Jilid 6, Isu 2 (2025)
e-ISSN: 2735-2234

Jurnal EVOLUSI

Please cite this article as: Raja Abdullah, R. R. (2025). English Monophthong: Assessing Malay Young Learners' Perception Accuracy . *Jurnal Evolusi* , *6*(2), 1–10. https://doi.org/10.61688/jev.v6i2.445

ENGLISH MONOPHTHONG: ASSESSING MALAY YOUNG LEARNERS' PERCEPTION ACCURACY

Raja Rosemawati binti Raja Abdullah

Institute of Graduate Studies, Universiti Poly-Tech Malaysia

Corresponding author: rosemawati@uptm.edu.my
Received 29 September 2024, Accepted xx June 2024, Available online xx May 2023

ABSTRACT

Intelligibility, the ability to differentiate speech sounds, is crucial for effective communication. While communicating in one's native language may be straightforward, challenges arise when speaking in a second language like English. This study investigates the accuracy of Malay young learners in Malaysia when distinguishing between three English monophthong pairs: /ɪ/ - /iː/, /ʌ/ - /ɑː/, and /ʊ/ - /uː/. Ten 11-year-old students from a semi-urban primary school in Kuala Lumpur participated in a quantitative research study using a descriptive-comparative design. Participants were selected through a Language Profile Test based on Birdsong et al. (2012) and completed a two-alternative forced-choice (2AFC) test using PsychoPy software. Each targeted vowel was presented in blocks and repeated three times. Results indicated that participants could distinguish short and long vowels with an overall accuracy of 73.9%. Notably, they performed best with the pair /ʊ/-/uː/ (98.3%), while distinguishing /ɪ/-/iː/ yielded only 53.3% accuracy, and /ʌ/-/ɑː/ achieved 70%. Additionally, participants showed better performance with short vowels (75.6% accuracy) compared to long vowels (72.7% accuracy), although the difference was not statistically significant ($t(89) = -.555$, $p = .580$). Overall, the findings demonstrate that participants possess a high level of intelligibility in perceiving and differentiating the targeted English monophthongs.

Keywords: *English in Malaysia, Malay young learners, English Monophthong, Accuracy Perception*

INTRODUCTION

English pronunciation remains a persistent challenge for Malaysian learners of English as a second language (ESL), despite an average of eleven years of formal English education. Numerous studies highlight that Malaysian students frequently excel in written examinations yet struggle with practical language application, particularly in oral communication and pronunciation (Kashinathan & Azlina, 2021; Raja Rosemawati, 2024). Mispronunciations create communication barriers, leading to misunderstandings, reduced confidence, and avoidance of English use (Cebrian et al., 2021; Yahaya et al., 2021). This gap between academic achievement and communicative competence is exacerbated by educational practices that prioritise examination performance over oral proficiency, contributing to graduates' limited employability in contexts where accurate and intelligible English speech is increasingly valued (Azhar et al., 2025; Raja Rosemawati, 2024). Psychological factors such as anxiety, low self-esteem, and fear of negative evaluation further hinder pronunciation development (Yahaya et al., 2021).

Within this broader context, the accurate perception and production of English vowels, particularly monophthongs, pose significant difficulties for Malay learners. Prior research shows that Malay speakers frequently mispronounce English vowels and diphthongs, often substituting or assimilating sounds absent from the Malay phonological system (Bello et al., 2020; Kamarudin & Ahmad Kamal, 2021). For example, Malaysian ESL speakers exhibit limited phonetic distance in high vowel pairs such as /ɪ–iː/ and /u–uː/, while showing variable success in other contrasts (Bello et al., 2020). Similar cross-linguistic influence has been observed in multilingual contexts, where Malay speakers learning German tend to collapse unfamiliar vowels to neighbouring L1 or L2 categories (Yunus & Pillai, 2020). These findings align with second language (L2) phonetic research demonstrating that non-native listeners rely on perceptual cue weighting strategies shaped by their first language (L1) phonemic inventory, which can hinder accurate discrimination of non-native vowel contrasts (Raja Rosemawati & Chiew, 2024).

The present study focuses on three English monophthong pairs, /iː/-/ɪ/, /ɒ/-/ɔː/, and /æ/-/ʌ/, which are known to be perceptually challenging for L2 learners due to their acoustic similarity and absence of direct equivalents in Malay phonology (Md. Jaharul Islam et al., 2023). Accurate perception of these contrasts is essential for achieving comprehensibility and intelligibility in English (Kabakoff et al., 2021; Yenkimaleki & van Heuven, 2021). While most previous studies have examined adult learners or production-based errors, there is limited research on the perceptual abilities of younger Malay learners, a demographic with greater neuroplasticity and potential for successful L2 phonetic acquisition (Chen & van de Weijer, 2022; Saloranta & Heikkola, 2023). Investigating this group provides valuable insights into the early stages of L2 phonetic development and opportunities for targeted intervention (Alshangiti et al., 2023).

Therefore, this study aims to evaluate the accuracy with which young Malay learners perceive and differentiate the selected English monophthong pairs. By comparing their performance with established benchmarks for native English speakers, the findings will contribute to a deeper understanding of how L1 phonological influence, educational practices, and developmental factors shape vowel perception in Malaysian ESL contexts. Such insights may inform more effective pronunciation pedagogy and early intervention strategies to enhance English intelligibility among young learners.

LITERATURE REVIEW

The Vowel System of Malay and English Language

Malay and English may appear to share similar vowel sets, but key structural differences remain. English contains around eleven monophthong vowels (/ɪ/, /ʊ/, /ʌ/, /ɒ/, /ə/, /e/, etc.) and five long vowels (/iː/, /uː/, /ɑː/, /ɔː/, /ɜː/), in addition to eight diphthongs: /aɪ/, /eɪ/, /əʊ/, /aʊ/, /eə/, /ɪə/, and /ɔɪ/ (Davenport & Hannahs, 2020; Roach, 2010) By contrast, Malay contains only six monophthongs (/i/, /u/, /a/, /ə/, /e/, /o/) and three diphthongs (/ai/, /au/, /oi/) (Ramli et al., 2020).

Recent research continues to show that Malay speakers' production of English vowels reveals systematic patterns of vowel merger and reduced contrasts compared to native English varieties. For example, Yap et al. (2010) found that Malay–

English bilinguals categorise English front vowels into only three contrastive categories, reflecting L1 patterns. Pillai et al. (2010) similarly demonstrated that Malaysian English vowels occupy a smaller vowel space than British English, with particularly weak contrasts in /iː–ɪ/, /e–æ/, and /ɔ–ɑː/ pairs. Kamarudin and Ahmad Kamal (2021) confirmed that Malay native speakers frequently mispronounce the monophthong /uː/ and diphthongs /əʊ/ and /ɪə/ due to assimilation and substitution of sounds absent from Malay phonology. The complexity increases for multilingual speakers: Acehnese–Indonesian learners, for instance, approximate English monophthongs to their existing vowel systems, producing English vowels with qualities resembling either Acehnese or Indonesian equivalents (Masykar et al., 2022). When the structure of a second language (L2) diverges from the first language (L1), learners experience persistent difficulty in perceiving or producing the target sounds (Raja Rosemawati, 2024). The absence of vowel-quality contrasts and length discrimination can therefore hinder effective communication.

Perceptual Assimilation Model (PAM)

The Perceptual Assimilation Model (PAM) and its second-language extension, PAM-L2, propose that the learner's L1 phonological categories shape the acquisition of an L2 sound system (Best et al., 2016; Millet et al., 2021; Tyler et al., 2014). According to PAM-L2, L2 phones are perceived through the filter of the native system and are assimilated into existing L1 categories based on their acoustic and articulatory similarity (Cebrian et al., 2021; Chen & van de Weijer, 2022). Mature learners can detect both similarities and discrepancies between native and non-native phones, but the degree of similarity determines whether discrimination is easy or difficult. When an L2 phoneme closely resembles an L1 phoneme, assimilation into the native category occurs, making discrimination challenging. Conversely, when an L2 phoneme differs significantly, the discrepancy is detected, allowing a new category to emerge (Raja Rosemawati, 2024).

Importantly, these assimilation patterns are not fixed; targeted exposure can shift perceptual mappings and improve discrimination (Cebrian et al., 2021). PAM identifies four assimilation patterns for non-native contrasts: Single Category (SC), where two L2 phones merge into one native category; Two Category (TC), where each L2 phone is mapped to a distinct L1 category; Category Goodness (CG), where both L2 phones map to the same L1 category but differ in perceived "goodness"; and Non-Assimilable (NA), where the L2 phones cannot be matched to any native category. These categories explain how non-native sounds are perceived and why perceptual similarity influences L2 sound learning, a framework confirmed by recent cross-linguistic studies (Cebrian et al., 2021; Chen & van de Weijer, 2022).

Research on English Vowel Perception

Foreign speakers often rely on duration cues when identifying English vowel minimal pairs, even when they do not produce apparent durational differences in their own speech (Ćavar et al., 2022; Md. Jaharul Islam et al., 2023; Yazawa et al., 2023). Recent work continues to show a positive relationship between the perceptual and temporal characteristics of vowels, as longer sounds are perceived and often produced with greater duration. Yazawa et al. (2020) demonstrated that Japanese learners of English rely heavily on duration when perceiving the /iː–ɪ/ contrast. However, they can shift toward spectral cues when explicitly primed to think in an "English mode." Ćavar et al. (2022) similarly found that learners whose L1 contains phonemic length distinctions (Croatian) use duration more strongly than learners without such distinctions (Polish), confirming the influence of native-language phonological structure.

Investigating the balance between duration and spectral cues, Md. Jaharul Islam et al. (2023) reported that Bangla speakers of English relied more on spectral (quality) cues than on duration when producing tense–lax vowel pairs, suggesting that L2 production may not always mirror perceptual strategies. Training studies show that classroom exposure improves the perception of duration contrasts while production lags, highlighting different developmental timelines for perception and production (Saloranta & Heikkola, 2023). The presence of a duration cue in the L1 system facilitates L2 vowel perception. For example, Croatian learners exploit duration more effectively than Polish learners (Ćavar et al., 2022). Similar L1 transfer effects appear elsewhere: Japanese speakers exhibit native-like spectral patterns only at higher proficiency, while duration patterns remain influenced by Japanese quantity distinctions (Yazawa et al., 2023). Cross-language assimilation further explains EFL learners' difficulties. Yang and Fox (2014) showed that Mandarin Chinese listeners assimilate English vowels to the closest native categories, relying on both duration and spectral cues; bilingual Chinese listeners performed more accurately, demonstrating the roles of experience and proficiency. More recent Malay data reveal similar

challenges: Malay speakers continue to assimilate English short–long contrasts to Malay categories, making it challenging to maintain a native-like distinction between /e/ and /ɛ/ (Yap et al., 2010). Overall, these findings reinforce that both L1 phonology and proficiency shape how L2 listeners weigh duration and spectral cues in perceiving English vowel contrasts.

METHODOLOGY

Participants

The participants for this research were 10 Year-5 female pupils from Sekolah Kebangsaan Desa Setapak, Kuala Lumpur. Participants were selected using the following criteria: i) Malay ethnicity, ii) English as a second language (ESL), iii) Year 5 level, and iv) Female gender. Only female students were included to control for the gender variable, as vocal-tract length and resonance differ between males and females, which can influence vowel formant frequencies and potentially affect perception or production results (Ćavar et al., 2022; van der Slik et al., 2015). A small, controlled sample size (n = 10) was chosen to manage the volume of acoustic data while maintaining experimental control (Saloranta & Heikkola, 2023). Ethical approval was obtained from the Bahagian Perancangan dan Penyelidikan Dasar Pendidikan (BPPDP), Ministry of Education Malaysia.

Methods

This study adopted a quantitative, descriptive–comparative design using survey and experimental techniques. The Language Profile Test was adapted from Birdsong et al. (2012) and updated following recent L2 background survey practices (Saito et al., 2020) to screen participants' linguistic histories. The questionnaire was prepared in Malay to ensure comprehension. As for the experimental technique, the Correct Identification Test is used. The perception task followed a Two-Alternative Forced Choice (2AFC) paradigm, adapted from Casillas (2015) and supported by recent phonetic perception studies showing its efficiency in reducing decision latency compared to ABX tasks (Flege, 1995; Saito & Poeteren, 2018). This psychophysical method is widely regarded as unbiased and accurate for vowel discrimination.

The stimuli (Table 1) were designed based on these criteria: i) familiar everyday objects, ii) concrete referents, and iii) inclusion of the target vowels within a CVC context. Instead of orthographic prompts, cartoon images were used to represent the words to minimise orthographic influence on auditory processing, consistent with findings that written word knowledge can bias speech perception (Macher et al., 2021).

Table 1:
Word List for Correct Identification Test

Short Vowel	Long Vowel
Bin	Bean
Book	Boot
Hut	Heart

Each target word was embedded in a carrier sentence:

_____, the word is _____

Recordings were produced by a highly proficient Malay speaker of English (PhD in English studies, >20 years of teaching experience) to reduce accent interference and ensure natural yet intelligible stimuli (Casillas & Simonet, 2016). Acoustic measurements of the target vowels, which include duration, F1, and F2, were obtained using Praat software (Boersma & Weenink, 2017) and are shown in Table 2. Three trained phoneticians independently rated the stimuli for perceptual reliability, confirming clear contrasts across vowel pairs.

Table 2:
The value of duration (msec), F1 and F2 (Bark) of the stimuli

Vowel	Duration (msec)		F1 (Bark)	F2 (Bark)
	Actual	Difference		
/ɪ/	54.75	106.62	3.739	13.645
/iː/	161.36		3.228	14.611
/ʊ/	93.56	136.94	4.207	8.795
/uː/	230.50		3.884	10.298
/ʌ/	60.40	206.24	7.818	11.296
/ɑː/	266.64		7.836	9.796

Stimuli were presented using PsychoPy v3 (Peirce, 2017), a platform widely used for phonetic perception experiments. The task comprised four parts: i) Practice trials (familiarisation with the interface), ii) /ɪ–iː/ contrast, iii) /ʌ–ɑː/ contrast, and iv) /ʊ–uː/ contrast. At the start of each trial, the carrier sentence appeared on the screen while the stimulus audio played. Two images representing the possible answers were displayed simultaneously at the bottom of the screen. Participants responded by pressing the LEFT or RIGHT key. The subsequent trial only appeared after a response was made, with a five-second break between each block to reduce fatigue. Each vowel contrast was repeated three times, producing 18 responses per participant (6 vowels × 3 repetitions). Across all 10 participants, a total of 180 tokens were collected.

FINDINGS AND DISCUSSION

Previous research has shown that L2 speakers often rely on duration cues when perceiving English vowel contrasts, aiding them in making accurate perceptual judgments even if their production does not reflect these distinctions (Arum Perwitasari et al., 2015; Casillas, 2015). This reliance on duration suggests that perceptual knowledge may develop independently of productive ability. Figure 1 presents the percentage of correct and incorrect identifications for each participant. Overall, the Malay primary pupils successfully distinguished English short and long vowels, achieving more than 50% correct identification across all contrasts. Participant P7 achieved the highest accuracy (94.4%), followed by P1 and P10, while P4 and P5 recorded the lowest (55.6%). P7 also showed perfect accuracy for short vowels (100%) and 88.9% for long vowels. Language background information from the Language Profile Test revealed that participants who self-rated higher in English listening fluency (e.g., P1 and P7) tended to perform better in the perception task. In contrast, those with lower self-ratings (e.g., P4 and P5) obtained weaker scores. This pattern supports previous findings that individual differences in exposure and proficiency influence L2 vowel perception (Cebrian et al., 2021; Yazawa et al., 2020).

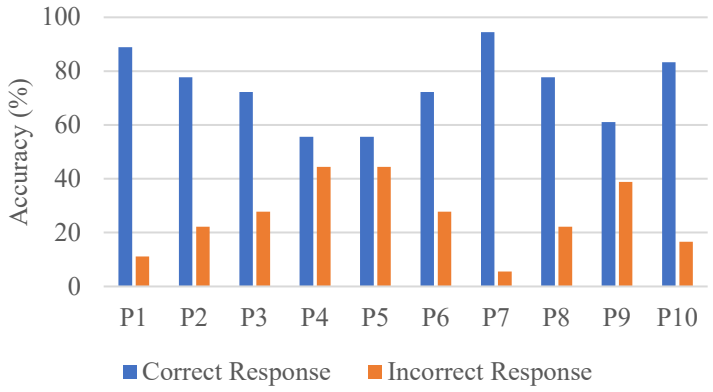

Figure 1
Percentage of Participants' Accuracy

A paired-sample t-test compared accuracy across vowel pairs (Figure 2). Participants identified the /ʊ/–/uː/ contrast most accurately (98.3%), followed by /ʌ/–/ɑː/ (70%), while /ɪ/–/iː/ was the most difficult (53.3%). The /ʊ/–/uː/ words (book vs. boot) were not strict minimal pairs, and the final consonants (/k/ vs. /t/) may have provided additional cues, facilitating perception. This aligns with evidence that redundant phonetic information can enhance discrimination even when duration differences are the primary cue (Ćavar et al., 2022; Saloranta & Heikkola, 2023). In contrast, the /ɪ/–/iː/ pair, with the smallest durational difference (≈106 ms), yielded no significant difference between correct and incorrect identification. Similar difficulty with close front vowels has been reported among other L2 groups, where limited spectral separation hampers the formation of distinct categories (Cebrian et al., 2021; Yazawa et al., 2020).

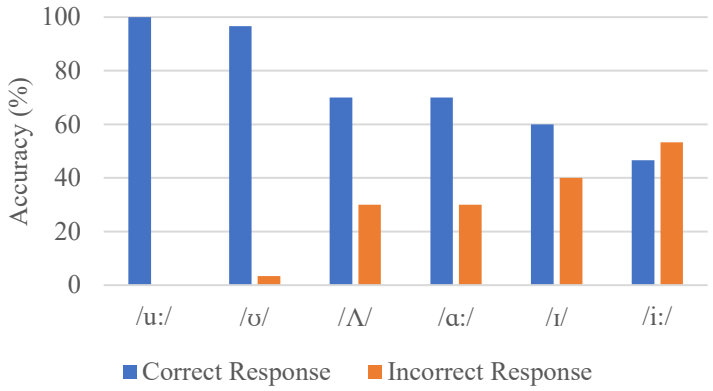

Figure 2
Targeted Vowels Accuracy (%)

For the /ʌ–ɑː/ pair, participants achieved 70% accuracy, with a significant difference observed for /ʌ/ ($t(29) = 2.35$, $p = .026$) but not for /ɑː/ ($t(29) = 1.49$, $p = .147$). Despite being minimal pairs, participants successfully distinguished the vowels. The duration difference for this pair (206.24 msec) was the largest among all stimuli, likely aiding perception. In contrast, the /ɪ–iː/ pair posed the greatest challenge, with no significant differences for either vowel (/ɪ/: $t(29) = 1.10$, $p = .281$; /iː/: $t(29) = –0.36$, $p = .722$). This pair also had the smallest duration difference (106.62 msec), compounding perceptual difficulty.

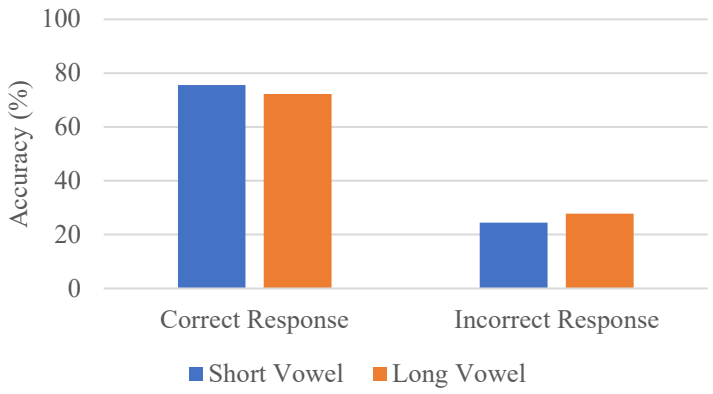

Figure 3
Short and Long Vowel Accuracy (%)

Figure 3 compares accuracy for short vs. long vowels across all pairs. Participants scored slightly higher on short vowels (75.6%) than long vowels (72.7%), but a paired-sample t-test showed no significant difference ($t(89) = –0.555$, $p = .580$). Similar non-significant patterns emerged when analysing individual pairs: /ɪ–iː/ ($t(29) = 1.68$, $p = .103$), /ʌ–ɑː/ ($t(29) = 0.81$, $p = .423$), and /ʊ–uː/ ($t(29) = –1.00$, $p = .326$). This pattern suggests that listeners relied on multiple cues beyond absolute duration, a finding consistent with reports that L2 listeners integrate both temporal and spectral information depending on cue salience and L1 experience (Yazawa et al., 2020; Ćavar et al., 2022)

These results can be interpreted within the Perceptual Assimilation Model (PAM; Best & Tyler, 2007). The /ɪ/–/iː/ contrast appears to follow a Single Category pattern, where both English vowels assimilate to a single Malay category, making discrimination particularly challenging. The /ʊ/–/uː/ and /ʌ/–/ɑː/ contrasts better fit a Category Goodness pattern, in which both L2 sounds map to the same L1 category but one is a poorer exemplar, allowing partial separation. Although Malay /a/ differs articulatorily from English /ʌ/ and /ɑː/, listeners still demonstrated 70% accuracy, indicating sensitivity to fine-grained acoustic cues despite the lack of a direct L1 equivalent (Pillai et al., 2010).

Table 3 shows the overall result of the Perception Accuracy Test. The /ɪ/ - /iː/ pair is the most difficult to perceived, follow by /Λ/ - /ɑː/ and /ʊ/ - /uː/. The findings show that duration remains a robust cue for Malay listeners, especially when combined with segmental or prosodic information, but subtle spectral distinctions like /ɪ/–/iː/ require more targeted instruction. Recent studies emphasise the importance of training that highlights both temporal and spectral dimensions to improve L2 vowel perception (Azhar et al., 2025; Raja Rosemawati & Chiew, 2024). This challenges earlier claims that Malay speakers neutralise these vowels into a single category (Pillai et al., 2010), suggesting that with sufficient input, Malay listeners can form distinct perceptual categories for these contrasts.

Table 3:
Result of Perception Test (%)[1]

Vowel		/ɪ/ - /iː/	/ʊ/ - /uː/	/Λ/ - /ɑː/
Accuracy Identification	Correct response	53.3	98.3	70
	Incorrect response	46.7	1.7	30
	Difference	6.7	96.7	40
Rank in Perception Level		1	3	2

[1] Ranked from most difficult (1) to easiest (3)

CONCLUSION

This study examined the ability of Malay primary pupils to perceive English short–long vowel contrasts. Consistent with recent evidence on L2 cue weighting (Yazawa et al., 2020a; Cebrian et al., 2021; Ćavar et al., 2022), the participants demonstrated overall success in distinguishing the targeted monophthongs, achieving above-chance accuracy across all pairs. The /ʊ/–/uː/ contrast was identified most accurately (98.3%), followed by /ʌ/–/ɑː/ (70%), while the /ɪ/–/iː/ pair remained the most challenging (53.3%). These outcomes suggest that duration cues and additional segmental information supported perception of back-vowel contrasts. In contrast, the limited spectral distance of front vowels hindered discrimination, in line with the Perceptual Assimilation Model's Single Category pattern (Best & Tyler, 2007).

The findings indicate that Malay children can exploit temporal and spectral cues to achieve a reliable perception of English vowels, even when their L1 lacks direct equivalents (Pillai et al., 2010; Yazawa et al., 2020b). Building on this evidence, future research should examine processing measures such as response latency or neural indices (Macher et al., 2021) and explore the impact of targeted perceptual training (Azhar et al., 2025; Raja Rosemawati & Chiew, 2024). Such work will inform early pronunciation pedagogy aimed at strengthening vowel perception and enhancing intelligibility among young Malaysian ESL learners.

ACKNOWLEDGEMENT

I would like to thank Universiti Malaya for providing me with the opportunity, and Dr Chiew Poh Shin for her assistance throughout the research journey. Not to forget Sekolah Kebangsaan Desa Setapak for their cooperation. I highly appreciate Bahagian Perancangan dan Penyelidikan Dasar Pendidikan (BPPDP), Kementerian Pendidikan Malaysia for granting permission to conduct my research at SKDS.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in the paper.

ETHICS STATEMENT

This study was conducted in accordance with the ethical standards of the relevant institution. All participants were informed about the purpose of the study and provided written informed consent prior to participation. Participants' privacy and confidentiality were maintained throughout the research, and all data collected were used solely for academic purposes.

REFERENCES

Alshangiti, W., Evans, B. G., & Wibrow, M. (2023). Investigating the effects of speaker variability on Arabic children's acquisition of English vowels. *Arab World English Journal, 14*(1), 3–27. https://doi.org/10.24093/awej/vol14no1.1

Arum Perwitasari, K., Klamer, M., & Schiller, N. O. (2015). Vowel duration in English as a second language among Javanese learners. In *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 1–5). https://doi.org/10.13140/RG.2.1.1752.5520

Azhar, S. B., Zaini, N., Atan, M. A., & Norwahi, N. A. (2025). Rethinking pronunciation pedagogy in Malaysian higher education: A contextualized conceptual framework. *International Journal of Research and Innovation in Social Science, 9*(4), 5913–5924. https://doi.org/10.47772/IJRISS.2025.90400422

Bello, H., Yap, N. T., Chan, M. Y., & Nimehchisalem, V. (2020). Intelligibility of English vowels produced by Nigerian and Malaysian speakers. *Journal of Research in Applied Linguistics, 11*(2), 81–94. https://doi.org/10.22055/RALS.2020.15948

Best, C. T., Tyler, M. D., Goldstein, L., & Nam, H. (2016). Articulating what infants attune to in native speech. *Ecological Psychology, 28*(4), 216–261. https://doi.org/10.1080/10407413.2016.1230372

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual Language Profile: An easy-to-use instrument to assess bilingualism*. COERLL, University of Texas at Austin. https://sites.la.utexas.edu/bilingual/

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer* (Version 6.0.37) [Computer software]. https://www.praat.org

Casillas, J. V. (2015). Production and perception of the /i/–/ɪ/ vowel contrast: The case of L2-dominant early learners of English. *Phonetica, 72*(2–3), 182–205. https://doi.org/10.1159/000431101

Casillas, J. V., & Simonet, M. (2016). Production and perception of the English /æ/–/ɑ/ contrast in switched-dominance speakers. *Second Language Research, 32*(2), 171–195. https://doi.org/10.1177/0267658315608912

Ćavar, M. E., Rudman, E. M., & Oštarić, A. (2022). Temporal versus spectral cues in L2 perception of vowels: A study with Polish and Croatian learners of English. *Journal of Slavic Linguistics, 30*(1), 85–107. https://doi.org/10.1353/jsl.2022.0002

Cebrian, J., Gorba, C., & Gavaldà, N. (2021). When the easy becomes difficult: Factors affecting the acquisition of the English /iː/–/ɪ/ contrast. *Frontiers in Communication, 6*, Article 660917. https://doi.org/10.3389/fcomm.2021.660917

Chen, W., & van de Weijer, J. (2022). The role of L1–L2 dissimilarity in L2 segment learning: Implications from the acquisition of English post-alveolar fricatives by Mandarin and Mandarin/Wu speakers. *Frontiers in Psychology, 13*, Article 1017724. https://doi.org/10.3389/fpsyg.2022.1017724

Davenport, M., & Hannahs, S. J. (2020). *Introducing phonetics and phonology* (4th ed.). Routledge. https://doi.org/10.4324/9781351042789

Flege, J. E. (1995). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics, 16*(4), 425–442. https://doi.org/10.1017/S0142716400066029

Kabakoff, H., Kharlamenko, J., Levy, E. S., & Levi, S. V. (2021). Differences in perceptual assimilation following training. *JASA Express Letters, 1*(4), Article 044401. https://doi.org/10.1121/10.0003863

Kamarudin, N. I., & Ahmad Kamal, M. A. (2021). Mispronunciation of English monophthongs and diphthongs among Malay native speakers. *International Journal of Academic Research in Business and Social Sciences, 11*(10). https://doi.org/10.6007/IJARBSS/v11-i10/11451

Kashinathan, S., & Azlina, A. A. (2021). ESL learners' challenges in speaking English in Malaysian classrooms. *International Journal of Academic Research in Progressive Education and Development, 10*(2). https://doi.org/10.6007/IJARPED/v10-i2/10355

Macher, N., Abdullah, B. M., Brouwer, H., & Klakow, D. (2021). Do we read what we hear? Modeling orthographic influences on spoken word recognition. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop* (pp. 16–22). https://doi.org/10.18653/v1/2021.eacl-srw.3

Masykar, T., Hassan, R. B., & Pillai, S. (2022). Perception of English vowel contrasts by Acehnese-Indonesian bilingual learners of English. *Indonesian Journal of Applied Linguistics, 11*(3), 718–728. https://doi.org/10.17509/ijal.v11i3.35086

Md. Jaharul Islam, Abdullah Al Masum, & Md. Sayeed Anwar. (2023). The role of temporal and spectral cues in non-native speech production: Bangla speakers' L2 English tense and lax vowels. *Crossing: A Journal of English Studies, 14*, 130–153.

Millet, J., Chitoran, I., & Dunbar, E. (2021). Predicting non-native speech perception using the perceptual assimilation model and acoustic models. In *Proceedings of the 25th Conference on Computational Natural Language Learning* (pp. 661–673). https://github.com/JAMJU/CONLL_2021_

Peirce, J. W. (2017). *PsychoPy* (Version 3.0.0b6) [Computer software]. https://www.psychopy.org

Pillai, S., Mohd. Don, Z., Knowles, G., & Tang, J. (2010). Malaysian English: An instrumental analysis of vowel contrasts. *World Englishes, 29*(2), 159–172.

Raja Rosemawati, R. A. (2024). Exploring English monophthong perception and production in Malaysia. *International Journal of Education, Psychology and Counselling, 9*(53), 520–537. https://doi.org/10.35631/IJEPC.953039

Raja Rosemawati, R. A., & Chiew, P. S. (2024). The production and perception of English monophthongs by young Malay learners in Malaysia. In W. A. Wan Aslynn, S. Pillai, & S. Y. Chew (Eds.), *Contemporary Malaysian English pronunciation* (pp. 42–73). IIUM Press.

Ramli, I., Jamil, N., & Ardi, N. (2020). Formant characteristics of Malay vowels. *International Journal of Evaluation and Research in Education, 9*(1), 144–152. https://doi.org/10.11591/ijere.v9i1.20421

Roach, P. (2010). *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.

Saito, K., & van Poeteren, K. (2018). The perception–production link revisited: The case of Japanese learners' English /r/ performance. *International Journal of Applied Linguistics, 28*(1), 3–17. https://doi.org/10.1111/ijal.12175

Saloranta, A., & Heikkola, L. M. (2023). Acquisition of non-native vowel duration contrasts through classroom education. *Journal of Second Language Pronunciation, 9*(2), 208–233. https://doi.org/10.1075/jslp.20040.sal

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica, 71*(1), 4–21. https://doi.org/10.1159/000356237

van der Slik, F. W. P., van Hout, R. W. N. M., & Schepens, J. J. (2015). The gender gap in second language acquisition: Gender differences in the acquisition of Dutch among immigrants from 88 countries with 49 mother tongues. *PLOS ONE, 10*(11), Article e0142056. https://doi.org/10.1371/journal.pone.0142056

Yahaya, R. S., Madzlan, N. A., & Muhammad, M. (2021). ESL learners' obstacles in learning speaking skills. *The English Teacher, 50*(3), 173–181. https://doi.org/10.52696/WYZY7207

Yang, J., & Fox, R. A. (2014). Acoustic properties of shared vowels in bilingual Mandarin–English children. In *Proceedings of Interspeech 2014* (pp. 1297–1301). https://doi.org/10.21437/Interspeech.2014-323

Yap, N. T., Wong, B. E., & Adi Yasran, A. A. (2010). Representation of English front vowels by Malay–English bilinguals. *Pertanika Journal of Social Science and Humanities, 18*(2), 379–389.

Yazawa, K., Konishi, T., Whang, J., Escudero, P., & Kondo, M. (2023). Spectral and temporal implementation of Japanese speakers' English vowel categories: A corpus-based study. *Laboratory Phonology, 24*(1), 1–33. https://doi.org/10.16995/labphon.6427

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

Yenkimaleki, M., & van Heuven, V. J. (2021). Effects of attention to segmental versus suprasegmental features on speech intelligibility and comprehensibility of EFL learners. *System, 100*, Article 102557. https://doi.org/10.1016/j.system.2021.102557

Yunus, Y., & Pillai, S. (2020). Pola sebutan monoftong bahasa Jerman oleh penutur bahasa Melayu. *Melayu: Jurnal Antarabangsa Dunia Melayu, 13*(2), 223–252. https://doi.org/10.37052/jm.13(2)no3