version)

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Supervisor XXX

同意书

项目名称: 英语元音感知实验

研究者: XXX

同意声明:

□ 我确认我已阅读此信息表中提供的信息。 并已经对其进行过思考，提出问题并得到满意的答复。
□ 我明白我的参与是自愿的，在实验过程中我可以随时退出而无需给出任何理由。 我也明白我的数据不能在实验结束时删除，因为数据是匿名的，而且不可能通过数据追溯参与者的身份。
□ 我了解牛津大学和监管机构的个人可能会查看研究期间收集的数据的相关部分。 我允许这些人访问我的记录。
□ 我同意将我的匿名（无法识别个人身份）数据存储在开放科学框架 (OSF) 网站上，全世界的研究人员都可以在该网站上访问我的匿名数据。
□我同意参加本研究。

Appendix 9: Language Background Questionnaire for Native English Speakers in Experiment 1

1. Age (Years):

2. Gender:
□ Female
□ Male
□ Prefer not to say
□ Other - please specify:

3. Please list all the languages you speak, and rank them in terms of the criteria listed.
(1-very poor; 2-Poor; 3-Fair; 4-Functional; 5-Good; 6-Very good; 7-Native like)

Language	Speaking	Listening	Reading	Writing

4. What English dialect(s) do you speak?
_____

Appendix 10: Language Background Questionnaire for L2 English Speakers in Experiment 1 (English version)

1. Age (Years):

2. Gender:

☐ Female

☐ Male

☐ Prefer not to say

☐ Other - please specify:

3. Please list all the languages you speak, and rank them in terms of the criteria listed.
(1-very poor; 2-Poor; 3-Fair; 4-Functional; 5-Good; 6-Very good; 7-Native like)

Language	Speaking	Listening	Reading	Writing

4. Do you have any experience of studying or living abroad in an English-speaking country or region? (6 months and above)

☐ Yes

☐ No

If so, please list:

1) these English speaking countries or regions where you worked/studied
2) the age of arrival
3) the respective time span.
e.g., UK; 15 years-old; 3 years.

_____

5. In your past experience of learning English, have you been explicitly instructed by your English teachers the difference between the /i:/ vowel (sheep) and the /ɪ/ vowel ?

☐ Yes

☐ No

If yes:

According to the instruction you received, what's the difference between the two vowels?

_____

Do you agree with the teacher's instruction and why?

_____

If no:

What do you think is the difference? Why do you think so?

_____

Appendix 11: Language Background Questionnaire for L2 English Speakers in Experiment 1 (Chinese version)

1. 年龄 （岁）：_____

2. 性别
□ 女
□ 男
□ 不方便透露
□ 其他（请展开说说）：_____

3. 请写出您会的所有语言 (包括母语)，并按照以下标准为每门语言‘听’、‘说’、’读‘、’写‘四项技能熟练程度打分。
(1-很差; 2-较差; 3-一般; 4-能用于日常交流; 5-好; 6-很好; 7-接近母语水平)

语言	说	听	读	写

4. 您有在英语国家或地区生活长达六个月及以上的经历吗?
□ 有
□ 没有

若有，请按照以下三个问题简要列举:
(1) 您生活过的英语国家或地区
(2) 您到达该国/地区时的年龄
(3) 您在这个国家/地区待了多长时间?
e.g.: 英国； 15 岁； 3 年
_____

5. 在学习英语的过程中，您的英语老师有明确教过你们/i:/（单词"sheep"中的元音）和/ɪ/（单词“ship”的元音）这两个元音的区别吗?
□ 有
□ 没有

若有的话:

根据您老师教你们的，/i:/（sheep）与/ɪ/ (ship)这两个元音有什么区别呢?
_____

关于这两个元音的区别，您同意老师所讲的吗？若不同意，请简要谈谈您自己的观点。
_____

若没有:

那么您认为/i:/（sheep）与/ɪ/ (ship)这两个元音的区别是什么呢？并且请再简要谈谈您为何会这么认为呢?
_____

Appendix 12: Treatment Materials for the Experimental Group in Experiment 2 (Chinese version)

1. 引入

当你最开始学习英语的时候，英语老师也许强调过，/i:/和/ɪ/这两个元音的区别在于音的长度，前者长而后者短。其实，这不完全正确。

/i:/和/ɪ/确实在长短上有区别，但这并不是它俩的主要区别。很多老师在教大家区分这两个原因的过程中忽略了一个最重要区别： /i:/和/ɪ/发音的时候嘴巴形状不同。

2. /i:/和 /ɪ/嘴形的不同

请看下面这幅图：

/ i: /



/ ɪ /

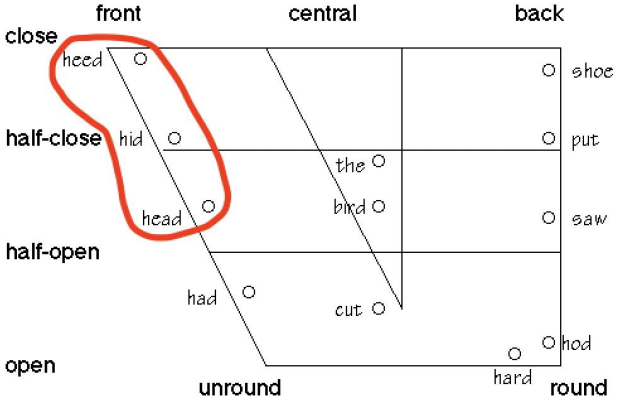


当你发/i:/这个音时，你的嘴巴咧得更开，唇部肌肉也更加的紧张。此时，上下牙之间的间隙会很小。相比之下，当你发/ɪ/这个音的时候，你的嘴唇咧得没那么开，因此更加的放松。这时，上下两排牙齿之间的距离会比发/i:/的时候更大。

（可以对着镜子自己试一试，观察一下）

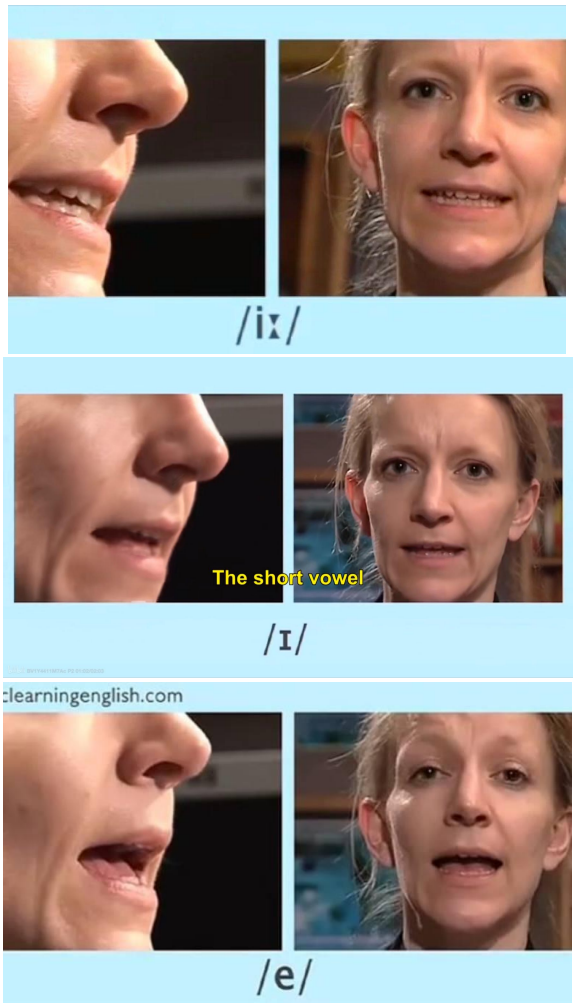
3. /i:/，/ɪ/和/e/嘴形的不同

下面这幅图展示的是英语中各元音发音是舌头的位置（横轴）和嘴巴的开合程度（纵轴）。我们现在主要来看看 /i:/(heed)，/ɪ/（hid）和/e/（head）这三个音在纵轴上的位置吧!



纵轴是元音发音时嘴巴开合程度的大小：图中越靠上的音嘴巴开口越小，越靠下开口越大。如图，发/ɪ/时的张口大小介于 /i:/ 和 /e/之间。

下面三张图更直观地展现了这种差别：



（图片来源：https://www.bbc.co.uk/learningenglish）

你可以将这种张口大小的区别理解为：/ɪ/这个音并不是单纯地将/i:/缩短，而是听上去介于/i:/和/e/之间的一个音。下面这个视频展示了/i:/，/ɪ/，/e/这三个音的区别。请仔细观看，尝试着去感受一下它们的区别。

（视频来源：https://www.bbc.co.uk/learningenglish）

4.如何更准确地区分/i:/和/ɪ/?

综上，/i:/和/ɪ/有两个主要区别：元音长短和发音时嘴巴形状不同。由于发音长短有很大的个体差异，即有的人发音普遍较长，有的人发音普遍偏短，因此仅仅通过长短来辨别

/i:/和/ɪ/很不可靠。事实上，大多数英语母语者并不是通过长短来区分/i:/和/ɪ/的。他们往往更倾向于通过发音时嘴形的不同所导致的不同音质来判断是/i:/还是/ɪ/。这意味着我们完全可以不依赖长短来分辨/i:/和/ɪ/。

举个例子：



你能分辨出哪个是 bead，哪个是 bid 吗？

	Bead	Bid
Word A	○	○
Word B	○	○

Appendix 13: Treatment Materials for the Experimental Group in Experiment 2 (English version)

1. Introduction

When you first started learning English vowels, your English teachers might have taught you that the difference between the vowel /i:/and /ɪ/ is that the former has a longer duration than the latter. However, this is not entirely true.

The vowel length indeed differentiates /i:/and /ɪ/, but it's not the whole story. Besides duration, another very important difference has often been ignored by English teachers when instructing the differences between /i:/and /ɪ/. This is the shapes of the mouth when pronouncing these two vowels.

2. Different mouth shapes of /i:/ and /ɪ/

Look at the graph:

/ iː /



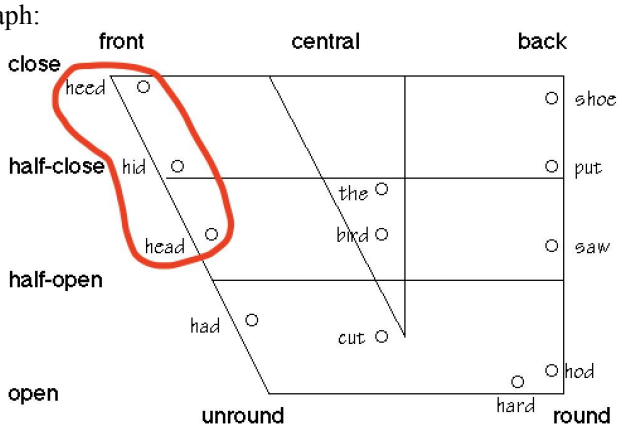
/ ɪ /



When you pronounce /i:/, your mouth is more stretched and tense. In this case, the opening between your upper and lower teeth is small. But when you pronounce /ɪ/, your mouth is less stretched and more relaxed compared to the /i:/ sound. Therefore, the opening between your upper and lower teeth is larger.

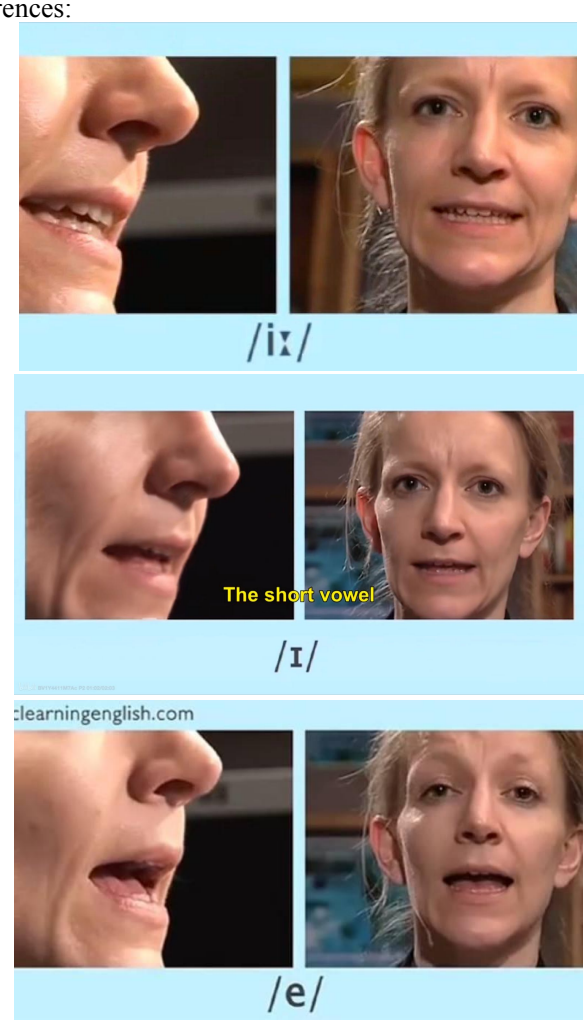
3. Different mouth shapes of /i:/, /ɪ/and /ɪ/

Now look at this graph:



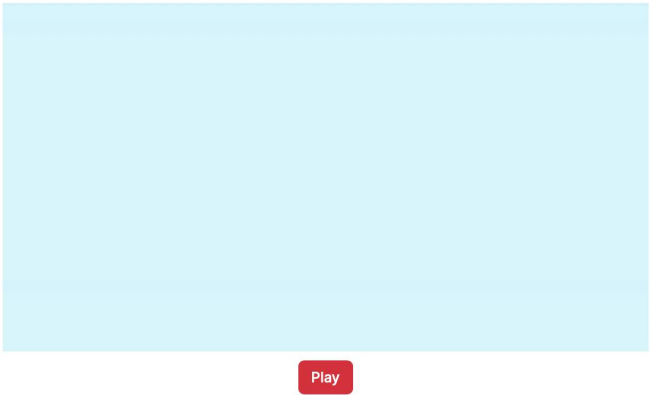
The vertical axis shows how wide open your mouth is when producing these vowels. As we can see, the mouth opening of /ɪ/ is between /i:/ and /e/. The following three pictures more clearly

demonstrate the differences:



（source of the pictures：https://www.bbc.co.uk/learningenglish）

So, you could understand this as: the property of /ɪ/ is between /i:/ and /e/, but more closed to /i:/. Here are some examples of the three vowels. Listen carefully and try to feel the difference:



（source of the video: https://www.bbc.co.uk/learningenglish）

4. How to differentiate /i:/和/ɪ/?

As we can see, there are two major differences between /i:/and /ɪ/: vowel duration and mouth

shape. Between the two differences, the latter is a more reliable way to distinguish /iː/and /ɪ/. This is because vowel duration is influenced by factors such as individual difference. In other words, some speakers may pronounce vowels longer in general, other speakers might pronounce vowels shorter. Therefore, to distinguish /iː/and /ɪ/, native English speakers tend to rely more on the difference of vowel quality caused by the mouth shape.

Now, listen to the following two words. Their duration is the same, but they differ in mouth opening. Can you try to distinguish which one is “bead” and which one is “bid”?

Word A

Word B

	Bead	Bid
Word A	○	○
Word B	○	○

Appendix 14: Examples of R Code

```
#heatmap

library(readxl)

NS_Bead <- read_excel("NS_Bead.xlsx")

NS_Bead$Dur<-factor(NS_Bead$Dur,levels=c(-2,-1,0,1,2),
                    labels=c("100","140","180","220","260"))

NS_Bead$Spec<-factor(NS_Bead$Spec,levels=c(-3,-2,-1,0,1,2,3),
      labels=c("488,2114","462,2190","436,2268","410,2349","387,2430","363,2514","340,2600"))

ggplot(NS_Bead,aes(x=Dur,y=Spec,fill=Percentage))
+geom_tile(color="#c3c7ca",lwd= 0.2)
+coord_fixed()
+scale_fill_gradientn(colours=c("white","#88cdf6","#015c92"),
                      limits=c(0, 1),breaks=seq(0,1, by=0.25))
+ylab("Spec (Hz)")+xlab("Dur (ms)")
```

```
#Mixed-Effect Logistic Regression Model

library(lme4)

Model_NSBead <- glmer(Response~Spec+Dur+(Spec+Dur|ParticipantID),data=NS_Bead,family=binomial,
            control=glmerControl(optimizer="bobyqa",optCtrl=list(maxfun=100000)))

summary(Model_NSBead)

coef(Model_NSBead)
```

Appendix 15: Individual weighting scores and transformed scores

Individual data for native speakers in Experiment 1 for the 'bead-bid' stimuli

ParticipantID	Spec	|Spec|	Dur	|Dur|	Weighting Score	Transformed Score
NS22	0.008	0.008	0.896	0.896	0.008	0.008
NS5	0.038	0.038	2.283	2.283	0.016	0.016
NS12	0.122	0.122	2.888	2.888	0.041	0.041
NS14	0.070	0.070	1.347	1.347	0.049	0.049
NS8	0.178	0.178	2.228	2.228	0.074	0.074
NS20	0.062	0.062	0.670	0.670	0.085	0.085
NS24	0.169	0.169	1.773	1.773	0.087	0.087
NS16	0.162	0.162	1.638	1.638	0.090	0.090
NS10	0.341	0.341	1.957	1.957	0.148	0.148
NS17	0.186	0.186	1.011	1.011	0.155	0.155
NS2	0.331	0.331	1.689	1.689	0.164	0.164
NS11	0.215	0.215	1.038	1.038	0.171	0.171
NS18	0.364	0.364	1.700	1.700	0.176	0.176
NS30	0.438	0.438	1.780	1.780	0.197	0.197
NS28	0.616	0.616	1.500	1.500	0.291	0.291
NS9	0.264	0.264	0.560	0.560	0.320	0.320
NS19	0.282	0.282	0.555	0.555	0.337	0.337
NS3	0.700	0.700	1.153	1.153	0.378	0.378
NS7	0.886	0.886	1.280	1.280	0.409	0.409
NS6	0.204	0.204	0.282	0.282	0.420	0.420
NS25	0.563	0.563	0.727	0.727	0.436	0.436
NS4	0.901	0.901	1.083	1.083	0.454	0.454
NS15	1.252	1.252	1.246	1.246	0.501	0.499
NS29	0.610	0.610	0.589	0.589	0.509	0.491
NS27	0.985	0.985	0.892	0.892	0.525	0.475
NS1	1.136	1.136	0.674	0.674	0.628	0.372
NS13	0.867	0.867	0.379	0.379	0.696	0.304
NS31	1.404	1.404	0.414	0.414	0.772	0.228
NS26	0.953	0.953	0.279	0.279	0.774	0.226

Individual data for L2 learners in Experiment 1 for the 'bead-bid' stimuli

ParticipantID	Spec	|Spec|	Dur	|Dur|	Weighting Score	Transformed Score
NNS2	0.007	0.007	1.786	1.786	0.004	0.004
NNS4	0.018	0.018	2.384	2.384	0.008	0.008
NNS29	0.017	0.017	1.651	1.651	0.010	0.010
NNS14	0.032	0.032	2.311	2.311	0.014	0.014
NNS11	-0.020	0.020	1.446	1.446	0.014	0.014
NNS13	-0.054	0.054	2.276	2.276	0.023	0.023
NNS7	0.015	0.015	0.579	0.579	0.026	0.026
NNS27	0.071	0.071	1.426	1.426	0.047	0.047
NNS23	0.057	0.057	1.001	1.001	0.054	0.054
NNS6	0.151	0.151	1.951	1.951	0.072	0.072
NNS18	0.139	0.139	1.779	1.779	0.073	0.073
NNS10	0.315	0.315	2.571	2.571	0.109	0.109
NNS26	0.131	0.131	0.917	0.917	0.125	0.125
NNS16	0.137	0.137	0.953	0.953	0.126	0.126
NNS19	0.231	0.231	0.978	0.978	0.191	0.191
NNS20	0.243	0.243	0.818	0.818	0.229	0.229

NNS12	0.460	0.460	1.259	1.259	0.268	0.268
NNS28	0.078	0.078	-0.091	0.091	0.460	0.460
NNS1	0.980	0.980	1.033	1.033	0.487	0.487
NNS5	0.327	0.327	0.327	0.327	0.500	0.500
NNS17	0.822	0.822	0.767	0.767	0.517	0.483
NNS25	-0.717	0.717	0.458	0.458	0.610	0.390
NNS8	0.221	0.221	0.133	0.133	0.624	0.376
NNS9	1.143	1.143	0.535	0.535	0.681	0.319
NNS24	1.275	1.275	0.231	0.231	0.847	0.153
NNS3	1.612	1.612	0.275	0.275	0.854	0.146
NNS15	1.086	1.086	0.117	0.117	0.903	0.097
NNS22	2.198	2.198	-0.225	0.225	0.907	0.093
NNS21	3.195	3.195	0.080	0.080	0.976	0.024

Individual data for native speakers in Experiment 1 for the ‘steet-stit’ stimuli

ParticipantID	Spec	|Spec|	Dur	|Dur|	Weighting Score	Transformed Score
NS11	0.017	0.017	0.361	0.361	0.044	0.044
NS9	0.010	0.010	0.174	0.174	0.057	0.057
NS8	0.113	0.113	1.237	1.237	0.084	0.084
NS10	0.157	0.157	1.181	1.181	0.117	0.117
NS17	0.188	0.188	1.249	1.249	0.131	0.131
NS24	0.183	0.183	1.067	1.067	0.146	0.146
NS2	0.329	0.329	1.889	1.889	0.148	0.148
NS5	0.154	0.154	-0.852	0.852	0.153	0.153
NS16	0.210	0.210	1.147	1.147	0.155	0.155
NS12	0.276	0.276	1.481	1.481	0.157	0.157
NS14	0.272	0.272	1.166	1.166	0.189	0.189
NS18	0.248	0.248	0.742	0.742	0.250	0.250
NS28	0.272	0.272	0.642	0.642	0.297	0.297
NS4	0.497	0.497	1.041	1.041	0.323	0.323
NS22	0.246	0.246	0.485	0.485	0.337	0.337
NS29	0.526	0.526	0.871	0.871	0.377	0.377
NS30	0.467	0.467	0.611	0.611	0.433	0.433
NS7	0.808	0.808	0.971	0.971	0.454	0.454
NS6	0.082	0.082	0.087	0.087	0.484	0.484
NS20	-0.058	0.058	0.059	0.059	0.495	0.495
NS27	0.362	0.362	0.301	0.301	0.546	0.454
NS3	0.400	0.400	0.327	0.327	0.550	0.450
NS19	0.751	0.751	0.573	0.573	0.567	0.433
NS1	1.086	1.086	0.596	0.596	0.646	0.354
NS15	1.079	1.079	0.551	0.551	0.662	0.338
NS25	0.674	0.674	0.338	0.338	0.666	0.334
NS26	0.811	0.811	0.230	0.230	0.779	0.221
NS31	1.492	1.492	0.380	0.380	0.797	0.203
NS13	1.245	1.245	-0.044	0.044	0.966	0.034

Individual data for L2 learners in Experiment 1 for the ‘steet-stit’ stimuli

ParticipantID	Spec	|Spec|	Dur	|Dur|	Weighting Score	Transformed Score
NNS27	0.005	0.005	0.822	0.822	0.006	0.006
NNS6	0.014	0.014	1.864	1.864	0.007	0.007
NNS16	0.021	0.021	1.001	1.001	0.020	0.020
NNS13	0.078	0.078	1.784	1.784	0.042	0.042
NNS18	0.041	0.041	0.910	0.910	0.043	0.043
NNS2	-0.060	0.060	1.045	1.045	0.054	0.054
NNS10	0.222	0.222	2.706	2.706	0.076	0.076
NNS14	-0.156	0.156	1.835	1.835	0.078	0.078
NNS12	0.089	0.089	0.869	0.869	0.093	0.093
NNS4	0.196	0.196	1.604	1.604	0.109	0.109
NNS26	-0.066	0.066	0.398	0.398	0.142	0.142
NNS25	-0.056	0.056	0.261	0.261	0.176	0.176
NNS29	0.181	0.181	0.810	0.810	0.183	0.183
NNS23	-0.183	0.183	0.808	0.808	0.184	0.184
NNS28	0.039	0.039	0.134	0.134	0.224	0.224
NNS19	0.437	0.437	1.207	1.207	0.266	0.266
NNS7	-0.250	0.250	0.560	0.560	0.308	0.308
NNS1	0.922	0.922	0.931	0.931	0.498	0.498
NNS20	0.346	0.346	0.335	0.335	0.508	0.492
NNS8	0.136	0.136	0.102	0.102	0.573	0.427
NNS17	1.007	1.007	0.679	0.679	0.597	0.403
NNS5	0.880	0.880	0.504	0.504	0.636	0.364
NNS3	1.494	1.494	0.538	0.538	0.735	0.265
NNS11	0.673	0.673	0.215	0.215	0.758	0.242
NNS24	1.783	1.783	0.204	0.204	0.897	0.103
NNS22	2.982	2.982	-0.318	0.318	0.904	0.096
NNS21	1.728	1.728	0.166	0.166	0.913	0.087
NNS15	1.894	1.894	-0.181	0.181	0.913	0.087
NNS9	1.693	1.693	0.110	0.110	0.939	0.061

Individual data for L2 learners in Experiment 2 for the pretest

Control group						
ParticipantID	Spec	|Spec|	Dur	|Dur|	Weighting Score	Transformed Score
CG10	0.001	0.001	1.845	1.845	0.001	0.001
CG21	-0.022	0.022	1.237	1.237	0.018	0.018
CG01	-0.086	0.086	2.872	2.872	0.029	0.029
CG22	0.054	0.054	1.353	1.353	0.039	0.039
CG18	0.031	0.031	0.690	0.690	0.043	0.043
CG09	0.100	0.100	1.366	1.366	0.068	0.068
CG12	0.040	0.040	0.511	0.511	0.072	0.072
CG11	0.189	0.189	1.809	1.809	0.095	0.095
CG20	0.200	0.200	1.616	1.616	0.110	0.110
CG24	-0.210	0.210	1.570	1.570	0.118	0.118
CG02	-0.224	0.224	1.142	1.142	0.164	0.164
CG23	-0.334	0.334	1.174	1.174	0.221	0.221
CG05	0.283	0.283	0.531	0.531	0.347	0.347
CG13	0.463	0.463	0.846	0.846	0.354	0.354
CG08	0.681	0.681	0.921	0.921	0.425	0.425
CG17	1.348	1.348	0.466	0.466	0.743	0.257
CG15	1.244	1.244	0.307	0.307	0.802	0.198
CG19	1.347	1.347	0.302	0.302	0.817	0.183

CG07	1.953	1.953	0.111	0.111	0.946	0.054
CG06	2.081	2.081	-0.067	0.067	0.969	0.031
			Experimental goup			
ParticipantID	Spec	|Spec|	Dur	|Dur|	Weighting Score	Transformed Score
EG25	-0.009	0.009	2.823	2.823	0.003	0.003
EG16	-0.021	0.021	1.971	1.971	0.011	0.011
EG22	-0.032	0.032	2.897	2.897	0.011	0.011
EG02	-0.042	0.042	3.117	3.117	0.013	0.013
EG01	0.056	0.056	2.197	2.197	0.025	0.025
EG17	-0.019	0.019	0.681	0.681	0.028	0.028
EG14	0.102	0.102	2.663	2.663	0.037	0.037
EG05	0.103	0.103	2.056	2.056	0.048	0.048
EG23	0.146	0.146	2.626	2.626	0.053	0.053
EG07	0.103	0.103	1.590	1.590	0.061	0.061
EG08	0.090	0.090	1.307	1.307	0.064	0.064
EG15	0.226	0.226	2.338	2.338	0.088	0.088
EG21	0.114	0.114	1.164	1.164	0.089	0.089
EG04	0.471	0.471	3.086	3.086	0.132	0.132
EG20	0.164	0.164	0.876	0.876	0.158	0.158
EG11	0.488	0.488	0.982	0.982	0.332	0.332
EG06	0.846	0.846	0.981	0.981	0.463	0.463
EG13	0.674	0.674	0.752	0.752	0.473	0.473
EG12	0.936	0.936	0.676	0.676	0.581	0.419
EG24	0.770	0.770	0.456	0.456	0.628	0.372
EG18	0.856	0.856	0.494	0.494	0.634	0.366
EG10	0.747	0.747	0.183	0.183	0.803	0.197
EG03	1.275	1.275	-0.104	0.104	0.924	0.076
EG09	2.064	2.064	0.132	0.132	0.940	0.060

Individual data for L2 learners in Experiment 2 for the post-test

			Control group			
ParticipantID	Spec	|Spec|	Dur	|Dur|	Weighting Score	Transformed Score
CG18	0.003	0.003	1.122	1.122	0.003	0.003
CG01	-0.077	0.077	2.850	2.850	0.026	0.026
CG24	0.069	0.069	2.506	2.506	0.027	0.027
CG02	-0.071	0.071	1.779	1.779	0.038	0.038
CG21	0.069	0.069	1.485	1.485	0.044	0.044
CG11	0.143	0.143	2.938	2.938	0.046	0.046
CG20	-0.158	0.158	2.286	2.286	0.065	0.065
CG10	0.208	0.208	2.164	2.164	0.088	0.088
CG22	0.125	0.125	1.271	1.271	0.090	0.090
CG12	0.139	0.139	0.822	0.822	0.145	0.145
CG09	0.347	0.347	1.408	1.408	0.198	0.198
CG23	0.919	0.919	1.125	1.125	0.450	0.450
CG07	0.974	0.974	1.122	1.122	0.465	0.465
CG13	0.464	0.464	0.461	0.461	0.502	0.498
CG05	0.976	0.976	-0.566	0.566	0.633	0.367
CG15	2.428	2.428	0.265	0.265	0.902	0.098
CG17	2.461	2.461	-0.102	0.102	0.960	0.040
CG08	3.183	3.183	0.090	0.090	0.972	0.028
CG19	3.570	3.570	-0.082	0.082	0.978	0.022

CG06	2.682	2.682	-0.030	0.030	0.989	0.011
			Experimental group			
ParticipantID	Spec	|Spec|	Dur	|Dur|	Weighting Score	Transformed Score
EG14	0.022	0.022	1.990	1.990	0.011	0.011
EG01	-0.037	0.037	1.879	1.879	0.020	0.020
EG16	0.068	0.068	1.481	1.481	0.044	0.044
EG23	0.046	0.046	0.900	0.900	0.048	0.048
EG17	-0.051	0.051	0.838	0.838	0.058	0.058
EG02	-0.164	0.164	2.343	2.343	0.065	0.065
EG20	0.138	0.138	1.475	1.475	0.086	0.086
EG07	0.224	0.224	1.905	1.905	0.105	0.105
EG05	0.417	0.417	1.642	1.642	0.203	0.203
EG21	0.295	0.295	0.959	0.959	0.235	0.235
EG04	0.575	0.575	1.095	1.095	0.344	0.344
EG15	0.569	0.569	1.035	1.035	0.355	0.355
EG25	1.022	1.022	1.631	1.631	0.385	0.385
EG22	0.753	0.753	1.069	1.069	0.413	0.413
EG08	0.727	0.727	0.746	0.746	0.494	0.494
EG18	0.876	0.876	0.313	0.313	0.737	0.263
EG06	1.537	1.537	0.470	0.470	0.766	0.234
EG12	2.702	2.702	-0.361	0.361	0.882	0.118
EG24	1.812	1.812	-0.176	0.176	0.912	0.088
EG09	1.777	1.777	0.167	0.167	0.914	0.086
EG03	2.629	2.629	0.168	0.168	0.940	0.060
EG11	2.486	2.486	-0.155	0.155	0.941	0.059
EG13	2.137	2.137	0.084	0.084	0.962	0.038
EG10	1.695	1.695	0.025	0.025	0.985	0.015