

# THE DIPHTHONGS: THE OBSTACLES FOR INDONESIAN SPEAKERS OF ENGLISH

Desri Maria Sumbayak

English Department, Faculty of Teacher Training and Education, University of Riau

---

---

## ABSTRACT

*Article investigated the difficulties of Indonesian speaker of English in producing diphthongs /eɪ/ and /oʊ/. Five postgraduate students and five spouses of students at University of Canberra participated in this study. The participants were recorded in pronouncing /eɪ/ and /oʊ/ by reading lists of words and a story. The data were analysed by two Australian native speakers. Interrater reliability was calculated by using Cohen's Kappa. The percentage was used to see the accurate diphthong realisations. The results showed that diphthong /oʊ/ was relatively more problematic than diphthong/eɪ/ and the students produced more diphthongs accurately than the spouses. The results also revealed that the ability to produce the diphthongs accurately was influenced by English proficiency and the type of tasks where diphthongs were pronounced.*

**Keywords:** *diphthong, Indonesian speaker, English pronunciation*

## ABSTRAK

*Artikel menjelaskan kesulitan penutur Indonesia yang mengucapkan diftong /eɪ/ and /oʊ/ dalam bahasa Inggris. Responden terdiri dari lima pasangan mahasiswa Pascasarjana dari Universitas Canberra. Responden direkam ketika melafalkan diftong /eɪ/ and /oʊ/ pada bacaan berbahasa Inggris, yang kemudian dianalisis oleh penutur asli Australia. Realibilitas Interrater dihitung dengan Cohen's kappa. Persentase digunakan untuk menyatakan ketepatan pengucapan diftong. Hasil penelitian menunjukkan bahwa pengucapan diftong /oʊ/ lebih bermasalah dari pada diftong /eɪ/, dan para mahasiswa lebih mampu mengucapkan diftong tersebut dari pada pasangan mereka. Hasil juga menunjukkan bahwa kemampuan untuk mengucapkan diftong secara tepat dipengaruhi oleh kemampuan berbahasa Inggris dan jenis tugas yang membutuhkan pengucapan kedua diftong tersebut.*

**Kata kunci:** *diftong, penutur bahasa Indonesia, Pengucapan bahasa Inggris.*

---

---

## INTRODUCTION

There are often learning and production problems when mother tongue (L1) differs from target language (L2). The degree of differences between a learner's native language and the target language can lead to greater difficulty (Lightbown and Spada, 2006). In an EFL teaching context, this issue has been discussed in almost all the world's languages (eq. Walter, Duguid in Swain & Smith (2001) The areas of characteristic problems which have been reviewed are phonology, grammar and vocabulary and these findings might help teachers to examine how these typical difficulties arise and anticipate the difficulties in their teaching contexts. Phonology is the distinctive area in which one's native language often interferes with one's attempt to acquire English as second language (Carr, 1999).

Kelly (2000) points out some typical difficulties of English learners in producing vowels and diphthongs. Diphthongs (centring and closing) seem to be the most difficult ones to be produced by many speakers of other languages. For Indonesian speakers, AGPS (1986) mentions that there are possibilities for the learners to have problems in diphthongs which are not found in Bahasa Indonesia: /ɪə/, /ʊə/, /eɪ/, /eə/, and /oʊ/. On the other hand, he argues that due to the familiarity to the sound, Indonesian speakers should not have difficulties with three diphthongs that appear in Bahasa Indonesia: /ɔɪ/, /aɪ/ and /aʊ/. Yong (2001) claims that for Indonesian speakers, English diphthongs are likely realised as pure vowels which might be uttered in the long and short forms. Since diphthong is one of the most problematic features in pronunciation for speakers of other languages, this study focused on the two English diphthongs that are not found in Indonesian. Diphthongs /eɪ/ and /oʊ/ were chosen due to their frequent use in English compared to other diphthongs: /ɪə/, /ʊə/, /eə/.

### Literature Review

Based on teaching experiences and observation Hart (1965) discovered that Malaysian students whose mother tongue is Malay or a southern Chinese dialect (Hokkien, Cantonese, or Hakka) had problems with pronunciation of English tense markers and lax vowels. He points out that these difficulties are due to the interferences of Malay, Hokkien and Cantonese in which lax vowels are considered as allophones not as phonemics as they are in English. Similarly, Yong (2001) claims that due to the different phonological system between Malay/Indonesia and English, there are serious problems that can cause confusions in pronunciation. For example: most vowels are pronounced with more or less comparable length: *bit/beat*, *pill/peal*, *full/fool*, *cot/caught*, conflation of /æ/ and /e/ into /e/, and diphthongs are likely to be realized as pure vowels which are pronounced long and short. Alip (2007) points out that Indonesian speakers tend to pronounce English diphthongs in a lax manner but diphthongs in English should be tense. However, these arguments are not supported by experimental research that confirm these difficulties are vis a vis the interferences of L1.

In the Indonesian context, Mathew (2005) carried out research on the mispronunciations of English consonants; voiced stops, voiceless stop, sibilants, affricates and interdental for learners whose first languages are Indonesian, Gayo and Acehese. She argued that transfer, developmental factors, spelling interference, general processes and communicative strategies were the factors leading to the mispronunciations. The data was taken in several universities in Banda Aceh, Indonesia by involving 24 students in which every ethnic group consists of four male and four female students. The instruments were designed in four different types of tasks; minimal pairs, word repetition task, a reading passage and an interview. The consonants appeared in three different environments of sound; initial, middle and final. The data from the second task was justified by a native speaker of English whereas the other was evaluated by the author.

This research is considered to be rich since it is taken by using four different types of tasks. But the author fails to explain why the errors are justified separately by two different evaluators. It is

difficult to see the consistent justification done by the evaluators because the data from the second task is analyzed by another native speaker and the other data is justified by the researcher. Apart from that, the comparison among Indonesian, Gayo and Acehese is arguable. Indonesian is a national language (Yong, 2001:279) which is used by all Indonesian people, whereas Gayo and Acehese are languages which are spoken by Acehese people. There is no justification in terms of ethnicity, that is, the participants who are considered to have Indonesian as L1. It is probably more valid if the participants are the students who speak Gayo, Alas, Jame or Tamiang (other languages in Aceh).

Some scholars have given serious attention to the area of vowel production. Deterding (2003) conducted an instrumental research on the quality and length of monophthong vowels of Singapore English (SgE). The research involved five male and five female Chinese Singaporean students of the National Institute of Education in Singapore who were training to be school teachers. The interview was recorded in the Phonetics Laboratory of the National Institute of Education. The data was transcribed and compared to the recorded data of British English (BrE) vowels of five male and five female BBC broadcasters. The measurement was done by using computer-based spectrograms. The result revealed that some of the vowel distinctions that were clear in BrE were not maintained in SgE. There was a neutralization between /i:/ and /ɪ/ and also between /e/ and /æ/ and there was not much difference between /ɔ:/ and /ɒ/ produced by Singaporean in English.

Deterding (2003) does not give explanations on how the neutralization occurs in SgE. There is no link between L1 interferences and L2 monophthong productions, yet he interprets the neutralization as the factors of speaking rate, degree of stress and the influence of neighboring consonants in a simplistic explanation. He cross-cultural communication. The comparison between the students' monophthongs and BBC broadcasters' monophthongs remains a questionable validity of comparison. The BBC broadcasters might have a high standard and artificial level of BrE pronunciation which does not represent the authentic pronunciation of BrE. The comparison seems to be unbalanced.

Based on similar methodology in his monophthong research, Deterding, Wong, and Kirkpatrick (2008) conducted an extensive study which investigated the pronunciation of Hong Kong English. Fifteen fourth-year undergraduates at the Hong Kong Institute of Education were interviewed by a native speaker to get the data of pronunciation of initial TH, initial and final consonant clusters, L-vocalisation, conflation between initial [ŋ] and [l], monophthong vowels, the vowels in FACE and GOAT, vowel reduction, rhythm and sentence stress. The data was transcribed and justified by two listeners and was assessed by using a digital spectrogram Praat software.

They discovered a few idiosyncratic features. Firstly, there was a conflation for initial [l] and [ŋ]. Even though Deterding et al (2008) approached the pronunciation realizations as the identity of Hong Kong English rather than as the interferences of L1, yet in their discussion they were tempted to mention that this conflation is due to the influence of Cantonese. The most salient feature in this study was the use of [f] for initial TH in content words and the use of diphthong for words FACE and GOAT which were found in BrE rather than in English of South-East Asia countries. Similarly, Chang in Swain & Smith (2001:225) points out that TH does not occur in Chinese, so the learners seem to replace this by /t/, /f/ or /s/. However, for diphthong realizations Chang has different argument with Deterding, Wong & Krikpark' findings. Chang mentions that for Chinese learners, the diphthongs are pronounced short with "not enough distinction between the two component vowels". Chinese diphthongs are usually pronounced with quicker and smaller tongue and lip movements than their English counterparts. On the other hand, Deterding et al (2008) argued that Hong Kong students were better able to realize diphthongs which were similar to Br.E than to English of other South-East Asian countries. This reveals that nowadays, the pronunciation of Hong Kong English is not influenced by L1 but by London and South of England pronunciation.

A study on an acoustic comparison of English monophthongs and diphthongs produced by Australian and Thai speakers has been done by Tsukada (2008). The participants were six Australian students of psychology and fifteen Thai native speakers who were mostly university students. By reading list of words for five times, the participants were recorded and the recording was digitized by using CoolEdit program. The vowel duration and quality were analyzed by using wide-band spectrograms and time domain waveforms.

In production of monophthongs, she discovered that there was similarity between Australian and Thai speakers in terms of duration and quality. However, there was a significant difference between Australian and Thai speakers in producing diphthongs. Thai speakers pronounced diphthongs as long monophthongs. The author argued that the four monophthongs and two diphthongs that were used in this research (/ɪ æ ʊ ʌ eɪ oʊ/) were not found in the Thai vowel system. Thai participants were able to produce monophthongs like the Australian native speakers. Yet in producing diphthongs, Thai speakers substituted the diphthongs for the vowels in Thai. Tsukada (2008) claimed that this difficulty was possibly due to numerous diphthongs that Thai had. Nevertheless, there is no satisfactory elaboration on how Thai speakers have difficulties in producing diphthongs even though they have many diphthongs in Thai and how /eɪ/ and /oʊ/ are substituted by vowel in Thai.

The comparison between Australian students and Thai students seems logical compared to Deterding (2003) in monophthongs. The data of Australian English is considered to be naturalistic since it is produced by Australian undergraduate students. However, the author does not give any justification on why the number of Australian and Thai students is different. The comparison of the quality of monophthong and diphthong production might be partial due to the unbalanced number of Australian and Thai students. So, the reason behind the different numbers of participants of Australian and Thai speakers might contribute to the understanding of the result.

The previous studies on diphthongs emphasized on the quality and length of the diphthongs. This study focused on the accuracy and inaccuracy of the diphthong productions. Different to Deterding (2003) and Tsukada (2008) that used one type of tasks in collecting the data, this study employed two different types of tasks; reading lists of words and reading a story. Another different aspect in this study was the use of two different groups of participants. The use of two different groups revealed how the difficulties in producing diphthongs were varied for the speakers with the same L1.

## **The Study**

The study aimed to investigate the accuracies and inaccuracies of diphthongs /eɪ/ and /oʊ/ which were produced by Indonesian speakers of English. The study also compared two groups of Indonesian speakers; students and spouses of the students. The following research questions guided the study: (1) Are the Indonesian speakers able to produce diphthongs /eɪ/ and /oʊ/ accurately?; (2) Is there a difference between students and spouses in producing diphthongs /eɪ/ and /oʊ/ accurately?

## **METHODS**

The study was conducted by qualitative method with the following conditions.

### **Participants**

The participants of the study were ten Indonesian people who were mixed in terms of ethnicity. Indonesian is the L1 of all participants. They were two groups of participants. Firstly, group of students (four females and one male) who were studying as postgraduate students at University of

Canberra. Secondly, group of spouses (four females and one male) who were joining their partners as postgraduate students at University of Canberra. The postgraduate students had advanced level of English proficiency with IELTS score between 6 and 6.5. The spouses were mostly graduated from universities in Indonesia with intermediate to upper-intermediate level of English proficiency. They had been living and working as part time workers in Australia for more than one year. The participants were categorized into two groups to discover whether any difference for the speakers in producing the diphthongs regarding to the proficiency level and academic environment.

## Method & Instrument

To collect the data, two types of reading styles were used. Firstly, careful style in which the participants were asked to read two lists of words. List 1 consisted of eight words with diphthong /eɪ/ and eight words as distracters (Attachment 1). List 2 consisted of eight words with diphthong /oʊ/ and eight words as distracters as well (Attachment 2). To obtain the most naturalistic data, the distracters were given to take the participants' attention away from the target diphthongs. Secondly, semi-careful style in which the participants were asked to read a story (Attachment 3). The story was written by the author in which eight words with diphthongs /eɪ/ and eight words containing diphthongs /oʊ/ in the lists of words were used. Labov (1972, 79-86) mentions that the structure of language production and pronunciation is different from replies to other types of questions. It means that the more types of tasks are used in a research the richer the data will be obtained. Two types of reading styles were used in this research to seek how the types of tasks might influence the speakers to produce accurate diphthongs.

## Data Collection and Analysis

The data was collected by recording the participants in reading the lists of words and a story. The diphthong productions were justified by two Australian native speakers as the raters. One of the raters was an English teacher at Canberra Institute of Technology and another rater was a postgraduate student at University of Canberra. One teacher and one non teacher were involved in this study to get a moderate justification between the two raters. To see the agreement between them, the interrater reliability was measured by using Cohen's Kappa.

For the data analysis, the justifications from the two raters were analysed separately. The percentage was employed to calculate the accurate pronunciations of the participants in producing diphthongs /eɪ/ and /oʊ/. From the percentage of the accurate productions, the difficulty level of the diphthongs and the difficulties of each group of participant could be seen clearly.

## RESULT AND DISCUSSIONS

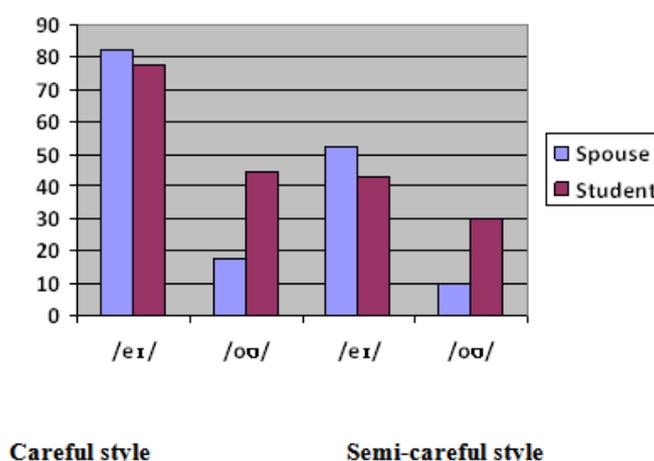
According to rater 1 as reflected in graph 1 (based on the data in attachment 4), it was interesting to find out that the spouses were able to produce more diphthong /eɪ/ accurately than the students. In careful style, of total 40 words containing diphthong /eɪ/, the spouses pronounced diphthong /eɪ/ accurately for 82.5%. In comparison, the students, in the case of same diphthong, were only able to produce 77.5%. Similarly, in semi-careful style, it also showed that the spouses produced more diphthong /eɪ/ accurately than students. Of total 40 words in semi-careful style, 52.5% were pronounced correctly by the spouse, whereas, the students were only able to produce 42,5% correctly.

Diphthong /oʊ/ seemed to be more difficult than diphthong /eɪ/ for all the participants because the highest percentage in pronouncing this diphthong correctly was less than 50% which produced by the spouses in careful style. In the case of diphthong /oʊ/, students produced more

diphthong /oʊ/ correctly than the spouse. There was a substantial difference between students and spouses in producing this diphthong. In careful style, of total 40 words containing diphthong /oʊ/, the students produced 45% accurate diphthong /oʊ/ whereas spouses only produced 17.5%. This represented a difference of 23.5%. In semi-careful style, the difference of diphthong /oʊ/ accuracy was similar, a difference of 20%.

From this description, it appeared that the spouses were able to produce more diphthong /eɪ/ accurately whereas for relatively difficult diphthong /oʊ/, students pronounced the diphthong more accurate than the spouses. Figure 1 also shows that consistently the participants pronounced correct diphthongs more in careful style rather than in semi-careful style.

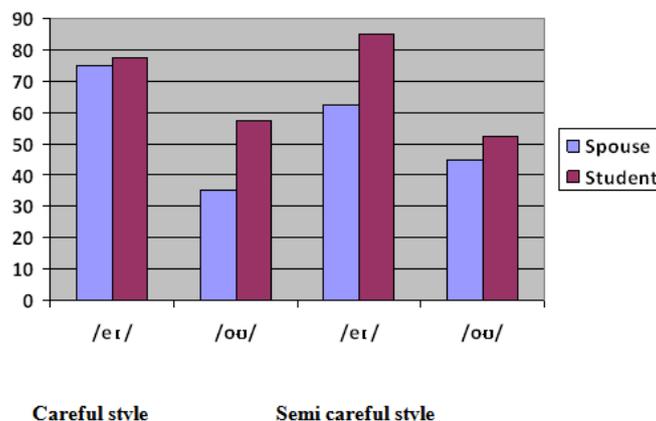
Figure 1 Percentage of the accurate production of diphthongs /eɪ/ and /oʊ/ in rater 1



In Figure 2, the justification from rater 2 was slightly different to rater 1. In general, Figure 2 shows that consistently the students produced more accurate diphthongs /eɪ/ and /oʊ/ than the spouses. In careful style, out of 40 words, the students produced 77% of diphthong /eɪ/ in the lists of words correctly whereas the spouses produced 75% accuracy. There is no significant difference between students and spouses in this case. Interestingly, in semi-careful style the students produced 85% correct diphthong /eɪ/ of 40 words, whereas in careful style the students produced 77% accuracy. The students produced more diphthong /eɪ/ correctly in semi-careful than in careful style. This is beyond the author's expectation since it is assumed that reading lists of words in careful style should give opportunities for the speakers to pronounce diphthongs correctly.

From this Figure, diphthong /oʊ/ also seemed to be more problematic than diphthong /eɪ/. The highest percentage for the participant in producing diphthong /oʊ/ out of 40 words was 57%. This high percentage was achieved by the group of student in careful style. There was a significant difference between the students and spouse in producing diphthong /oʊ/ in careful style, a difference of 22%. However, in semi-careful style there was no substantial difference between these two groups, even the spouses produced more diphthong /oʊ/ correctly than in careful style. Again, this result is beyond the author's expectation as aforementioned. The result from rater 2 revealed that in some cases, reading lists of words in careful style did not consistently give opportunities for the speakers to produce more diphthongs accurately.

Figure 2 Percentage of the accurate production of diphthongs /eɪ/ and /oʊ/ in rater 2



## Interrater reliability

Interrater reliability was calculated to find out the agreement and disagreement between the raters. Mackey and Gass (2005:243) mention that there are two main ways to get the interrater reliability; simple percentage agreement and Cohen's Kappa. In this study Cohen's Kappa was used because Cohen's Kappa requires the raters to justify forms as targetlike and non-targetlike (Mackey and Gass, 2005:243). Cohen's Kappa seemed to be appropriate in this study since the data of this study was analysed by two raters as correct or incorrect compared to the target like; the pronunciation of Australian native speakers. By using Cohen's Kappa, the interrater reliability between rater 1 and rater 2 in this study was 0.39933 which was considered as a *fair agreement* between two raters. The complete calculation of the interrater reliability can be seen in attachment 5.

The reason that could explain the fair agreement and disagreement between the raters was the different background of them. Rater 1 was an English teacher whereas rater 2 was a post-graduate student at university. From their justification, rater 1 seemed to be more tough and consistent than rater 2. Rater 1 gave less correct justification in overall productions (146 out of 320). But she seemed to be more consistent in justifying the data in terms of style that was the participants produced more diphthongs accurately in careful style than in semi-careful style. In comparison, rater 2 seemed to be more generous in justifying correct pronunciation of total productions (196 out of 320). Yet he seemed to be more inconsistent, since his justification seemed to be more varied than rater 1. There was inconsistency of his analysis in semi-careful style twice; diphthong /eɪ/ for the group of students and diphthong /oʊ/ for spouses. The rater's attitudes and experiences towards English might be different. As a teacher, rater 1 was probably more sensitive to listen to the correct pronunciation than rater2. The interrater reliability might reach a strong agreement and disagreement if the raters were teachers with similar background of experiences. In addition, to avoid of human errors, the justification could also be done by using software program as carried out by Deterding (2003) and Tsukada (2008).

## DISCUSSION

The study confirmed that there were difficulties for Indonesian speakers to pronounce English diphthongs accurately that are not found in Indonesian (AGPS, 1986; Yong, 2001). The similar inaccuracies were also shown in producing monophthongs and diphthongs of English in other South-East Asian countries (Singapore, Thai and Hong Kong), as have been discussed by Deterding and

Tsukada (2008). Yong (2001) claims that for Indonesian speakers, English diphthongs are likely realised as pure vowels which might be uttered in the long and short forms. Based on the interview with the raters, it was found that the participants tend to realise diphthongs /eɪ/ and /oʊ/ as long /e/ or /o/. Instead of producing the diphthongs, the participants substitute the diphthongs to vowels which are common in Indonesian. It is undeniable that participants might be influenced by Indonesian in producing English diphthongs.

For Indonesian speakers, diphthong /oʊ/ seemed to be more problematic than diphthong /eɪ/. There are three diphthongs in Indonesian e.g. /ɔɪ/, /aɪ/, /aʊ/ (AGPS, 1986). Diphthong /eɪ/ seems to be more similar to diphthong /aɪ/ which is used most frequently in Indonesian. Besides, in many cases, there are tendencies to substitute diphthong /aɪ/ to vowel /e/ or sometimes with unclear /eɪ/. Thus, it is probably more familiar and easier for Indonesian speakers to pronounce /eɪ/ rather than /oʊ/.

Students were able to produce accurate diphthongs more than the spouse. But in spite of this, for diphthong /eɪ/ spouses and students do not show significant difference in pronouncing accurate diphthongs. Even in the case of the justification from rater 1, spouses were able to produce diphthong /eɪ/ slightly more than the students. This revealed that there was a difference of difficulty for the speakers of same L1. For relatively difficult diphthong /oʊ/, students significantly produce accurate productions more than the spouses. It can be inferred that English proficiency and academic involvement might influence the pronunciations. The students' use of complex words in their academic environment might enable them to produce difficult diphthongs. The pronunciation difficulties in L2 speakers are not fossilised. Native speaker like pronunciation can be achieved by L2 speakers as long as the speakers have intensive interaction to the language inputs, particularly in the academic environment.

The different context where diphthongs were pronounced also influences the accuracy of pronunciation. Two different styles which were used in this study shows that in most cases, speakers tend to produce more diphthongs accurately in careful style (list of words) rather than in semi-careful style (reading a story). Labov (1972) points out that different types of tasks will result in different language production and pronunciation. Similarly, Littlewood (1984:82) says that the type of situation or task is one of the factors that influence the variability of learners' pronunciation. Furthermore, he argues that "...as learners devote more conscious attention to their speech, they come closer to target norms". The use of lists of words as careful style in this study gives the participants more consciousness to pronounce diphthongs slowly, carefully and accurately. On the other hand, when the words that containing diphthongs were incorporated in a sentence and text level, the participants tended to have less consciousness to produce accurate pronunciation.

## CONCLUSION

There were difficulties for Indonesian speakers to pronounce English diphthongs /eɪ/ and /oʊ/ accurately. They tended to substitute the diphthongs to long vowels /e/ and /o/. The students were able to produce the diphthongs more accurately than the spouses particularly in producing relatively difficult diphthong /oʊ/. The ability to pronounce diphthongs was influenced by English proficiency and the use of English in academic environment. Producing diphthongs in careful style gave more opportunity for the participants to produce accurate diphthongs rather than in semi-careful style.

## REFERENCES

- AGPS. (1986). *Asian Language Notes-Some Likely Areas of Difficulty for Asian Learners of English, No.3. Indonesian/Malay. 2<sup>nd</sup> edition.* Canberra: Australian Government Publication Service.
- Alip, F.B. (2007). Localization of English Phonology in the Indonesian Context. *Phenomena, Journal of Language and Literature*, 10/2, 159-167.
- Carr, P. (1999). *English Phonetic and Phonology: An Introduction.* Oxford: Blackwell.
- Chang, J. (2001). Chinese Speakers. In Swain, M. & Smith, B. (Eds). *Learners English: A Teacher's Guide to Interference and Other Problems.* (pp.279-295). Cambridge: Cambridge University Press.
- Deterding, D. (2003). An Instrumental Study of the Monophthong Vowels of Singapore English. *English World Wide*, 24/1, 1-16.
- Deterding, D., Wong, J., Kirkpatrick, A. (2008). The Pronunciation of Hongkong English. *English World-Wide*, 29/2, 148-175.
- Hart, C.D. (1969). Some English Pronunciation Difficulties in Malaysia. *ELT Journal*, 23/3, 270-273.
- Kelly, G. (2000). *How to teach pronunciation.* Essex: Longman.
- Labov, W. (1972). *Sociolinguistic Patterns.* Oxford: Blackwell.
- Lightbown, P.M. & Spada, N. (2006). *How Languages are Learned.* Oxford: Oxford University Press.
- Littlewood, W.T. (1984). *Foreign and Second Language Learning.* Cambridge: Cambridge University Press.
- Mathew, I. (2005). Errors in pronunciation of consonants by learners of English as a foreign language whose first languages are Indonesian, Gayo and Acehese. *Monash University Linguistics Papers*, 3/2, 29-44.
- Mackey, A., and Gass, S.M. (2005). *Second Language Research: Methodology and Design.* New Jersey: Erlbaum.
- Swan, M. and Smith, B. (Eds.) (1987) *Learner English: A teacher's guide to interference and other problems.* New York: Cambridge University Press.
- Tsukada, K., (2008) "An acoustic comparison of English monophthongs and diphthongs produced by Australian and Thai speakers", *English World-Wide* 29, pp.194-211(18).
- Yong, J. Y. (2001) Malay/Indonesian speakers. In M. Swain and B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (pp. 279-295). New York: Cambridge University Press.