

Pronunciation Features of Indonesian-Accented English

Syam, AR, Gardner, S & Cribb, M

Published PDF deposited in Coventry University's Repository

Original citation:

Syam, AR, Gardner, S & Cribb, M 2024, 'Pronunciation Features of Indonesian-Accented English', *Languages*, vol. 9, no. 6

<https://doi.org/10.3390/languages9060222>

DOI 10.3390/languages9060222

ISSN 2226-471X

Publisher: MDPI

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Pronunciation Features of Indonesian-Accented English

Abdi Rahmat Syam , Sheena Gardner  and Michael Cribb

Centre for Arts Memory and Communities, Coventry University, Coventry CV1 2NE, UK;
ab1248@coventry.ac.uk (S.G.); aa6177@coventry.ac.uk (M.C.)

* Correspondence: abdisyam@outlook.com

Abstract: English as a Lingua Franca is emerging in Indonesia, but it is not a well-documented variety. This paper aims to describe the pronunciation features of Indonesian-Accented English (IAE). Fifty educated Indonesians who were regular users of English were recorded reading two texts. The phonological features of consonants, clusters, and vowels were investigated through acoustic analysis and spectrographic observation. The results show that IAE is not predictable from contrastive Indonesian English analyses; that IAE may confuse listeners (e.g., if ‘she’ is realised as [si:]); and that speakers may regularly produce sounds at the beginning of words that they do not produce at the ends of words.

Keywords: English as a second language; phonology; accented English; Indonesia

1. Introduction

With the increasing interconnectedness of the world through technology, trade, and international communication, English has become the most widely used language in global diplomacy, business, science, and the Internet. Since the Age of Discovery to the modern day, Indonesia has not been immune to the influence of English. According to [Smith \(1991\)](#), the first Englishman to set foot in Indonesia was Francis Drake when he purchased cloves on the island of Ternate in 1580. The first English trading post was established in 1602 by James Lancaster in Banten not far from present-day Jakarta. Despite English having minimal influence during the three centuries of Dutch colonial rule, after Indonesia declared independence in 1945, the provisional Indonesian government issued a decree to adopt English as its first foreign language. This decision was influenced by the fact that Dutch was associated with the former colonial rulers and was considered the language of the enemy. Additionally, English was deemed more valuable for global communication. Since then, the role of English has become even more instrumental in many aspects of Indonesian society. Propagated by the government, sought after by employers, broadcasted through media, mandated in schools, and endorsed by parents, it is no wonder that English holds a significant place for numerous young Indonesians, surpassing its mere instrumental value in everyday life ([Lamb and Coleman 2008](#)).

With a population of 270 million people and a language policy that favours the spread of English, the number of Indonesian speakers of English might be increasing every year, and a new variety of English seems to be emerging, as is happening in many other countries, e.g., China ([Wang 2023](#)). While English communication involving different varieties of English could become increasingly common, the considerable diversity in phonetic and phonological aspects within English as a Lingua Franca (ELF) communication often raises concern over mutual intelligibility ([Thir 2020](#)). [Jenkins \(2000\)](#) introduced the Lingua Franca Core (LFC) to maintain mutual intelligibility in ELF communication. The LFC comprises pronunciation priorities designed to enhance intelligibility, focusing on specific consonants (e.g., maintaining the quality of all consonants except /θ, ð/), consonant clusters (e.g., avoiding the omission of consonants in word-initial clusters), prioritising the distinction



Citation: Syam, Abdi Rahmat, Sheena Gardner, and Michael Cribb. 2024. Pronunciation Features of Indonesian-Accented English. *Languages* 9: 222. <https://doi.org/10.3390/languages9060222>

Academic Editors: Martha Young-Scholten, Alex Ho-Cheong Leung and Natalia Pavlovskaya

Received: 25 September 2023

Revised: 22 May 2024

Accepted: 7 June 2024

Published: 18 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

between long and short vowels over vowel quality, and emphasising primary stress (see [Dauer \(2005\)](#) for a summary of the LFC list).

[Jayanti and Norahmi \(2014\)](#) contend that an ELF approach is more appropriate in an Indonesian context, as Indonesian speakers interact with non-native English speakers more often than native English speakers. In teaching, the LFC helps teachers use limited class time to prioritise pronunciation features that are essential for intelligibility. This paper aims to explore a subset of Indonesian-Accented English informed by the LFC ([Jenkins 2000](#)) and the extent to which it has stabilised as a distinct variety of English.

The paper will begin with a review of the existing literature, covering previous research on Indonesian-Accented English, as well as relevant studies on Indonesian phonology. Section 3 will provide an overview of this current study, including the rationale for the chosen methodology and instrumentation. Section 4 will detail the findings, and Section 5 will offer a discussion of these results.

2. Previous Literature

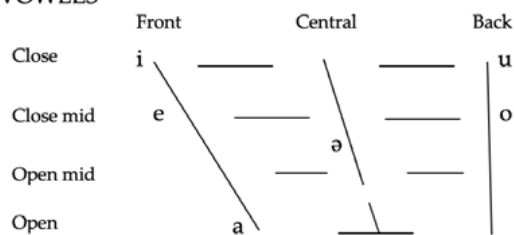
Indonesian possesses a relatively simple phonemic inventory, as indicated in Figure 1. While [Weinberger \(2015\)](#) noted that these sounds are found in most native Indonesian dialects, [Soderberg and Olson \(2008\)](#) suggested that bilabial and palatal approximants (i.e., /w/ and /j/), which are absent in Figure 1, are also part of the Indonesian phonemic inventory.

CONSONANTS

	Bilabial	Labiodental	Palatal	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d			c ɟ	k g			
Nasal		m			n		ɲ	ŋ			
Trill				r							
Tap/flap											
Fricative				s							h
Affricate											
Lateral fricative											
Approximant											
Lateral approximant				l							

Where symbols appear in pairs, the one on the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

Figure 1. Phonemic inventory of Indonesian ([Weinberger 2015](#)).

Previous research on Indonesian English phonology has tended to take a contrastive approach by comparing the phonological systems of Indonesian and English. In terms of consonants, [Tiono and Yostanto \(2008\)](#) proposed that Indonesian learners of English may encounter challenges in pronunciation due to the absence of certain English phonemes in Indonesian. Specifically, the phonemes /v/, /θ/, /ð/, /ʒ/, /dʒ/, and /tʃ/ were identified as potential sources of difficulty. The study highlighted that all six consonants presented pronunciation challenges, with the voiced postalveolar fricative /ʒ/ being particularly problematic. However, the specific reasons why /ʒ/ presented more difficulties compared to the others were not investigated. [Chaira’s \(2015\)](#) more recent analysis also considered vowels and allophonic variation, and, although many of the experimental details (e.g.,

number of participants, data elicitation, and analysis method) were not reported, the study suggested that the following sounds are challenging for Indonesian learners of English: initial [p^h, t^h, k^h] as well as [f, v, ʃ, i:, u:, æ]. There is, therefore, little consistency or clarity in these studies, which makes this study even more pressing.

The main studies on consonant clusters took a different approach. Yuliarti (2014) focused on the simplification of final consonant clusters and its potential impact on intelligibility. It is worth noting that the study did not analyse actual data, instead summarising existing research on consonant clusters and proposing hypotheses based on an L1 transfer and the LFC (Jenkins 2000) for intelligibility. Without concrete data, the study's claims are challenging to support. Nonetheless, Yuliarti did offer valuable insights into final consonant clusters.

Suyanto et al. (2016) categorised Indonesian as a simple syllabic language with a prevalence of CV syllables. Their study revealed that among 50,000 words from the Great Dictionary of the Indonesian Language, 50.63% were CV syllables, whereas English had a relatively lower number of CV syllables, approximately 35%.

Yuliarti (2014) suggested that Indonesian allows for two consonants in the onset position, and onset clusters involving /s/ followed by a non-liquid consonant could be problematic for Indonesian speakers. She explained that sibilant-plosive combinations such as /sk/ and /st/ are not found in the Indonesian language. Thus, in sibilant-plosive clusters such as (/st/), Indonesian speakers of English often insert an epenthetic vowel. (e.g., [sətæmp] instead of [stæmp] for 'stamp').

There is considerable variation in descriptions of Indonesian syllable structure (see Adisasmito 1993; Dardjowidjojo 2009), but there seems to be agreement that in Indonesian words, only a few consonant clusters are allowed in the word-initial and word-final positions, and a common syllable pattern in basic Indonesian words is CVC. This means that learners are inclined to simplify clusters in English words through deletion, saying 'san' for 'sand' and 'talk' for 'talked', or by inserting a schwa vowel in numerous loan words, such as 'sekerip' for 'script', 'filem' for 'film', and 'kelinik' for 'clinic' (Yong 2001). Yuliarti (2014) and Yong (2001) centred their studies predominantly on production aspects. However, Leung et al. (2023) propose that examining consonant clusters from a perceptual standpoint could provide valuable insights into the production of English as a second language (L2) speakers. Their research suggests that speakers may align their perception of L2 consonant clusters with the syllable structure of their native language (L1).

In terms of vowels, some of the studies on L2 vowels have been limited to the investigation of the role of durational and spectral cues in L2 learners' perception of tense and lax vowels in English (Stoel-Gammon et al. 1995; Ylinen et al. 2010). While Stoel-Gammon et al. (1995) investigated the reliance on intrinsic cues (i.e., spectral and durational cues) and extrinsic cues (i.e., phonetics features of the following consonant) on the production of English vowel duration by children, Ylinen et al. (2010) compared the reliance between two intrinsic cues (spectral and durational cues) in perceiving tense and lax vowels by Finnish and English speakers.

In relation to the distinctions between lax and tense vowels, there have been examinations of the phonological characteristics of Indonesian vowels. These studies indicate that the distribution of tense and lax vowels in Indonesian differs from that in English. While tense and lax vowels are distinct phonemes in English, in Indonesian, lax vowels serve as allophones of the tense phonemes. According to research by Andi-Pallawa and Alam (2013) and Verhaar (1996), the allophones [i] and [u] are found in open syllables, while the lax variants [ɪ] and [ʊ] are present in closed syllables (e.g., *bisik* [bi.sɪk] 'whisper'). However, this analysis appears incomplete, as Wijana (2003) argues that while tense vowel phonemes occur in various positions, including closed syllables in non-final positions, their lax allophones are limited to only occurring in the final closed syllable, as demonstrated in (1).

- | | | |
|-----|---------|---------|
| (1) | sin.dir | 'tease' |
| | bun.tot | 'tail' |
| | bim.bɪŋ | 'guide' |
| | lum.pʊr | 'mud' |

In addition, lax variants [ɪ] and [ʊ] also occur in monosyllabic words ending with a consonant (i.e., CVC) such as in (2).

- | | | |
|-----|-----|----------|
| (2) | bɪs | 'bus' |
| | bʊŋ | 'fellow' |

(Wijana 2003)

In a comparative analysis investigating the phonological systems of both English and Indonesian, [Andi-Pallawa and Alam \(2013\)](#) conducted an examination of how students pronounced English vowels. Their results suggested that students faced challenges in understanding the distinctions between the pronunciations of vowel pairs, such as /i/ and /ɪ/, as they tended to overlook the differentiation in pronouncing words like 'bead' and 'bid'. The researchers proposed that the observed result could be attributed to the students' L1 and the limited competence in the vowel system of the L2. The research by [Andi-Pallawa and Alam \(2013\)](#), taking into account [Wijana's \(2003\)](#) study, suggested that Indonesian speakers might rely on extrinsic cues (i.e., syllable structure and position) in their production of English tense and lax vowels with consideration.

The majority of studies on the topic have primarily adopted a contrastive approach, which judges the pronunciation of non-native speakers as lacking when measured against that of native speakers. They have identified issues with certain consonants, with consonant clusters, and with vowels. This study aims to examine a subset of Indonesian-Accented English pronunciation informed by the LFC ([Jenkins 2000](#)) from an empirical perspective. It aims to describe a typical Indonesian 'accent' and to identify areas where miscommunication might occur and, therefore, where teachers should focus their attention.

3. Materials and Methods

This study aims to investigate the characteristics of a subset of Indonesian-Accented English pronunciation. It focuses on the LFC features proposed by [Jenkins \(2000\)](#), which are deemed important as the minimal requirement in preserving intelligibility in ELF teaching contexts. This focus on a subset of LFC features makes sense in the context of English language teaching in Indonesia where time is restricted and English is used more often with other non-native speakers than with native speakers.

In particular, this study seeks to answer the following research questions:

How is the subset of LFC pronunciation features realised in Indonesian-Accented English?

Is there uniformity in the realisation of LFC features in Indonesian-Accented English?

3.1. Participants

Indonesian English speakers with both high and intermediate proficiency levels were chosen as participants. However, this research did not specifically address potential variations between different levels of proficiency since learners' pronunciation could be intelligible despite being accented ([Munro and Derwing 1995](#)). The inclusion of proficient speakers served two purposes—to control for other linguistic aspects such as grammar and lexis and to provide a pedagogical model for ELF teaching in Indonesia. The selection process involved assessing the participants' language proficiency through tests like IELTS, TOEFL, or TOEIC. Those who scored between 5.5 and 6.5 were considered intermediate-level speakers, while those who scored 7.0 or above in IELTS or equivalent in other tests were classified as highly proficient. In cases where language proficiency test scores were unavailable, participants' self-ratings were used. Those whose English proficiency was rated as low intermediate or lower were excluded from this study.

Nineteen participants fell within the 21–30 age group, twenty-seven were aged 31–40, five were 18–20, and one was 41. Of these, 86% were academics or students, while 14%

worked in international trading and hospitality. In terms of language proficiency, 76.5% were multilingual, with the remaining 23.5% being bilingual with English. Eighty-eight percent spoke Indonesian as their first language, while 12% spoke a regional language as their primary language. Slightly over half of the participants identified English as their second or third language, while the rest mentioned a regional or other foreign language. Javanese was the most spoken regional language, reported by 26% of the respondents, followed by Makassarese, Sundanese, Buginese, and Malay. Finally, 13% reported Arabic, Mandarin, or French as their second or third languages.

3.2. Data Collection

The collection of speech data for this study was conducted in major cities in four provinces in Indonesia, namely Jakarta, East Java, Central Java, and South Sulawesi. Participants were invited to volunteer in the study through an online research invitation, which was sent to individuals and gatekeepers from institutions in the researcher's contact. These institutions included universities, an export–import company, a hospitality business, and a non-governmental organisation.

There were 52 participants involved in the speech data collection—12 males and 40 females—but only 50 were used because two data sets were of poor quality. The recruitment of participants used purposive sampling, as the participants needed to meet certain inclusion criteria (e.g., fluent speaker and frequent user of English) to participate in the study. The majority of the participants had never met the researcher prior to data collection.

The recording of participants' speech data occurred in settings such as language labs and meeting rooms where consistent efforts to meet the requirement for a quiet environment were made; none of the recordings took place in a noisy outdoor setting. The recording of speech data was conducted using a Rode NT-USB Microphone with a maximum sound pressure level of 110 dB.

3.3. Speech Elicitation Task

Reading tasks were used to elicit speech data from the participants in this study. The selection of reading tasks as a method to elicit speech data is based on several considerations of its advantages and appropriateness for the purpose of this study. Compared to the speech elicitation strategy using spontaneous speech, reading tasks allow the speakers to focus more on the pronunciation aspect. In the context of L2 learners, this is particularly relevant, as they are often very conscious about their grammar accuracy and sacrifice their fluency. Another advantage is that using the same reading tasks enables a comparative analysis across the speakers' data.

A further consideration in selecting the reading material for the speech elicitation tasks was the authenticity of the text. There were some texts that are often used in phonetics and phonological research, such as 'The North Wind and the Sun' (NWS), which is a well-known fable. Deterding (2006) points out that the passage in the NWS has substantial limitations, such as the absence of certain sounds, lexical repetition, and a lack of occurrences of certain sounds in a particular position (e.g., word-initial/medial), which might be unsuitable for different purposes of studies concerning the description of varieties of English. Although Deterding (2006) proposes the use of the passage 'The Boy who Cried Wolf' (BCW), which is longer than the NWS but has fewer lexical repetitions, the BCW text was modified specifically for phonetic research in a way that the flow of the story might be unnatural.

Although this current study used an experimental design, it aims to have a real-world application of the results. Thus, speech authenticity was one of the main aspirations when conducting this study. Although speech data elicited from reading tasks, which are modified to target specific sounds, are often seen as detached from the category of authentic speech, in some contexts, reading a text is a natural form of communication, such as reading a story or a report to an audience. The use of original texts can preserve a certain level of authenticity and, at the same time, allows the researcher to capture comparable speech content from various speakers. In order to achieve this, the originality

of the reading material used for speech elicitation in the study must be retained. However, finding a suitable original text for research could be challenging, as some texts may contain unfamiliar words for the readers, or the sentence length might be inappropriate, which could result in mispronunciation and monotonous reading speech.

In order to mitigate the possibility of mispronunciation, dysfluency, and unnatural prosody, the readability of the original text was the main consideration for selection. The reading task used in this study consists of two passages; the first passage is a segment taken from a short story called 'The Man in the Brown Coat', while the second passage is taken from a radio broadcast script called 'Reluctant Spirit'. The readability of the texts was analysed using online readability analysers ([Datayze n.d.](#)). In terms of Flesch–Kincaid grade level, the first passage has a score level equal to grade 3 and 4, and the grade level of the second passage is between 4 and 5. Both passages score above 70 in Flesch reading ease, which shows that the text can be read by a 9th grader or younger. The results of Dale–Chall readability analysis for the first and second passage are 4.94 and 5.48, respectively. Dale–Chall's Readability Index uses a method that counts the percentage of words that are out of the Dale–Chall Word List and determines the level of word difficulty of a text ([Yan et al. 2006](#)). The word list consists of 3000 words that are assumed to be known by most fourth-grade readers. In terms of the number of words per sentence, the first passage has an average of 10 words per sentence, while the second passage has a significantly higher number of words with an average of 14.5 words per sentence. Although the difference in sentence length between the two passages might affect the outcome of the speech data to some extent, the overall score of the readability analysis indicated that both texts appeared to be readable for native-speaking readers well below college level. It is important to recognize that the criteria for assessing the readability of the reading materials in this study were based on the standards of native speakers. Consequently, a set of texts was subjected to a preliminary examination to explore their influence on the reading fluency and pronunciation of L2 readers. Following this pilot study, it was expected that the chosen texts for this research would elicit authentic data reflecting a natural and fluent speech pattern.

3.4. Data Analysis

Since the three types of features investigated in this study are different in nature, different approaches were employed to analyse consonants, consonant clusters, and vowel duration. While consonants and consonant clusters were measured using categorical data, vowel quantity was measured involving numerical data. Thus, the former were analysed using descriptive statistics, whereas the latter was analysed using inferential statistics.

3.4.1. Consonants

In terms of consonants, [Jenkins \(2000\)](#) suggests that all consonantal sounds in the English phonemic inventory are important, with the exception of labiodental fricatives and the lateral approximant (i.e., /θ/, /ð/, pre-consonantal and syllabic /l/). She also adds that some phonetic requirements such as aspirated voiceless plosives are also important. While most English consonants have equivalents or near equivalents in Indonesian consonants, /f v θ ð z ʃ ʒ/ are consonants that might be difficult for Indonesian learners of English to pronounce ([Yong 2001](#)). From the seven consonants she identified, /θ ð ʒ/ are excluded from analysis in this study; /θ/ and /ð/ are exempted in [Jenkins' \(2000\)](#) LFC, while /ʒ/ is considered to carry a low functional load because it is rarely found in English words, particularly in initial position ([Cruttenden 2014](#), p. 336).

Regarding the phonetic requirements in [Jenkins' LFC](#), [Yong \(2001\)](#) suggests that /p, t, k/ in Indonesian are always unaspirated, which might cause them to sound similar to their voiced counterparts, /b, d, g/, to an English ear in particular. With the LFC-specific consonants and phonetic requirements taken into consideration, there were seven consonant phonemes of English investigated in this study, namely /f, v, z, ʃ, p, t, k/. Altogether, 68 tokens were analysed from each of the 50 speakers, comprising ten tokens each for /f, z, ʃ, p, t, k/ and eight tokens for /v/.

In this study, several acoustic parameters were used to classify the phonological features being examined. One of the acoustic measurements is voice onset time (VOT), which is typically used to differentiate plosives in terms of voicing and aspiration. [Lisker and Abramson \(1964\)](#) explain that VOT can be measured by marking off the interval between the stop release and the voicing onset of the following vowel. In their study, the point of voicing onset can be located in the first of regularly spaced spectral striations, that is, a row of vertical 'lines' in a spectrogram, which represent a single pulse in the vocal folds ([Hagiwara 2009](#)).

Another acoustic measurement utilised in this study was spectral peak location or peak frequency, which helps determine the place of articulation of consonants, particularly the fricatives. According to [Liu et al. \(2000\)](#), the primary acoustic cues for identifying the place of articulations of plosives can be found in the peak frequency in the release burst. They indicate that bilabial plosives typically have a peak frequency between 0.5 and 1.5 kHz, whereas velar plosives have a peak frequency ranging from 1.5 kHz to 4 kHz; a peak frequency higher than 4000 Hz usually indicates an alveolar plosive. Regarding fricatives, [Jongman et al. \(2000\)](#) suggest that a mid-frequency spectral peak at around 2.5 to 3 kHz typically signifies postalveolar fricatives, often corresponding to the following vowel's F3. Conversely, higher primary spectral peak frequencies at around 4 to 5 kHz generally indicate alveolar fricatives. For non-sibilant fricatives, such as those in bilabial and labiodental positions, a relatively flat pattern without any significant peak is normally shown in the spectrogram.

3.4.2. Consonant Clusters

The LFC proposes three guidelines for consonant clusters. The first one is that no deletion is allowed in word-initial clusters. The second is that only in accordance with Inner Circle English standards is deletion permitted in medial and final clusters. Finally, insertion is preferred over deletion (e.g., product as [pər'ʌdʌktə], not [pʌdʌk]) ([Dauer 2005](#)).

To investigate the consonant clusters, the reading tasks consist of 61 word-initial and 64 word-final CC clusters, as well as a few CCC clusters (i.e., two word-initial and three-word final CCC clusters). The limited number of three consonant clusters in the reading was due to the selection of the reading material, which focused on the readability of the reading task. Adapting [Hansen Edwards \(2006\)](#), the consonant clusters produced by the Indonesian speakers were coded as (a) well-formed; (b) deletion; (c) epenthesis; or (d) mispronounced. The term well-formed refers to target-like production without any addition or deletion to a member of the consonant clusters. Any changes in features to the members in a cluster (e.g., /sɒbz/ → [sɒbs]) were still categorised as well-formed. Deletion refers to the absence of a member within a cluster, while epenthesis involves the addition of a segment to the cluster. Mispronounced indicates a situation where the sounds cannot be identified as belonging to any particular word. Unclear cases such as cases involving identical consonants in adjacent words (e.g., books standing [-s s-]) were excluded from the analysis, following [Osburne \(1996\)](#).

Regarding insertion, short periodic energy in a waveform or a vocalic element in a spectrogram within consonant clusters indicates an epenthetic vowel. In [Ramírez's \(2006\)](#) study, epenthetic vowels in Spanish clusters are 32% shorter than the average vowel duration in a non-cluster situation. The mean length of an epenthetic vowel is around 26.98 ms, while the full vowel length constitutes an average of 85.61 ms. For consonant cluster deletion, inspection of a waveform and a spectrographic display on the absence of a spectral sign in one of the expected consonants in a cluster is the main method to determine deletion. Parameters such as stop burst and friction noise in the spectrogram are the primary cues for identifying consonants.

3.4.3. Vowels

Regarding vowels, the LFC distinguishes vowel features between quality and duration. While vowel quality is excluded from the LFC, vowel quantity (i.e., long- and short-vowel

contrast) appeared to be essential in order to ensure that pronunciation did not obstruct intelligibility in ELF interactions. Therefore, this current study aims to investigate the tense and lax¹ vowels in Indonesian-Accented English, particularly if there was a significant difference between the two vowels in terms of relative duration or length.

In addition, based on this information on the phonological system of Indonesian vowels, this study also analyses if allophonic variation due to the effect of the syllable structure occurred in the vowel duration of Indonesian-Accented English. Three pairs of tense and lax vowels were analysed to investigate the vowel duration contrast in Indonesian-Accented English. They are the high front vowels /i/ and /ɪ/, high back vowels /u/ and /ʊ/, and low back vowels /ɔ/ and /ɒ/. Other English vowel pairs were excluded from the analysis due to insufficient speech data containing the vowels. Furthermore, in the case of syllable structure, only the pair /i/ and /ɪ/ and pair /u/ and /ʊ/ met the required data (i.e., occurring in both open and closed syllables) for the analysis. The token analysed for open syllables were 'see' and 'mu.sic' (tense) and 'vi.sit' and 'wo.men'(lax), while the token for closed syllables were 'feed' and 'spoon' (tense) and 'vi.sit' and 'books' (lax). Based on the findings of Wijana (2003), one might anticipate that Indonesian speakers would produce shorter vowel durations for supposedly long vowels in words such as 'feed' and 'spoon'.

Two statistical tests were used to test the tense and lax vowel duration contrasts and the two-level factors of syllable type (CV, CVC) and tenseness (tense, lax) on vowel duration. Firstly, a paired sample *t*-test was administered to test the hypothesis that there were no significant differences between the tense and lax vowels. Secondly, a two-way repeated measures ANOVA was carried out to examine whether vowel duration was influenced by syllable type, vowel tenseness, or a combination of both factors.

Overall, descriptive statistics were utilised to analyse the trend of speech production. In total, an 80% frequency of occurrence of a particular pattern/feature was considered the rule, while any pattern below this percentage was considered the exception or a personal variation in a speaker's speech production. This percentage was adapted from Cancino et al.'s (1975) study on the acquisition of English auxiliaries by Spanish speakers. In their study, the auxiliary is considered acquired when it appeared at a rate of 80% in at least three consecutive utterances. For the purposes of this study, this percentage was used as a threshold for inclusion criteria. The features that appeared consistently across the speech data above the threshold of 80% frequency of total occurrences were attributed to Indonesian-Accented English.

4. Results

This section presents the results of four LFC core features of Indonesian-Accented English. The features are consonants (Section 4.1), aspiration of plosives (Section 4.2), consonant clusters (Section 4.3), and vowel quantity (Section 4.4). The data presented in these results were obtained from the data of 50 speakers.

4.1. Consonants

As explained above, seven consonants of English were investigated in this study, namely /f, v, z, ʃ, p, t, k/. Samples of words (henceforth tokens) where the seven consonants occurred were investigated from the speakers' data.

While content words were preferable for the analysis because they are more resistant to reduced articulation (Bell et al. 2009; Johnson 2004), some phonemes might only occur in a function word. However, the use of function words in the analysis was kept minimal in order to minimise the factor of reduced articulation. It is suggested that frequent words in connected speech tend to have a variety of lenited characteristics, and function words are more frequent and predictable compared to content words (Bell et al. 2009).

The primary analysis for determining the realisation of the consonant was an auditory perception analysis. In addition, analysis of voicing, burst or peak frequency (Shadle 1985, p. 179), and voicing length through observation of a spectrographic image and the waveform of sound was also used in order to enhance the analysis of the consonant type.

4.1.1. Voiceless Labiodental Fricative /f/

For voiceless labiodental fricatives, there were ten tokens of words to be analysed. Five tokens with the phoneme /f/ in syllable-onset positions were *father*, *frighten*, *floats*, *feels*, and *infant*. In the coda positions, the selected tokens were *wife*, *afternoon*, *itself*, *shift*, and *myself*. In these words, the phoneme /f/ occurs in both simple and complex syllable structures.

Table 1 shows the realisation of phoneme /f/ in the ten tokens by 50 speakers. In each onset or coda positions, there were five tokens containing the target sound. The mean score of the data and the standard deviation were obtained from the number of [f] sounds produced from the five tokens by the 50 speakers. The data suggest that Indonesian speakers in this study seem consistent in their realisations of a voiceless labiodental fricative. Nearly 100% of the tokens containing /f/ were pronounced as [f] regardless of their position in a syllable. The other 1% where /f/ was pronounced differently in the onset position accounts for one token where /f/ was realised as [v] by a speaker in the word *infant*, in which the voicing of the preceding nasal consonant and/or of the following vowel might be assimilated to the labiodental fricative. Regardless, the speaker showed a consistent realisation of /f/ as [f] in the other four words. This result contradicts the previous literature (Tiono and Yostanto 2008; Yong 2001) that suggest that /f/ could be challenging for Indonesian speakers to pronounce.

Table 1. Realisation of /f/.

/f/	N *	n/s **	Realisation	Mean	SD	Percentage
Onset	250	5	[f]	4.96	0.19	99%
Coda	250	5	[f]	5	0	100%

* Overall tokens. ** Number of tokens per speaker.

4.1.2. Voiced Labiodental Fricative /v/

There were eight tokens of words selected to be analysed for voiced labiodental fricatives. Four tokens had the voiced labiodental fricative /v/ in the word-initial positions or the syllable-onset positions, while the other four tokens contained the /v/ word in the final or in the coda position of a syllable. These words were *visit*, *veg*, *over*, *even*, *leaves*, *above*, *have*, and *shelves*.

Figure 2 shows the samples of spectrograms taken from two speakers pronouncing the word *visit* (top) and *leaves* (bottom). A spectrogram is a sound's spectro-temporal representation; a spectrogram's horizontal direction denotes time, while its vertical direction denotes frequency (Odden 2005, p. 10). The spectrographic images in this study were generated using the speech analysis software Praat version 6.1.42. (Boersma and Wennink 2021). It can be seen from Figure 2 that there is variation in the realisation of the phoneme /v/ between the two speakers. While speaker A pronounced the phoneme /v/ as [f], which is voiceless, speaker B maintained the voicing feature of the phoneme. Based on the spectrographic image, a clear voicing bar can be observed in the first formant of the speech of speaker B (indicated in red), whereas the speech of speaker A does not seem to have a voicing bar in the F1 where phoneme /v/ is located (indicated in blue). Moreover, the harmonic-like pattern in waveforms (indicated by the arrows) clearly indicates that the /v/ sound in Speaker B's data is voiced, while the friction noise pattern in the waveform of Speaker A suggests that the sound is voiceless. This variation does not reflect the variation in different speakers' groups from different regions, as some speakers often pronounce phoneme /v/ as [v] or [f] inconsistently.

Table 2 shows the overall mean, standard deviation, and percentage of realisations of phoneme /v/. There are two major phonetic variations in the phoneme /v/ that are found in the data (i.e., [v] and [f]). The data in Table 2 show that the mean score of [v] in word-initial positions or syllable-onset positions is 1.92, with a standard deviation of 1.04. This means that each speaker has the phoneme /v/ realised as [v] in only one or two of

the total four tokens on average. Thus, less than half of the total 200 tokens that contain phoneme /v/ are realised as [v], while 52% of them are realised as [f].

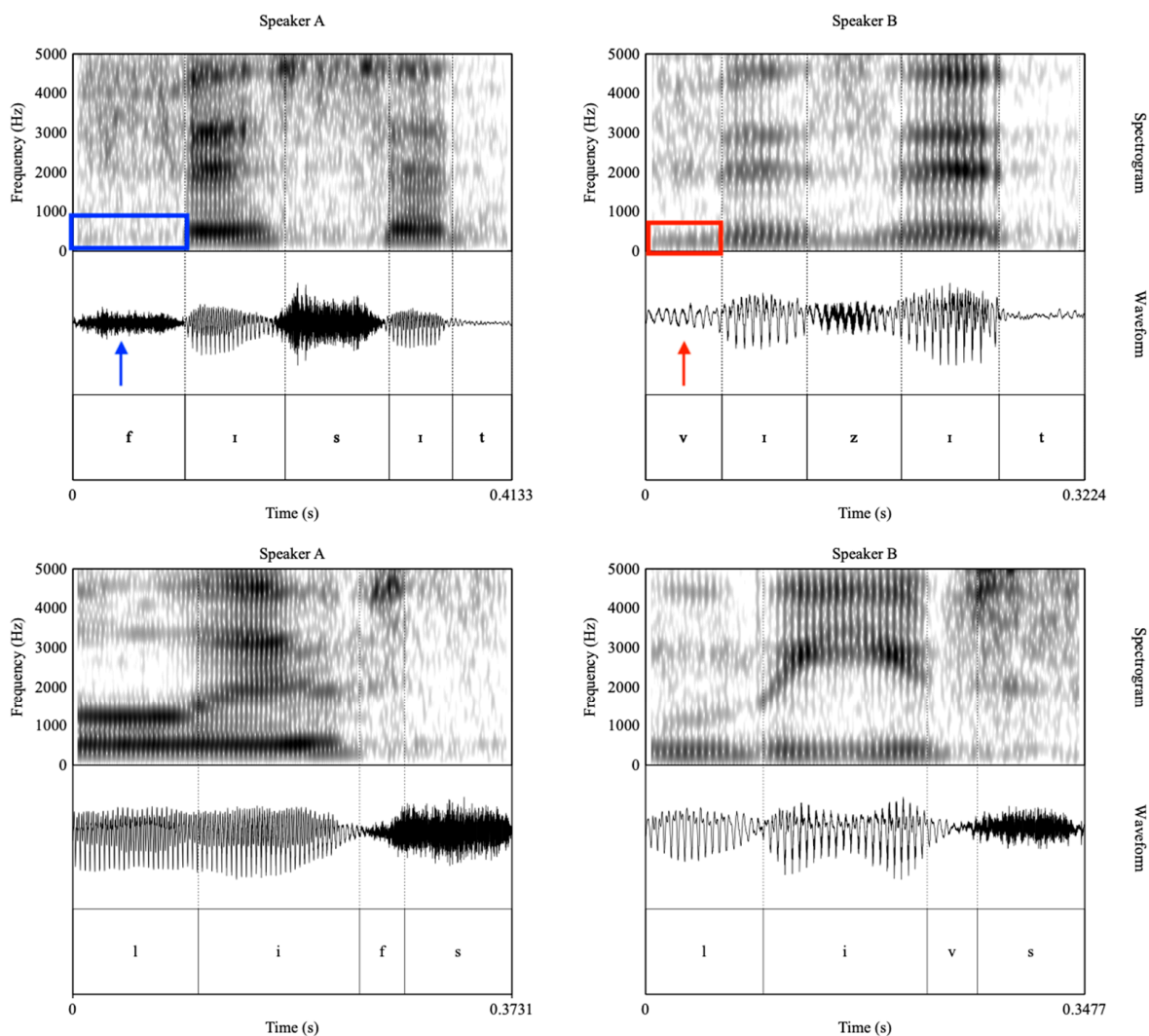


Figure 2. Sample of a spectrographic image of individual differences of the realisation of /v/; speaker A (left); speaker B (right).

Table 2. Realisation of phoneme /v/.

/v/	N *	n/s **	Realisation	Mean	SD	Percentage
Onset	200	4	[v]	1.92	1.04	48%
			[f]	2.06	1.08	52%
Coda	200	4	[v]	2.18	1.58	55%
			[f]	1.8	1.60	45%

* Overall tokens. ** Number of tokens per speaker.

Compared to the onset position, the speakers showed a more consistent use of either [v] or [f] in their speech in the coda position. There were 14 speakers, or about 28% of them, who realised the phoneme /v/ consistently as [v], while 26% of the total speakers realised it as [f]. In the onset position, only 4% and 10% of the speakers consistently pronounced the phoneme /v/ as [v] and [f], respectively. More than 30% of the speakers exhibited inconsistency in the onset position, pronouncing /v/ as [v] in two of the four tokens and as [f] in the other two tokens. Although there was a more consistent pattern of realisation in the coda position, the difference in the overall frequency of realisation between [v] and [f] was only defined by a small margin.

4.1.3. Voiced Alveolar Fricative /z/

In order to investigate the voiced alveolar fricatives produced by Indonesian speakers, ten tokens containing phoneme /z/, five each in the onset and coda positions, were analysed. Although the words that had the phoneme /z/ in the word-initial position were limited in the reading texts, intervocalic consonant /z/, which is generally found in the word-medial position, will be considered a syllable-onset consonant of the following vowel based on the maximal onset principle in phonological theory (i.e., /visit/ → [vi.zit] as opposed to [viz.it]) (Fallows 1981). The tokens with the phoneme /z/ in the onset position were *thousand*, *visit*, *music*, *realises*, and *opposite*. For the coda position, some words with the phoneme /z/ in this position could be found in the reading texts; the tokens selected for analysis were *rise*, *closed*, *supposed*, *because*, and *sometimes*.

Figure 3 shows samples of the spectrograms taken from two speakers pronouncing the word *thousand*. Similar to the earlier analysis of /v/, the speakers in Figure 3 showed variation in terms of the realisation of /z/. While the peak frequencies of the fricatives appear to be similar at above 4000 Hz for both speakers, as indicated in the yellow area of the spectrogram, which is an indication of an alveolar fricative (see Section 4.1.4 for further explanation), they differed in terms of voicing. The spectrographic picture of speaker A in Figure 3 shows the absence of a voicing feature in the F1 where phoneme /z/ is located (indicated in the blue rectangle area). Another indication of the absence of voicing for this phoneme could also be observed in the waveform (indicated by the blue arrow), which seems to show a friction noise-like pattern. On the other hand, Speaker B appears to have a voicing feature as indicated in red at the bottom of the spectrogram and a harmonic-like pattern in the wave form as pointed out by the red arrow. This spectrographic and wave form analysis suggests that speaker A appeared to realise phoneme /z/ as [s], while speaker B realised it as [z].

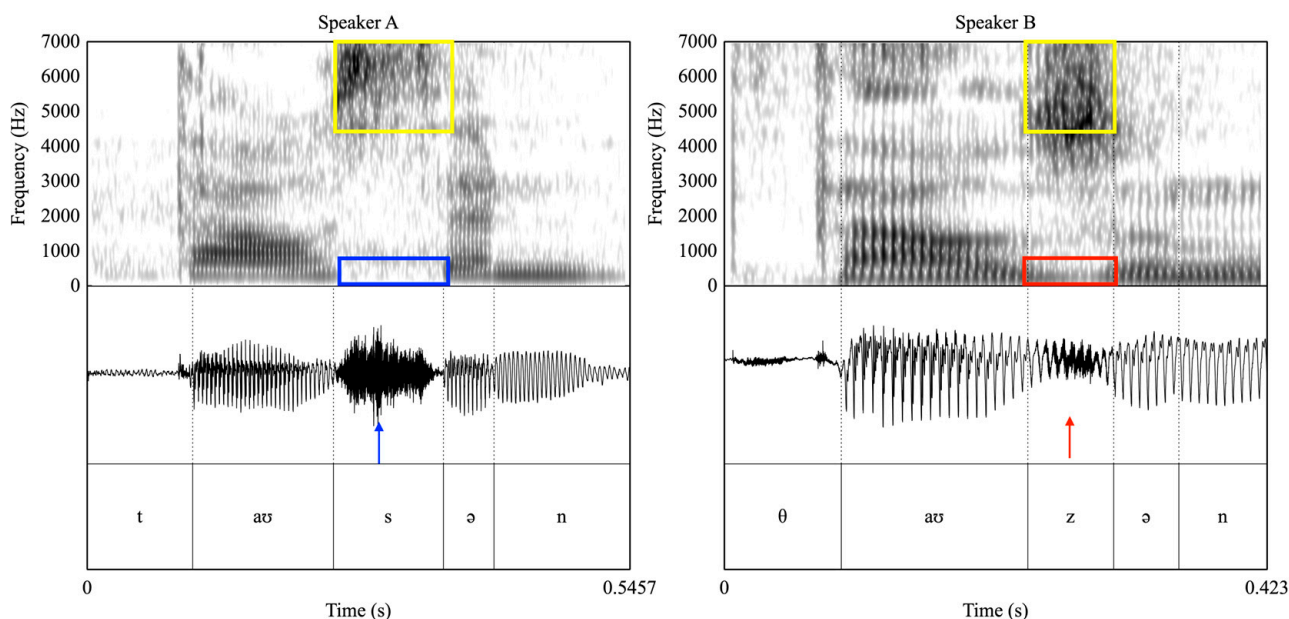


Figure 3. Sample of a spectrographic and waveform image of the realisation of /z/ in the word ‘thousand’; speaker A (left); speaker B (right).

Table 3 shows the realisation of phoneme /z/ in the onset and coda positions. It appears that the Indonesian speakers in this study realised phoneme /z/ as [z] and [s]. While phone [z] seems to occur marginally, more in onset positions in intervocalic environments, phone [s] appears to dominate significantly in the word-final position, as far as phonetic variations are concerned. In the onset position, 56% of the speakers pronounced the five

tokens that contained phoneme /z/ as [z] consistently, while almost the same number of speakers realised the phoneme as [s] in five out of five tokens.

Table 3. Realisation of phoneme /z/.

/z/	N *	n/s **	Realisation	Mean	SD	Percentage
Onset	250	5	[z]	2.78	1.69	56%
			[s]	2.16	1.68	43%
Coda	250	5	[z]	0.36	0.63	7%
			[s]	4.62	0.64	92%

* Overall tokens. ** Number of tokens per speaker.

In the coda position, the mean score of the realisation of /z/ as [z] was 0.36, whereas the mean score of the realisation of /z/ as [s] was 4.62. With a standard deviation of 0.6, there seemed to be more regularity among the speakers where 92% of phonemes /z/ in the 250 tokens were realised as [s].

4.1.4. Voiceless Post Alveolar Fricative /ʃ/

There were eight tokens selected from the reading to investigate a voiceless alveolar fricative in Indonesian-Accented English. Due to the limitation of the reading text, only the phoneme in a syllable-onset position in word-initial and -medial positions were analysed. The selected tokens that contained the phoneme /ʃ/ word initially were *shelves, shoulders, shy, she, shift,* and *sure*, and the other tokens that contained the phoneme /ʃ/ in word-medial positions were *nation* and *communication*.

A sample of acoustic analysis of phoneme /ʃ/ from two speakers is given in Figure 4. The frequency in which the peak amplitude of the friction appears provides information as to whether a fricative is a post alveolar or an alveolar one (Bjorndahl 2022). A post alveolar fricative is associated with a lower peak frequency, whereas an alveolar fricative has a higher peak frequency, which often occurs above the F2 of the surrounding vowel (Shadle 1985). In Figure 2 above, speaker A has the peak amplitude of friction noise at 6000 Hz (shown in blue), indicating that the fricative is articulated as an alveolar fricative. On the other hand, speaker B has a peak noise below 4000 Hz (shown in red), or relatively at the same frequency of the following high front vowel, which indicates that it is a post alveolar fricative. This acoustic information was used to support the auditory analysis of phoneme /ʃ/.

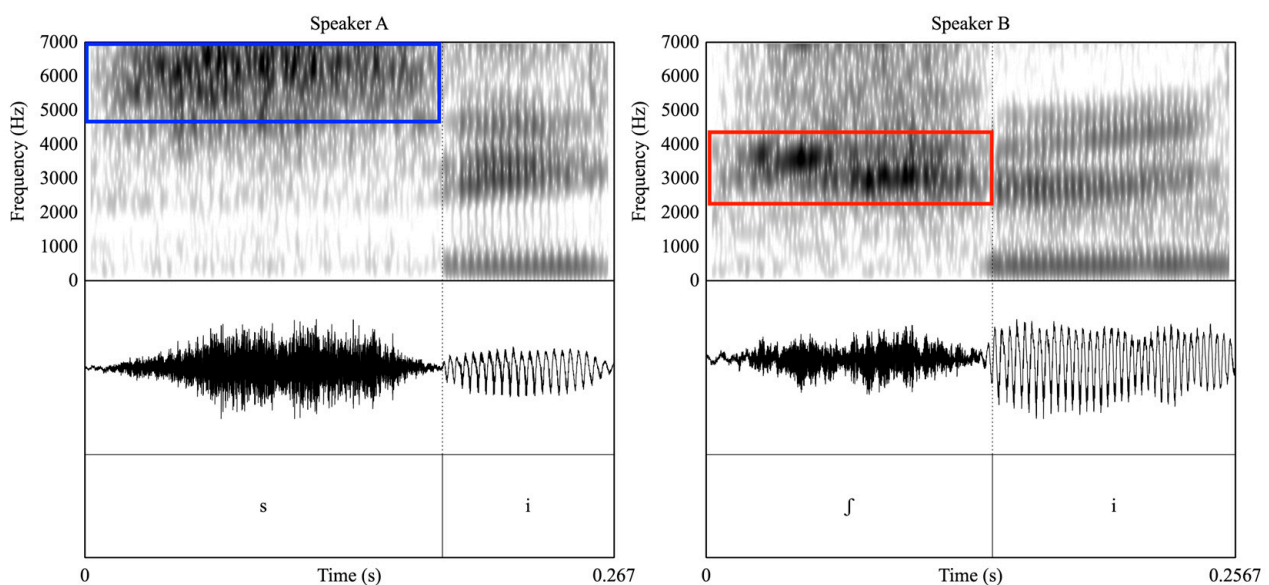


Figure 4. Sample of a spectrographic and waveform image of the realisation of /ʃ/ in the word ‘she’; speaker A (left); speaker B (right).

Table 4 demonstrates that there were two recurring phonetic variations in phoneme /ʃ/ in the speakers’ speech in this study. The mean scores show the average number of each of the two variations from the total eight tokens produced by each speaker. While the mean score for phoneme /ʃ/, which was realised as [ʃ], was 5.8, the mean score for the phoneme /ʃ/ realised as [s] was 2.2. With a standard deviation of 2.1, the data suggest that some speakers tended to realise the phoneme /ʃ/ as either [ʃ] or [s] more consistently throughout the eight tokens, although a few others might have had an equal number of both phonetic variations as the realisation of /ʃ/ (i.e four [ʃ] and four [s]). According to the data, 60% of the speakers realised /ʃ/ as [ʃ] in six tokens or more out of the total eight tokens, and only two speakers showed a consistent use of [s] in their pronunciation of the phoneme, whereas 32% of the total speakers had a rather balanced combination of the use of [ʃ] and [s] in their production of the eight tokens that contained the voiceless post alveolar fricative.

Table 4. Realisation of phoneme /ʃ/.

/ʃ/	N *	n/s **	Realisation	Mean	SD	Percentage
Onset	400	8	[ʃ]	5.82	2.09	73%
			[s]	2.2	2.1	27%

* Overall tokens. ** number of tokens per speaker.

4.2. Aspiration of Plosives

In addition to consonants, the aspiration of voiceless plosives in the word-initial position is also important in LFC (Jenkins 2000). This is important because unaspirated voiceless plosives are often confused with voiced plosives in this position. Figure 5 shows the percentage of the realisation of /p, t, k/ in Indonesian-Accented English speech data. There were two common realisations of the three plosives in the word-initial position in the data (i.e., aspirated and unaspirated).

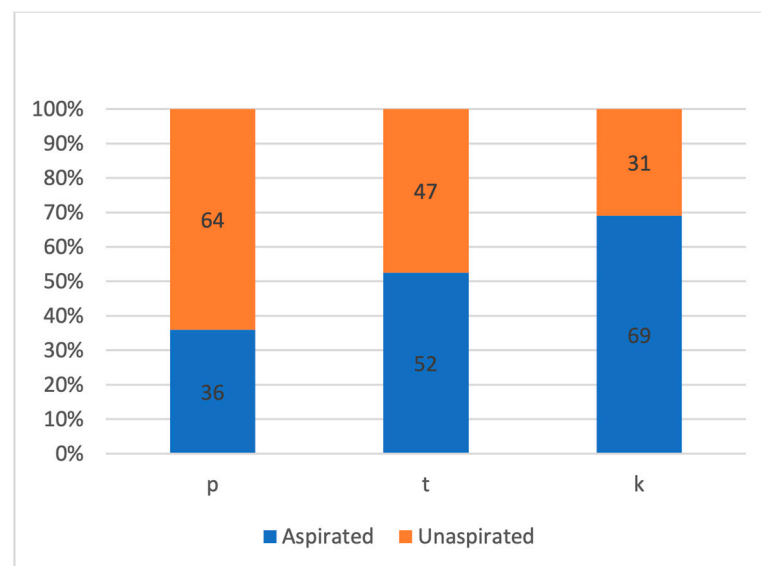


Figure 5. Percentage of aspiration of /ptk/ in the word-initial position.

For the word-initial /p/, the tokens investigated were *passes*, *plastic*, *pureed*, *pull*, *pain*, two tokens of the word *people*, and three tokens of the word *picture*. The findings indicate that, on average, the 50 speakers predominantly produced the voiceless bilabial plosive as an unaspirated [p], with a mean score of 6.4 out of 10 tokens. Conversely, the remaining 3.56 mean score corresponds to instances where the plosive was realised with aspiration [p^h]. With a standard deviation of 2.4, the data indicate a notable level of variation. While there were only five speakers who realised /p/ as [p^h] in more than 6 of the 10 tokens, 28 speakers realised the phoneme as [p] in 7/10 to 10/10 tokens. In addition, there were

five speakers who realised /p/ as [p^h] in half of the total tokens and as [p] in the other half. Thus, the data suggest that [p] is the predominant realisation of phoneme /p/, as opposed to [p^h], in the syllable-onset position.

For /t/, the five tokens with the word-initial /t/ investigated were *turning*, *town*, *told*, *twist*, and *takes*. The overall distribution of [t^h] and [t] as the realisations of /t/ in the onset position seems comparable, even though there is a scant tendency towards realising /t/ as [t^h]. With the total of 5 tokens investigated in each of the 50 speakers, the mean score for [t^h] was 2.6 while the mean score for [t] was 2.36, with a standard deviation of 1.29. In terms of consistency, there were 14 speakers who realised /t/ as [t^h] in at least four out of five tokens, whereas nine speakers realised /t/ as [t] in 4/5 tokens or more. Thus, half of the total speakers had rather equal distributions of [t^h] and [t] as far as the five tokens are concerned.

For voiceless velar plosives, 5 tokens containing /k/ at the initial position of the word (i.e., *came*, *closed*, *crying*, *contact*, and *communication*) were analysed. Based on the data in Figure 5, the aspirated plosive [k^h] seems to be the predominant realisation of the phoneme /k/ compared to its unaspirated counterpart. According to the data, the mean score for the realisation of /k/ as [k^h] was 3.46, while the mean score for that of [k] was 1.5, with standard deviations of 1.23 and 1.4, respectively. As far as a consistent pattern of realisation is concerned, there were 15 speakers who pronounced the phoneme /k/ as [k^h] in all of the 5 tokens, and another 10 speakers had the same realisation in 4 out of 5 tokens, which constitutes half of the total speakers combined. On the other hand, the number of speakers who realised /k/ as [k] in at least four out of five tokens was limited to five speakers, while 40% of the speakers showed a less evident pattern of realisation.

Besides the realisation of plosives, it is also interesting to observe the overall average of Voice Onset Time (VOT) of each plosive. VOT is often used to determine whether a plosive is aspirated or not (Abramson and Whalen 2017).

In terms of VOT of the three English plosives, Figure 6 indicates that the average length of VOT increased as the articulation of plosives moved from the anterior to the posterior. As can be seen from Figure 6, the mean VOT length of /p/ is shortest among the three plosives, and the mean VOT of /t/ is shorter than the mean for /k/. The mean VOT of /p/ is 21 milliseconds with a standard deviation of 14 ms. The mean VOT of /t/ is significantly longer than that of /p/ at 42 milliseconds with a standard deviation of 17 ms. Lastly, the mean VOT of /k/ is 54 milliseconds with a standard deviation of 15 ms. The relatively large standard of error suggested that some speakers had significantly shorter or longer VOT than others, which is consistent with the non-uniformity of the realisation of voiceless plosives in the perceptual analysis.

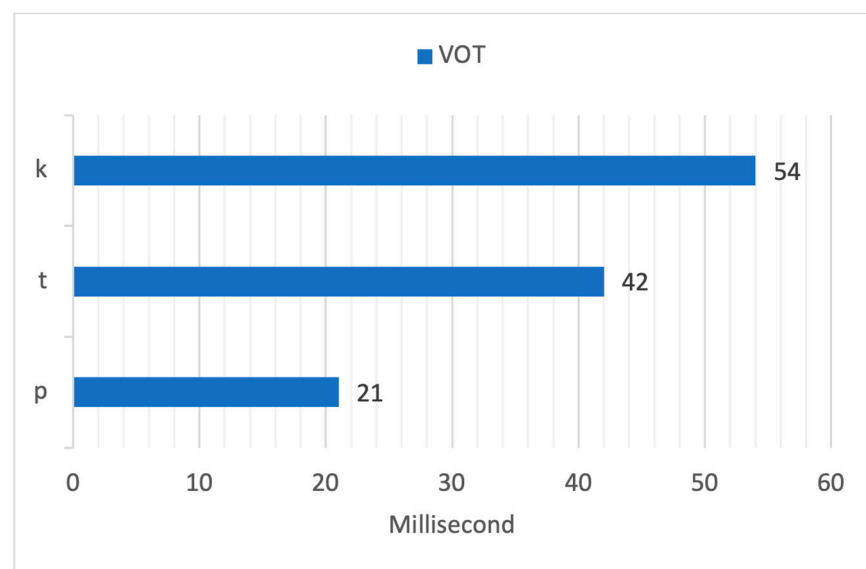


Figure 6. Overall mean VOT of voiceless plosives.

4.3. Consonant Clusters

This study examined a total of 63 instances of consonant clusters at the beginning of words. Among these, 61 clusters consisted of two consonants, while 2 clusters included three consonants. As outlined in Table 5, there were minimal variations observed in the production of both two-consonant (CC) and three-consonant (CCC) clusters. It is important to acknowledge that although the speakers managed to produce CCC clusters correctly 100% of the time, it is worth noting that this study’s exploration of CCC clusters was limited in quantity. Nevertheless, the findings demonstrated that all speakers were capable of accurately articulating the ‘str’ cluster.

Table 5. Descriptive statistics of word-initial clusters.

N = 2	CCC			
	Well-Formed	Deletion	Epenthesis	Mispronunciation
Average	2	0.00	0.00	0.00
SD	0	0.00	0.00	0.00
%	100%	0.0%	0.0%	0.0%
N = 61	CC			
	Well-Formed	Deletion	Epenthesis	Mispronunciation
Average	60.77	0.08	0.04	0
SD	0.5	0.3	2	0
%	99.7%	0.1%	0.1%	0%

Regarding CC clusters, the number of well-formed clusters produced by the Indonesian speakers was numerically significantly higher at 99.7% than the modified-formed clusters. The percentage of clusters that underwent deletion and epenthetic processes (Côté 2000) were both at 0.1%. The results of this study suggest that the Indonesian speakers in this study were able to produce a wide range of well-formed word-initial CC consonant clusters.

Unlike the consonant clusters in word-initial positions, the data in Table 6 show that the production of consonant clusters in word-final positions displays a certain level of variability rather than exhibiting a distinct pattern. Regarding clusters with three consecutive consonants (CCC), the data illustrate that the participants in this study opted for deletion in slightly more than 60% of cases when confronted with the four word-final CCC clusters considered in this research. The remaining consonant clusters were pronounced correctly at a rate of 37.5%. Upon analysing the data, it was noted that over 90% of the consonant clusters in the word ‘next’ (/nekst/) underwent the omission of the final consonant [t], while the most retained cluster was found in the word ‘shelves’ (i.e., ‘lvs’), with only 28% of speakers omitting the final consonant [z].

Table 6. Descriptive statistics of word-final clusters.

N = 3	CCC			
	Well-Formed	Deletion	Epenthesis	Mispronunciation
Average	1.5	2.5	0	0
SD	0.71	0.71	0	0
%	37.5%	62.5%	0%	0%
N = 64	CC			
	Well-Formed	Deletion	Epenthesis	Mispronunciation
Average	42.00	21.69	0.00	0.31
SD	8.80	8.35	0.00	0.97
%	65.6%	33.9%	0.0%	0.5%

In the case of CC clusters at the end position of a word, over 60% of the clusters were well formed, while 34% of them went through a deletion process. There were 0.5% instances of mispronunciation found in the data (e.g., ‘muscles’ pronounced as [muskel]). It was observed in the data that a plosive was the more frequently deleted consonant in the final coda clusters. The data show that some speakers pronounced words such as *worked*, *twist*, *don’t*, and *think* without the final plosives. Regarding the inflectional /s/ sound within word-final clusters, an observation was made that more proficient speakers tended to retain the final /s/, while intermediate speakers exhibited inconsistent production of the inflectional /s/.

In summary, the data indicate that Indonesian speakers display greater consistency in producing consonant clusters at the beginning of words, showing a preference for maintaining the standard form of the cluster. On the other hand, there appears to be a lack of consistency in the production of coda clusters, with no form of production reaching 80% consistency, whether it involves well-formed clusters or otherwise. In the case where the target form was not maintained by the speakers, consonant deletion was the predominant strategy applied in their production compared to the epenthetic process.

4.4. Vowel Quantity

Jenkins (2000) suggests that vowel quantity or, in other words, the contrast between long and short vowels is important to preserve intelligibility in communication. Although long- and short-vowel contrast exists in both English and Indonesian, the relationship between the types of vowels is different in the two languages. While long and short vowels in English are different phonemes, in Indonesian, the short vowel is an allophone of the long vowel, which is affected by the syllable structure (Wijana 2003). The following results show whether the vowel contrast in Indonesian-Accented English is affected by the syllable structure as the result of an allophonic relation or if it is associated more with the laxness and tenseness of the vowels.

4.4.1. Tense and Lax Vowel Contrast

Tense vowels were associated with an average vowel duration $M = 0.094$ ($SD = 0.01$). By comparison, lax vowels were associated with a numerically shorter vowel duration $M = 0.063$ ($SD = 0.01$). To test the hypothesis that tense vowels and lax vowels were associated with statistically significantly different mean durations, a paired sample *t*-test was performed. For the purpose of conducting a *t*-test, Table 7 shows that the mean value of tense and lax vowels was sufficiently normal (i.e., skew < 2.0 and kurtosis < 9.0; (Schmider et al. 2010)). In addition, a Shapiro–Wilk test showed no significant departure from normality ($W(50) = 0.93, p = 370$).

Table 7. Descriptive statistics associated with vowel durations.

Category	Mean	SD	Skew	Kurtosis
Tense	0.094	0.006	−0.37	−0.77
Lax	0.063	0.006	0.19	0.86

The result of the paired sample *t*-test was associated with a statistically significant effect, $t(50) = 11.69, p < 0.001$. Thus, the tense vowels were associated with a statistically significantly longer duration than the lax vowels. Cohen’s *d* was estimated at 3.01, which is a large effect based on Cohen’s (1992) guidelines.

4.4.2. Syllable Structure and Tenseness Effect

A graphical representation of the mean duration of vowels concerning syllable type and tenseness factors is displayed in Figure 7. It can be seen that tense vowels were associated with numerically longer vowel duration ($M_{open} = 0.082$ and $M_{closed} = 0.083$) compared to lax vowels in either open or closed syllables ($M_{open} = 0.064$ and $M_{closed} = 0.062$).

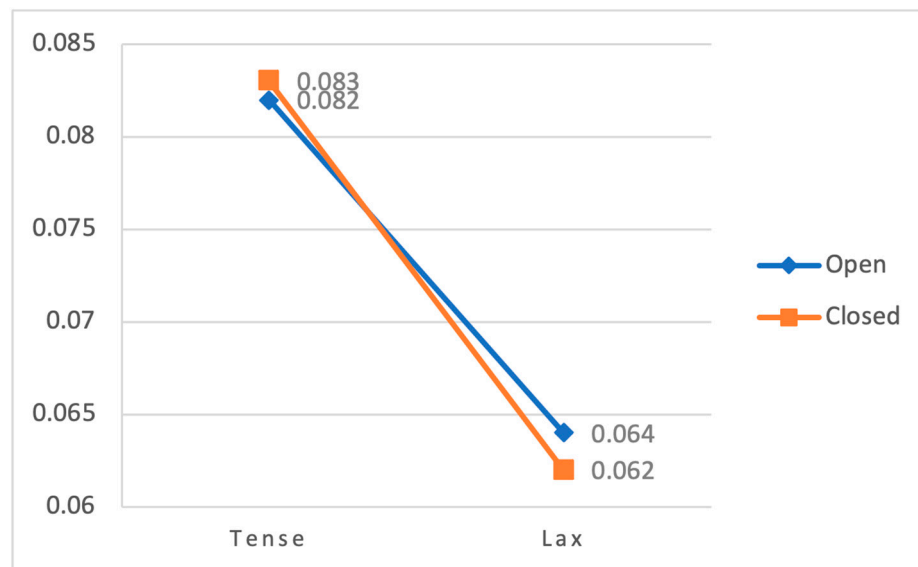


Figure 7. The mean duration (in ms) of tense and lax vowels in open and closed syllables.

In order to test the hypothesis that syllable type and tenseness of the vowel had a statistically significant effect on vowel duration, a two-way repeated measures ANOVA was performed. Before conducting the ANOVA, an assumption of normality was evaluated using the Shapiro–Wilk test, as well as the skewness and kurtosis values. Based on the Shapiro–Wilk test, the assumption of normality of the four combinations of subject factors was determined to be satisfied, as the *p*-values of the four combinations were over alpha 0.05 (i.e tense: $W(50) = 0.96, p = 0.32$ and lax: $W(50) = 0.96, p = 0.39$ for open syllables; tense = $W(50) = 0.97, p = 0.66$, and lax: $W(50) = 0.97, p = 0.62$ for closed syllables). Additionally, the four combinations’ distributions were associated with skew and kurtosis vales of less than 2.0 and 9.0, respectively (see Table 8).

Table 8. Descriptive statistics associated with two levels of within-subject factors.

Category	Mean	SD	Skew	Kurtosis
Tense-open	0.082	0.03	0.49	−0.26
Lax-open	0.064	0.01	0.25	−0.05
Tense-closed	0.083	0.03	0.43	0.06
Lax-closed	0.062	0.14	−0.150	−0.03

The result of the ANOVA indicates that the main effect of syllable type was not statistically significant ($F(1, 49) = 0.003, p = 0.956, \text{partial } \eta^2 < 0.001$). On the other hand, the main effect of tenseness yielded a statistically significant effect ($F(1, 49) = 34.52, p < 0.001$), with an effect size of 0.54, indicating that 54% of the variance in the vowel durations was accounted for by tenseness. The interaction between the two factors was not statistically significant ($F(1, 49) = 0.229, p = 0.636, \text{partial } \eta^2 = 0.008$). This means that the effects of syllable types and tenseness are independent of each other.

5. Discussion

The results show that consonants, consonant clusters, and vowel duration in the Indonesian-Accented English speech data vary in terms of the regularity of speech production. In the case of consonants, the results indicate that there was a regularity in the production of consonant /f/, which was realised as [f] by all speakers. Similarly, despite being below 80%, which was considered the rule (Cancino et al. 1975), the realisation of phoneme /f/ also indicated a high degree of regularity, where over 70% of the speakers maintained the sound quality. This is contrary to what was predicted by Chaira (2015) and Yong (2001, p. 281), who indicated that consonants /f/ and /f/ might cause pronunciation

problems for Indonesian learners, as the sound is not in their L1 phonemic inventory. One potential explanation is that a significant proportion of Indonesians, particularly those practicing Islam, have received some level of Arabic training for religious purposes. This exposure to Arabic could have familiarised them with the /f/ and /ʃ/ consonants, which are present in the Arabic sound system. Ortega (2008, p. 48) suggested that knowledge of two (or more) languages can accelerate the learning of an additional one.

Another possible factor is the incorporation of the /f/ sound into the Indonesian phonological system, which might have been influenced by loanwords from other languages such as English. The adoption of numerous English terms containing the /f/ consonant, like 'infrastruktur' (infrastructure), 'informasi' (information), 'inflasi' (inflation), and 'farmasi' (pharmacy), has become increasingly prevalent in Indonesian. However, there remain instances where the borrowed English words in Indonesian realise the /f/ phoneme with other sounds, such as [p] in 'telepon' (telephone) (Abdurrahim and Jalil 2020). Nevertheless, it can be argued that the trend in Indonesian loan words from English containing the phoneme /f/ is to preserve the original sound. On the other hand, the sound /ʃ/ in most Indonesian loan words from English is changed to [s], as indicated in the examples above (i.e., information and inflation).

While some degree of regularity was found in /f/ and /ʃ/, both of which occur in IAE, more variations were found in the consonants /v/ and /z/. It appears that roughly 50% of the speakers can accurately produce these consonant sounds, while the remaining half tend to alter the target sounds by modifying certain distinctive features, such as voicing. For instance, some speakers tend to devoice /v/ and /z/, making them sound more like /f/ and /s/, respectively.

In relation to the discussion of the effect of L1 and L2 on L3, this evidence presents some complications in the case of /z/. While /v/ is absent in both the Indonesian and Arabic phonemic inventories, /z/ is one of the sounds found in Arabic. Furthermore, in terms of loan words, there are several words containing /v/ and /z/ sounds from English that are loaned to Indonesian such as 'vaksin' (vaccine), 'variabel' (variable), 'zona' (zone), and 'zombi' (zombie). While consonant 'v' in Indonesian loan words is often pronounced as [f] (i.e., /vaksin/ → [faksin]), phoneme /z/ in word 'zona' and 'zombi' are frequently pronounced with the voiced sibilant in Indonesian. Thus, as far as language background and loan words are concerned, one would expect a certain level of consistency in pronouncing the phoneme /z/. However, this was not observed in the data, and such analysis is somewhat speculative, so further research will be welcome. This inconsistent evidence further amplifies claims such as the one pointed out by Ortega (2008) that transfer is a highly complex phenomenon, and it cannot explain all phenomena in interlanguage development. The variable realisations of /v/ and /z/ could be considered features of Indonesian-Accented English.

Concerning the aspiration of plosives in the beginning of a word, the VOT for /p/ was notably shorter at 21 ms compared to /t/ or /k/, which had VOTs of 42 ms and 54 ms, respectively. Lisker and Abramson (1964) propose that VOT values ranging from 0 to 25 ms indicate unaspirated plosives, suggesting that plosive [p] in the data appear to be unaspirated, as the VOT is below 25 ms. In addition, compared to that of English native speakers, VOTs of voiceless plosives produced by Indonesian speakers are relatively shorter. The mean VOTs of /p, t, k/ in Lisker and Abramson's (1964) study are 58, 70, and 80 ms, respectively, whereas the mean VOTs of /p, t, k/ produced by the Indonesian speakers in the study are 21, 42, and 54 ms, respectively. The tendency of Indonesian speakers to use unaspirated plosives in the word-initial position in their English could be associated not only with L1 transfer but also markedness. Markedness refers to the inherent tendencies or preferences across languages for specific forms or features such as voiceless over voiced sounds (Hansen Edwards and Zampini 2008). Indonesian phonology does not have allophonic variation in plosives in the word-initial position. In addition, it is generally assumed that aspirated voiceless plosives are more marked than the plain

voiceless plosives (Jakobson and Halle 2002; Rice 2007). These unaspirated plosives are a regular feature of Indonesian-Accented English.

In terms of consonant clusters, this study's results indicate that Indonesian speakers generally maintain standard consonant clusters at the beginning of words but display variability in the production of consonant clusters at the end of words. Regarding word-initial clusters, this study's findings contradict traditional beliefs about how Indonesian speakers handle consonant clusters. Previous studies in the literature (Dardjowidjojo 2009; Yong 2001; Yuliarti 2014) suggested that Indonesian speakers often insert schwa vowels within onset clusters. However, the data from this study did not reveal widespread instances of added epenthetic vowels within consonant clusters in words like 'speak', 'spectrum', 'state', or 'stop'. It should be noted that the data collection involved reading tasks, allowing participants to view the orthography, which may have facilitated their pronunciation.

In the data, Indonesian speakers consistently omitted the final /t/ in a three-consonant coda cluster /kst/ (as in 'next') and in some words in two-consonant clusters (i.e., 'direct'), which was traditionally not regarded as an extraprosodic element. Conversely, the final inflectional /s/ consonants in CCC coda clusters like /lvs/ and /lts/ in words such as 'shelves' and 'adults' were often retained, with only a few speakers exhibiting instances of deletion. These results indicate that speakers tend to preserve clusters in inflectional morphemes more than those in lexical stems. This contrasts with Abrahamsson's (2003) findings, where /r/ codas are pronounced more correctly in lexical stems than in inflectional morphemes (p. 342). However, considering the proficiency level of the participants, which ranged between intermediate and high, this study's outcomes seem to align with Abrahamsson's (2003) claim that higher accuracy rates in coda production are found at more advanced stages of proficiency. However, this assertion cannot be verified in this study due to methodological constraints related to the participants' language backgrounds.

Although this might imply a language transfer phenomenon where CC clusters represent the maximal coda consonant cluster in Indonesian speakers' production, the deletion of plosive consonants even in CC coda clusters in the data suggests that the deletion process might not be solely triggered by syllable structure. Instead, it could be attributed to a phonological constraint related to plosives in final coda clusters. However, it is also likely that this phenomenon is due to the existence of a multiple syllabification system, as proposed by Adisasmito (1993), within the language system of Indonesian speakers. Hence, there are instances where a speaker retains the syllable structure on one occasion and omits one segment of the cluster on another. Nevertheless, in terms of consonant clusters, the result showed that IAE speakers show a tendency to preserve the consonant clusters structure in word-initial positions while showing variation in dealing with the word-final clusters. They tended to delete the final plosives in a word-final cluster if they could not preserve the original structure of the consonant cluster.

The lack of faithfulness to consonant clusters in the coda position, as compared to the onset position, suggests that the markedness principle might play a dominant role in the production of Indonesian speakers. Vennemann (1987) argued that a syllable's head (i.e., syllable-onset) is more preferred when the number of speech sounds is closer to one, while a smaller number of speech sounds in the coda is more preferred (p. 13). Carlisle (2006) summarises these principles, stating that CV is the most unmarked syllable in languages with a single consonant as the optimal onset and a zero consonant as the optimal coda (p. 107). The clusters data reflected these principles particularly in the case of coda clusters, as it is more marked with two margins than the onset cluster with only one margin. Furthermore, the ratio of deletion in coda position between CCC and CC clusters was 2:1, indicating that a less marked cluster tended to preserve more than the more marked one.

Regarding vowel duration, the results indicate a difference in the duration of vowels between tense and lax vowels in the speech of Indonesian participants. The statistical analysis provides evidence that tense vowels were pronounced with a longer duration

compared to lax vowels. This result is expected, as tense and lax vowels exist in the Indonesian phonological systems, although they are in allophonic distributions, unlike in English where the tense and lax vowels are distinctive phonemes (Andi-Pallawa and Alam 2013).

Regarding the influence of allophonic variations in Indonesian in relation to syllable structures, the findings show that there seems to be no significant effect of syllable structures on vowel duration; instead, vowel duration is largely associated with the tenseness of the vowel. This result indicates that Indonesian speakers did not seem to shorten vowels in words such as 'feed' or 'spoon', even though these are closed monosyllabic words. Additionally, the results also show that there was no significant interaction effect of syllable structure and tenseness on vowel duration in Indonesian-Accented English. These findings contradict prior research by Andi-Pallawa and Alam (2013), which suggested that Indonesian speakers tend not to differentiate between tense and lax vowels. It is possible that participants in this study have developed a better understanding of the tense and lax vowel distinction in English, while actively reducing the influence of their native language (L1) in their speech. Thus, IAE maintains a distinction between tense and lax vowels.

Lastly, as far as the LFC is concerned, the findings indicate that Indonesian-Accented English appears to agree with the LFC's guidelines regarding consonant clusters, vowel length contrast, and some consonants. The results show that IAE speakers tend to maintain consonant clusters in the word-initial position and appear to show contrast between tense and lax vowels, which are important in the LFC. However, IAE shows variability in terms of consonants, particularly the production of /v/ and /z/, where some speakers appear to devoice these consonants. Some speakers also seem to fail to aspirate the word-initial voiceless plosives, particularly the bilabial plosives.

Although the LFC was used to guide the methodology of this paper, the aim was to identify features of IAE, not specifically to consider intelligibility, which is a focus of the LFC. Therefore, to conclusively confirm the effect of these IAE features on intelligibility, additional research is needed. Nevertheless, these results might be useful for teachers in choosing which area of pronunciation to focus on in their classes.

6. Conclusions

This paper presents an empirical study of 50 Indonesian speakers of English. Informed by previous contrastive and theoretical studies and the LFC, it aims to describe Indonesian-Accented English.

We first investigated seven consonants identified in earlier contrastive studies as problematic for Indonesian learners of English and identified as crucial for comprehensibility by Jenkins' LFC. Contrary to earlier contrastive studies, our participants showed no difficulty in pronouncing /f/. By contrast, the English /v/ phoneme was realised in about half the instances in both the onset and coda positions as [v] and about half as [f]. This can be identified as a feature of Indonesian-Accented English. Other features of Indonesian-Accented English include /z/ in the onset position, which can be realised as [s]; /ʃ/, which is occasionally realised as [s]; and initial /p, t, k/, which are regularly unaspirated. In terms of the consonant clusters, the data suggest that final consonant clusters are prone to consonant deletion in Indonesian-Accented English. In terms of vowels, the results indicate that there is a contrast between tense and lax vowels in terms of the duration, regardless of the syllable structure.

As far as the LFC proposed by Jenkins (2000) is concerned, IAE speakers could approximate the standard pronunciation of English consonants listed in the LFC with potential variations on a group of sibilant sounds (i.e., /ʃ, z, and v/). Consonant clusters at the beginning of words tend to be preserved, while word-final clusters might be simplified through deletion. Regarding the contrast between long and short vowels, IAE speakers were able to produce contrastive length between lax and tense vowels.

These findings are important for several reasons. They represent the analysis of a substantial data set (the speech of 50 participants) where the analysis combines a perceptual analysis with spectrometry. The empirical findings are therefore robust, and contrast with

earlier studies based on small samples, contrastive analysis, or introspection. Secondly, the findings provide a nuanced set of results that could be translated into a focus for teaching. They identify not only phonemes that tend to be realised in ways that are distinctive of Indonesian-Accented English but also the positions in which this occurs. Such information is useful for other speakers wishing to adjust to Indonesian-Accented English and to teachers wishing to teach learners who may wish to lose their Indonesian-Accented English in favour of a British or American standard pronunciation. As far as pronunciation goals are concerned, Munro and Derwing (1995) noted that speech can be accented while remaining intelligible, and as early as 1949, pronunciation experts have stressed improved intelligibility as the most important goal in pronunciation teaching. However, their more recent observations suggest a significant emphasis in classrooms on accent reduction, with achieving native-like production as the desired goal (Munro and Derwing 2020). Future research might address some inevitable limitations of this current study. For instance, it might involve empirical studies of the pronunciation of less advanced speakers, employ a different reading task or spontaneous speech, and focus on features beyond the LFC. It is important to remind readers here that our participants are advanced speakers, because we assume that Indonesian-Accented English is not the pronunciation of lower-level learners but rather that of mature English users, which is relatively stabilised and has proved functional in Indonesia.

Author Contributions: Conception, A.R.S., S.G. and M.C.; Investigation and data curation, A.R.S.; writing—original draft preparation, A.R.S.; writing—review and editing, S.G. and M.C.; supervision, S.G. and M.C.; funding acquisition, A.R.S. All authors have read and agreed to the published version of the manuscript.

Funding: The work described in this paper was supported by a doctoral research grant awarded to the corresponding author by the Ministry of Religious Affairs (MORA) Indonesia scholarship.

Institutional Review Board Statement: The research for this paper was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Coventry University (Project reference number: P98501; date of approval: 19 December 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: The data presented in this study are unavailable due to privacy restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

Note

¹ Even though vowel lax/tense-ness and length are concurrent features in English, the notions of length and lax/tense-ness are not identical constructs.

References

- Abdurrahim, Abdurrahim, and Syahrir Jalil. 2020. Phonological replacement of loan words used in Indonesian. *Journal of Applied Studies* 4: 160–76. [CrossRef]
- Abrahamsson, Niclas. 2003. Development and recoverability of L2 codas. *Studies in Second Language Acquisition* 25: 313–49. [CrossRef]
- Abramson, Arthur S., and Douglas H. Whalen. 2017. Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics* 63: 75–86. [CrossRef] [PubMed]
- Adisasmito, Niken. 1993. Syllable structure and the nature of schwa in Indonesian. *Studies in Linguistic Sciences* 23: 1–19.
- Andi-Pallawa, Baso, and Andi Fiptar Abdi Alam. 2013. A comparative analysis between English and Indonesian phonological systems. *International Journal of English Language Education* 1: 103–29. [CrossRef]
- Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60: 92–111. [CrossRef]
- Bjorndahl, Christina. 2022. Voicing and frication at the phonetics-phonology interface: An acoustic study of Greek, Serbian, Russian, and English. *Journal of Phonetics* 92: 1–26. [CrossRef]
- Boersma, Paul, and David Wennink. 2021. Praat: Doing Phonetics by Computer [Computer Program]. Version 6.1.42. Available online: <https://www.fon.hum.uva.nl/praat/> (accessed on 16 December 2021).
- Cancino, Herlinda, Ellen J. Rosansky, and John H. Schumann. 1975. The acquisition of the English auxiliary by native Spanish speakers. *TESOL Quarterly* 9: 421–30. [CrossRef]

- Carlisle, Robert Stephen. 2006. The sonority cycle and the acquisition of complex onsets. *Studies in Bilingualism* 31: 105–37. [CrossRef]
- Chaira, Salwa. 2015. Interference of first language in pronunciation of English segmental sounds. *English Education Journal (EE)* 6: 469–83.
- Cohen, Jacob. 1992. A Power Primer. *Psychological Bulletin* 112: 155–59. [CrossRef] [PubMed]
- Côté, Marie-Hélène. 2000. *Consonant Cluster Phonotactics: A Perceptual Approach*. Cambridge: Massachusetts Institute of Technology.
- Cruttenden, Alan. 2014. Gimson's Pronunciation of English. In *Gimson's Pronunciation of English*. Abingdon: Taylor and Francis Inc. [CrossRef]
- Dardjowidjojo, Soenjono. 2009. *English Phonetics and Phonology for Indonesians*. Jakarta: Yayasan Pustaka Obor Indonesia.
- Datayze. n.d. Readability Analyser. Available online: <https://datayze.com/readability-analyzer> (accessed on 25 November 2020).
- Dauer, Rebecca. M. 2005. The Lingua Franca Core: A New Model for Pronunciation Instruction? *TESOL Quarterly* 39: 543–550. [CrossRef]
- Deterding, David. 2006. The North Wind versus a Wolf: Short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association* 36: 187–96. [CrossRef]
- Fallows, Deborah. 1981. Experimental evidence for English syllabification and syllable structure. *Journal of Linguistics* 17: 309–17. [CrossRef]
- Hagiwara, Rob. 2009. November 19. How to Read a Spectrogram. Available online: <https://home.cc.umanitoba.ca/~robh/howto.html> (accessed on 10 January 2022).
- Hansen Edwards, Jette. G. 2006. *Acquiring a Non-Native Phonology: Linguistic Constraints and Social Barriers*. London: Bloomsbury Publishing Plc.
- Hansen Edwards, Jette. G., and Mary. L. Zampini. 2008. *Phonology and Second Language Acquisition*. Amsterdam: John Benjamins Publishing Company.
- Jakobson, Roman, and Morris Halle. 2002. *Fundamentals of Language*. Berlin: Mouton de Gruyter.
- Jayanti, Fernandita Gusweni, and Maida Norahmi. 2014. EFL: Revisiting ELT practices in Indonesia. *Journal on English as a Foreign Language* 4: 5–14. [CrossRef]
- Jenkins, Jennifer. 2000. *The Phonology of English as an International Language*. Oxford: Oxford University Press.
- Johnson, Keith. 2004. Massive reduction in conversational American English. In *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*. Edited by Kiyoko Yoneyama and Kikuo Maekawa. Tokyo: The National International Institute for Japanese Language, pp. 29–54.
- Jongman, Allard, Ratee Wayland, and Serena Wong. 2000. Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America* 108: 1252–63. [CrossRef] [PubMed]
- Lamb, Martin, and Hywel Coleman. 2008. Literacy in English and Transformation of Self and Society in Post-Soeharto Indonesia. *International Journal of Bilingual Education and Bilingualism* 11: 189–205. [CrossRef]
- Leung, Alex, Martha Young-Scholten, Wael Almurashi, Saleh Ghadanfari, Chloe Nash, and Olivia Outhwaite. 2023. (Mis) perception of consonant clusters and short vowels in English as a foreign language. *International Review of Applied Linguistics in Language Teaching* 61: 731–64. [CrossRef]
- Lisker, Leigh, and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20: 384–422. [CrossRef]
- Liu, Hwei-Mei, Chin-Hsing Tseng, and Feng-Ming Tsao. 2000. Perceptual and acoustic analysis of speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Clinical Linguistics and Phonetics* 14: 447–64. [CrossRef]
- Munro, Murray J., and Tracey M. Derwing. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of Second language learners. *Language Learning* 45: 73–97. [CrossRef]
- Munro, Murray J., and Tracey M. Derwing. 2020. Foreign accent, comprehensibility and intelligibility, redux. *Journal of Second Language Pronunciation* 6: 283–309. [CrossRef]
- Odden, David. 2005. *Introducing Phonology*. Cambridge: Cambridge University Press.
- Ortega, Lourdes. 2008. *Understanding Second Language Acquisition*, 1st ed. Abingdon: Routledge. [CrossRef]
- Osborne, Andrea G. 1996. Final cluster reduction in English L2 speech: A case study of a Vietnamese speaker. *Applied Linguistics* 17: 164–81. [CrossRef]
- Ramírez, Carlos. Julio. 2006. Acoustic and perceptual characterization of the epenthetic vowel between the clusters formed by consonant+ liquid in Spanish. In *Selected Proceedings of the Second Conference on Laboratory Approaches to Spanish Phonetics and Phonology*. Somerville: Cascadilla Proceedings Project, pp. 48–61.
- Rice, Keren. 2007. Markedness in phonology. In *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press, pp. 79–97.
- Schmider, Emanuel, Matthias Ziegler, Erik Danay, Luzi Beyer, and Markus Bühner. 2010. Is It Really Robust?: Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology* 6: 147–51. [CrossRef]
- Shadle, Christina Helen. 1985. *The Acoustics of Fricative Consonants*. Cambridge: Massachusetts Institute of Technology.
- Smith, Brian D. 1991. English in Indonesia. *English Today* 7: 39–43. [CrossRef]
- Soderberg, Craig. D., and Kenneth S. Olson. 2008. Indonesian. *Journal of the International Phonetic Association* 38: 209–13. [CrossRef]

- Stoel-Gammon, Carol, Eugene. H. Buder, and Margaret. M. Kehoe. 1995. Acquisition of vowel duration: A comparison of Swedish and English. Paper presented at XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden, August 13–19, vol. 4, pp. 30–36.
- Suyanto, Suyanto, Sri Hartati, Agus Harjoko, and Dirk Van Compernelle. 2016. Indonesian syllabification using a pseudo nearest neighbour rule and phonotactic knowledge. *Speech Communication* 85: 109–18. [CrossRef]
- Thir, Veronika. 2020. International intelligibility revisited; L2 realizations of NURSE and TRAP and functional load. *Journal of Second Language Pronunciation* 6: 458–82. [CrossRef]
- Tiono, Nani Indrajani, and Arlene Maria Yostanto. 2008. A study of English phonological errors produced by English department students. *K@ta* 10: 79–112. [CrossRef]
- Vennemann, Theo. 1987. *Preference Laws for Syllable Structure: And the Explanation of Sound Change with Special Reference to German, Germanic, Italian, and Latin*. Berlin: De Gruyter Mouton. [CrossRef]
- Verhaar, John W. M. 1996. *Asas-Asas Linguistik Umum*. Yogyakarta: Gadjah Mada University Press.
- Wang, Qian. 2023. An instrumental investigation of the vowel inventory of China English. *Asian Englishes* 25: 111–32. [CrossRef]
- Weinberger, Steven. 2015. *Speech Accent Archive*. Fairfax: George Mason University. Available online: <https://accent.gmu.edu> (accessed on 6 February 2023).
- Wijana, I Dewa Putu. 2003. Indonesian Vowels and Their Allophones. *Humaniora* 15: 39–42.
- Yan, Xin, Dawei Song, and Xue Li. 2006. Concept-based Document Readability in Domain Specific Information Retrieval. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management (CIKM '06)*. New York: Association for Computing Machinery, pp. 540–9. [CrossRef]
- Ylinen, Sari, Maria Uther, Antti Latvala, Sara Vepsäläinen, Paul Iverson, Reiko Akahane-Yamada, and Risto Näätänen. 2010. Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *Journal of Cognitive Neuroscience* 22: 1319–32. [CrossRef] [PubMed]
- Yong, Janet. Y. 2001. Malay/Indonesian Speakers. In *Learner English: A Teacher's Guide to Interference and Other Problems*, 2nd ed. Edited by Michael Swan and Bernard Smith. Cambridge: Cambridge University Press, pp. 279–95.
- Yuliarti, Yuliarti. 2014. Final Consonant Clusters Simplification by Indonesian learners of English and its intelligibility in international context. *International Journal of Social Science and Humanity* 4: 513–7. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.