Perception and Categorisation of Vowels in German and English by Chinese University Learners

Yihan Li

MSc in Applied Linguistics and Second Language Acquisition, 2024

DECLARATION BY THE CANDIDATE AS AUTHOR OF THE DISSERTATION

DEPARTMENT OF EDUCATION

UNIVERSITY OF OXFORD

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]:	李艺涵 LI Yihan.
Date:	12 August 2024

Table of Contents

Abstract

Accurate vowel perception is essential for effective communication and is a fundamental aspect of language learning. Vowels are important in any language because they carry important phonetic and phonological information that is essential for distinguishing between words and meanings. This research highlights the pivotal role of acoustic cues - particularly duration and formant cues - in the perception of vowels. It also explores how different factors, such as the learning environment and prior phonetic training, affect vowel perception. Understanding how learners perceive vowels and identifying the factors that influence them is crucial for improving communication skills and refining teaching methods. This dissertation examines the perceptual patterns of Chinese learners of German and/or English, focusing on their responses to English and German vowel categorisation and discrimination in three perceptual tasks. The study takes into account various background factors, including multilingualism, language proficiency, learning environment, and prior knowledge of phonetics and phonology, to explore possible correlations between learners' language learning backgrounds and their vowel perception abilities. The results of these tasks highlight the importance of both durational and formant cues in vowel perception. However, they also reveal a notable preference for durational cues among learners. The results also show that phonetic training and immersive language environments significantly increase learners' sensitivity to subtle variations in vowel duration and formant properties. Based on these insights, the dissertation argues for tailoring language teaching to learners' backgrounds. This approach involves exploiting strengths such as the processing of durational cues and incorporating explicit phonetic training and immersive learning experiences into language curricula. Such adaptations are proposed to optimise language learning and teaching by facilitating the more effective acquisition of the phonetic skills that are crucial for multilingual communication.

Acknowledgement

As I reflect on the days I spent living in the serene embrace of St Hugh's College and studying alongside so many brilliant souls in ALSLA, I find no trace of regret in my decision to come to Oxford. Everything I have encountered here has been nothing short of beautiful. This year has been a turning point in my life, a time when my research idea blossomed from a mere seed into its full, mature form - much like myself, growing more independent and brave with each passing day.

There are many people to whom I owe my deepest gratitude. First and foremost, I would like to thank my family for their unwavering support, both emotional and material. They have always been by my side. I am also deeply grateful to the friends - both old and new - I have made along the way, especially those in the Department of Education. I feel incredibly lucky to have had them by my side throughout this challenging year of study.

Special thanks go to my supervisor, Dr Faidra Faitaki. To have her as my supervisor is the best thing that happened to me at Oxford. Faidra is an exceptionally thoughtful and dedicated supervisor. She gave me a systematic understanding of academic writing and nurtured my fledgling research idea with great generosity and gentleness. I am forever grateful for her meaningful and challenging suggestions. I will always cherish the moments we spent exchanging ideas, sparking inspiration and witnessing the collision of thoughts that fuelled my academic journey.

List of Tables

List of Figures

1. Introduction

Speech perception is the complex process of transforming acoustic signals into meaningful linguistic units (Vouloumanos, 2015). This process plays a crucial role in language learning. Accurate perception of phonetic features is essential for effective communication and language comprehension. How learners' perceptual patterns affect their language learning has been studied extensively, particularly in the field of language teaching and second language acquisition. A key area of interest within this research, driven by the belief that improving perception can positively impact learners' production abilities, is the relationship between the perception and production of phonetic features in language learners. To date, this relationship remains not fully understood. Some studies have reported a positive connection between L2 production and perception (e.g., Cheng & Zheng, 2015; Flege et al., 1999). Conversely, other studies have found weak or no correlation (Kartushina & Frauenfelder, 2014; Levy & Law, 2010; Peperkamp & Bouchon, 2011).

Regardless of the perception-production relationship, it is undeniable that accurate perception is crucial for effective communication in everyday interactions. As highlighted by Zimmerer and Trouvain (2015), misperception of phonemes in words can lead to confusion between everyday vocabulary with very different meanings. For instance, in German, confusing the vowels in "Stahl" (steel) and "Stall" (stable) can drastically alter the meaning of a sentence, potentially leading to serious misunderstandings. Vowels, a critical component of speech, carry significant phonetic and phonological information that is essential for distinguishing between words and meanings in any language (Derwing et al., 2012). Therefore, understanding how vocalic phonemes are perceived by language learners merits further investigation due to their critical function in distinguishing linguistic meaning and can provide valuable insights for improving communication skills and classroom practice.

Central to this investigation is the complex interaction of acoustic and articulatory settings that define vowel systems. These include factors such as vowel length, tongue height, tongue position, lip rounding, and tenseness, all of which correspond to acoustic markers such as the first and second formant frequencies (F1 and F2). Understanding how these features contribute to vowel perception is crucial. Studies by scholars such as Chládková et al. (2015) and Yazawa et al. (2023) support a feature-based approach, in which listeners associate these acoustic signals with phonological categories such as vowel height and duration. Conversely, studies such as that of Llompart & Reinisch (2018) emphasise that acoustic cues, rather than phonological features, are the primary drivers of vowel perception, especially in languages with complex phonetic systems such as German.

This divergence in research perspectives highlights the need for a more nuanced understanding of how vowels are perceived, particularly through the lens of durational and spectral cues. Recent studies by Gao et al. (2019) and Yang et al. (2016) have highlighted the paramount importance of these cues in vowel discrimination. Durational cues, which refer to the temporal length of vowel articulation, are crucial for distinguishing between short and long vowels in different languages. Spectral cues, on the other hand, deal primarily with the frequency characteristics of vowels, in particular the values of F1 and F2, which are essential for identifying vowel quality.

However, the integration of these cues is complex and their impact on vowel perception can vary significantly depending on the linguistic and acoustic context. For example, in multilingual environments, second language (L2) learners may prioritise these cues differently depending on the phonetic characteristics of their native language. For example, research suggests that Spanish L2 learners may rely more on durational cues when discriminating Dutch vowels, whereas Dutch listeners rely more on spectral cues (Lipski et al., 2012). Similarly, studies have found that Chinese learners tend to focus more on durational rather than spectral properties in vowel identification tasks, demonstrating a perceptual strategy that differs from native English speakers (Yang et al., 2016).

Existing research has significantly advanced our understanding of how durational and spectral cues influence vowel perception in multilingual contexts. Nevertheless, there are notable gaps that require further investigation to advance our understanding in this area. First and foremost, current research focuses predominantly on L2 acquisition. The intricate dynamics of learning an additional language beyond the second are underexplored, particularly how different phonetic and phonological systems from previously acquired languages influence the perception and production of new phonetic elements in a third language (L3) setting. This oversight limits our understanding of the complexities and challenges faced by L3 learners. In addition, most of the conclusions drawn from existing studies are specific to particular language backgrounds, focusing primarily on Indo-European languages. This focus limits the applicability of findings across language families and fails to take into account the diverse phonetic landscapes of languages outside this group. There is an urgent need for research that includes a wider range of language families and learners with diverse linguistic backgrounds to ensure the generalisability and relevance of studies of phonetic perception.

More importantly, while much emphasis has been placed on the interference of a learner's previously acquired languages (L1, L2, L3, etc.), certain elements of their linguistic background, such as the environment in which they learned a language, can also play a significant role in shaping their perception and processing of these phonetic cues. However, these factors have often been overlooked in existing research frameworks.

Immersive learning environments, for example, provide a rich context for language acquisition, giving learners extensive exposure to the target language in real-life contexts. This exposure can significantly enhance the perceptual skills needed to identify subtle phonetic differences. Research by Zimmerer and Trouvain (2015) shows that exposure to immersive environments can improve the accuracy with which French speakers perceive German vowels. The benefits of such environments, as highlighted in research by Derwing et al. (2012), suggest that prolonged immersion can lead to more native-like pronunciation and better phoneme discrimination, which are crucial for effective communication.

Another approach to achieving these results is through explicit phonetic training which can provide learners with systematic and targeted instruction in the specific phonetic elements of the target language. This pedagogical approach is carefully designed to increase learners' awareness of phonetic and phonological distinctions that may be absent from their mother tongue repertoire. Research conducted by Bohn (1995) and Escudero & Boersma (2004) shows that such specialised training significantly improves the perceptual accuracy of L2 learners, equipping them with the ability to recognise and articulate phonetic sounds that are often difficult for speakers of their mother tongue.

Despite the proven benefits of certain learning backgrounds, current studies of vowel perception often fail to fully account for their effects. This oversight points to a significant gap in our understanding of how different learning backgrounds influence language learners' perceptual abilities.

The current study seeks to fill these gaps. It focuses on the perceptual abilities of Chinese learners of English and German and includes a wider range of linguistic backgrounds. It aims to thoroughly investigate whether the perceptual patterns identified in previous research are consistent with the findings in this under-investigated research context, and to examine how different aspects of linguistic background influence learners' ability to perceive phonetic elements.

The present study begins with an introductory chapter that provides the theoretical and contextual basis for the study. Chapter 2 follows with a comprehensive literature review that explores the use of acoustic cues by language learners and the factors that influence their perceptual patterns. This review pays particular attention to elements such as multilingualism, language proficiency, explicit phonetic training, and immersive learning environments. It also examines the vowel systems of Chinese, English, and German, establishing a specific linguistic framework for the subsequent analysis. The chapter concludes with an outline of the research questions that guide the study. Chapters 3 and 4 are devoted to the experimental part of the study, detailing the methods used and presenting the experimental results. Chapter 5 then summarises the results of the experiments in a general discussion and revisits the original research objectives, effectively linking the experimental results. Before concluding, Chapter 5 also considers the pedagogical implications and limitations of the research. This section provides valuable insights into the applicability of the findings to language teaching and openly acknowledges the limitations encountered during the study. The final chapter provides a summary of the main findings and discusses the contributions of the study to the fields of language learning and teaching. It also suggests possible directions for future research and possible improvements to the methodology and scope of the study.

2. Literature Review

This chapter presents a thorough review of the literature on vowel perception, highlighting several critical areas. It begins by examining durational and spectral cues, elucidating their nature and role in vowel differentiation. It then considers various factors influencing vowel perception, such as multilingualism, language proficiency, explicit training and immersive environments, by reviewing relevant studies, their findings and limitations. The chapter then compares the vowel systems of English, German and Chinese to provide a linguistic context for the current study. Finally, it identifies existing research gaps, outlines how the current study aims to address these gaps, and presents the research questions that will guide the investigation.

2.1 Durational and Spectral Cues in Vowel Perception

Phonological features describe vowel systems through combinations of acoustic or articulatory qualities that vowels share, such as vowel length, tongue height, tongue position, or tenseness. Vowel length, or duration, refers to the amount of time a vowel is articulated during pronunciation. In addition to duration, formants, which are the resonant frequencies of the vocal tract, are also key elements in shaping the acoustic characteristics of the vowel. Specifically, the first formant (F1) and the second formant (F2) are crucial to vowel perception. The F1 is strongly correlated with tongue height; a higher tongue position results in a lower F1 frequency, while a lower tongue position results in a higher F1 frequency. The F2, on the other hand, reflects both horizontal and vertical tongue positions; a frontal tongue position results in a higher F2 frequency, while a retracted tongue position results in a lower F2 frequency (Lee et al., 2016).

While some studies, such as Chládková et al. (2015) and Yazawa et al. (2023), support a feature-based approach in which listeners relate acoustic signals to phonological categories such as vowel height and duration, other studies emphasise the dominance of acoustic cues. For example, Llompart & Reinisch (2018) found that acoustic cues, rather than phonological features, drive perception, especially for complex features such as the tenseness of German vowels. In other words, when listening to vowels, people rely more on precise sound details (acoustic cues) than on general sound patterns (phonological features). To be more specific, rather than using broad and abstract categories such as 'high' or 'low' sounds, listeners pay close attention to the acoustic aspects of vowel sounds in order to process them.

Among these acoustic cues, recent research (e.g. Gao et al., 2019; Yang et al., 2016) on vowel perception has focused primarily on two cues: durational and spectral cues. Durational cues pertain to how long a vowel is pronounced. These cues are particularly salient for distinguishing between short and long vowels in different languages. In contrast, spectral cues deal with the acoustic frequency characteristics of the vowel, in particular F1 and F2. Research shows that these phonetic attributes have a significant impact on vowel perception. Listeners rely on these acoustic properties, considering both the duration (quantitative aspect) and the frequency characteristics (qualitative aspect), to categorise and discriminate vowels (Boersma & Chládková, 2011; Chládková et al., 2015; Gao et al., 2019).

Although both durational and spectral cues are essential, they may not contribute equally to vowel perception when multiple cues are involved in acoustic differences in a single contrast (Holt & Lotto, 2006; Nittrouer, 2004). Research in second language acquisition has extensively investigated this phenomenon in a number of languages, including English, Dutch, Spanish, etc. (e.g. Escudero & Boersma, 2004; Gao et al., 2019; Hillenbrand et al., 2000). It has been found that L2 learners and native speakers tend to attach different importance to the two cues. To be more specific, Lipski et al. (2012) found that Spanish L2 listeners rely heavily on duration to discriminate Dutch vowels, whereas native Dutch listeners primarily use spectral cues. Similarly, Yang et al. (2016) also found that Chinese participants relied more on vowel duration than spectral properties in vowel identification tasks in English, unlike native English listeners.

Researchers have attempted to unravel the mystery of L2 learners' preference for durational cues and, according to some studies, this phenomenon can be partly explained by the concept of perceptual salience (Kajouj & Kager, 2019). The perceptual salience of cues highlights the discriminability of the cues per se, which influences L2 perceptual behaviour regardless of the listener's prior language background. It has been suggested that the human auditory system processes some acoustic properties, such as the length of vowels, in a more robust manner, making them more discriminable (Holt & Lotto, 2006; Pisoni, 1977). Consequently, listeners tend to assign greater perceptual importance to these cues when discriminating between phonetic contrasts in L2.

However, concerns have been raised about cross-linguistic influence, particularly when the native languages of some language learners share similar durational and spectral properties with the target language. This influence may confound the dominant role of perceptual salience, as phonological similarities between the learners' native language and the target languages may lead to a preference for durational cues (Grenon et al., 2019). In such cases, it becomes difficult to distinguish whether learners' reliance on particular cues is due to their perceptual salience or simply a transfer from their native phonological system.

To address these concerns, evidence from research involving native Chinese L2 learners has been particularly illuminating. Chinese learners of English have been found to rely more on durational cues than spectral cues in a variety of perceptual tasks (Bohn, 1995; Wang & Munro, 1999). This reliance is noteworthy because Mandarin, the native language of these learners, does not use duration or spectrum as prominent cues for vowel differentiation. Mandarin vowels are not differentiated by duration, and the vowel space lacks significant spectral variety, particularly in the high frontal range. This scenario mitigates the possibility of cross-linguistic transfer influencing the preference for durational cues. Instead, it highlights the inherent accessibility and robustness of durational cues, making them easier for learners to acquire, even in the absence of similar features in their native language.

Although this evidence highlights the importance of duration cues, perceptual salience alone does not account for a comprehensive understanding of vowel perception. Many of the above studies (e.g., Bohn, 1995; Yang et al., 2016) have focused on a pair of minimal dimensional contrasts, limiting the applicability of their findings. In light of this, studies on tense-lax vowel perception, which include both quantitative and qualitative cues within vowel changes, show that vowel quality significantly enhances the perceptual contrast between short and long vowels (Tomaschek et al., 2011). To clarify, tense-lax vowels are pairs of vowels that differ in both duration and quality. For example, in Germanic languages, tense vowels are typically longer (quantitative) and more peripheral in the vowel space (qualitative), whereas lax vowels are shorter (quantitative) and more centralised (qualitative) (Lindau-Webb, 1987). Quality and quantity are not isolated aspects, but work together to optimise vowel perception.

However, the primacy of durational cues in vowel perception is not universally accepted. Research has shown that while duration is an important factor, it does not always take precedence over other acoustic features (Bennett, 1986; Reinisch & Sjerps, 2013). Bennett's (1968) study, for instance, used a broad range of vowel pairs and focused on native speakers of English and German. In this study, perceptual experiments were conducted on vowel pairs from both German and English, as well as a pair of backward, unrounded vowels ([ɯː] and [ɤ]) not found in either language. The aim of the study was to assess the relative importance of durational and spectral cues. Synthetic speech was used. Contrary to the finding that durational cues are more salient, the results showed that spectral form is generally more important than duration for vowel recognition in the pairs tested. The importance of duration cues in discriminating between vowels only becomes more pronounced when the vowels have very similar qualities, i.e. pitch, loudness and other specific articulatory features such as tongue position and mouth shape. Recent research resonates with this idea, supporting that the uptake of spectral cues slightly precedes durational cues in vowel perception, and that both are rapidly influenced by context (Reinisch & Sjerps, 2013). The relative importance of durational and spectral cues may vary depending on the specific phonetic context and language-specific phonological structures, indicating a more nuanced understanding of vowel perception.

Another important observation from Bennett's (1968) research is the contrasting perceptual strategies that English and German speakers use when encountering unfamiliar vowel pairs. Specifically, German speakers predominantly used spectral cues, whereas English speakers tended to draw on durational cues. Despite these differences, both groups exhibited similar perceptual patterns when recognising German and English vowels. The researcher suggests that the mixed results could be due to several possibilities, including inherent nuances in the German and English phonological systems, and individual differences such as auditory sensitivity and attentional control. Although the study did not investigate the causes of variability between different groups of language learners, it points to the importance of understanding such different perceptual patterns across linguistic contexts and between learners from various linguistic backgrounds. This finding therefore encourages further investigation into factors beyond perceptual salience and cross-linguistic influence. For example, it is of particular interest to explore how individual learner variables, such as previous exposure to multiple languages, interact with phonetic and phonological training to shape vowel perception. In addition, investigating the impact of immersive language environments on the perceptual strategies of learners from different linguistic backgrounds could provide deeper insights into the adaptive mechanisms of vowel perception in multilingual contexts.

2.2 Factors Influencing Vowel Perception

2.2.1 Multilingualism and Language Proficiency

Research from a broad perspective on language learning has investigated how multilingualism affects perceptual skills. Empirical studies have shown that a larger phonological repertoire can improve perceptual accuracy. A study by Rabinovitch and Parver (1966) examined how well monolingual and bilingual young adults could discriminate phoneme sequences in a foreign language that was unfamiliar to them. The results showed that bilinguals had an enhanced ability to discriminate phonemes in foreign languages, supporting the notion that bilingualism positively impacts the learning of additional languages. Consistent with these findings, later studies (e.g., Cohen et al., 1967; Enomoto, 1994) further confirmed that multilingual and bilingual individuals have perceptual advantages over monolinguals by comparing their perceptual patterns. However, it is important to recognise that although bilinguals often outperform monolinguals in the experiments, not all phonetic studies report statistically significant differences between monolinguals and bilinguals (Gallardo Del Puerto, 2007). This suggests that the perceived advantages of bilingualism in phonetic perception may not be as unequivocal as some studies suggest. Furthermore, acquiring additional languages, particularly beyond the second, may involve a complex interplay of a variety of factors and requires more in-depth research.

In order to enrich the understanding of how multilingualism affects phonetic perception, attempts have been made to incorporate proficiency levels in the analyses. Some research (e.g., Cenoz & Valencia, 1994; Onishi, 2016; Sanz, 2000) suggests that proficiency in previously acquired languages significantly enhances overall competence in a target language. For example, Onishi (2016) conducted a study with native Korean speakers who had English as their L2 and Japanese as their L3. Participants were asked to perform perceptual tasks involving both Japanese and English contrasts. Correlations between participants' performance in their L2 and L3 and phonological perception in the L3 are analysed. The results showed that, in general, the more proficient participants were in their L2, the better they performed in their L3, albeit on a limited range of tasks. More specifically, the positive correlation is observed for a few specific contrasts and may not generalise across all phonological tasks or linguistic contexts. This selective task limitation suggests that further research is needed to determine whether the findings hold more broadly, or whether they are restricted to particular phonological features and task types.

Similarly, Ding and Jokisch (2012) found that language proficiency did not consistently show a positive correlation with perceptual performance. In their study, they presented 12 Chinese learners with German words containing vowel pairs and asked them to identify whether the vowels were tense or lax in order to assess their ability to discern acoustic differences. They discovered that the learners' proficiency level did not significantly affect their ability to distinguish between these tense and lax vowels. They concluded from the results that having a larger vocabulary did not significantly improve the perceptual ability to discriminate between nuanced acoustic features of these vowels. Based on the demographic and background information of the participants, Ding and Jokisch (2012) found that other factors might have influenced the perceptual process, creating a complex scenario. In particular, they found that students with more years of learning experience did not show improved perceptual skills in the absence of specific phonetic training. Notably, their study had only 12 participants, which means that their analyses may have been underpowered and significant effects may have been missed due to the small sample size. Nevertheless, their findings suggest valuable directions for future research, suggesting that mere exposure to a language and general language learning over time are not sufficient to improve specific phonetic perception skills. Instead, explicit phonetic instruction can serve as a crucial complement to general language learning and play an essential role in the effective development of these phonetic perception abilities.

2.2.2 Explicit training

Over the past three decades, explicit instruction in target language phonetics and phonology has increasingly become a focus of academic research (Bryan, 2019). Several studies provide compelling evidence of how targeted training can improve phonetic discrimination and change perceptual patterns. For example, Kingston (2003) investigated how native English speakers perceive German vowels and demonstrated that perceptual training can effectively address the challenges English L1 listeners face in acquiring German vowel sounds. Empirical evidence from Dlaska and Krekeler (2008) also supports the importance of explicit phonetic training. In their study, 45 advanced L2 students were trained on German phonemes and were asked to record and compare their pronunciations with those of native speakers. Initially, without instructor feedback, the participants accurately identified less than half of their mispronunciations. However, after receiving direct feedback, their accuracy increased dramatically to 89%.

Miller (2012) interprets these findings by suggesting that learners can only integrate unfamiliar phonetic features into their phonological system if they become aware of them. This awareness is typically achieved through explicit instruction and focused training, highlighting the critical need for deliberate and structured phonetic education. More specifically, it has been found that students without phonetic training (i.e., non-linguists) assume that long and short vowels differ only in duration, unaware of the qualitative differences due to different acoustic properties (Ding & Jokisch, 2012).

It is noteworthy that even short-term phonetic training can lead to substantial perceptual improvements, achieving a qualitative leap from initial inability to more accurate phonetic discrimination. In Wang and Munro's (1999) study, they used the pretest-training-posttest paradigm to investigate how L1-Mandarin English learners and native English speakers perceived the English vowel contrasts. The training sessions were brief but intensive, and continued until the participants achieved 95% correct identification in a block of 24 tokens, with each block consisting of the extreme tokens of the continuum presented four times in random order. Despite the brevity of the training, which involved the identification of stimuli varying along the spectral dimension and immediate feedback, participants' perceptual performance improved significantly, shifting from a duration-based to a spectrum-based perceptual focus, in line with the patterns of native English speakers.

Given the promising results from various studies, some researchers advocate explicit phonetic training, claiming its effectiveness in improving speech accuracy and listener comprehension (Derwing et al., 2012; Olson, 2014). However, it is important to approach these claims with caution. As discussed above, the relationship between perception and production is complex and not yet fully understood. Therefore, the validity of the claim that explicit training unambiguously leads to more accurate speech should be further tested. Recent studies that have examined the relationship between perception and production from the perspective of training studies have produced mixed results. For example, Carlet and Kivisto de Souza (2018) found that their instructional treatment improved learners' perception but did not significantly affect their production. Conversely, Sakai and Moorman (2018) demonstrated that perception training can indeed induce positive changes in production. Given these mixed results, researchers and educators should be cautious in interpreting the effects of such training, recognising that improvements in perceptual abilities do not always translate directly into improved production.

2.2.3 Immersive Environment

In contrast to explicit training, immersive learning environments offer a more implicit opportunity that can significantly influence learners' phonetic perception. Longer exposure to a second language, particularly through study abroad experiences, can lead to improved perception of non-native vowels (Flege et al., 1997; Schoonmaker-Gates, 2024). Research with native Japanese speakers of English demonstrates that adults living in an immersive language environment can learn to accurately perceive English consonants that are absent in their native language, such as /r/ and /l/ (Yamada et al., 1994). Their perception of these sounds improves with English language experience in an immersive language environment. Similar findings apply to research on English vowels; non-native speakers with more extensive English language experience perceive English vowels more accurately than their less experienced counterparts (Flege et al., 1997). A commonly accepted interpretation of the positive correlation is that having ample opportunities to engage with the language, through frequent use and immersive exposure, enhances L2 learners' perception of segmental contrasts in a foreign language (Park, 2018).

A major limitation of research into the effects of immersive environments is the predominant focus on the English language and English-speaking environments. According to DeKeyser (2007), language learning is strongly influenced by the specific linguistic features and cultural contexts of the target language. The linguistic and cultural specificity of English-speaking environments may not capture the diverse challenges faced by learners of other languages. In order to increase the generalisability and applicability of research on immersive environments, it is essential to broaden the scope of research and provide more nuanced insights into how immersive learning environments can be optimised for different linguistic backgrounds.

2.3 Vowels in English, German and Chinese

Building on the understanding that perception varies across linguistic contexts, this section compares the relevant vowels of the languages involved in the study, highlighting their phonetic and phonological characteristics.

In both German and English, vowel duration and vowel quality, specifically differences in height, frontness and rounding, play an important role in distinguishing different vowels. The vowel systems of both languages contain several vowel pairs, sets of vowels that are related or compared on the basis of certain linguistic features. For example, long and short vowel pairs differ in length, which can change the meaning of a word. In German, pairs such as /a:/ in 'Haar' and /a/ in 'Hut' illustrate this feature, with /a:/ and /a/ showing a purely durational contrast (Tomaschek et al., 2011). In addition, in some pairs, vowel quantity interacts with aspects of vowel quality. For example, the high vowel contrasts /iː/ and /ɪ/ in both languages involve a distinction between long and tense vowels and short and lax vowels, known as tense and lax pairs. The ability to differentiate vowel contrasts is crucial for L2 learners, as these contrasts play a significant role in distinguishing a wide range of high-frequency word pairs (Souza et al., 2017). Mastering these distinctions is essential for learners to distinguish between similar-sounding words and thus achieve native-like comprehension in a foreign language.

The use and importance of durational and spectral cues in vowel perception differ significantly between English and German. In English, vowel discrimination often relies heavily on spectral rather than durational cues. For example, the English vowels /i:/ and /ɪ/ are distinguished primarily by their spectral properties, specifically F1 and F2, rather than by length (Cohen, 1971). Although English distinguishes between lax and tense vowels, which may influence vowel length, this feature is not as contrastive as spectral differences (Crystal & House, 1988). In contrast to English, the German vowel system places more emphasis on both spectral and durational cues. Vowel differentiation, especially with regard to the tense-lax contrast, is characterised by a greater difference in duration than in English. For example, German lax vowels are about half as long as their tense counterparts, whereas the tense-lax ratio for English is about 0.70 (Antoniadis & Strube, 1984; House, 1961). This leads to a greater reliance on duration cues in German.

Compared to English and German, the Chinese vowel system shows significant differences, particularly in the absence of tense-lax and long-short distinctions. The Chinese vowel inventory consists of tense-only vowels, lacking the tense-lax differentiation present in both English and German. Chinese speakers have developed an internalised template of tense-only vowels that serves as a perceptual filter, meaning that they interpret vowel sounds based on the vowel system of their native language. This filter potentially leads to difficulties in distinguishing between tense and lax contrasts (Brown, 1998; Ding & Jokisch, 2012). This discrepancy highlights the additional challenges Chinese learners face in accurately perceiving and producing vowel sounds in languages that rely heavily on these distinctions.

Interestingly, the acquisition of an L2 with tense-lax vowel contrasts may mitigate the perceptual challenges Chinese learners face when they later acquire additional languages with similar contrasts. As outlined by Ivaska and Siitonen (2017), constructional similarity can trigger CLI on a construction-by-construction basis, regardless of general similarity or genealogical categorisation. Constructional proximity can sometimes override typological similarity in L3 acquisition, with CLI occurring cumulatively from both previously acquired languages (Kolb et al., 2022). In other words, the process of acquiring subsequent languages draws on references from the learner's existing linguistic knowledge, emphasising that both the L1 and L2 influence the learning of an L3. Building on this framework, Wu (2022) studied Chinese-native learners who speak English as their L2 and are learning German as their L3. She found that exposure to tense-lax distinctions in the L2 helps to eliminate the tense-only filter and the perceptual obstacles presented by the learners' native Mandarin. This results in a compromise effect, where the influence of both L1 and L2 is balanced. The learners' perception strategies are adjusted by integrating features from both their L1 and L2, which facilitates more accurate vowel perception in the L3 and helps them navigate the complex vowel systems of additional languages by leveraging the strengths of previously acquired linguistic frameworks.

2.4 Research Gap and Current Study

To sum up, although there has been extensive research on the role of durational and spectral cues in vowel perception in different languages, there remain some significant gaps.

The current body of research predominantly focuses on L2 acquisition, with only a limited number of studies addressing L3 acquisition. This imbalance is evident in the extensive literature on factors influencing perception in L2 contexts (e.g., Ding & Jokisch, 2012; Kingston, 2003; Flege et al., 1996), while the complexities of L3 learning remain less explored. Furthermore, the diversity of languages included generally tends to be skewed towards Indo-European languages, with very limited research on other language families (e.g., Wang & Munro, 1999; Wu, 2022). This highlights the need for more comprehensive investigations of L3 acquisition and the inclusion of a wider range of languages to better understand phonetic perception across linguistic backgrounds.

For instance, the role of durational versus spectral cues in vowel perception is debated. Some studies prioritise spectral cues (e.g., Bennett, 1968), while others emphasise durational cues, especially in languages such as German where vowel length is crucial (e.g., Ding & Jokisch, 2012). Further research is needed to clarify these priorities under different conditions. As another example, the relationship between language proficiency and perceptual performance shows mixed results. While some studies find a positive relationship (e.g., Cenoz & Valencia, 1994; Onishi, 2016; Sanz, 2000), others do not (e.g., Ding & Jokisch, 2012). Onishi (2016) found that higher L2 proficiency correlated with better L3 perception for some contrasts but not for others, suggesting variability in outcomes even under similar experimental conditions and within the same task. These mixed results underscore the need for more robust and comprehensive evidence to support the literature.

In addition, some studies with less significant or conflicting results acknowledge that unconsidered factors, such as individual differences in learning environments, may influence L2 and L3 vowel perception (e.g., Ding & Jokisch, 2012; Onishi, 2016). There is a lack of comprehensive research that considers multiple influencing factors. In particular, there is an urgent need for studies that provide a more comprehensive understanding of phonetic perception by addressing the complex interplay of different elements, which is the aim of the present study.

In this context, the following research questions arise:

RQ1: How do Chinese learners of English and German use quantitative (durational) and qualitative (spectral) cues to perceive vowels in foreign languages?
RQ2: How do Chinese learners of English and German prioritise quantitative and qualitative cues in vowel perception?
RQ3: Does language learning background, such as immersive learning, language proficiency, prior phonetic and phonological training, etc., influence learners' perceptual patterns?

Based on the literature presented, several predictions can be made regarding the research questions. For RQ1, it is anticipated that Chinese learners will be able to perceive vowels and discriminate vowel differences based on acoustic features. Despite the lack of tense-lax and long-short distinctions in their native language, they are expected to develop native-like perceptual patterns through their experience of learning the relevant structures in foreign languages. However, this process may be influenced by individual background. For RQ2, it is predicted that Chinese learners will generally prioritise quantitative cues (duration). Students with immersive learning backgrounds, higher proficiency, and specific phonetic training are expected to rely more on qualitative cues (spectral properties) compared to their counterparts. Finally, for RQ3, it is expected that learners' perceptual patterns will be significantly influenced by their language learning background. Those with immersive learning experiences, higher language proficiency, and prior phonetic and phonological training are predicted to show more accurate and nuanced vowel perception.

3. Methodology

The methodology chapter outlines the experimental design aimed at investigating language learners' perceptual patterns in response to durational and formant cues. It includes detailed descriptions of the stimulus design and materials, including a demographic questionnaire, a language proficiency pretest in English and German, and three specific perceptual tasks designed to target different vowel contrasts. The recruitment of participants and the experimental procedure are also explained.

The experiment aims to investigate language learners' perceptual patterns in response to durational and formant cues. The study includes a demographic questionnaire, a language proficiency test in English and German, and three perceptual tasks, each targeting specific vowel contrasts.

The first task examines how learners discriminate between short and long vowels, focusing on the /a:/-/a/ contrast. The second task assesses learners' sensitivity to both durational and formant variations in tense-lax vowel contrast, focusing on the /i:/-/ɪ/ distinction. The third task investigates learners' perception of formant differences alone, analysing their responses to the /ɛ/-/ø/ contrast in German and the /ɛ/-/æ/ contrast in English.

Learners' performance on the three tasks will be analysed in conjunction with their language background, which will be explored through a demographic questionnaire and a language proficiency test. This approach aims to provide a more comprehensive understanding of how language experience influences the perceptual processing of vowel contrasts.

3.1 Participants

Two groups of Chinese learners of English and German were recruited for the study. The first group consisted of Chinese university students who had learned English as their L2 and German as their L3. The second group consisted of Chinese university students who had learned English as their L2 but had no prior knowledge of German. The recruitment criteria were as follows: (1) participants should not have studied any language other than Chinese, English and German; (2) participants should not be bilingual in two of the languages studied; and (3) participants should not have any hearing impairment.

To ensure that the study was adequately powered to detect meaningful effects, a power analysis was conducted using G*Power (Faul et al., 2007). The aim was to determine the minimum sample size required for a multiple linear regression analysis that included five predictors: pre-test scores and learners' language learning background (study abroad experience, multilingualism and prior knowledge of phonetics and phonology). These predictors were identified on the basis of the initial literature review and research hypotheses. The parameters were set as follows: an effect size f^2 of 0.15, an alpha level of 0.05 and a desired power of 0.80. The alpha level of 0.05 and the desired power of 0.80 are widely accepted standards in statistical analysis. The effect size f^2 of 0.15 was chosen based on Cohen's (1988) guidelines for a medium effect size. The power analysis indicated that a sample size of 92 participants was required.

Given the requirements of the analysis, we initially recruited a total of 130 participants to account for potential drop-outs and to ensure sufficient data for robust statistical analyses. However, after checking their activity against the background data provided by Gorilla, it was observed that some participants made random choices without clicking the 'play' button to listen to the recordings. These participants were excluded from the analysis. In addition, those who failed to respond correctly to at least half of the attention checks were excluded due to potential distraction during the experiment. In the end, responses from 110 valid participants were included in the final analysis. Of these participants, 75 were learners of both German and English, while 35 were learners of English only.

3.2 Materials

The materials used in this study consisted of several key components: a demographic questionnaire, a language pre-test and three perceptual tasks. Each component was carefully designed to collect comprehensive data on the participants' linguistic background and perceptual abilities. All materials were accessed and administered via the experiment builder platform 'Gorilla' (https://gorilla.sc/) (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2019).

3.2.1 Questionnaire

The demographic questionnaire administered in this study was designed to collect detailed information about the participants' backgrounds and language learning experiences. It included sections on age, academic year, first language, monolingual status, and prior training in phonetics or phonology. In addition, participants provided information about their English language learning history, including the age at which they started learning English and any experience of living or studying in English-speaking countries. For participants who were German learners, an additional section collected data on their German learning background, including the age at which they started learning German, standardised German test scores, and experience of living in German-speaking countries (see Appendix 4 and Appendix 5 for further details).

3.2.2 Language Pre-test

For the language pre-test, the LexTale tests for English and German (Lemhöfer & Boersma, 2012) were integrated into the Gorilla platform and administered to assess the participants' level of proficiency in these languages. The LexTale is a lexical decision task designed to measure vocabulary knowledge and, by extension, general language proficiency. In the test, participants are presented with a series of letter strings and must decide whether each string is a real word or a non-word in the target language. The LexTale score is the percentage of correct responses, corrected for the unequal proportion of words and non-words in the test by averaging the percentages correct for these two item types. It is calculated as follows:

*((number of words correct/40*100) + (number of nonwords correct/20*100)) / 2*

3.2.3 Perceptual tasks

The three perceptual tasks consisted mainly of the test items and the attention checks. The attention checks were included to ensure that participants were listening attentively to the recordings, thereby validating their responses for subsequent analysis. The test items and the attention checks were presented in a randomised order.

The stimuli for the perceptual tasks consisted of four sets of synthesised trochaic words in which the first vowel varied in duration, formant values, or both. The trochaic foot was chosen because it is a common metrical structure in German (Féry, 1996) and the study primarily accommodates the learners' L3, German, while also considering their English and Chinese language backgrounds. For the /a:/-/a/ contrast, the vowels were embedded in the 'words' /'ta:tən/ - /'tatən/ and /'fa:dən/ - /'fadən/. For the /i:/-/ɪ/ contrast, /'fiːdən/ - /'fɪdən/ was used. For the /ɛ/-/ø/ and /ɛ/-/æ/ contrasts, /ˈbɛtən/ - /ˈbøtən/ - /ˈbætən/ was selected. In the perceptual tasks, words were presented in phonetic transcription to avoid a 'word frequency effect' (Dufour et al., 2013), which suggests that high-frequency real words are recognised more quickly and easily (Fox, 1984). Although most of the words are pseudowords, many German verbs and plural nouns end in /-ən/. Simple trochaic words such as /taten/ correspond to the real German word Taten, the plural of Tat (act). This approach also minimises the chance that listeners, especially German learners, will associate the synthesised word pairs with real words (e.g. /'ta:tən/ - /'tatən/ with Taten) and develop a preference for a particular vowel category, such as a preference for the long vowel /a:/ in Taten.

The original recordings used to create the stimuli were synthesised from Anton Vasetenkov's project, IPA Reader (Vasetenkov, n.d.). This programme was chosen over others because of its high quality and accurate synthesis of phonetic transcriptions, which is essential to ensure the consistency of the stimuli. Crucially, the words were synthesised to maintain precise control over the acoustic properties of the stimuli, which is necessary to systematically manipulate and test specific phonetic contrasts without the variability inherent in human speech recordings. The stimuli were manipulated using PRAAT (Boersma & Weenink, 2024), a software tool widely used in phonetics research for the analysis, synthesis and manipulation of speech. It precisely controls and adjusts various acoustic parameters, thus ensuring the reliability and accuracy of the experimental stimuli.

The first task consisted of 16 questions and was designed to examine participants' ability to judge vowel length. Of the 16 questions, 14 were target stimuli with different vowel durations. Two additional questions with extremely long vowels are included as attention checks. In this task, the /a:/ - /a/ contrasts in the stimuli were manipulated only in terms of duration. First, the vowel /a:/ was cut from the synthesised sounds /'ta:tən/ and /'fa:dən/. Using the 'To Manipulation' function in PRAAT, the duration of the vowel segment was modified. By applying different multiplication factors in the 'Multiply Duration' function, the duration of the vowel segment was modified to 50ms, 75ms, 100ms, 125ms, 150ms, 175ms and 200ms, providing a range of vowel lengths that span the perceptual boundary between short and long vowels and accommodate both English and German vowels (Crystal & House, 1988; Tomaschek et al., 2011). The formant values remained constant across all stimuli. The seven vowels with duration differences were then resynthesised into the original /'ta:tən/ and /'fa:dən/ contexts, replacing the original vowel segments.

The second task was designed to assess participants' ability to perceive differences in vowel duration and formant values. This task consisted of 26 items, each of which presented pairs of sounds for comparison. Participants were asked to determine whether the vowels in each pair were different. The vowel pairs were divided into four types: 6 pairs differed only in terms of vowel duration; 6 pairs differed only in their formant values; 8 pairs differed in both duration and formant values; and 6 pairs served as attention checks, with a significant duration difference of 150ms to ensure that participants were listening attentively. To create the stimuli, both the duration and the formant values of the /i:/-/ɪ/ contrast were manipulated. The formant values were adjusted to three levels, while the duration was varied in five levels. The vowel continuum included the original formant values of /iː/ and /ɪ/ as well as an intermediate value, thus creating two units of formant difference along the continuum from /iː/ to /ɪ/. To manipulate the formant values of the intermediate variants, the vowels /iː/ and /ɪ/ were cut from the original synthesised sounds /'fiːdən/ and /'fɪdən/ produced by the IPA reader. Before editing the formant values, the duration of both vowels was manipulated to the same length. The formants were then edited using the Manipulation Editor in PRAAT. A detailed description of the manipulation steps and the relevant data can be found in Appendix 6. After editing the formant values, the duration of the three vowels was manipulated using the same method applied in Task 1. The durations were set to 100ms, 150ms, 200ms, 250ms, and 300ms. In total, the manipulation yielded 15 vowel variants, resulting from the combination of three formant steps and five duration variations. These variants were then resynthesised into the original sounds, replacing the source vowels.

The third task was designed to assess participants' ability to perceive differences in vowel quality along the vowel continua. This task consisted of 18 items, each of which presented pairs of sounds for comparison in a format similar to Task 2. Participants were asked to determine whether the vowels in each pair were different. Of the 18 questions, 6 pairs involved vowels from the German vowel continuum /ɛ/-/ø/, with differences ranging from 1 to 3 units of formant change; 6 pairs involved vowels from the English vowel continuum /ɛ/-/æ/, with differences also ranging from 1 to 3 formant units. The remaining 6 pairs served as attention checks, pairing a word with a German vowel with a word with an English vowel to demonstrate obvious differences and ensure that participants were listening attentively. The duration of all vowels in task 3 is set to 150 ms. The formant values of these stimuli were manipulated in the same way as described in Task 2 (see Appendix 7).

3.3 Procedure

The study has received ethical approval from the Central University Research Ethics Committee, University of Oxford, with reference code [C1B-24HT-Educ-020]. The experiment was conducted using the Gorilla platform, with eligible participants accessing it via a provided link. All participants were instructed to complete the experiment in a quiet environment using headphones. First, they completed an informed consent form (see Appendix 2). They then answered a branching question to determine their language learning background, which divided them into two groups: German learners and English learners. English learners completed a demographic questionnaire focusing on their English learning experience, followed by the standard LexTale pre-test in English. German learners completed the same demographic questionnaire, along with additional questions about their German learning experience and pre-tests in both English and German. After the pre-tests, all participants proceeded through the same sequence of three perceptual tasks. In these tasks, participants listened to recordings and answered specific questions presented on the screen, which included the transcription of the word and two response options. Before their responses were considered valid, sample questions demonstrated how to complete the tasks. A fixation screen was displayed between each question to prevent misclicks. At the end of the experiment, participants had the opportunity to provide their email address for a chance to enter a prize draw.

4. Results

This chapter presents a comprehensive analysis of the experimental data collected to investigate language learners' perceptual patterns in response to durational and formant cues. It includes visualisations of the data and inferential statistical analyses. Specifically, the chapter covers

1. Task 1: Long-short vowel perception, including chi-squared test results and insights into the factors influencing vowel length perception.
2. Task 2: Tense-lax vowel perception, focusing on durational and formant perceptual differences, and the relationships between background variables and perceptual ability.
3. Task 3: Formant perception, detailing participants' ability to perceive formant differences in English and German vowel pairs, supported by statistical analyses to validate the observations.

According to the information provided in the questionnaire, 53 participants reported having some prior knowledge of phonetics and phonology, including familiarity with the phonetic alphabet and professional training in these areas. The remaining 57 participants indicated that they had no prior knowledge of phonetics and phonology. Additionally, 35 participants had studied abroad in English-speaking or German-speaking countries, or both. Considering that almost all participants (107 out of 110) started learning English in primary school and the majority of German learners, except for 10 out of 75, had started learning German at university, the 'age of onset' variable shows minimal variability. This lack of variability, especially for English, means that the 'age of onset' variable would not contribute significantly to distinguishing between participants in the analysis. Therefore, "age of onset" was excluded from the independent variables.

Based on the questionnaire responses, three categorical variables were identified concerning the participants' language background: "branch", indicating whether they were English-only learners or learners of both German and English; "abroad", indicating whether they had experience of studying abroad in English-speaking or German-speaking countries, or both; and "prior knowledge", indicating whether they had previously learned anything specifically about the sounds being studied in this research. These variables, along with the language proficiency test scores in English and German, are considered independent variables in the following analysis to determine whether these factors influence the participants' task performance.

The pre-test scores for English and German were collected to assess the participants' proficiency levels. For English-only learners, the English test scores ranged from 40.00 to 76.25, with a mean score of 59.84 (SD = 11.68). For English and German learners, the English test scores ranged from 38.75 to 73.75, with a mean score of 63.2 (SD = 9.58), while the German test scores ranged from 25 to 85, with a mean score of 56.62 (SD = 11.33). In general, the English scores represented a higher level of proficiency. The statistical significance of the differences is confirmed by a paired samples t-test. For the English and German learners, the t-test results were $t(74) = 5.627$, $p < .001$, with a correlation coefficient of 0.612, indicating a strong positive relationship between the scores. The effect size was Cohen's d = 0.650 (95% CI: 0.399 to 0.897), indicating a medium effect size.

4.1 Task 1: Long-short vowel perception

Task 1 was to determine whether participants perceived vowels as short or long. The results for the two sets of stimuli are visualised in Figures 1 and 2, which show bar graphs indicating the number of participants who perceived each vowel as either long or short. These visualisations show clear trends in participants' perception of vowel duration.

In Set 1, which captures the perception of the vowel contrast /aː/-/a/ in /'taːtən/ - /'tatən/, participants' responses to seven different vowel durations (75 ms to 225 ms) show clear patterns (see Figure 1). For the first four durations (75 ms, 100 ms, 125 ms, and 150 ms), the vast majority of participants identified the vowels as short. This is shown by the high bars representing short vowels, with only minimal responses indicating long vowels.

However, as the vowel duration increased, particularly from 175 ms (items 1.5 to 1.7), the number of participants identifying the vowels as short decreased significantly. For durations of 175 ms, 200 ms and 225 ms there was a noticeable shift in perception. The majority of participants identified these vowels as long, with a sharp increase in responses for long vowels. Durations of 200 ms and 225 ms showed almost unanimous identification of long vowels, indicating that participants consistently perceived these durations as long.

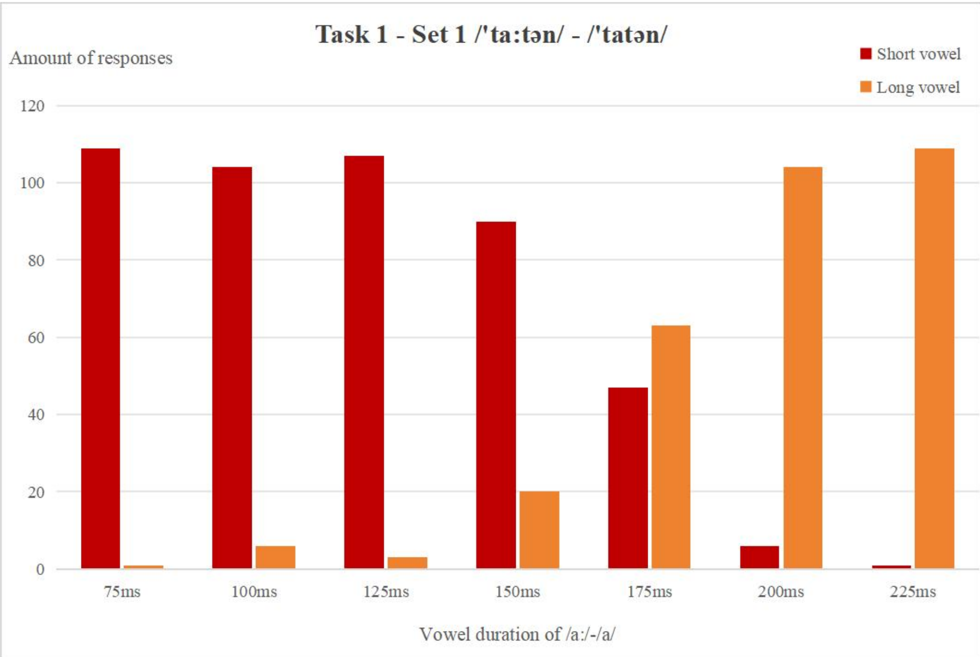

Figure 1 Participant Responses to Vowel Duration in Task 1 - Set 1

The trends observed in Set 2 (see Figure 2) are consistent with those observed in Set 1. For the word pairs /'fa:dən/ - /'fadən/ and the first three durations (75 ms, 100 ms and 125 ms), participants predominantly identified the vowels as short, as shown by the high bars for short vowels and minimal responses for long vowels. However, the identification of vowels as short decreased significantly for durations of 150 ms and beyond. From 150 ms to 225 ms the majority of participants identified the vowels as long, with high bars for long vowels, similar to the trend observed in Set 1. Durations of 200 ms and 225 ms showed almost unanimous identification as long vowels, indicating a strong consensus among participants regarding these durations.

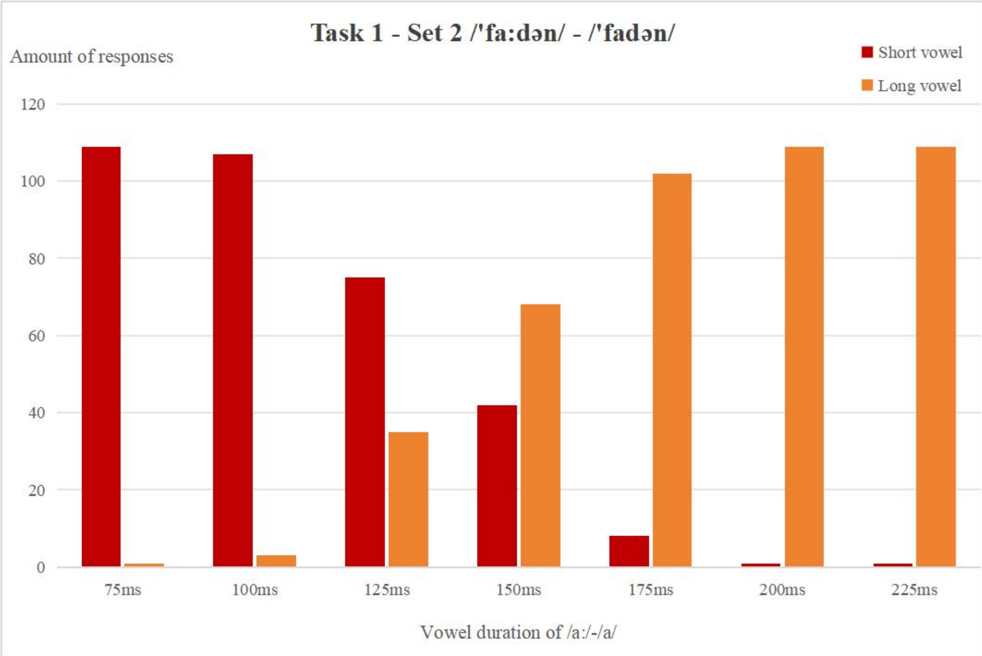

Figure 2 Participant Responses to Vowel Duration in Task 1 - Set 2

To examine the associations between the background variables and the participants' responses in Task 1, chi-squared tests were conducted. These revealed several significant associations. In particular, there were significant associations between the variables and the perception of durations around 150 ms.

Significant associations were found between participants' branch (English vs. German) and vowel duration perception at 175 ms and 150 ms. Chi-Square tests indicated significant differences for 175 ms in Set 1, $\chi^2(1, N = 110) = 6.072$, $p = .014$, and for 150 ms in Set 2, $\chi^2(1, N = 110) = 12.419$, $p < .001$, suggesting that German learners were more likely to categorise these vowels as short than English learners. Similarly, participants' prior phonetic knowledge was significantly associated with vowel duration perception at 125 ms and 150 ms in Set 1, $\chi^2(1, N = 110) = 6.320$, $p = .012$ for 125 ms, and $\chi^2(1, N = 110) = 4.230$, $p = .040$ for 150 ms. Participants with relevant prior knowledge were less likely to perceive vowels around 150 ms as long vowels. In contrast, participants' perception of the long-short contrast did not differ by learning environment (see Appendix 8 for full results).

4.2 Task 2: Tense-lax vowel perception

Task 2 is designed to assess participants' ability to perceive differences in vowel duration and formant values. The data collected from Task 2 are visualised in three bar graphs, each illustrating the participants' perception of differences under specific conditions.

The bar chart in Figure 3 illustrates the participants' responses to word pairs of different durations, categorised into two primary duration differences: 50 ms and 100 ms. Each vowel contrast is labelled '1', '2' and '3' to distinguish them. Focusing primarily on the duration differences reveals several key trends in participants' perceptions. For the 50 ms duration difference, participants overwhelmingly perceived no difference for contrast 1 and 3. Contrast 2 showed a more balanced distribution, with around 60 participants perceiving no difference and around 40 participants perceiving a difference. A different pattern emerged for the 100 ms duration difference. For contrast '1', a significant majority of participants (about 80) perceived the difference, while about 40 participants perceived no difference. Contrast '2' and '3' showed a more balanced response. The 100 ms duration difference resulted in a higher detection of differences in all contrasts compared to the 50 ms duration difference.

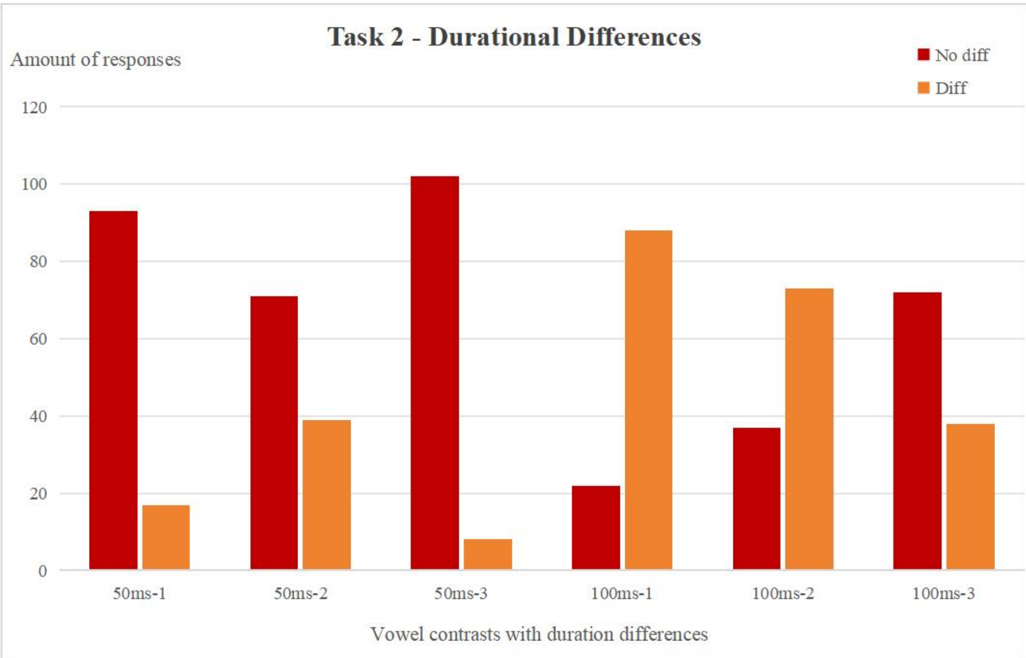

Figure 3 Participants' Perception of Durational Differences in Task 2

The graph shows the number of participants who perceived differences ("Diff") and those who perceived no differences ("No diff") for each word pair. The terms '1', '2' and '3' distinguish between vowel contrasts with equivalent differences. Durational and formant differences are represented as 1 or 2 fd (units of formant difference) and 50 or 100 ms (milliseconds of durational difference), respectively. For instance, "1fd-50ms-short" indicates that the vowels in the word pair have 1 unit of formant difference, 50 ms of durational difference, and that these changes occur in two relatively short vowels.

The second bar chart (see Figure 4) illustrates participants' perception of formant differences. The word pairs are categorised into two levels of formant difference (1 fd and 2 fd) and further divided into short, mid, and long. At 1 fd, participants predominantly identified no difference across all environments. At 2 fd, there was a notable increase in the recognition of differences, particularly in the 'short' and 'mid' environments. Overall, participants demonstrated a greater ability to perceive formant differences at higher levels (2 fd).

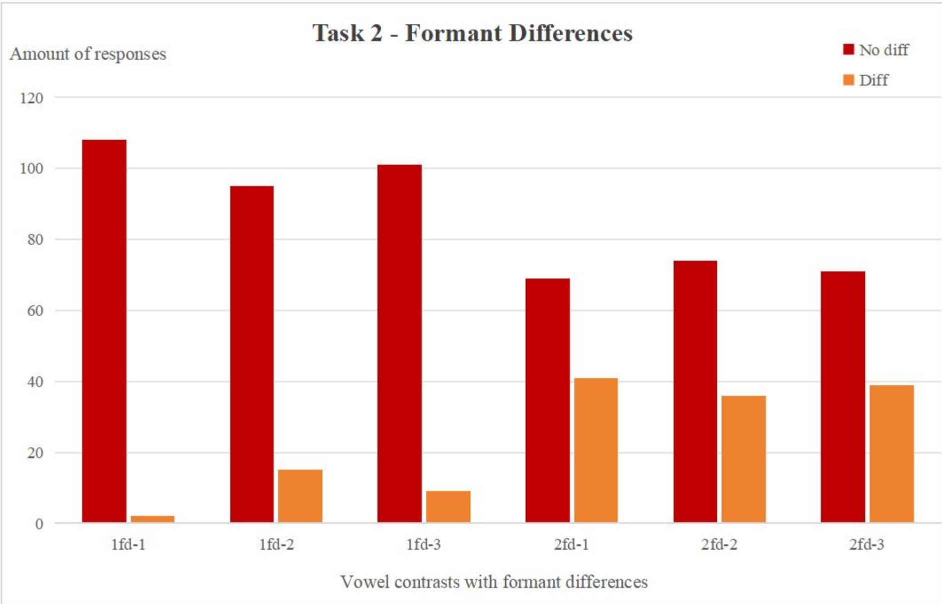

Figure 4 Participants' Perception of Formant Differences in Task 2

Shifting the focus from the detailed analysis of individual differences, Figure 5 presents participants' perceptions of variations that include both formant and duration changes.

For pairs with a formant difference of 1 unit (1fd) and a durational difference of 50 ms for short vowels, the majority of participants did not perceive the differences, with approximately 80 participants reporting no difference and about 20 participants perceiving the difference. This trend remained consistent when the formant difference was increased to 2 units, while the durational difference was kept at 50 ms across environments.

For pairs with a duration difference of 100 ms, a significant increase in difference perception is observed across different formant differences and vowel environments. For example, in pairs with 1 formant difference and a 100 ms duration difference in short vowels, about 60 participants perceived a difference compared to about 40 who did not.

The highest detection rates were observed in pairs with 2 formant differences and a 100 ms duration difference in long vowels, where about 90 participants detected the difference compared to 20 who did not. This set of mixed differences was also highly perceptible in short vowel environments.

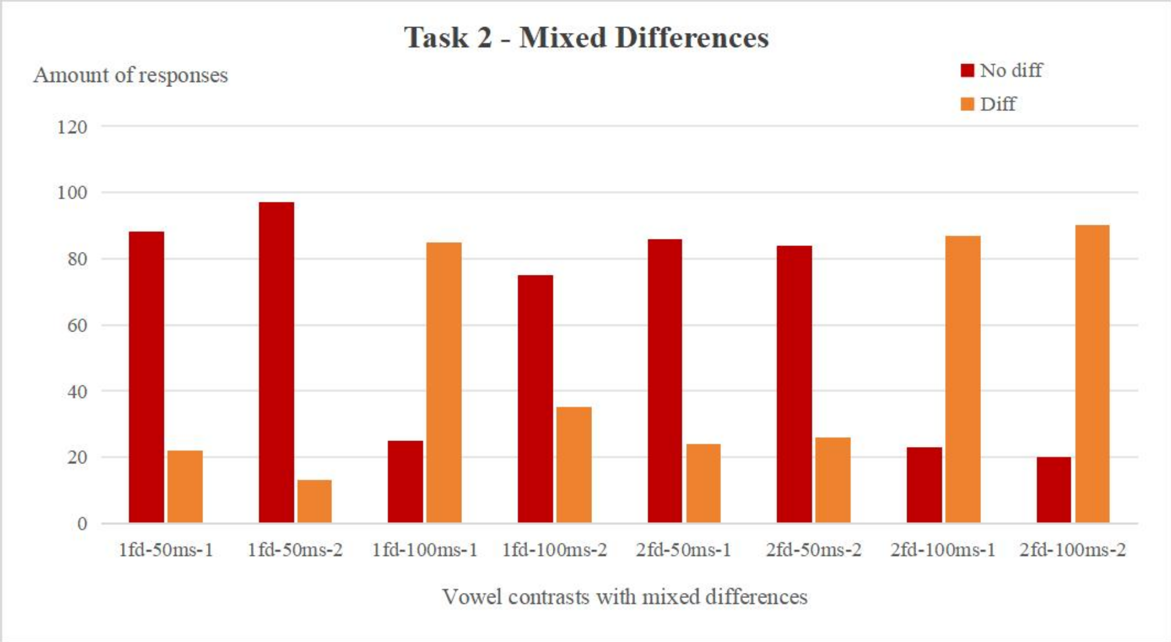

Figure 5 Participants' Perception of Mixed Differences in Task 2

To gain deeper insights into the participants' perceptual patterns and their relationship to language background, inferential statistical analyses were conducted. These included chi-square tests, correlation tests, independent samples t-tests and regression analyses. These analyses focused on the number of perceived differences, which were treated as continuous variables. Table 1 provides a summary of the continuous variables used in the analyses, representing the number of various types of differences perceived in Task 2. The detailed results are presented below.

Table 1 *Description of Continuous Variables Used in Task 2 Analyses*

Variable	Description
task2_dd	Number of perceived durational differences in Task 2
task2_fd	Number of perceived formant differences in Task 2
task2_md	Number of perceived mixed differences in Task 2
task2_total	Total number of perceived differences in Task 2

4.2.1 Chi-square Tests

Chi-squared tests were performed to determine whether there were significant associations between the background variables and participants' responses to each question in Task 2. The results showed that only prior knowledge of phonetics and phonology was significantly related to participants' perception of vowel differences in 5 out of 20 cases (see Table 2). Specifically, participants with prior phonetic knowledge were significantly more likely to perceive durational differences at 50 ms and 100 ms, as well as in the perception of formant differences at the level of 2 units.

Table 2 *Significant Chi-Square Test Results for Prior Phonetic Knowledge and Perceived Differences in Task 2*

Difference	χ^2	*df*	*N*	*p*
50 ms	9.658	1	110	.002
100 ms -1	9.321	1	110	.002
100 ms -2	8.392	1	110	.004
100 ms -3	6.410	1	110	.011
2 units of formant difference	12.296	1	110	< .001

4.2.2 Correlation

Correlation tests were used to examine the relationship between pre-test scores and the number of perceived differences. The results for German learning participants, as shown in Table 3, indicate a significant relationship between English pre-test scores and the number of perceived differences. These results suggest that higher proficiency in English, as measured by pre-test scores, is significantly associated with a greater ability to perceive differences in vowel duration and formant values. Notably, this correlation was not observed for German pre-test scores.

Table 3 *Correlation Coefficients for Amount of Perceived Differences and Language Pre-Test Scores in Task 2*

Variable		task2_dd	task2_fd	task2_md	task2_total
English pre-test scores	Pearson Correlation	.223	.245	.114	.221
	p	.019	.010	.238	.020
	N	110	110	110	110
German pre-test scores	Pearson Correlation	.254	.068	.210	.232
	p	.058	.560	.070	.105
	N	75	75	75	75

4.2.3 Repeated Measures ANOVA

A repeated measures ANOVA was conducted to compare the effects of the three types of acoustic cues (duration, formant, and mixed) on perceived differences while controlling for the effects of background variables, including the pre-test scores, branch of study, prior phonetic knowledge and learning environment.

The main effect of cues was significant, $F(2, 99) = 4.495$, $p = .012$, partial $\eta^2 = .043$, indicating a significant difference in perceived differences across the three types of acoustic cues. Specifically, the estimated marginal means showed that the highest mean perceived differences were observed for mixed cues ($M = 2.869$, $SE = 0.162$), followed by durational cues ($M = 2.724$, $SE = 0.191$), and the lowest for formant cues ($M = 1.293$, $SE = 0.210$). Bonferroni adjusted pairwise comparisons further clarified these differences. There was a significant difference between durational cues and formant cues, with durational cues having higher perceived differences (mean difference = 1.432, $SE = 0.246$, $p < .001$). Similarly, mixed cues had significantly higher perceived differences than formant cues (mean difference = -1.576, $SE = 0.210$, $p < .001$). However, there was no significant difference between durational cues and mixed cues (mean difference = -0.144, $SE = 0.205$, $p = 1.000$).

The analysis also showed that prior phonetic knowledge had a significant effect on perceived differences in acoustic cues on average, $F(1, 100) = 9.848$, $p = .002$, partial $\eta^2 = .090$. Participants with prior knowledge of phonetics perceived acoustic differences significantly more than those without such knowledge. Additionally, study abroad experience had a significant effect on perceived differences, $F(1, 100) = 6.078$, $p = .015$, partial $\eta^2 = .057$. Participants who had studied abroad were more accurate in perceiving acoustic differences compared to those who had not.

The analysis also revealed that several background variables did not significantly affect the perceived differences in acoustic cues. Participants' proficiency in English, as measured by pre-test scores, did not significantly affect their perception of acoustic differences, $F(1, 100) = 1.797$, $p = .183$, partial $\eta^2 = .018$. Similarly, proficiency in German also did not have a significant effect, $F(1, 100) = 1.379$, $p = .243$, partial $\eta^2 = .014$. This indicates that participants' skills in English and German did not influence their ability to perceive differences in the acoustic cues. Furthermore, the branch of study alone did not have a significant effect on their perception of acoustic differences, $F(1, 100) = 2.569$, $p = .112$, partial $\eta^2 = .025$. Notably, there was a significant interaction between branch of study and study abroad experience, $F(1, 100) = 4.028$, $p = .047$, partial η^2

= .039, suggesting that these factors jointly influenced participants' perception of acoustic cues. Given the small effect size, this result is reported for completeness but is not the focus of further analysis in this study. The full results of the ANOVA can be found in Appendix 9.

4.2.4 Regression

Chi-squared, correlation and independent samples t-tests for Task 2 revealed several significant and consistent relationships between the background variables and their ability to perceive differences in vowel duration and formant values. This analysis provides a robust and consistent basis for further investigation using regression. Regression analyses further investigate the predictive power of these background variables on participants' perceptual patterns in Task 2. Examining the influence of multiple predictors simultaneously provides a comprehensive understanding of the relative contribution of each factor.

The regression analyses provide robust evidence that prior phonetic knowledge and English pre-test scores are significant predictors of participants' ability to perceive vowel differences, particularly for perceived differences in overall and specific durational and formant differences. The current model includes the following predictors: prior knowledge, study environment, proficiency in English and proficiency in German. As the branch of study did not show any significant associations with perceptions of acoustic differences in previous analyses, it was not included in this model. Specifically, the model for Task 2 total differences was a good fit ($R = .500$, $R^2 = .250$, $F(4, 105) = 8.741$, $p < .001$). It highlighted prior knowledge ($\beta = .366$, $p < .001$) and English pre-test scores ($\beta = .247$, $p = .008$) as significant predictors. Similarly, the model for Task 2 durational differences was also a good fit ($R = .497$, $R^2 = .247$, $F(4, 105) = 8.621$, $p < .001$), and revealed prior knowledge ($\beta = .407$, $p < .001$) and English pre-test scores ($\beta = .203$, $p = .030$) to be significant predictors. The model for Task 2 formant differences was also significant but explained less of the variance ($R = .396$, $R^2 = .157$, $F(4, 105) = 4.887$, $p = .001$). It also showed prior knowledge ($\beta = .241$, $p = .012$) and English pre-test scores ($\beta = .248$, $p = .013$) as significant predictors of distinguishing formant differences.

However, for the mixed differences in Task 2, the regression model was not a good fit for the data, ($R = .260$, $R^2 = .068$, $F(4, 105) = 1.909$, $p = .114$) and did not yield significant predictors, suggesting the need for further investigation of other potential factors influencing this perceptual ability.

4.3 Task 3: Vowel Formant Perception

Task 3 was designed to assess participants' ability to perceive formant differences in vowel pairs. Participants listened to either German vowel pairs (ε-ø) or English vowel pairs (ε-æ). These pairs varied in formant differences, ranging from 1 unit to 3 units. The data collected in Task 3 are visualised in two bar charts (see Figure 6 and Figure 7).

For English vowels, the results for items with a 1 unit formant difference (1fd) show that the majority of participants did not perceive the formant difference. For each item, approximately 100 participants indicated that there was no difference between the pairs they heard, and only a small number perceived the difference. A similar pattern is observed for the 2fd items, with almost all participants reporting no difference. When the formant difference was increased to 3 units (3fd), there was a noticeable increase in the number of participants who perceived the difference, although responses of no difference still dominate.

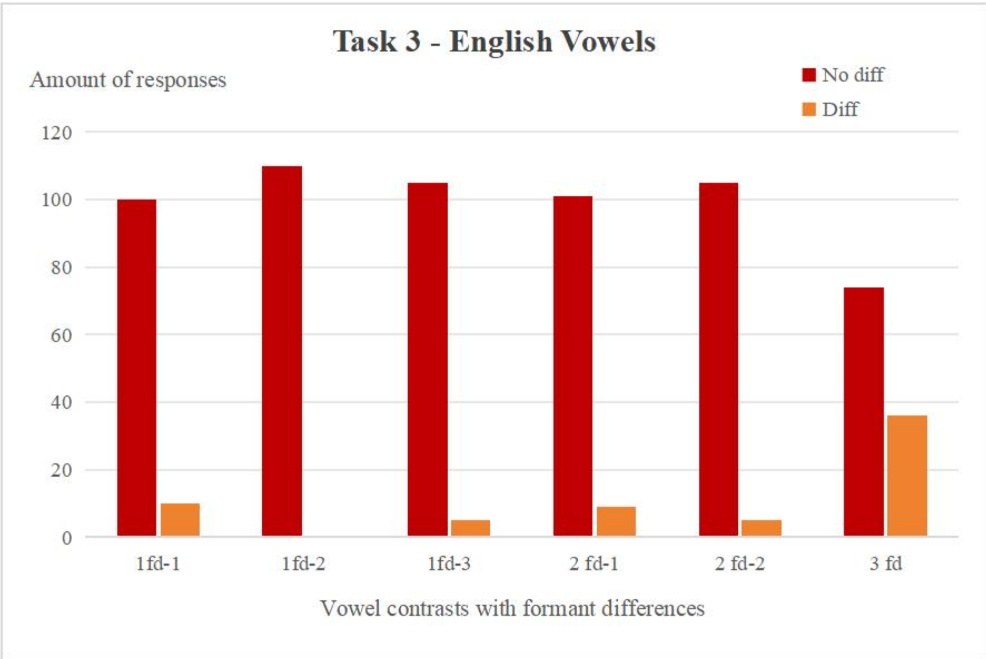

Figure 6 Participants' Perception of English Vowel Differences in Task 3

Figure 6 shows the participants' perception of English vowel differences, while Figure 7 shows their perception of German vowel differences. Both graphs report the number of participants who perceived differences ("Diff") and those who perceived no differences ("No diff") for each level of formant difference. The labels '-1', '-2' and '-3' were used to differentiate between the three items and have no intrinsic or practical meaning.

For German vowels, the 1fd items demonstrate that the majority of participants did not detect the difference. However, compared to the English vowels, there was a slightly more even response pattern, especially in the 1fd-3 condition. In the 2fd items, participants' responses were more balanced. With a 3-unit formant difference (3fd), there came a significant increase in the number of participants recognising the difference, yielding a more balanced distribution compared to the smaller formant differences.

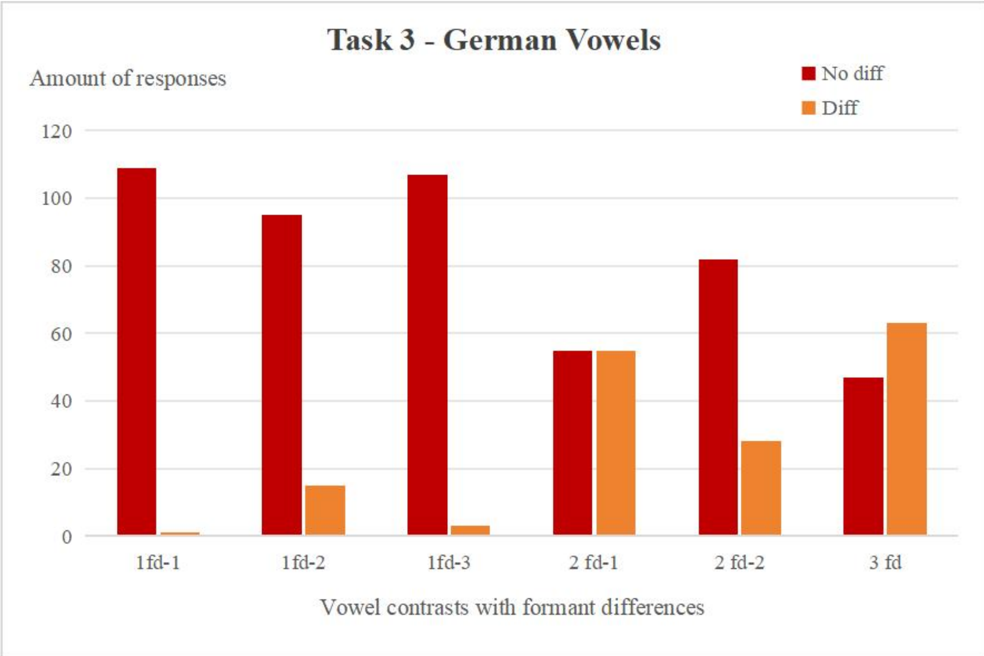

Figure 7 Participants' Perception of German Vowel Differences in Task 3

Following the descriptive analysis of participants' responses in Task 3, further inferential statistical analyses were carried out to explore the data in more depth. These analyses include chi-squared tests, correlation tests and repeated measures ANOVA. The results of these analyses are presented in detail in the following sections.

4.3.1 Chi-square Tests

The chi-squared test results for Task 3 show an inconsistent pattern regarding the relationship between the background variables and the participants' ability to perceive formant differences in vowel pairs.

For study abroad experience, there were significant associations with the perception of formant differences in certain items. Specifically, a significant relationship was found for the German ε-ø pair with a 1 unit formant difference, $\chi^2(1, N = 110) = 6.358$, $p = .012$, and for the English ε-æ pair with a 2 unit formant difference, $\chi^2(1, N = 110) = 5.487$, $p = .019$. Participants with study abroad experience reported a greater chance of perceiving formant differences. However, these significant associations were not observed consistently across all items.

For prior knowledge, a significant relationship with the perception of formant differences was observed in the German ɛ-ø pair with a 3-unit formant difference, $\chi^2(1, N = 110) = 4.266$, $p = .039$. This indicates that participants with prior phonetic knowledge were more likely to perceive formant differences in this particular vowel pair. Similar to the results reported for study abroad experience, this significant association was not found consistently across all items in the prior knowledge analysis.

In terms of language branch, the analysis did not yield significant results overall. The full Chi-square results for Task 3 are provided in Appendix 8.

4.3.2 Correlation

The correlation test results for Task 3 show a significant positive correlation between English pre-test scores and the amount of perceived formant differences in English vowels ($r = .203$, $p = .034$). This indicates that higher English pre-test scores are associated with a greater ability to perceive formant differences in English vowels. No significant correlations were found between pre-test scores (English or German) and the amount of perceived formant differences in German vowels, or the total amount of perceived differences in Task 3.

4.3.3 Repeated Measures ANOVA

A repeated measures ANOVA was conducted to compare the effects of different types of acoustic cues (English and German vowels) on participants' perceived differences while considering the influence of several background variables, including pre-test scores, branch of study, prior phonetic knowledge, and study abroad experience.

The main effect of cue type was significant, indicating a notable difference in the perception of English and German vowels, $F(1, 100) = 12.185$, $p < .001$, partial $\eta^2 = .109$. Specifically, participants perceived more differences in German vowel contrasts ($M = 1.640$, $SE = 0.197$) compared to English vowel contrasts ($M = 0.682$, $SE = 0.133$).

The effects of background variables were analysed to understand their impact on perception. The effect of study abroad experience was significant, $F(1, 100) = 7.456$, $p = .007$, partial $\eta^2 = .069$. Participants who had studied abroad showed higher accuracy in perceiving differences. Other background variables showed no significant influence (see Appendix 9 for full results).

4.3.4 Regression

Analysis of the Task 3 data using chi-squared tests, correlation tests and independent samples t-tests revealed inconsistent and sporadic results regarding the relationships between participants' background variables and their ability to perceive formant differences in vowel pairs. While there were instances of significant associations, such as the relationship between study abroad experience and the perception of formant differences in specific items, these significant findings were not consistent across conditions.

Given the inconsistent nature of these results, the relationships between background variables and perceptual abilities appear to be weak and influenced by factors not captured in the current data set. This variability suggests that any significant findings may be due to chance or uninvestigated variables rather than robust patterns. Considering these points, it may not be necessary to proceed with regression testing at this stage. The current data do not provide a strong basis for such modelling, and conducting regression analysis in this context could lead to overfitting, where the model captures random noise rather than meaningful relationships.

5. Discussion

This chapter discusses the results of the experimental tasks and interprets the findings in the context of the existing literature. The discussion is structured around three main tasks and will highlight how study abroad experiences, language learning branches and prior phonetic knowledge influence vowel perception. In addition, this chapter provides a comprehensive discussion of the overall findings by revisiting the research questions. It considers the pedagogical implications of the findings, highlighting their significance for language learning and teaching practice. The chapter also critically examines the limitations of the current study, discussing potential methodological issues and the limitations of the data collection process.

5.1 Discussion by Task

5.1.1 Task 1

Overall, the data from both sets show a consistent pattern in participants' perception of vowel duration. Vowels with shorter durations (75 ms to 125 ms) are consistently perceived as short by the majority of participants. In contrast, vowels with longer durations (175 ms to 225 ms) are predominantly perceived as long, with a sharp transition observed around the midpoint of 150 ms. However, there are subtle differences between the two sets in terms of the perceptual boundary. In Set 1, the shift from short to long vowel identification begins markedly at 150 ms (Item 1.5), whereas in Set 2 this shift is slightly more gradual, beginning at around 125 ms (Item 2.3). This suggests that while participants are highly sensitive to changes in vowel duration and can reliably discriminate between short and long vowels, the perceptual boundary may vary slightly depending on the specific word pairs used in the stimuli. These variations in perceptual boundaries can be further understood through Bennett's (1968) findings, which indicate that vowel perception is closely related to the specific phonetic context and language-specific phonological structures. In this study, the pre- and postvocalic consonants differed between the two sets, with /t/ and /t/ in Set 1 and /f/ and /d/ in Set 2, potentially causing variations in vowel perception. This observation supports the view that the phonetic context, including the surrounding consonants, correlates with how vowels are perceived and identified (van Son, 1999).

The chi-squared test results provide important insights into the factors that influence vowel length perception in different groups of language learners. The current results indicate that study abroad experience, language learning branch and prior phonetic knowledge significantly influence participants' vowel perception. In particular, those who have studied abroad, are learning German, and have prior phonetic knowledge tend to perceive vowels in the mid-length range as short.

One possible explanation for these findings is the greater stability and distinctiveness of German long vowels. In both English and German, long vowels generally have more stable and distinct vowel spaces compared to short vowels (Muehlbauer, 2012; Rosenhouse et al., 2014). Stability refers to the consistency of the acoustic characteristics of a vowel across different instances of pronunciation. For long vowels, this implies a relatively unchanging articulatory position, resulting in a more uniform sound. Distinctiveness refers to the clear differentiation of a vowel from other vowels, with long vowels occupying a more defined and isolated position in the vowel space. This reduces acoustic overlap and makes long vowels easier to distinguish. However, as noted in the literature above, German long-short vowels have a more obvious contrast than their English counterparts (Antoniadis & Strube, 1984; House, 1961), making them more distinctive and stable for recognition. Furthermore, because the stimuli in the tasks simulated the form of German words, German learners might be more experienced in perceiving these words. Thus, German learners may be more adept at identifying the clear distinctions of long vowels than English-only learners. Thus, they might also be more likely to perceive ambiguous durations as short due to the relative instability and variability of short vowels.

Furthermore, learners who have studied abroad in native-speaking environments are exposed to naturalistic language use, which enhances their perceptual abilities in vowel duration, as highlighted by Stevens (2011). Stevens' study suggests that English learners of L2 Spanish demonstrate more naturalistic use of their L2 in immersive learning environments, particularly in production. The current study extends this claim by showing that learners also develop more native-like perceptual abilities, becoming more accurate and salient in their perception of formant differences. As native speakers of English and German rely more on formant cues (Bennett, 1968; Yang et al., 2016), the study abroad experience helps learners to align their perceptual abilities with those of native speakers.

Prior phonetic knowledge provides these learners with a deeper understanding of acoustic properties and phonological structures, enabling them to make more informed perceptual decisions about vowel duration (Aliaga-García, 2009; Saloranta et al., 2020). This background allows them to better exploit the distinctiveness and stability of long vowels.

Another important finding from this task was that the majority of significant associations are observed primarily around critical transition periods, specifically for vowels with durations between 125 ms and 175 ms. This finding suggests that these vowel durations may be inherently controversial or ambiguous. Focusing on items around these midpoints is essential, as they represent the transition zone where participants' perception shifts from short to long vowels. Understanding this zone is central to capturing the perceptual boundary and sensitivity to changes in duration, which is fundamental to developing accurate models of vowel perception.

5.1.2 Task 2

The descriptive data from Task 2 reveal several insights into the participants' ability to perceive durational and formant differences in vowels. The ability to detect differences appeared to be influenced by the magnitude of the variation, with participants showing greater sensitivity to larger changes in both duration and formant values. The findings from Lidestam (2009) and Lijzenga (1997) provide a detailed understanding of learners' sensitivity to variations in vowel sounds. Lidestam's research provides a robust indicator of perceptual sensitivity to vowel durations as small as 33-66 milliseconds, with a 75% just-noticeable difference (JND) falling within this range under most conditions. Specifically, the JND refers to the smallest detectable difference in vowel duration that participants can reliably perceive. The 75% JND represents the duration difference at which participants achieve 75% accuracy in discrimination, ensuring that language learners can consistently detect subtle changes in vowel length. On the other hand, Lijzenga's study shows that just-noticeable differences in the formant frequencies of synthetic vowels can range from 20-100 Hz, depending on the specific formant and frequency range. In particular, the minimum duration gap of 50 milliseconds and the differences in formant values in Task 2 fall within the range of just-noticeable differences. This means that participants can reliably perceive such small differences in duration. However, these changes are not always immediately apparent in perception and still present considerable difficulty.

Furthermore, the results of the study illustrate a strong preference of the participants to discriminate duration differences over formant variations in speech sounds. Analysis of perceptual results from multiple sets of items revealed that the majority of participants were more effective at perceiving durational differences than formant differences in word pairs. Specifically, when both types of differences were present, an increase in durational changes significantly improved perceptual accuracy, whereas formant changes did not show a similar effect. Statistical analyses confirmed the prominence of durational cues in perception. This observation is supported by Kajouj and Kager (2019), who found that certain auditory cues, such as vowel durations, are inherently more accessible and thus more easily used by L2 listeners, regardless of their prior language experience. Additionally, the influence of cross-linguistic factors is suggested, as studies such as Fu et al. (2019) show that, while native Mandarin speakers do not typically rely heavily on durational and spectral cues for vowel discrimination, they do show a preference for durational cues under certain conditions, such as in Mandarin concurrent vowel identification. This preference may extend to more complex situations such as foreign language learning, although further research is needed to explore this potential transfer effect.

Analysis of the descriptive data and perceptual accuracy also shows that mixed cues - combining duration and formant differences - elicit more accurate responses than formant cues alone. This suggests that the integration of multiple acoustic cues is a more effective tool for discriminating between different sounds. Specifically, adding durational cues to spectral cues increases perceptual salience, whereas adding formant cues to durational cues does not yield a significant increase in perceptual detection. This is consistent with Steffman & Zhang (2023), who found that duration not only serves as an intrinsic vowel cue, but also enhances the perception of formant cues under conditions of prominence. Although they focused on the fundamental formant (F0), representing the pitch of the sound, rather than the F1 and F2 of vowels used in the present study, their findings provide valuable insights for further research in this area.

Notably, the distribution of responses in task 2 showed significant variability across different vowel environments, such as short and long vowel contexts. Perception was not consistent across these environments, with some showing clear numerical differences in participants' responses. This suggests that perception may vary across different phonetic environments, influenced not only by the consonantal context surrounding vowels, but also by the vowel environments per se. The environment was not the main focus of the study: rather, it was used to discriminate items and avoid repetition. Nevertheless, these results highlight the complexity of vowel perception, where contextual factors may be a potential factor influencing vowel perception.

Further analyses provide a comprehensive understanding of the factors that influence vowel perception. Several common themes and findings emerged from the analyses. A consistent theme throughout the literature and empirical studies, including the one at hand, is that prior phonetic knowledge significantly improves perceptual accuracy in vowel identification, with trained participants showing increased sensitivity to vowel duration and formant variation. For example, this is consistent with Kingston's (2003) findings that perceptual training significantly assists L2 speakers in acquiring vowel sounds in the target language, and Dlaska and Krekeler's (2008) findings that highlight the transformative effect of explicit phonetic training on the perceptual abilities of advanced L2 learners. However, it is important to note that the current study found that learners without prior knowledge or training did not completely miss spectral cues; they just perceived them to a lesser extent. This contrasts with Ding and Jokisch's (2012) findings, which suggest that while training improves perception, the absence of training doesn't completely negate the perception of spectral differences, but indicates reduced sensitivity.

Another recurring theme in the data was the relationship between language proficiency and perceptual ability. A significant finding of the current study is that higher proficiency in L2 English correlates with better vowel discrimination performance for all participants. This notion is evident in research suggesting that higher L2 proficiency leads to more target-like phonological acquisition in L3 due to enhanced meta-linguistic competence (Garcia, 2013). However, the influence of language proficiency varies depending on the order of language acquisition; while L2 proficiency significantly impacts vowel discrimination, L3 proficiency does not. This pattern is partly explained by the finding that lower proficiency in one language allows the more proficient language to dominate L3 phonological acquisition (Cal & Sypiańska; 2020), as reported L3 learners' pre-test scores are significantly lower in their L3 than in their L2.

Other factors did not appear to be consistently related, showing only occasional or no positive correlations with participants' perceptual performance. The inconsistent performance across participants suggests a need for further discussion. For example, the finding that the number of languages studied—whether learners were bilingual (English only) or multilingual (English and German)—did not significantly predict perceptual accuracy across contrasts and tasks. This result may seem to contradict previous research suggesting that multilingualism enhances phonetic discrimination (e.g., Cohen et al., 1967; Enomoto, 1994). Besides, while studying abroad often leads to better performance in certain linguistic contrasts, this is not the case for all task types and contrasts. This observation ties with the findings of Alshangiti and Evans (2015), who noted that learners in immersion environments primarily improve in production rather than perception. Furthermore, Flege and MacKay (2004) highlighted the need to consider additional conditions related to the immersive learning environment, such as the age of arrival and the extent of L1 usage. These factors are found to have a significant impact on vowel perception, with earlier arriving learners and those who use their L1 less frequently generally exhibiting better L2 vowel perception. These suggest that while immersion through study abroad can improve certain language skills, such as production, its effects on perception are complex and influenced by the broader context.

5.1.3 Task 3

The data from Task 3 provided several insights into participants' ability to perceive formant differences in vowel pairs. A key observation from the visualised responses was the participants' increased sensitivity to larger formant differences. Vowel contrasts with greater differences in formant values were generally perceived more accurately by the participants. The analysis from Task 3 corroborates the trends observed in Task 2, showing a consistent pattern in the perception of acoustic changes. Participants showed increased accuracy when perceiving larger formant differences in vowel pairs.

Despite similar trends, an additional phenomenon was observed in this task: the formant differences of German vowels elicited more correct responses than those of the English vowels, especially when subjected to significant formant variations (2 and 3 steps of formant differences). The statistical analyses confirmed these descriptive findings, revealing a moderate difference in the perception of English versus German vowel pairs. This suggests that language-specific characteristics of German vowels may inherently support better perceptual accuracy. This trend was consistent across different participant backgrounds, suggesting that these effects are likely to be influenced by the intrinsic qualities of German vowel sounds rather than participant-specific factors. In order to understand why German vowels received more accurate responses, particularly given that greater formant variation leads to greater accuracy, the F1 and F2 values used to create the stimuli were re-examined. This examination confirmed that the formant differences between /ɛ/ and /ø/ used to create the German stimuli were greater than those between /ɛ/ and /æ/ used for the English stimuli. These larger gaps provided learners with more reliable perceptual cues for discriminating between /ɛ/ and /ø/. This discrepancy may have influenced the results by introducing a slight bias in the perception of formant differences. Future studies should address this oversight by ensuring that formant values are controlled and consistent across different language stimuli to avoid such confounding factors. Yet, the finding supports the important role that phonological features play in vowel discrimination (Boersma & Chládková, 2011) and suggests that vowel perception is guided more by phonetically based phonological features than by phonemic representations.

In the analysis of factors affecting perception, Task 3 showed limited significant results, with only the correlations between English proficiency and perceptual performance standing out as significant. The results of Task 3 extend the findings of Task 2 by confirming the importance of language proficiency in the perceptual accuracy of vowel differences. The correlation between English pre-test scores and the ability to discriminate formant differences in English vowels underscores the impact of language proficiency on perceptual sensitivity. Yet, this relationship did not emerge in the German vowel tests, echoing the suggestion from the previous discussion that the complex interaction between L2 and L3 proficiency warrants further investigation in future studies (Onishi, 2016). One possible improvement could be the use of more proficiency-specific language assessments. General language proficiency tests measure general language knowledge, which does not necessarily correlate with participants' perceptual abilities (Ding & Jokisch, 2012). Instead, the use of a standardised listening test may be more relevant in the context of investigating perceptual skills.

In contrast, the chi-squared tests and correlation analyses revealed an inconsistent pattern regarding the influence of other background variables on vowel perception. While there were significant associations between study abroad experience and prior phonetic knowledge with the perception of specific formant differences, these associations were not observed consistently across items and conditions. This variability potentially suggests that significant results may be due to chance or confounded by unexamined factors.

Overall, the results from Task 3 underscore the necessity of a multifaceted approach to understanding vowel perception. Apart from English pre-test scores, the other variables provided limited actionable insights for analysis. The significant, yet sporadic, impact of background variables suggests that vowel perception is a complex interplay of acoustic and contextual factors.

5.2 Discussion by Research Question

5.2.1 Research Question 1

In response to Research Question 1, which asked how Chinese learners of English and German use quantitative (durational) and qualitative (spectral) cues to perceive vowels, it is first and foremost clear that these learners rely primarily on acoustic features. It is worth noting that the study did not use phonemic representations but measurable aspects of sound, such as duration in milliseconds and frequency in Hertz. The results suggest that these aspects significantly influence vowel perception. In particular, it was shown that changes in duration and formant values that were within the range of just noticeable differences are perceptible, but can sometimes lead to ambiguous interpretations. When these measurable differences increase in magnitude, they lead to a significant improvement in perceptual accuracy.

Furthermore, the Chinese learners effectively used duration cues to discriminate between long and short vowels, establishing a clear boundary between the two categories. This pattern approximates the categorical perception observed in native speakers of languages with a contrast between long and short vowels, where there is a clear and sharp boundary (Tomaschek et al., 2011; Wilson et al., 2005). Even in the absence of corresponding phonological features in their native language, Chinese learners in the current study were able to achieve native-like phonetic perception. Such abilities demonstrate the learners' considerable potential to adapt and refine their auditory skills, highlighting their capacity for phonetic learning and acquisition.

Nevertheless, the influence of specific linguistic environments, including consonantal and vocalic contexts, plays a crucial role in shaping these perceptual outcomes observed in this study. These environments can alter the perception of vowels, highlighting the complex interaction between acoustic signals and the linguistic context in which they are embedded.

5.2.2 Research Question 2

Research Question 2 investigated how learners prioritise quantitative and qualitative cues in vowel perception. The behaviour of the Chinese participants in this study offered revealing insights. Specifically, these learners tended to focus on differences in the duration of speech sounds rather than spectral variations influenced by formant values. This preference for durational cues, which are likely to be more perceptually salient and thus more easily registered by listeners' auditory systems regardless of the constraints of specific listening conditions, highlights their importance in vowel discrimination. However, the potential influence of cross-linguistic interference cannot be overlooked, as Mandarin speakers show a similar preference for duration cues when discriminating between two vowels presented simultaneously in their native language (Fu et al., 2019). Given this precedent, it is plausible that native Chinese speakers might also develop a duration-biassed perceptual pattern. This suggests a complex interaction between inherent perceptual strategies and linguistic background in the processing of phonetic cues.

While durational cues play an important role, they alone do not provide a complete picture of vowel perception in this study. The results have shown that a combination of acoustic cues, particularly when formant frequencies are considered alongside duration, is more effective in discriminating between different sounds than relying on a single type of cue. This supports an integrated approach to perceptual patterns, in which both durational and qualitative cues are considered simultaneously and work together to enhance perception.

Nevertheless, this study found no significant difference between the effectiveness of mixed cues (both durational and spectral) and durational cues alone. It could be that the learners tested in the present study may have had high efficiency in using durational cues, which may have in turn reduced the perceptual benefits of using additional spectral cues. This efficiency in processing durational cues could mean that, when both types of cues are available, learners may preferentially rely on the more salient durational cues, which alone provide sufficient phonetic information for accurate vowel discrimination. This selective reliance on durational cues, which renders additional cues redundant in some contexts, highlights the need for further research into how different acoustic properties are prioritised in phonetic perception with more learners, with different language combinations and in different listening environments. Understanding these perceptual strategies could lead to improved language teaching methods and more effective communication strategies in multilingual settings. For example, language teachers could design a curriculum that emphasises training in both durational and spectral cues, particularly for learners whose native language relies heavily on only one type of cue. This targeted training can help learners to become more versatile in their phonetic perception, thereby improving their ability to understand and produce sounds accurately in a foreign language. As a result, learners have a richer knowledge repertoire and use different types or combinations of cues to discriminate specific contrasts.

5.2.3 Research Question 3

Research Question 3 investigated the impact of language learning background – including immersive learning experiences, multilingualism, language proficiency, and prior phonetic and phonological training – on learners' perceptual patterns. This inquiry delved into how such broad background factors influence the nuances of auditory perception, leading to distinct differences in the processing and interpretation of sounds. It was found that learners who have had immersive learning experiences tended to make more effective and accurate use of the acoustic properties of vowels. This improved perception is likely to be facilitated by the rich linguistic input and heightened meta-linguistic awareness that contributes to refined phonetic discrimination skills (Carrera-Sabaté, 2014; Pratiwi, 2023). In addition, German L3 learners often develop perceptual patterns that are particularly attuned to the phonetic nuances of the German language, distinguishing them from English-only learners. Taken together, these findings suggest that the learners' linguistic backgrounds significantly influence perception — not only in terms of overall perceptual ability but also in relation to specific acoustic features of individual vowels.

Besides, it was found that learners with more extensive learning experience, including those who have undergone phonetic and phonological training or are more proficient in their language studies, generally perform better in vowel perception. Such expertise seems to significantly enhance the learners’ perceptual accuracy, enabling them to access and utilise the spectral characteristics of vowels and to recognise subtle phonetic nuances. This suggests that prior systematic phonetic training has a profound effect on learners' ability to process and understand phonetic cues. Although the current study did not directly test the effectiveness of phonetic training, prior phonetic training did appear to be a predictor of performance on the tasks. In addition, the correlation between proficiency and perceptual ability found in the tasks highlights the complexity of phonetic perception. Higher levels of proficiency are typically correlated with improved perceptual skills that enhance proficiency in subsequent languages, although this relationship may vary between L2 and L3 acquisition, possibly due to different degrees of linguistic interference and the cognitive strategies adopted by multilingual speakers. It may also be the case that the learners in the current study are generally more proficient in English than in German. Findings in L3 production have suggested a possible explanation: the interplay between L2 and L3 proficiency levels affects L3 capacities, with lower proficiency in one language allowing the more proficient language to exert a greater influence (Cal & Sypiańska, 2020). Thus, in the present case, English exerted a greater influence and yielded significant results.

It should be highlighted that not all background factors had significant correlations with all outcome measures, suggesting that certain background factors may enhance learning but do not universally guarantee improved perceptual outcomes. For example, the role of multilingualism was not found to be a significant predictor of perception. These factors, at least from a more general perspective, do not negatively impact perceptual outcomes. This inconsistency highlights the need to include a wider range of research variables. For example, Flege and MacKay (2004) highlight the importance of considering additional factors associated with immersive learning environments, such as age of arrival and L1 use. These variables can have a profound effect on how phonetic cues are perceived and processed. Thus, adopting a multifaceted approach ensures a deeper insight into how diverse linguistic backgrounds influence phonetic perception and the dynamics involved.

5.3 Pedagogical Implications

The research findings provide valuable insights for designing effective language learning strategies and improving pedagogical methods. First and foremost, the integration of explicit phonetic training into the language classroom seems to be crucial for improving learners' phonetic perception and overall language proficiency. Explicit phonetic training involves the systematic teaching of the sounds of a language through focused practice and detailed explanations of how sounds are physically produced (articulation) and perceived (auditory characteristics). It should be noted that the specific form, content and intensity of the training that the learners had undergone was not taken into account in the present study, so no specific conclusions can be drawn about how the nature of the training affects the learners' perception. However, the results showed that the completion of some form of training significantly differentiated the participants who received it from their untrained counterparts and increased their sensitivity to important phonetic cues such as vowel durational and spectral characteristics. In this respect, phonetic training seems to be important for the development of auditory perception, which is essential for language learning. High variability phonetic training (HVPT) is among the effective methods, in which learners are exposed to different pronunciations of phonetic sounds produced by different speakers in different contexts. A study by Lengeris (2019) evaluated the effectiveness of HVPT on the perception and production of Southern British English vowels by Greek native speakers. Participants in this study were exposed to a variety of English vowel sounds, resulting in significant improvements in their ability to perceive and produce the target vowels. This suggests that such training not only improves phonetic discrimination, but also aids accurate vowel production. Therefore, systematic phonetic training, such as the HVPT, could be incorporated into language teaching curricula.

In terms of pedagogical recommendations for training, combining the effective use of durational cues with the opportunity to use spectral cues can optimise phonetic training, particularly for learners from backgrounds such as Chinese who demonstrate a strong ability to process durational cues. Language teaching could initially emphasise these durational cues through exercises that focus on vowel length discrimination, which is critical in languages with significant vowel length differences. However, while durational cues alone are highly effective, the integration of spectral cues, such as formant frequencies, can further provide learners with further potential for perceptual improvement. Educational programmes should therefore take a step-by-step approach, starting with the more perceptible durational cues and gradually incorporating spectral cues. This method could include complex listening exercises that gradually expose the learners to less perceptually salient cues, thus comprehensively expanding their auditory skills. By strategically combining these cues in training, educators can better equip learners with the nuanced auditory skills needed to master a new language, and fully exploit the potential of both durational and spectral cues in phonetic instruction.

To increase the effectiveness of phonetic training, language courses need to be tailored to students' linguistic backgrounds and previous learning experiences. Courses should be designed to capitalise on the strengths of students' previous immersion experience by providing advanced phonetic discrimination exercises to challenge their refined skills. In contrast, beginners or learners from non-immersion backgrounds and those with lower levels of proficiency should receive basic phonetic input along with additional support such as increased linguistic exposure and input. This approach could help bridge learning gaps and improve their learning trajectories.

Finally, it's important to recognise that the field of research is characterised by ongoing debates and evolving research findings. The complexity of the interplay between individual learners' characteristics, their linguistic backgrounds and the phonetic features of the target languages requires a dynamic approach to curriculum development. Educational researchers are therefore encouraged to continually refine teaching methods on the basis of the latest empirical findings. By incorporating evidence from different linguistic and cultural contexts, language teaching can be tailored to be both inclusive and effective.

5.4 Limitations

This section discusses several limitations of the current research, mainly related to the design and structure of the stimuli used and the methodology of data collection. These limitations are important to consider as they affect the interpretation and generalisability of the research findings.

One limitation of the current research pertains to the design and structure of the stimuli used in the study. The vowel pairs in the stimuli were minimal vowel pairs or close in vowel space and were designed to have relatively similar features. However, based on Bennett's (1968) finding, the importance of duration cues in vowel discrimination becomes more pronounced when the vowels have very similar qualities. Although the specific quantitative features that make vowels 'similar' were not explicitly detailed in Bennett's study, this finding introduces a potential flaw in our analyses that might have biassed the participants’ prioritisation of cues. In other words, this similarity between vowels could potentially lead to an increased prominence of duration cues within the word pairs used in our experiments.

Another challenging issue in stimuli design was the isolation of vowels in order to study how they were perceived in different consonantal environments. While consonants significantly influence vowel perception, this study maintained a consistent consonantal context within each task, but not across tasks. This approach, while beneficial for internal validity within the task, may limit the generalisability of the findings to other tasks. Maintaining a consistent consonantal context ensures that the observed effects on vowel perception are not confounded by variations in consonantal environments. However, this methodological choice also means that the results may not fully capture how vowel perception works in more varied, naturalistic settings where consonantal contexts vary. Furthermore, in natural speech, the articulatory movements required to produce a sound may overlap with those required to produce adjacent sounds, resulting in a modification of the acoustic properties of the original sound. The perception of sounds is also based on the sounds that come before or after them (Harrington et al., 2008). By using synthesised sounds to keep vowels 'clean', this study may have sacrificed some ecological validity. The use of synthesised sounds ensured that vowels were not influenced by surrounding consonants, but this is not entirely realistic when considering how vowels are perceived in natural speech. Therefore, the results of this study are highly reliable within the specific experimental conditions, but do not necessarily extend to other linguistic tasks or real-world scenarios. Future research should rigorously investigate the effects of different consonantal contexts and thoughtfully incorporate these variations into experimental designs. This will allow for a more comprehensive understanding of their effects on vowel perception, thereby increasing the robustness and real-world applicability of the research findings.

Other limitations of this research relate to the data collection process, in particular the homogeneity of participants' educational backgrounds and the terminology used in the surveys. Firstly, due to the relatively uniform educational trajectory in China, the age of onset (AoO) of English and German learning among participants was predominantly concentrated around the primary school and university stages, respectively. This lack of variability in AoO meant that it could not be effectively analysed as an independent variable. The uniformity of AoO across the sample may have limited our ability to fully explore how different starting ages might affect language acquisition and phonetic perception. Previous research has shown that AoO is a crucial factor in L2 vowel perception. For example, Stölten et al. (2014) found that early L2 learners of Swedish showed near-native perceptual categorisation, whereas late learners did not. Similarly, Stölten (2009) showed a general age effect on category boundary placement in near-native L2 speakers' perception of Swedish voicing contrasts. These findings highlight the importance of AoO and suggest that it should be considered as a significant variable in future research.

Secondly, the survey asked participants to report on their language learning experiences, including whether they had received phonetic and phonological training. However, the terms 'phonetic' and 'phonological' are linguistic terms that some participants may not have fully understood. They may have perceived such training as a more formalised or advanced process than what they experienced, leading them to report no such training when they may have been exposed to it. In a future replication, researchers could use simpler language or provide examples to clarify what constitutes phonetic and phonological training. While this limitation casts some doubt on the data collected from the present study, it also highlights the inherent limitations of self-reported data, where participants' understanding and willingness to disclose information can affect the accuracy and reliability of the data collected.

6. Conclusion

This exploratory study systematically investigated the perceptual patterns of Chinese language learners' use of quantitative and qualitative cues in vowel perception. By investigating three research questions, this study has elucidated how learners prioritise these cues and how various factors such as language learning background and prior phonetic training influence their perceptual patterns.

The results highlight a pronounced reliance on durational cues over spectral cues among Chinese learners of English and German. This preference could be attributed to possible transfer from the L1 and the greater perceptual salience of durational information in vowel differentiation. Furthermore, learners' backgrounds, including factors such as immersive learning experiences and prior phonetic training, were found to significantly enhance their ability to perceive and discriminate vowels. Learners with such educational experiences demonstrate improved perceptual accuracy, highlighting the critical role of language exposure and explicit phonetic training in language education. These findings not only contribute to our understanding of the use of phonetic cues, but also highlight the influence of educational and experiential background in shaping phonetic perception.

In general, the study suggests several directions for future research to further explore and extend the current findings. One recommendation for broadening the scope of the field is to conduct studies that include previously unexamined or understudied factors, such as the age of onset (AoO) of language learning and other nuanced variables that fall under broader categories. By broadening the focus to include these variables, future research can provide a more comprehensive view of the interplay between linguistic background and phonetic learning, thereby enhancing our understanding of language acquisition processes. Longitudinal studies are also proposed to track changes in cue prioritisation as learners progress in language proficiency, providing insights into the dynamic nature of phonetic perception over time. In addition, investigating the effects of specific phonetic training interventions could help to determine their impact on improving perceptual accuracy and refining language teaching methods. These future avenues of research aim to deepen our understanding of phonetic perception and contribute to more effective language acquisition strategies.

References

Aliaga-Garcia, C. (2009). Effects of audiovisual auditory and articulatory training on second-language (L2) vowel category learning. *The Journal of the Acoustical Society of America*, *125*(4_Supplement), 2765-2765. https://doi.org/10.1121/1.4784692

Antoniadis, Z., & Strube, H. W. (1984). Untersuchungen zur spezifischen Dauer deutscher Vokale. *Phonetica*, *41*(2), 72-87. https://doi.org/10.1159/000261713

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

Bennett, D. C. (1968). Spectral form and duration as cues in the recognition of English and German vowels. *Language and speech*, *11*(2), 65-85. https://doi.org/10.1177/002383096801100201

Boersma, P., & Chladkova, K. (2011). Asymmetries between speech perception and production reveal phonological structure. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 328-331). Retrieved from https://hdl.handle.net/11245/1.346141

Boersma, P. & Weenink, D. (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.15, retrieved 27 July 2024 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.

Brown, C. A. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, *14*(2), 136-193. https://doi.org/10.1191/026765898669508401

Bryan, A. D. (2019). *The Effectiveness of Explicit Pedagogical Intervention in the L2 Perception and Production of German Vowels* (Order No. 30502523). Available from ProQuest Dissertations & Theses Global. (2827702040). https://proxy1.library.jhu.edu/login?url=https://www.proquest.com/dissertations-theses/effectiveness-explicit-pedagogical-intervention/docview/2827702040/se-2

Cal, Z., & Sypiańska, J. (2020). The interaction of L2 and L3 levels of proficiency in third language acquisition. *Poznan Studies in Contemporary Linguistics*, *56*(4), 577-603. https://doi.org/10.1515/psicl-2020-0019

Carrera-Sabaté, J. (2014). Does meta-linguistic awareness play any role at the beginning of an ongoing sound change? The case of some vowel-ended verbs in Catalan. *Sociolinguistic Studies*, *8*, 193-221. https://doi.org/10.1558/SOLS.V8I2.193

Cenoz, J., & Valencia, J. F. (1994). Additive trilingualism: Evidence from the Basque country. *Applied psycholinguistics*, *15*(2), 195-207. https://doi.org/10.1017/S0142716400005324

Chládková, K., Boersma, P., & Benders, T. (2015). The perceptual basis of the feature vowel height. In *International congress of phonetic sciences* (18th: 2015) (pp. 1-5). University of Glasgow. Retrieved from https://hdl.handle.net/11245/1.485964

Cheng, B., & Zhang, Y. (2015). Syllable structure universals and native language interference in second language perception and production: Positional asymmetry and perceptual links to accentedness. *Frontiers in psychology*, *6*(1801), 1-17. https://doi.org/10.3389/fpsyg.2015.01801

Cohen, S. P., Tucker, G. R., & Lambert, W. E. (1967). The comparative skills of monolinguals and bilinguals in perceiving phoneme sequences. *Language and speech*, *10*(3), 159-168. https://doi.org/10.1177/002383096701000303

Cohen, A. (1971). The vowel phonemes of English. In *The phonemes of English* (pp. 75-107). Springer, Dordrecht. https://doi.org/10.1007/978-94-010-2969-8_4

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

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

DeKeyser, R. M. (ed.) (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge: Cambridge University Press.

Derwing, T. M., Thomson, R. I., Foote, J. A., & Munro, M. J. (2012). A longitudinal study of listening perception in adult learners of English: Implications for teachers. *Canadian modern language review*, *68*(3), 247-266. https://doi.org/10.3138/cmlr.1215

Ding, H., & Jokisch, O. (2012). A Preliminary Investigation of Tense-Lax Contrast of German Vowels by Chinese Learners. In *Konferenz Elektronische Sprachsignalverarbeitung* (pp. 153-158). Dresden: TUDpress.

Dlaska, A., & Krekeler, C. (2008). Self-assessment of pronunciation. *System*, *36*(4), 506-516. https://doi.org/10.1016/j.system.2008.03.003

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. https://doi.org/10.1111/cogs.12015

Enomoto, K. (1994). L2 Perceptual Acquisition: The Effect of Multilingual Linguistic Experience on the Perception of a 'less novel' Contrast. *Edinburgh Working Papers in Applied Linguistics*, *5*, 15-29. Retrieved from https://eric.ed.gov/?id=ED373553

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

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191. https://doi.org/10.3758/BF03193146

Féry, C. (1996). German foot and word stress in OT. In P. Bye, O. Lorentz & C. Rice (Eds.), *Papers from the 2nd workshop on comparative Germanic phonology*, (pp. 63-96). Universitet i Tromsø. Institutt for språk og litteratur.

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

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*(5), 2973-2987. https://doi.org/10.1121/1.428116

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

Fu, Z., Wu, X., Chen, J., & Perception, M. (2019). Effects of Spectral and Temporal Cues to Mandarin Concurrent-Vowels Identification for Normal-Hearing and Hearing-Impaired Listeners. In *Interspeech* (pp. 3083-3087). https://doi.org/10.21437/interspeech.2019-3209

Gallardo del Puerto, F. (2007). Is L3 phonological competence affected by the learner's level of bilingualism?. *International Journal of Multilingualism*, *4*(1), 1-16. https://doi.org/10.2167/ijm042.0

Gao, Y., Ding, H., Birkholz, P., Jäckel, R., & Lin, Y. (2019). Perception of German tense and lax vowel contrast by Chinese learners. *Studientexte zur Sprachkommunikation:*

Elektronische Sprachsignalverarbeitung 2019, 25-32. Retrieved from https://www.essv.de/paper.php?id=57

Garcia, A. (2013). The effects of L2 proficiency on L3 phonological acquisition: A preliminary test of the L2 proficiency hypothesis. In *Selected proceedings of the 2011 second language research forum: Converging theory and practice* (pp. 173-186). Retrieved from https://www.lingref.com/cpp/slrf/2011/abstract2915.html

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

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: an acoustic and perceptual study. *The Journal of the Acoustical Society of America*, *123*(5), 2825-2835. https://doi.org/10.1121/1.2897042

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

House, A. S. (1961). On vowel duration in English. *The Journal of the Acoustical Society of America*, *33*(9), 1174-1178. https://doi.org/10.1121/1.1908941

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

Ivaska, I., & Siitonen, K. (2017). Learner language morphology as a window to crosslinguistic influences: A key structure analysis. *Nordic Journal of Linguistics*, *40*, 225-253. https://doi.org/10.1017/S0332586517000129

Kajouj, F., & Kager, R. (2019). Effects of bilingualism on cue weighting: How do bilingual children perceive the Dutch [ɑ]-[a:] contrast?. *International Journal of Bilingualism*, *23*(2), 509-524. https://doi.org/10.1177/1367006917745696

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in psychology*, *5*, 1246. https://doi.org/10.3389/fpsyg.2014.01246

Kingston, J. (2003). Learning foreign vowels. *Language and Speech*, *46*(2-3), 295-348. https://doi.org/10.1177/00238309030460020201

Kolb, N., Mitrofanova, N., & Westergaard, M. (2022). Crosslinguistic influence in child L3 English: An empirical study on Russian-German heritage bilinguals.

International Journal of Bilingualism, *26*, 476-501. https://doi.org/10.1177/13670069211054891

Lee, J., Shaiman, S., & Weismer, G. (2016). Relationship between tongue positions and formant frequencies in female speakers. *The Journal of the Acoustical Society of America*, *139*(1), 426-440. https://doi.org/10.1121/1.4939894

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*, 325-343. https://doi.org/10.3758/s13428-011-0146-0

Lengeris, A. (2008, August). The effectiveness of auditory phonetic training on Greek native speakers' perception and production of southern British English vowels. In *ExLing* (pp. 133-136). https://doi.org/10.36505/exling-2008/02/0034/000093

Levy, E. S., & Law, F. F. (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America*, *128*(3), 1290-1305. https://doi.org/10.1121/1.3466879

Lidestam, B. (2009). Visual discrimination of vowel duration. *Scandinavian journal of psychology*, *50*(5), 427-435. https://doi.org/10.1111/j.1467-9450.2009.00746.x

Lindau-Webb, M. (1987). Acoustic correlates of TENSE-LAX vowels and ATR vowels. *The Journal of the Acoustical Society of America*, *82*, 116. https://doi.org/10.1121/1.2024605

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

Lijzenga, J. (1997). *Discrimination of simplified vowel spectra* [Master's thesis, Rijksuniversiteit Groningen]. Groningen, The Netherlands. Retrieved from https://pure.rug.nl/ws/portalfiles/portal/14535988/c3.pdf

Llompart, M., & Reinisch, E. (2018). Acoustic cues, not phonological features, drive vowel perception: Evidence from height, position and tenseness contrasts in German vowels. *Journal of Phonetics*, *67*, 34-48. https://doi.org/10.1016/j.wocn.2017.12.001

Miller, J. S. (2012). Teaching French pronunciation with phonetics in a college-level beginner French course. *NECTFL Review*, *69*, 47-68. Retrieved from https://www.iastatedigitalpress.com/pslt/article/id/15181/

Muehlbauer, J.D. (2012). Vowel spaces in Plains Cree. *Journal of the International Phonetic Association*, *42*, 91-105. https://doi.org/10.1017/S0025100311000302

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

Olson, D. J. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania*, *97*(1), 47-68. https://doi.org/10.1353/hpn.2014.0030.

Onishi, H. (2016). The effects of L2 experience on L3 perception. *International Journal of Multilingualism*, *13*, 459-475. https://doi.org/10.1080/14790718.2016.1217604

Park, M. (2018). A Study on Factors Affecting the Perceptual Accuracy of L2 Vowels in an FL Context. *Studies in Linguistics*, 81-100. Retrieved from https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE07303340

Peperkamp, S., & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. In *Proceedings of the International Speech Communication Association*, Florence, Italy, 161-164. https://doi.org/10.21437/Interspeech.2011-72

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

Pratiwi, W. R. (2023). Investigating immersive environment's contributing factors for learners' metacognitive strategies in an Indonesian English village. *Lentera Pendidikan: Jurnal Ilmu Tarbiyah dan Keguruan*, *26*(1), 200-215. https://doi.org/10.24252/lp.2023v26n1i14

Rabinovitch, M. S., & Parver, L. M. (1966, June). Auditory discrimination in monolinguals and polyglots. Paper presented at the meeting of the Canadian Psychological Association, Montreal, Canada.

Reinisch, E., & Sjerps, M.J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics*, *41*, 101-116. https://doi.org/10.1016/j.wocn.2013.01.002

Rosenhouse, J., Amir, N., & Amir, O. (2014). Vowel systems of quantity languages compared: Arabic dialects and other languages. *Proceedings of Meetings on Acoustics*, *21*(1), 1-10. https://doi.org/10.1121/1.4880205

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187-224. https://doi.org/10.1017/S0142716417000418

Saloranta, A., Alku, P., & Peltola, M. S. (2020). Listen-and-repeat training improves perception of second language vowel duration: Evidence from mismatch negativity (MMN) and N1 responses and behavioral discrimination. *International Journal of Psychophysiology*, *147*, 72-82. https://doi.org/10.1016/j.ijpsycho.2019.11.005

Sanz, C. (2000). Bilingual education enhances third language acquisition: Evidence from Catalonia. *Applied psycholinguistics*, *21*(1), 23-44. https://doi.org/10.1017/S0142716400001028

Schoonmaker-Gates, E. (2024). Dialect-specificity and learners' L2 perceptual resilience after study abroad. Journal of Second Language Pronunciation. *Journal of Second Language Pronunciation*, *9*(3), 362-383. https://doi.org/10.1075/jslp.23004.sch

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

Stevens, J. J. (2011). Vowel duration in second language Spanish vowels: Study abroad versus at-home learners. *Journal of Second Language Acquisition and Teaching*, *18*, 77-103. Retrieved from https://journals.uair.arizona.edu/index.php/AZSLAT/article/view/21234

Tomaschek, F., Truckenbrodt, H., & Hertrich, I. (2011). Processing German vowel quantity: Categorical perception or perceptual magnet effect? In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 2002-2005). Retrieved from https://web.archive.org/web/20220430075004id_/https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Tomaschek/Tomaschek.pdf

van Son, R. J., & Pols, L. C. (1999). Perisegmental speech improves consonant and vowel identification. *Speech Communication*, *29*(1), 1-22. https://doi.org/10.1016/S0167-6393(99)00024-2

Vasetenkov, A. (n.d.). IPA Reader. Retrieved from https://www.antvaset.com/ipa-to-speech

Vouloumanos, A. (2015). Speech Perception. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 1-7. https://doi.org/10.1002/9781118900772.etrds0315

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). Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_0125.pdf

Wilson, A., Kato, H., & Tajima, K. (2005). Native and non-native perception of phonemic length contrasts in Japanese: Effects of speaking rate and presentation context. *The Journal of the Acoustical Society of America*, *117*, 2425-2425. https://doi.org/10.1121/1.4809405

Wu, J. (2022, October). The perception of L3 Quebec French (QF) tense and lax vowels contrasts by L1 Mandarin-L2 English learners. In *International Symposium on Monolingual and Bilingual Speech 2022* (p. 85). Retrieved from https://www.researchgate.net/publication/364305810

Yamada, R., Strange, W., Magnuson, J., Pruitt, J. & Clark, W. III. (1994). The intelligibility of Japanese speakers' production of American English /r/, /l/, and /w/, as evaluated by native speakers of American English. *International Conference on Spoken Language Processing*, *94*, 2023-2026. https://doi.org/10.21437/ICSLP.1994-508

Yazawa, K., Whang, J., & Escudero, P. (2023). Australian English listeners' perception of Japanese vowel length reveals underlying phonological knowledge. *Frontiers in Psychology*, *14*, 1-14. https://doi.org/10.3389/fpsyg.2023.1122471

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

Zimmerer, F., & Trouvain, J. (2015). Perception of French speakers' German vowels. *Proc. Interspeech*, Dresden, pp. 1720-1724. https://doi.org/10.21437/Interspeech.2015-395

Appendix 1 CUREC Approval Letter

Subject: CUREC 1B: Perception and Categorisation of Vowels in German and English by Chinese Learners

From: Student CUREC <student.curec@education.ox.ac.uk>

Dear Yihan,

I am writing to acknowledge receipt of your CUREC 1B application entitled 'Perception and Categorisation of Vowels in German and English by Chinese Learners'. The application was reviewed and approved by Dr Faidra Faitaki, your supervisor. No further approval from the Education DREC is required for applications reviewed under the CUREC 1B process. As such, the project will not receive a formal letter of ethical approval from the SSH IDREC.

The ethics reference for your application is C1B-24HT-Educ-020. Please add this reference to your CUREC 1B form and include it on documents for the research participants such as the participant information sheet.

Please note that this is contingent on the research project adhering to the criteria set out in the CUREC 1B guidance. Please ensure, therefore, that you comply with the conditions of this process and, should anything change in the course of the project, you should discuss this with your supervisor to determine whether this requires further review and approval by the Education DREC.

Please don't hesitate to get in touch if you have any questions.

All the best for your research – we hope it goes well.

Appendix 2 Information Sheet and Consent Form

General Information

The aim of this research is to investigate how learners with different language backgrounds differ when perceiving German and English vowels and whether these differences correlate with the amount of their linguistic experience.

We appreciate your interest in participating in this online task. You have been invited to participate as you are a Chinese adult learner in English (and German) at university level. Please read through this information before agreeing to participate (if you wish to) by ticking the 'yes' box below. You may ask any questions before deciding to take part by contacting the researcher (details below).

You will be asked to fill in a questionnaire on your background and language learning experience, to complete a pre-test on language proficiency, and to take a perceptual test on some vowels in English and German. The whole process should take about 30 minutes. No background knowledge is required. Your answers to the perceptual test will be collected. We will analyse these data to explore perceptual patterns from participants with different language backgrounds. The research data will be fully anonymous, and stored according to Oxford University guidelines.

Do I have to take part?

No. Please note that participation is voluntary. If you do decide to take part, you may withdraw at any point for any reason before submitting your answers by pressing the 'Exit' button/ closing the browser. However, participants who want to enter the lottery for a chance to win £50 vouchers (or an equivalent amount) will be given the option to provide their email address at the end of the experiment.

How will my data be used?

All data will be completely anonymous.

Your IP address will not be stored. We will take all reasonable measures to ensure that data remain confidential. The responses you provide will be stored in a password-protected electronic file on University of Oxford secure servers and may be used in academic publications and conference presentations. If you do provide your email address in order to participate in the lottery, note that it will be stored securely in the main researcher's protected university account and deleted as soon as it is no longer required for the research.

Research data will be stored for 3 years after publication or public release of the work of the research.

The data that we collect from you may be transferred to, stored and/ or processed at a destination outside the UK and the European Economic Area. By submitting your personal data, you agree to this transfer, storing or processing.

Who will have access to my data?

The University of Oxford is the data controller with respect to your personal data and, as such, will determine how your personal data is used in the research. The University will process your personal data for the purpose of the research outlined above. Research is a task that we perform in the public interest. Further information about your rights with respect to your personal data is available from https://compliance.admin.ox.ac.uk/individual-rights.

The data you provide may be shared with Gorilla, a software for online behavioural experiments.

The results will be written up for an MSc degree.

Who has reviewed this research?

The application was reviewed and approved by my supervisor on behalf of the Department of Education's Research Ethics Committee. (Ethics reference: C1B-24HT-Educ-020)

Who do I contact if I have a concern or I wish to complain?

If you have a concern about any aspect of this research, please speak to the principle researcher (yihan.li@st-hughs.ox.ac.uk), or the supervisor (faidra.faitaki@education.ox.ac.uk), and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how itwill be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

Education Departmental Research Ethics Committee (DREC);

Email: student.curec@education.ox.ac.uk; Address: Department of Education, University of Oxford, 15 Norham Gardens, Oxford OX2 6PY

Please note that you may only participate in this survey if you are 18 years of age or over.

If you have read the information above and agree to participate with the understanding that the data (including any personal data) you submit will be processed accordingly, please tick the box below to start.

基本信息

本研究旨在调查不同语言背景的学习者在感知德语和英语元音时的差异，以及这些差异是否与他们的语言背景相关。

非常感谢您对本在线测试的兴趣。本项测试将关注大学阶段学习英语（和德语）的成年中文母语者，如果您您满足此项，我们诚挚邀请您参与测试。请仔细阅读此信息，如果您想继续参与测试，请在最下方的 “是” 框内进行勾选。在决定参与之前，如果您有任何问题，请联系研究人员（详情见下文）。

您将被要求填写一份关于您的背景和语言学习经历的调查问卷，完成一项语言水平预测试，并参加一项关于英语和德语元音的感知测试。整个过程大约需要 30 分钟，且无需任何背景知识。我们将收集您在感知测试中的答案，将对这些数据进行分析，以探索不同语言背景的参与者的感知模式。研究数据将完全匿名，并按照牛津大学的指导原则进行存储。

我必须参与吗？

不是，参与是自愿的。如果您决定参加，您可以在提交答案之前的任何时候以任何理由退出，只要按下 "退出 "按钮或关闭浏览器即可，但仅有完成测试的参与者有机会在测试后的抽奖中获得赢得 50 英镑代金券（或同等金额报酬）。参与者仅需在实验结束时提供自己的电子邮件地址参与抽奖。

我的数据将如何被使用？

所有数据都是完全匿名的。

您的 IP 地址不会被保存。我们将采取一切合理措施确保数据保密。您提供的回复将被保存在牛津大学安全服务器上受密码保护的电子文档中，并可能用于学术出版物和会议演示。如果您为了参加抽奖而提供了自己的电子邮件地址，该地址将被安全地保存在主要研究人员受保护的大学账户中，一旦研究不再需要该地址，将立即被删除。

研究数据将在研究成果发表或公开发布后保存 3 年。

我们向您收集的数据可能会被传输、存储和/或处理到英国和欧洲经济区以外的目的地。提交您的个人数据即表示您同意这种传输、存储或处理。

谁有权访问我的数据？

牛津大学是您个人数据的数据控制者，因此将决定在研究中如何使用您的个人数据。牛津大学将出于上述研究目的处理您的个人数据。研究是我们为公共利益而执行的任务。有关您个人数据权利的更多信息，请访问 https://compliance.admin.ox.ac.uk/individual-rights。

您提供的数据可能会与在线行为实验软件 Gorilla 共享。

研究结果将撰写成硕士学位论文。

谁审查了本项研究？

我的导师代表教育系研究伦理委员会审查并批准了该申请。(伦理编号：C1B-24HT-Educ-020）

如果我有疑虑或想投诉，该与谁联系？

如果您对本研究的任何方面有疑虑，请与主要研究人员(yihan.li@st-hughs.ox.ac.uk)或其导师(faidra.faitaki@education.ox.ac.uk)联系，我们将尽力回答您的问题。我们会在 10 个工作日内确认您的问题，并向您说明将如何处理。如果您仍然不满意或希望提出正式投诉，请联系牛津大学研究伦理委员会主席，他们会尽快解决此事：

Education Departmental Research Ethics Committee (DREC);

Email: student.curec@education.ox.ac.uk; Address: Department of Education, University of Oxford, 15 Norham Gardens, Oxford OX2 6PY

请您注意，只有年满 18 周岁方可参与本调查

如果您已阅读上述信息并同意参与调查，同时了解您提交的数据（包括所有个人数据）也将得到相应处理，请勾选以下方框开始测试。

I certify that I am 18 years of age or over 我证明我已年满 18 岁

Yes, I agree to take part 是，我同意参加

Appendix 3 Recruitment Poster

DEPARTMENT OF EDUCATION
University of Oxford
15 Norham Gardens
Oxford, OX2 6PY

PARTICIPANTS NEEDED FOR VOWEL PERCEPTION TESTS!

Perception and categorisation of vowels in German and English by Chinese university learners
Ethics Approval Reference: [C1B-24HT-Educ-020]

Vowel perception is a significant challenge in phonological language acquisition. Many empirical studies have shown that language learning experience and prior language knowledge can influence learners' L2/L3 phonological perception. In order to further explore the cross-linguistic influence in this regard, the present study focuses on the perception of long and short vowel contrast and vowel categorisation by Chinese learners.

We are looking for **Chinese university students who are 1) learning English as their L2, or 2) learning English as their L2 and German as their L3**, to conduct a vowel perception test. You are invited to participate in an online study, which would take under 30 minutes of your time.

Eligible participants would be asked to fill in a questionnaire on your background and language learning experience, to complete a pre-test on language proficiency, and to complete a perceptual test on some vowels in English and German.

Participants who finish the experiment will be given the option to provide their email address at the end of the experiment to enter the lottery for a chance to win a **£50** voucher (or an equivalent amount).

You are unfortunately **not eligible** to participate if you:

1. Have learnt languages other than Chinese, English and German
2. Are a bilingual in any two of the languages studied
3. Have any hearing impairment

If you are interested and would like more information please contact yihan.li@st-hughs.ox.ac.uk.

Thank you!

Appendix 4 Questionnaire for English-only Learners

1. Your Age 你的年龄

2. Your Grade 你的年级

    c Freshman 大一

    Sophomore 大二

    Junior 大三

    Senior 大四

3. Is your mother tongue Chinese? 你的母语是汉语吗？

    Yes 是

    No 不是

4. Are you a monolingual? 你是单语母语者吗？

    Yes 是

    No 不是

5. Have you ever had any kind of training in phonetics or phonology before? If yes, please specify what you have been taught. 您以前是否接受过任何形式的语音学或音韵学训练？如果回答为 "是"，请说明您接受过哪些训练。

6. When did you start to learn English? 你是什么时候开始学习英语的？

    Primary school 小学

    Middle school 初中

    High school 高中

    University 大学

7. What is your score for CET-4? 你的大学英语四级考试成绩为？

8. What is your score for listening section in CET-4? 你在大学英语四级考试中的听力部分得分为

9. What is your score for CET-6? 你的大学英语六级考试成绩为?

10. What is your score for listening section in CET-6? 你在大学英语六级考试中的听力部分得分为?

11. Have you ever lived or studied in an English-speaking country? If yes, in which country and for how long? 你曾在英语国家生活或学习过吗？如果回答为 "是"，那么是在哪个国家以及生活了多久?

Appendix 5 Questionnaire for English and German Learners

1. Your Age 你的年龄

2. Your Grade 你的年级

    Freshman 大一

    Sophomore 大二

    Junior 大三

    Senior 大四

3. Is your mother tongue Chinese? 你的母语是汉语吗？

    Yes 是

    No 不是

4. Are you a monolingual? 你是单语母语者吗？

    Yes 是

    No 不是

5. Have you ever had any kind of training in phonetics or phonology before? If yes, please specify what you have been taught. 您以前是否接受过任何形式的语音学或音韵学训练？如果回答为 "是"，请说明您接受过哪些训练。

6. When did you start to learn English? 你是什么时候开始学习英语的？

    Primary school 小学

    Middle school 初中

    High school 高中

    University 大学

7. What is your score for CET-4? 你的大学英语四级考试成绩为？

8. What is your score for listening section in CET-4? 你在大学英语四级考试中的听力部分得分为

9. What is your score for CET-6? 你的大学英语六级考试成绩为？

10. What is your score for listening section in CET-6? 你在大学英语六级考试中的听力部分得分为？

11. Have you ever lived or studied in an English-speaking country? If yes, in which country and for how long? 你曾在英语国家生活或学习过吗？如果回答为 "是"，那么是在哪个国家以及生活了多久？

12. When did you start to learn German? 你是什么时候开始学习德语的？

Primary school 小学

Middle school 初中

High school 高中

University 大学

13. Have you taken any of the following German language tests? 你是否参加过以下德语语言测试？

PGG (Prüfung für das Germanistik-Grundstudium)

PGH (Prüfung für das Germanistik-Hauptstudium)

Test DaF (Test Deutsch als Fremdsprache)

Goethe Institute Deutsch-Prüfungen A1-C2

PHD-4

PHD-6

I haven’t taken part in any of the tests 我没有参加过这些测试中的任何一个

Other - please specify:

14. If you have taken any kind of standardised German language test (including the tests mentioned above and others specified in the 'other' option), please enter your score. And if the test has a listening test that is marked separately, please also enter your listening

score (the overall and listening scores can be separated by a "/"). 如果您参加过任何形式的标准化德语测试（包括上述测试和 "其他 "选项中指定的其他测试），请输入您的分数。如果该考试有单独评分的听力测试，也请输入您的听力分数（总分和听力分数之间可以用"/"隔开）。

15. Have you ever lived or studied in an German-speaking country? If yes, in which country and for how long? 你曾在德语国家生活或学习过吗？如果回答为 "是"，那么是在哪个国家以及生活了多久？

Appendix 6 Manipulation of Stimuli in Task 2

First, the formant values of the original sound recordings were analysed to obtain the formant data. The formant frequencies were then adjusted manually using the Manipulation Editor in PRAAT. Within the manipulation window, 10 formant control points were placed along the F1 and F2 tracks to allow for precise and detailed adjustments over time. After positioning these points, the formant values at each point were extracted, and the mean values of F1 and F2 across the two original sounds were calculated to generate an intermediate sound variant. Based on these mean values, the formants were systematically adjusted to create the desired intermediate variant, as detailed in the tables below.

F1 values for task 2 stimuli at 10 control points (cp)

	cp 1	cp 2	cp 3	cp 4	cp 5	cp 6	cp 7	cp 8	cp 9	cp 10
/ɪ/	273	378	415	415	412	401	259	184	854	797
Intermediate	422	323	355	366	371	359	295	262	557	1384
/i:/	570	267	295	317	329	316	331	340	260	1971

F2 values for task 2 stimuli at 10 control points (cp)

	cp 1	cp 2	cp 3	cp 4	cp 5	cp 6	cp 7	cp 8	cp 9	cp 10
/ɪ/	2143	2096	2112	2188	2140	2058	2033	1934	1943	1923
Intermediate	2287	2398	2480	2563	2478	2455	2418	2300	2144	2269
/i:/	2431	2699	2847	2937	2816	2852	2803	2665	2344	2615

Appendix 7 Manipulation of Stimuli in Task 3

The original vowels used in Task 3 were first analysed. The formant values of the two intermediate variants were calculated, resulting in four sounds for each continuum (/ɛ/ to /ø/ and /ɛ/ to /æ/). The formant values of the two variants were manually manipulated in the same way as described in Task 2, ensuring that the differences between each pair of adjacent variants were equal. The tables below provide detailed information on the formant values at 10 control points used to create the variants.

F1 values for /ɛ/ to /ø/ stimuli at 10 control points

	cp 1	cp 2	cp 3	cp 4	cp 5	cp 6	cp 7	cp 8	cp 9	cp 10
/ɛ/	599	439	502	536	538	515	483	449	385	457
Variant 1	759	415	473	497	491	476	454	428	363	457
Variant 2	919	391	443	457	443	437	425	406	342	458
/ø/	1079	367	414	418	396	398	396	385	320	458

F2 values for /ɛ/ to /ø/ stimuli at 10 control points

	cp 1	cp 2	cp 3	cp 4	cp 5	cp 6	cp 7	cp 8	cp 9	cp 10
/ɛ/	2054	1958	2007	2059	2070	2058	2059	2056	2032	2158
Variant 1	2148	1849	1877	1907	1922	1938	1984	2034	2017	2189
Variant 2	2241	1739	1748	1754	1774	1819	1910	2013	2002	2221
/ø/	2335	1630	1618	1602	1626	1699	1835	1991	1987	2252

F1 values for /ɛ/ to /æ/ stimuli at 10 control points

	cp 1	cp 2	cp 3	cp 4	cp 5	cp 6	cp 7	cp 8	cp 9	cp 10
/ɛ/	752	790	788	793	783	762	716	615	478	337
Variant 1	772	828	826	834	819	804	736	648	512	382
Variant 2	792	865	864	876	854	845	755	680	546	426
/æ/	812	903	902	917	890	887	775	713	580	471

F2 values for /ɛ/ to /æ/ stimuli at 10 control points

	cp 1	cp 2	cp 3	cp 4	cp 5	cp 6	cp 7	cp 8	cp 9	cp 10
/ɛ/	2044	2065	2056	2051	2005	2002	1889	1997	2232	2294
Variant 1	2003	2028	2000	1983	1971	1949	1742	2036	2191	2291
Variant 2	1961	1992	1945	1915	1938	1897	1596	2076	2151	2288
/æ/	1920	1955	1889	1847	1904	1844	1449	2115	2110	2285

Appendix 8 Results of Chi-square tests

Results for Task 1

Chi-Square Analysis of the Relationship Between Branch of Study and Learner Performance in Task 1

Tested Variables	Chi-square Value (χ²)	df	P-value
Set 1 - 75 ms	2.163	1	.141
Set 1 - 100 ms	.672	1	.413
Set 1 - 125 ms	.003	1	.954
Set 1 - 150 ms	.037	1	.847
Set 1 - 175 ms	6.072	1	.014
Set 1 - 200 ms	6.72	1	.413
Set 1 - 225 ms	2.163	1	.141
Set 2 - 75 ms	.471	1	.493
Set 2 - 100 ms	1.439	1	.230
Set 2 - 125 ms	2.883	1	.089
Set 2 - 150 ms	12.419	1	<.001
Set 2 - 175 ms	.128	1	.720
Set 2 - 200 ms	.471	1	.493
Set 2 - 225 ms	.471	1	.493

Chi-Square Analysis of the Relationship Between Prior Phonetic Traning and Learner Performance in Task 1

Tested Variables	Chi-square Value (χ²)	df	P-value
Set 1 - 75 ms	1.085	1	.298
Set 1 - 100 ms	.869	1	.351
Set 1 - 125 ms	.272	1	.602
Set 1 - 150 ms	2.769	1	.096
Set 1 - 175 ms	3.211	1	.073
Set 1 - 200 ms	.560	1	.454
Set 1 - 225 ms	1.085	1	.298
Set 2 - 75 ms	.938	1	.333
Set 2 - 100 ms	.272	1	.602
Set 2 - 125 ms	6.320	1	.012
Set 2 - 150 ms	4.230	1	.040
Set 2 - 175 ms	.708	1	.400
Set 2 - 200 ms	.938	1	.333
Set 2 - 225 ms	1.085	1	.298

Chi-Square Analysis of the Relationship Between Learning Environment and Learner Performance in Task 1

Tested Variables	Chi-square Value (χ^2)	df	P-value
Set 1 - 75 ms	.471	1	.493
Set 1 - 100 ms	.672	1	.413
Set 1 - 125 ms	1.439	1	.230
Set 1 - 150 ms	.114	1	.736
Set 1 - 175 ms	.716	1	.397
Set 1 - 200 ms	2.962	1	.085
Set 1 - 225 ms	.471	1	.493
Set 2 - 75 ms	2.163	1	.141
Set 2 - 100 ms	1.726	1	.189
Set 2 - 125 ms	.144	1	.704
Set 2 - 150 ms	3.816	1	.051
Set 2 - 175 ms	.185	1	.667
Set 2 - 200 ms	.471	1	.493
Set 2 - 225 ms	2.163	1	.141

Results for Task 2

Chi-Square Analysis of the Relationship Between Branch of Study and Learner

Tested Variables	Chi-square Value (χ^2)	df	P-value
50ms-1	.054	1	.817
50ms-2	.364	1	.547
50ms-3	.128	1	.720
100ms-1	.000	1	1.000
100ms-2	1.955	1	.162
100ms-3	.153	1	.696
1fd-1	.310	1	.577
1fd-2	.416	1	.519
1fd-3	.018	1	.892
2fd-1	.163	1	.686
2fd-2	.403	1	.526
2fd-3	3.560	1	.059
1fd-50ms-1	.007	1	.931
1fd-50ms-2	1.048	1	.306
1fd-100ms-1	.912	1	.340
1fd-100ms-2	1.584	1	.208
2fd-50ms-1	.099	1	.752
2fd-50ms-2	.376	1	.540
2fd-100ms-1	.717	1	.397
2fd-100ms-2	3.725	1	.054

Performance in Task 2

Chi-Square Analysis of the Relationship Between Prior Phonetic Traning and Learner Performance in Task 2

Tested Variables	Chi-square Value (χ^2)	df	P-value
50ms-1	1.338	1	.247
50ms-2	9.658	1	.002
50ms-3	.011	1	.915
100ms-1	9.321	1	.002
100ms-2	8.392	1	.004
100ms-3	6.410	1	.011
1fd-1	.003	1	.959
1fd-2	.866	1	.352
1fd-3	3.220	1	.073
2fd-1	1.182	1	.277
2fd-2	3.123	1	.077
2fd-3	12.296	1	<.001
1fd-50ms-1	1.790	1	.181
1fd-50ms-2	.036	1	.849
1fd-100ms-1	.792	1	.373
1fd-100ms-2	.003	1	.955
2fd-50ms-1	1.403	1	.236
2fd-50ms-2	1.288	1	.256
2fd-100ms-1	1.875	1	.171
2fd-100ms-2	2.769	1	.096

Chi-Square Analysis of the Relationship Between Learning Environment and Learner Performance in Task 2

Tested Variables	Chi-square Value (χ^2)	df	P-value
50ms-1	2.153	1	.142
50ms-2	3.859	1	.059
50ms-3	.128	1	.720
100ms-1	2.357	1	.125
100ms-2	1.443	1	.230
100ms-3	.153	1	.696
1fd-1	.310	1	.577
1fd-2	2.546	1	.111
1fd-3	3.706	1	.054
2fd-1	1.565	1	.211
2fd-2	.455	1	.500
2fd-3	3.859	1	.054
1fd-50ms-1	3.297	1	.069
1fd-50ms-2	.262	1	.609

1fd-100ms-1	3.731	1	.053
1fd-100ms-2	.004	1	.952
2fd-50ms-1	.457	1	.499
2fd-50ms-2	.693	1	.405
2fd-100ms-1	.717	1	.397
2fd-100ms-2	.037	1	.847

Results for Task 3

Chi-Square Analysis of the Relationship Between Branch of Study and Learner Performance in Task 3

Tested Variables	Chi-square Value (χ^2)	df	P-value
English-1fd-1	.339	1	.560
English-1fd-2*	-	-	-
English-1fd-3	.337	1	.561
English-2fd-1	.720	1	.396
English-2fd-2	.162	1	.688
English-3fd	.057	1	.812
German-1fd-1	2.163	1	.141
German-1fd-2	1.765	1	.184
German-1fd-3	.003	1	.954
German-2fd-1	.377	1	.539
German-2fd-2	3.696	1	.055
German-3fd	.000	1	.985

*No statistics are computed because the variable English-1fd-2 is a constant.

Chi-Square Analysis of the Relationship Between Prior Phonetic Traning and Learner Performance in Task 3

Tested Variables	Chi-square Value (χ^2)	df	P-value
English-1fd-1	.295	1	.587
English-1fd-2*	-	-	-
English-1fd-3	1.666	1	.197
English-2fd-1	.866	1	.352
English-2fd-2	.293	1	.588
English-3fd	.299	1	.584
German-1fd-1	1.085	1	.298
German-1fd-2	.185	1	.667
German-1fd-3	.422	1	.516
German-2fd-1	1.784	1	.182
German-2fd-2	.046	1	.830
German-3fd	4.266	1	.039

*No statistics are computed because the variable English-1fd-2 is a constant.

Chi-Square Analysis of the Relationship Between Learning Environment and Learner Performance in Task 3

Tested Variables	Chi-square Value (χ^2)	df	P-value
English-1fd-1	1.676	1	.195
English-1fd-2*	-	-	-
English-1fd-3	.337	1	.561
English-2fd-1	5.487	1	.019
English-2fd-2	.162	1	.688
English-3fd	.455	1	.500
German-1fd-1	.471	1	.493
German-1fd-2	6.358	1	.012
German-1fd-3	1.439	1	.230
German-2fd-1	2.053	1	.152
German-2fd-2	2.110	1	.146
German-3fd	1.495	1	.221

*No statistics are computed because the variable English-1fd-2 is a constant.

Appendix 9 Results of Repeated Measures ANOVA

Results for Task 2

Within-Subjects Factor

Factors	Dependent Variables	Description
Cues	Task2_dd	Number of perceived durational differences in Task 2
	Task2_fd	Number of perceived formant differences in Task 2
	Task2_md	Number of perceived mixed differences in Task 2 (corrected)

Descriptive Statistics

Cues	Mean	Std. Error	N
Task2_dd	2.724	.191	110
Task2_fd	1.293	.210	110
Task2_md	2.869	.162	110

Multivariate Tests

Effect	Value	F	Hypothesis df	Error df	Sig.
Cues	.628	29.265	2	99	<.001

Tests of Within-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Cues	7.539	2	3.769	4.495	.012	.043

Pairwise Comparisons

(I) Cues	(J) Cues	Mean Difference (I-J)	Std. Error	Sig.
Task2_dd	Task2_fd	1.432	.246	<.001
Task2_fd	Task2_md	-1.576	.210	<.001
Task2_md	Task2_dd	.144	.205	1.000

Between-Subjects Factor

Factors	Value	Value Label	N
Branch	0	English-only	35
	1	German and English	75
Prior phonetic knowledge	0	No prior knowledge	53
	1	With prior knowledge	57
Abroad	0	No experience of studying abroad	75
	1	Have experience of studying abroad	35

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	9.922	1	9.922	4.968	.028	.047
Pre_Eng	3.590	1	3.590	1.797	.183	.018
Pre_Ger	2.755	1	2.755	1.379	.243	.014
Branch	5.132	1	5.132	2.569	.112	.025
Prior Knowledge	19.670	1	19.670	9.848	.002	.090
Abroad	12.140	1	12.140	6.078	.015	.057
Branch * Prior Knowledge	2.479	1	2.479	1.241	.268	.012
Branch * Abroad	8.044	1	8.044	4.028	.047	.039
Prior Knowledge * Abroad	.565	1	.565	.283	.596	.003
Branch * Prior Knowledge * Abroad	.338	1	.338	.169	.682	.002
Error	199.730	100	1.997			

Results for Task 3

Within-Subjects Factor

Factors	Dependent Variables	Description
Cues	Task3_English	Number of perceived formant differences of English vowel contrasts in Task 3
	Task3_German	Number of perceived formant differences of German vowel contrasts in Task 3

Descriptive Statistics

Cues	Mean	Std. Error	N
Task3_English	.682	.133	110
Task3_German	1.640	.197	110

Multivariate Tests

Effect	Value	F	Hypothesis df	Error df	Sig.
Cues	.843	18.554	1	100	<.001

Tests of Within-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Cues	10.364	1	10.364	12.185	<.001	.109

Pairwise Comparisons

(I) Cues	(J) Cues	Mean Difference (I-J)	Std. Error	Sig.
Task3_English	Task3_German	-.958	.222	<001

Between-Subjects Factor

Factors	Value	Value Label	N
Branch	0	English-only	35
	1	German and English	75
Prior phonetic knowledge	0	No prior knowledge	53
	1	With prior knowledge	57
Abroad	0	No experience of studying abroad	75
	1	Have experience of studying abroad	35

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	5.099	1	5.099	4.696	.033	.045
Pre_Eng	.042	1	.042	.038	.845	.000
Pre_Ger	.041	1	.041	.037	.847	.000
Branch	1.248	1	1.248	1.150	.286	.001
Prior Knowledge	1.482	1	1.482	1.365	.245	.013
Abroad	8.096	1	8.096	7.456	.007	.069
Branch * Prior Knowledge	.536	1	.536	.494	.484	.005
Branch * Abroad	2.794	1	2.794	2.573	.112	.025
Prior Knowledge * Abroad	.369	1	.369	.340	.561	.003
Branch * Prior Knowledge * Abroad	.294	1	.294	.271	.604	.003
Error	108.583	100	1.086			