4

English Vowels Produced by Javanese and Sundanese Speakers

A preliminary version of this chapter appeared as:

Perwitasari, A., Klamer, M., & Schiller, N. O. (2016). Formant frequencies and vowel space area in Javanese and Sundanese English language learners. *3L: The Southeast Asian Journal of English Language Studies, 22*(3), p141-152. DOI: 10.17576/3L-2016-2203-10.

Perwitasari, A., Klamer, M., Witteman, J., & Schiller, N.O. (2015). Vowel duration in English as a second language among Javanese learners. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: The University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0392.1-4. Retrieved from http://www.icphs2015.info/pdfs/Papers/ICPHS03.

Abstract

First language (L1) vowel systems play an important role in the vowel production of a second language (L2). In this study, the focus is specifically on Javanese and Sundanese - two of the most widely spoken Indonesian local languages. This present study investigated how the Javanese and Sundanese speakers produce ten English vowels. Forty Javanese and Sundanese speakers and ten native English speakers participated in the experiment. According to the Speech Learning Model (SLM), highly advanced Javanese and Sundanese speakers of English should continue have trouble producing vowels that are *similar*, such as /iː/ and /uː/, but should no longer exhibit native-language interference with *new* L2 vowels, such as /ɪ, ɛ, ʊ, æː, ɑː, ɔː, ʌ, ɜ˞/. In contrast, the Second Language Linguistic Perception (L2LP) model predicts that the production of *new* L2 vowels is more difficult than of *similar* L2 vowels - as long as the L2 acquision process has not been completed. The results show that that the Javanese and Sundanese speakers have more difficulty with the *new* than with the *similar* vowels in English, which indicates that the L2 acquisition process has not been completed. Moreover, the members of English tense-lax vowel pairs are poorly contrasted by spectral parameters while the use of duration is relatively adequate.

Keywords: *phonetics*, *bilingualism*, *second language production*, *second language acquisition*

4.1 Introduction

Producing the English vowels would be problematic for the Javanese and Sundanese learners of English. Cross-linguistic studies reveal the effects of an L1 vowel system with L2 vowel production. The L2 learners are predicted to use the categories from their L1 vowel system and apply them to L2 production. This situation may present advantages if the L1 has a complex vowel system. Speakers with a large L1 vowel system may be more successful in approximating the vowel categories of an L2 with a small vowel inventory by substituting the nearest category from their L1 and changing the L1 category representation to match the L2 vowels by creating mergers or compromise categories (Flege, 2003; MacKay, Meador, & Flege, 2001). For instance, McAllister et al. (2002) found that English speakers, with a large L1 vowel system, performed well when producing the five vowels of Spanish. Conversely, L2 vowel production is going to be problematic for L2 learners whose L1 vowel system has a smaller inventory than the target language. Iverson and Evans (2007) revealed that Germans and Norwegians, who have a complex L1 vowel system, were more accurate recognizing English vowels than French and Spanish speakers who have smaller L1 vowel systems.

To predict the difficulties that will be experienced by adult foreign- language learners, the vowel quality (i.e. formant frequencies) and vowel quantity (i.e. duration) of the L2 speech sounds produced by the learners should be investigated.

4.1.1 Formant Frequencies

Generally, the phonetic quality of vowel sounds can be expressed in terms of acoustic properties by measuring the center frequencies of the lowest two or three resonances of a speaker's vocal tract. The articulatory dimension of vowel height correlates very well with the lowest resonance of the vocal tract, also called the first formant (F_1). The constriction place of a vowel along the articulatory front-back dimension together with the degree of lip rounding (co-defining the length of the oral cavity), correlate very well with the second lowest resonance, called the second formant (F_2) (Fant, 1973; Gimson, 1980; Cruttenden, 2001). Formant frequencies (correlating with vowel quality) affect a speaker's intelligibility and can be used to assess the accuracy of pronunciation, as well as the naturalness of speech (Peterson & Barney, 1952; Hillenbrand & Nearey, 1999). L1 speakers

with a large F_1 range were found to have higher intelligibility scores than the L1 speakers with a narrow F_1 range (Bradlow et al., 1996; Hazan & Markham, 2004). The F_1 range is correlated with the intelligibility of words (Hazan and Markham, 2004), not so much with the understanding of sentences (Bradlow et al., 1996).

The effects of first language (L1) vowel systems on second language (L2) acquisition have been cross-linguistically assessed. In production tasks, if an L1 has a complex vowel system, the spectral vowel space is predicted to be crowded (Flege, 1995, 2003). Such a crowded vowel space leaves less room for new a vowel category needed in an L2 and brings disadvantages in learning L2 vowels. This prediction, however, seems to be unresolved for an L1 vowel system with a small number of categories (Meunier et al., 2003). If an L1 has a small vowel system, the spectral vowel space will be less crowded but if two or more different L2 vowel sounds partially overlap with the same category in the learner's L1, they may all assimilate to this single L1 category (Iverson & Evans, 2007) and the learner will be unaware of the contrast in the L2. The current study seeks to contribute to this area of study by examining the use of vowel quality (as evidenced by formant frequencies F_1 and F_2) and vowel duration in the production of English vowel sounds by learners from Indonesia with different regional languages with relatively sparse vowel systems, namely, Javanese - with six vowel phonemes - and Sundanese - with seven vowels (see Chapter 2 for more information on, and an acoustic study of, these vowel systems).

4.1.2 Vowel Duration

There are some languages where vowels are distinguished only by duration. For instance, Danish distinguishes long and short vowels and Estonian has short, long, and super long vowels (Lehiste, 1970). Similarly, German has 14 vowels which are organized in seven pairs the members of which differ only in duration while the vowel quality difference within the pairs is negligible (Strange et al., 2004). In Spanish, the duration is not used at a phonological level (e.g. McAllister et al., 2002; Escudero & Boersma, 2004). Thus, a cross-linguistic difference in vowel quality and duration between languages may be noticeable in the production of second language vowels (Iverson & Evans, 2007; Bent, Bradlow, & Smith, 2008).

Previous research on second language learners has shown that the production of English vowels by L2 learners also varies depending on the native language in terms of duration. For instance, Bohn and Flege (1990, 1992) found that the vowel duration of the English vowels

/ɛ/ and /æ/ produced by experienced and inexperienced German English learners differed significantly from that of native English speakers. The inexperienced L1 German speakers produced a shorter /ɛ/ than their native English counterparts, while the experienced L1 German speakers matched the duration of native English speakers. Interestingly, for the vowel /æ/ which has no counterpart in their L1, both L1 German groups displayed shorter durations than the native English speakers. In related research, Munro (1993) compared the speech production pattern of Arabic and English vowels. The study found that the L1 Arabic speakers exaggerated the difference in duration between the English tense and lax vowel pairs as Arabic speakers contrast long and short vowels in their first language.

Bohn (1995) reported that Spanish and bilingual Catalan-Spanish English learners rely on the duration of vowel production to a greater extent than native English speakers, even though vowel length distinction does not exist in L1 Catalan and Spanish. In another study, Makarova (2010), who examined native Russian speakers who studied English, concluded that L2 learners went through a higher degree of overreliance on duration in /ɛ - æ/ than in /i - ɪ/. The stage of overreliance was the greatest for the low vowel pair /ɛ - æ/ and the least for the back-vowel pair /u - ʊ/. Likewise, Lin (2013) found that L1 Mandarin speakers relied simply on duration cues to distinguish the English /ɪ/ from the neighboring /iː/. Earlier, the prominent use of duration over spectral properties in distinguishing the ten monophthongs of American English by Chinese learners was demonstrated by Wang (2007) and Wang and Van Heuven (2006). These studies have claimed that non-native speakers relied on L2 contrastive length even though their first language did not exploit the length feature.

The current study explores the formant frequencies (F_1 and F_2 values) and vowel duration produced by Javanese and Sundanese English language learners. This chapter first reviews the relevant literature. It will then focus on the methods, results, and report a speech production experiment. L2 speech production problems are investigated using traditional statistical analysis and Linear Discriminant Analysis (see method section).

4.2 L2 Learning Models

Language interference is a phenomenon that describes the transfer of L1 rules to the learning of L2 (Flege, 1995). The L1 rules happen to be the language filter in the learning process which may either facilitate or inhibit the L2 learning. If the L1 and L2 rules have some similarities, L1

norms facilitate the learning through positive transfer. Conversely, if the L1 and L2 rules have some differences, the negative transfer often takes place. Krashen (1981) mentioned that only the negative transfer is called interference.

Below we discuss two different L2 learning models. First, L2 speech production models are introduced to underpin the present investigation of the formant frequencies. Second, feature dependent models are discussed, which may specifically shed light on L1 interference in terms of the use of duration.

4.2.1 L2 Speech Production Models

Two prominent models of L2 learning are Flege's Speech Learning Model (SLM) (1995, 1999, 2002) and the Second Language Linguistic Perception (L2LP) model (Escudero, 2005, 2009). According to SLM, L2 learners can accurately produce L2 sounds if they have an accurate understanding of L2 sound properties and the phonetic distance between L1 and L2 sounds. SLM hypothesizes that L2 learners will always be relatively unsuccessful in learning L2 sounds that are *similar* to L1 sounds, typically represented by the same IPA base symbol and differing in diacritics only (Flege, 1997). The reason is that the similarity between the L1 and L2 sounds would block the formation of a new phonetic category but instead lead to the formation of a single supercategory that comprises the L1 and L2 tokens - at the cost of a shift of the category prototype. In contrast, L2 learners would ultimately (that is, after the completion or fossilization of the learning process) experience no challenges in perceiving *new* sounds (Flege, 1997). *New* L2 sounds, which are different from L1 categories and have no phonetic counterpart in the L1, enable the learners over time to develop new L2 categories (Flege, 1997). SLM does not address the early stages of L2 acquisition and makes no prediction of the relative difficulty of similar versus new sounds in the earlier stages of L2 acquisition.

L2LP postulates that, as long as the acquisition process is still ongoing, L2 learners will face production problems specifically in acquiring *new* sounds. The reason is that the L2 learners not only have to create new categories and perceptual mappings, but also integrate the already categorized dimensions with the newly categorized dimensions (Escudero, 2005). In contrast, L2LP predicts that the L2 learners will easily produce reasonably acceptable *similar* L2 sounds since the L1 and L2 sound categories would be phonologically equivalent (Escudero, 2005).

4.2.2 Feature Dependant Models

Current theories of non-native vowel production are often referred to as the Feature Hypothesis (McAllister et al., 2002) and the Desensitization Hypothesis (Bohn, 1995). McAllister's Feature Hypothesis claims that L2 features that are not contrastive in L1 are difficult to acquire. The difficulty in producing phonetic features will be reflected in a low production accuracy of these features in L2 speech production. McAllister et al. (2002) argued that re-attunement to duration can be difficult for second language learners with an L1 that does not exploit this feature phonologically. Additionally, the hypothesis predicts that these L2 learners may be better able to attune their phonological systems to the spectral, rather than the durational, cues of the L2. This is because the spectral cues are used in the L1 phonological contrasts while the durational cues are not.

A contrasting idea is proposed by Bohn's Linguistic Desensitization Hypothesis, which states that L2 learners are sensitive to durational cues when acquiring L2 vowels. According to Bohn (1995), whenever spectral differences are insufficient to differentiate vowel contrasts because previous linguistic experience did not sensitize speakers to these spectral differences, duration differences will be used to differentiate the non-native vowel contrast (see also Chapter 1, section 1.3.4.2). Because vowel duration is easy to access and is salient, the hypothesis predicts that L2 learners will employ durational information, which is contrastive in the L1 (Bohn, 1995). Bohn argues that late learners can detect temporal differences between the members of an unfamiliar English tense-lax contrast more readily than the spectral differences even if the learner's native language has no length contrast at all.

4.3 Javanese, Sundanese and English Vowel System

The phonetic and acoustic category differences of L1 Javanese, Sundanese, and English motivated the present investigation. As explained in Chapter 2, Javanese vowels are grouped into six phonemes, /a, ə, i, u, e, o/, Sundanese has seven vowels /i, ɨ, u, e, ə, o, a/. American English has a complex vowel system with ten monophthongs, /iː, ɪ, ɛ, æː, ɑː, ɔː, ʊ, uː ʌ, ɝ/ (Ladefoged, 2001, 2006). The vowel diagrams of Javanese, Sundanese, and English are shown in Figure 4.1. In contrast to Javanese and Sundanese, vowel duration in English plays a major role in its phonological system (Jones, 1957). Moreover, durational and

spectral information is used to categorize vowels as tense or lax (Bohn & Flege, 1992). Long vowels in English are claimed to be tense and short vowel are pronounced lax (Chomsky & Halle, 1968). This reflects the fact that the short vowels are articulated with less muscular tension (Gimson & Cruttenden, 1994).

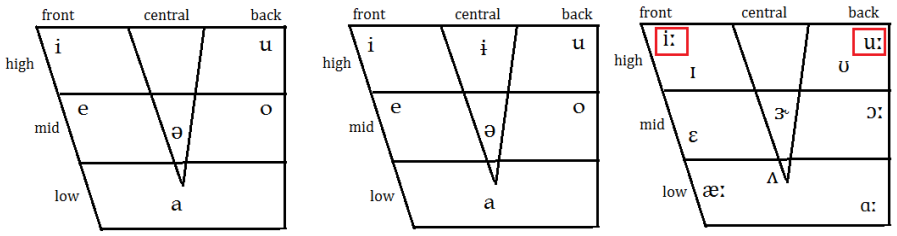

Figure 4.1 Articulatory plots for vowels of Javanese (left), Sundanese (middle) and English (right). The vowels in squares are considered similar L2 vowels for Indonesian learners of English.

We set two hypotheses based on two different types of L2 learning models mentioned in the previous subchapter.

4.3.1 L2 Speech Production Hypothesis

Inspection of the plot reveals that the position of English /iː/ and /uː/ lie roughly in the same location in the vowel space as the Javanese and Sundanese sounds /i/ and /u/ sounds. However, English /iː/ and /uː/ are tense and long, and accordingly differ from their closest counterpart on the Indonesian language in the diacritic length mark only. This makes them strong candidates for the status of similar vowels. The English /ɪ, ɛ, ʊ, æː, ɑː, ɔː, ʌ, ɜˑ/, on the other hand, find themselves in locations that are not used phonemically in the Indonesian languages; they are transcribed with IPA base symbols that are not used for any Javanese or Sundanese vowel phoneme, and should therefore be considered new vowels in SLM terminology.

To examine the production patterns of English vowels by Javanese and Sundanese English language learners, we address two research questions regarding formant frequencies: would Javanese and Sundanese learners show the same formant frequencies producing new L2 vowels /ɪ, ɛ, ʊ, æː, ɑː/, /ɔː, ʌ, ɜˑ/ in an authentic, native-like way? Would the production of similar L2 vowels /iː/ and /uː/ show subtly different formant frequencies between the Indonesian L2 learners and the L1 speakers of English?

According to SLM, the Javanese and Sundanese speakers may show different formant frequencies as compared to English speakers in producing the similar L2 vowels /iː/ and /uː/. However, Javanese and Sundanese speakers would be capable of producing the same formant frequencies with the new L2 vowels /ɪ, ɛ, ʊ, æː, ɑː, ɔː, ʌ, ɝ/. In contrast, according to L2LP the Javanese and Sundanese speakers will have different formant frequencies than native speakers when producing the new L2 vowels and will not produce different formant frequencies when pronouncing the *similar* L2 vowels /iː/ and /uː/.

4.3.2 Feature-dependant Hypothesis

Based on the Feature-dependent Hypothesis, the L2 learners will have difficulty in producing the target vowel duration if the duration cue is not exploited in the L1. Specifically, Javanese and Sundanese learners of English are predicted to have difficulties in producing the L2 vowels /iː, ɝ, ɑː, ɔː, uː/ and possibly /æː/ with longer duration and are expected to pronounce L2 short sounds /ɪ, ɛ, ʌ, ʊ/ successfully.[3]

In contrast, according to the Desensitization Hypothesis, Javanese and Sundanese speakers will have no difficulty in pronouncing the long and short vowels of English with a clear contrast in duration as they will generally be sensitive to length differences in any language (including English). The durational cues are predicted to be available for the Javanese and Sundanese speakers, even though the information is not found in their first language.

4.4 Method

4.4.1 Participants

A total of fifty participants took part in the speech production experiment. Based on their first language background, they were divided into three groups: English speakers, Javanese speakers (JE), and Sundanese speakers (SE).

The JE and SE participants all came from Central and West Java. They were students from various universities in Yogyakarta. The mean age for the twenty Javanese speakers was 21.9 (10 female, 10 male, age

[3] In American (but not in British) English /æː/ is generally considered (and shown to behave like) a phonetically tense and long vowel (Strange et al., 2004; Wang & Van Heuven, 2006). See also section 1.2, pp. 12-13.

range: 20-30 years, SD = 1.16) and for the twenty Sundanese speakers, it was 22 (10 female, 10 male, age range: 21-32 years, SD = 2.41). The average age at which the Javanese and Sundanese speakers began learning English was 8.7 (Javanese: SD = 1.5; Sundanese: SD = 2.5). The Javanese and Sundanese participants had no history of traveling abroad. They were proficient in their first language in the sense that they were still using their L1 in their daily lives. The Javanese and Sundanese participants had learned English for a minimum of 9 years in formal education (JE: M= 11.75 years, SD = 2.2; SE: M= 9.8 years, SD = 2.4).

The ten English speakers had come from the central and western areas of the United States. Their mean age was 26.2 (5 female, 5 male, age range: 23-36 years, SD = 1.75). All participants were tested at either Gadjah Mada University in Yogyakarta or Padjajaran University in West Java - both in Indonesia. Participants signed an informed consent form.

4.4.2 Stimuli

In this study, only the American-English monophthongs were taken into consideration. The Javanese, Sundanese, and American English speakers produced the monophthongs, /iː, ɪ, ɛ, æː, ɝ, ʌ, ɑː, ɔː, ʊ, uː/. The vowels were embedded in two different consonantal contexts - namely in /bVd/ and /hVd/ syllables. The ten /bVd/ items were *bead*, *bid*, *bed*, *bad*, *bird*, *bud*, *body*, *bawd*, *Buddhist*, and *booed*. In the /hVd/ context, the items *heed*, *hid*, *head*, *had*, *heard*, *hudd*, *hod*, *hawed*, *hood*, and *who'd* (Peterson & Barney, 1952; Ladefoged, 2001). All target items were embedded in the sentence frame "*I say (bVd)/(hVd) again*". Sentences were presented in print on a computer screen. During the recording, subjects produced the sentences twice. The stimuli were digitized and loaded into Stimuli Experiment 1.0 (Figure 4.2).

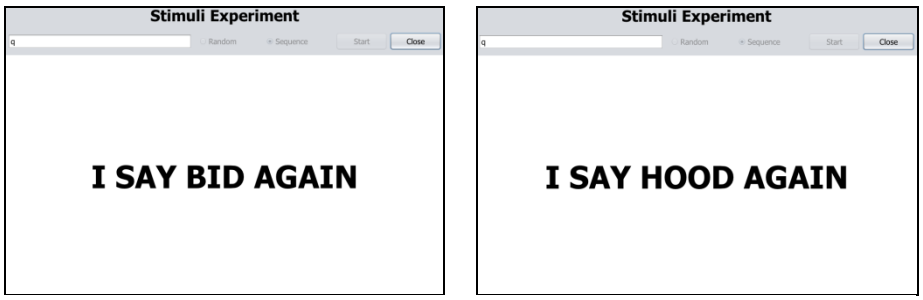

Figure 4.2 Examples of two stimulus strings as displayed on the computer screen.

4.4.3 Procedure

Before the production experiment took place, the researchers gave the explanation about the procedures. Then, all participants completed a brief sociolinguistic questionnaire and a consent form. The first part of the questionnaire elicited demographic information and inquired about experiences with both their native and second/foreign languages. The subjects reported their parents' first language and how often they used their native language. They stated their choice of language at home and at school. The second part of the questionnaire explored the subjects' background in the second language. The non-native speaker subjects reported their ages when they began learning English and how long they had been studying English. They listed any second or foreign language that they had studied and their competence level in each language. The Javanese and Sundanese subjects mostly listed Arabic and Japanese as other non-native languages that they had studied. They shared how long they had studied English and they confirmed that they had never lived in any English-speaking country.

After completing the questionnaire, participants took first and second language proficiency tests in order to measure the L1 and L2 competence among the Javanese or Sundanese learners. The first language test was taken from the Indonesian national exam, either in Javanese or Sundanese. The second language test was taken from a written English vocabulary test by Meara (2010). They had to read through the list of words carefully. If they knew what a word meant, they wrote Y (for YES) in the box and if they did not know what it meant, or if they were not sure, they wrote N (for NO) in the box.

All participants then received a short introduction monologue which contained words which would later be used as stimuli for the recording. This introductory text was presented on a computer screen. Next, the researcher explained the experiment and the recording procedures that would be involved. The subjects were given as much time as they needed and were encouraged to ask and comment at any point during the explanation. During the recording stage, each of the subjects sat in front of a computer display. Once the stimuli appeared on the screen, subjects produced the sentence, for instance "*I say bad again*". All of the stimuli were presented twice in random and sequenced orders.

All participants were recorded in a sound-attenuated room. Items were digitized (44.1 kHz, 16-bit sampling) using a digital audio recorder (H4N Zoom) and an adjustable microphone headset (Sennheiser PC

141). The distance that the microphone was set away from the speaker's mouth was approximately 3 cm to create a constant recording level for the entire session for every subject. After the completion of each experiment, subjects were given a post-experiment questionnaire. This questionnaire was given to obtain information with regard to the subjects' experiences in producing the stimuli. Afterwards, they received their compensation and were allowed to share any concerns about the experiment in a written form.

4.4.4 Analysis

The study utilized Praat 5.3.56 (Boersma & Weenink, 2013) for annotating speech. The Javanese and Sundanese groups each produced 800 English vowels (20 speakers × 2 contexts × 10 vowels × 2 repetitions), and the American-English group produced 400 vowels (10 speakers × 2 contexts × 10 vowels × 2 repetitions). The total corpus of data amounts to 2000 vowels.

The target vowels were segmented from their spoken context using the waveform and spectrogram representation in Praat. The beginning of the vowel was defined as the first glottal pulse with no visible noise due to either the preceding /b/-burst or breathiness of the prevocalic /h/. The end of the vowel was located at the earliest point in time where the formants had disappeared from the spectrogram and only the voice bar remained visible. Formant frequencies of the participants' speech were estimated using the Burg algorithm implemented in Praat. To assist the researcher, formant tracks were superimposed on the wideband spectrogram. Whenever there was a visual mismatched between the formant tracks and the spectrogram, the upper frequency of the analysis band or the model order of the LPC analysis was adjusted by trial and error until a satisfactory match was obtained for at least F_1 and F_2. Using a Praat script, the vowel duration and the mean values of F_1 and F_2 were extracted for the steady state portion of the vowel between 25 and 75% of the vowel duration and written to disk for later off-line data analysis.

We tested the hypotheses using two separate types of analysis. First, we performed a series of repeated measures ANOVAs using SPSS 22.0 (IBM, 2013). The RM-ANOVA was conducted to examine the effect of the independent variables L1 (Javanese, Sundanese, English), VOWEL (/iː, ɪ, ɛ, æː, ɜ˞, ʌ, ɑː, ɔː, ʊ, uː/). The dependent variables were the F_1 and F_2 values and vowel duration. As the present thesis is not concerned with co-articulatory effects of consonants and is only concerned with vowel

acquisition *per se*, main and interaction effects of the consonantal context will not be reported and discussed in the thesis. The interested reader is referred to Appendix B to inspect effects of consonantal context. To follow up the differences between Javanese vs. English and Sundanese vs. English, we conducted independent sample Mann-Whitney U tests to determine whether there was a statistically significant difference in the means of F_1 and F_2 between the two groups (Javanese vs. English and Sundanese vs. English). Bonferroni-corrected statistical thresholds are reported when applicable.

Second, we used LDA (Linear Discriminant Analysis, see Klecka, 1980; Weenink, 2006) to test what vowels are different between L1 and L2. Using LDA an algorithm can be trained to optimally categorize vowels based on the first two formant values and duration of native speakers. This algorithm can then be viewed as the machine (algorithmic-) equivalent of a native listener. Subsequently, the non-native (Javanese and Sundanese) productions of the various American English vowels can be fed to the classifier and the pattern of (mis-)classification by the classifier can then inform us about what vowels are particularly difficult for the non-native speakers to produce (see e.g. Strange et al., 2004; Wang & Van Heuven, 2006; Wang, 2007; Van Heuven & Gooskens, 2017).

The first two formants (F_1, F_2) were first transformed to the psychophysical Bark scale using Traunmüller's (1990) formula. Furthermore, to subtract out inter-individual differences in the morphology of the speech production apparatus both Bark transformed formant values and duration were z-normalized (Lobanov, 1971). Linear discriminant analyses were subsequently performed twice for every speaker group: once with F_1 and F_2 as predictors and once with F_1, F_2 and duration as predictors to test to what extent spectral and temporal (i.e. duration) parameters are used by each group to distinguish American English vowels. Furthermore, and most interestingly for the central question of the current thesis (what vowels of American English are particularly difficult for Javanese and Sundanese learners to acquire and why), the vowel productions from the two non-native groups were fed into the LDA algorithm *trained on the native speakers' productions* to observe in the two non-native groups which particular intended American English vowels were most frequently misclassified by the algorithm (indicating which particular American English vowels are hard to produce correctly by the two non-native speaker groups).

4.5 Results

4.5.1 Formant Frequencies

4.5.1.1 Javanese vs. English Speakers

An RM-ANOVA with a Greenhouse-Geisser correction determined that the first formant frequency (F_1) was significantly affected by the factor Vowel [$F(4.5, 126.8) = 84.7$, $p < .001$]. Significant interaction effects were noted for Vowel × Group, [$F(4.5, 126.8) = 5.4$, $p < .001$]. No other effects and interactions were found. The second formant frequency (F_2) was signific-antly affected by Vowel [$F(3.2, 90.9) = 88.5$, $p < .001$]. No other significant main effects and interactions were found.

We followed up the Vowel × Group interaction with Mann-Whitney U tests that were conducted to compare the F_1 between the Javanese and English groups for each English vowel (Table 4.1). As can be seen in Table 4.1, a significant difference between the Javanese and English speakers on F_1 values occurs in the English new L2 vowels /ɑː, ɪ, æː/ and in the similar L2 vowel /iː/.

Table 4.1 Mann-Whitney U tests comparing Javanese and American English speakers on first formant (F_1) of the ten English vowels, Mdn = median, * = p < .05, ** = p < .005 (Bonferroni corrected significance threshold).

English vowel	F_1 (Hz)					
	Javanese		English		U	*p*
	Mdn	SD	Mdn	SD		
New L2 vowel						
/ɑː/	671	170.6	794.3	77.2	39	.006 *
/ɝ/	541	109.3	497.2	47.9	136	.140
/ɔː/	700	144.8	742	95.3	64	.120
/ʌ/	632.9	182.3	620	80.8	99	.983
/æː/	683.4	158.1	726.5	121.5	52	.035 *
/ɛ/	627.4	150.1	593.6	67.9	95	.846
/ɪ/	420.2	150	475.2	48.9	45	.015 *
/ʊ/	401.4	174.5	425.7	49.3	74	.267
Similar L2 vowel						
/iː/	328	102.5	335	50.0	155	.015 *
/uː/	400.8	73.0	349	60.4	140	.082
English vowel	F_2 (Hz)					
	Javanese		English		U	*p*
	Mdn	SD	Mdn	SD		
New L2 vowel						
/ɑː/	1470	221.4	1386	166.2	128	.231
/ɝ/	1577.9	183.5	1652	121.8	88	.619
/ɔː/	1260.2	220	1260.5	194.8	105	.846
/ʌ/	1672.5	241.9	1608	144.9	72	.231
/æː/	1886.61	254.9	1748	190	114	.559
/ɛ/	1937	247.7	1807	228	122	.350
/ɪ/	2263.9	350.4	1840.6	240	144	.055
/ʊ/	1270.3	203.2	1458.8	111.9	68	.169
Similar L2 vowel						
/iː/	2362.5	305.7	2298.2	286	86	.559
/uː/	1121.2	298.2	1277.9	177.2	85	.539

4.5.1.2 Sundanese vs. English Speakers

The F_1 values differed significantly for the Vowel factor [$F(3.85, 107.9) = 93.9$, $p < 0.001$]. Furthermore, there was a Vowel × Group interaction [$F(3.85, 107.9) = 4.7$, $p < 0.05$]. For the F_2 values there was a significant effect of Vowel [$F(3.2, 89.9) = 78.4$, $p < 0.001$]. Furthermore, there was a significant Vowel × Group interaction [$F(3.2, 89.9) = 3.6$, $p < 0.05$].

Mann-Whitney U tests were conducted to examine the difference between the Sundanese and American English speakers on F_1 and F_2

values for each vowel (Table 4.2). The F_1 value of the Sundanese was lower than the English speakers for the English *new* L2 vowels /ɑː/, /ɪ/, and /æː/. The F_2 value of the English vowel /ɪ/ by the Sundanese speakers was significantly higher than the English speakers. The F_1 values of the vowels /ɑː/ and /ɪ/ and the F_2 value of the vowel /ɪ/ survived the Bonferroni correction.

Table 4.2 Mann-Whitney U tests comparing Sundanese and American English speakers on first formant (F_1) and second formant (F_2) frequencies of the ten English vowels, Mdn = median, * = $p < .05$, ** = $p < .005$ (Bonferroni corrected significance threshold).

English Vowel	F_1 (Hz)				U	*p*
	Sundanese		English			
	Mdn	SD	Mdn	SD		
New L2 vowel						
/ɑː/	664	85.4	794.2	77.2	30	.001 **
/ɝ/	488	80.6	497.2	47.9	90	.681
/ɔː/	649	141	742	95.3	58	.067
/ʌ/	699.1	141.9	620	80.8	120	.397
/æː/	644.7	134.7	726.5	121.5	44	.013 *
/ɛ/	561.7	113.3	593.6	67.9	72	.231
/ɪ/	384.8	71.9	475.2	48.9	27	.001 **
/ʊ/	425.2	82.3	425.7	49.3	88	.619
Similar L2 vowel						
/iː/	359	77	335	50	136	.120
/uː/	359.8	103	349	60.4	110	.681
English Vowel	F_2 (Hz)				U	*p*
	Sundanese		English			
	Mdn	SD	Mdn	SD		
New L2 vowel						
/ɑː/	1470	221.4	1386	166.2	131	.183
/ɝ/	1577.9	183.5	1652	121.8	83	.475
/ɔː/	1260.2	220	1260.5	194.8	91	.713
/ʌ/	1672.5	241.9	1608	144.9	113	.588
/æː/	1886.61	254.9	1748	190	137	.109
/ɛ/	1937	247.7	1807	228	132	.169
/ɪ/	2263.9	350.4	1840.6	240	172	.001 **
/ʊ/	1270.3	203.2	1458.8	111.9	65	.131
Similar L2 vowel						
/iː/	2362.5	305.7	2298.2	286	114	.559
/uː/	1121.2	298.2	1277.9	177.2	74	.267

The realization of the vowel quality of the ten English monophthongs as pronounced by the three groups of speakers is summarized in Figure 4.3. In the figure the centroid of each vowel category is plotted in the two-dimensional vowel space as the IPA base symbol that conventionally represents the vowel type. In order to abstract away from linguistically irrelevant differences between individual speakers, depending on vocal tract size and individual differences in the habitual setting of the articulators, the formant frequencies as measured in hertz were subjected to Lobanov normalization. This is done by subtracting from the F_1 or F_2 value of a vowel token the mean F_1 (or F_2) value determined for that particular speaker and then dividing the result by the speaker's standard deviation. The center of the vowel space will then be at the F_1-by-F_2 coordinates of 0, 0. High (or close) vowels will have negative z-values for F_1 while lower (more open) vowels have increasingly more positive z-values. Back vowels will have negative z-values for F_2, which will become more positive as the constriction place is more fronted. A prerequisite to performing this type of normalization is that the same set of vowels is available for each speaker, and, preferably, that the corner vowels /i, a, u/ are included in the set. The result of the z-transformation is that distances between pairs of vowels are proportionally scaled within each individual speaker, and that the speakers' vowel configurations will be moved (geometrically 'translated') so as to have the same overall means for F_1 and F_2 and that the vowel categories occupy approximately the same area within the vowel space.

It is customary in this type of plot to also provide an indication of the dispersion of the tokens around their centroids. This is done by drawing a so-called spreading ellipse around the centroid. The orientation of the ellipse is determined by computing the first principal component (PC1) of the scatter cloud of vowel points around the centroid, i.e. the line which optimally captures the directionality of the scatter cloud. The second principal component (PC2) is then drawn at right angles to PC1 intersecting at the centroid. The scatter points are then projected onto PC1 and PC2, after which the standard deviation of the projected points can be computed. The ellipses in Figure 4.3 were drawn at plus and minus one standard deviation away from the centroid along each PC1 and PC2. Assuming a two-dimensional normal distribution, these ellipses theoretically envelop the central-most 46% of the vowel tokens that belong the category, i.e. roughly the most typical half of the exemplars of the category. Lack of contrast between two vowel categories is seen in the amount of overlap between the

ellipses associated with the categories. The optimal boundary (which yields the least number of classification errors) between the categories in the two-dimensional space is drawn through the two intersection points of the overlapping ellipses.

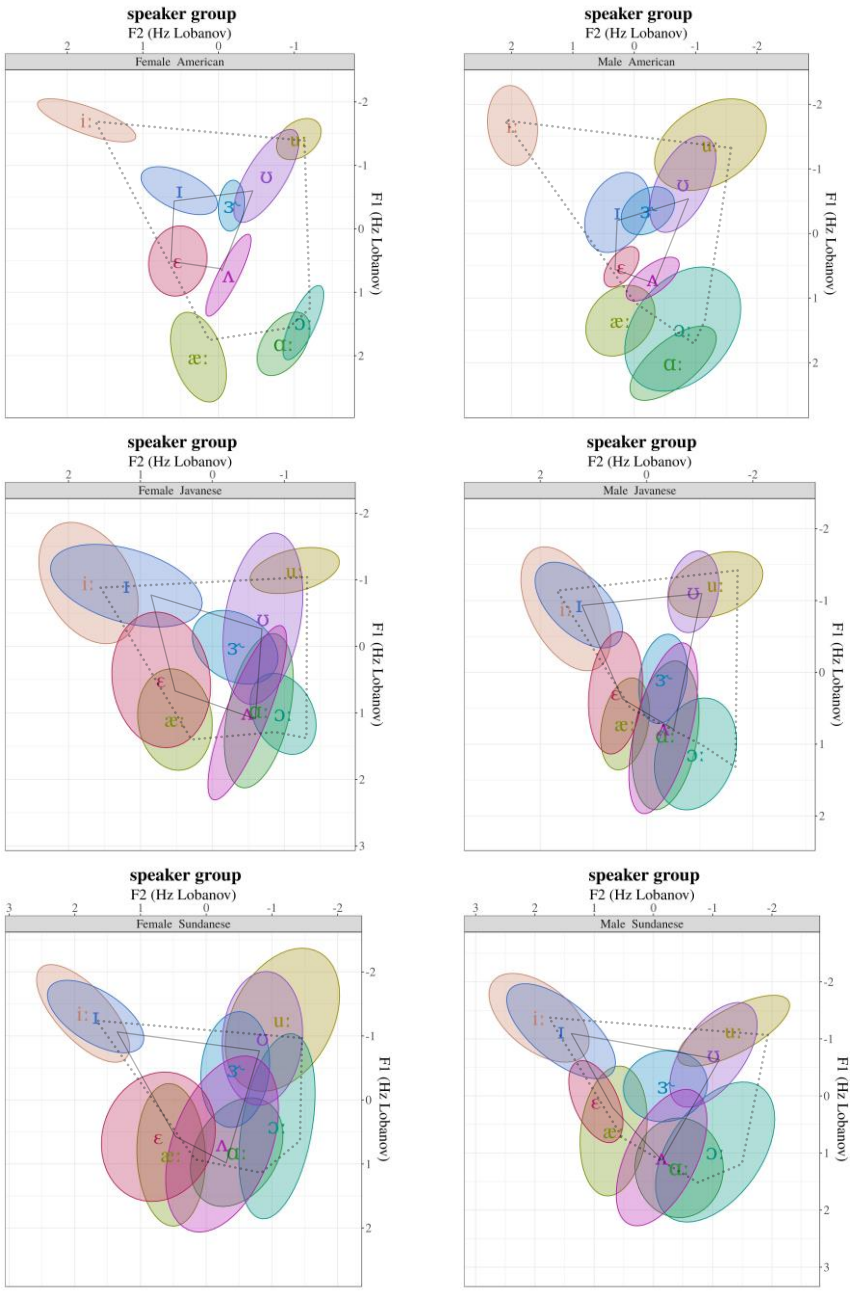

Figure 4.3 Centroids and ellipses of F_1 and F_2 values (z-normalized), for 10 English vowels produced by American English, Javanese and Sundanese female and male speakers. Dotted and solid polygons join tense and lax vowels, respectively.

In Figure 4.3, the vowel configurations are shown separately for male and female speakers. Although, theoretically, the within-speaker z-normalization should be sufficient to abstract away from the overall difference in size of the vocal tract, even across genders, inspection of our results reveals rather large differences between the centroids and especially the sizes of the dispersion ellipses of the male and female American speakers

What immediately strikes the eye is that the vowel categories of the American native speakers are much more narrowly defined than their counterparts in the L2 speaker groups. For instance, there is no overlap at all in the ellipses defining the /iː - ɪ/, /ɛ - æː/ and /ʌ - ɑː/ contrasts, and relatively little in the /ʊ - uː/ contrast. In the Indonesian speaker groups these contrasts are very poorly maintained, for two reasons. First, the centroids of the pair of categories are quite close to one another, so that even small-size ellipses would considerably overlap. Second, the dispersion ellipses themselves are much larger than those in the L1 plots, which is caused by large between-speaker differences (even after normalization) in the realization of the vowels concerned.

In the American L1 speaker group it is easy to see that the vowel system is characterized by an outer 'ring' of tense vowels and a much more centralized polygon formed by the four lax vowels (indicated in the figure). Similar inner polygons for the Indonesian speaker groups are substantially larger and approximate the convex hull (outer polygon, joining the tense vowels, dotted lines).

The relative positions of /iː, ɪ, ɛ, æː, ʌ, ɜ˞, ɑː, ɔː, ʊ, uː/ are quite similar in the Javanese and Sundanese male speakers. For Javanese and Sundanese speakers, /ʌ/ is lower than that of American English speakers. For Sundanese speakers, /ɪ/ is higher than that of Javanese and English speakers. Note, once more, how close together the members are of the pairs /æː - ɛ/ and /iː - ɪ/ in the Javanese and Sundanese speakers compared to American English speakers.

4.5.2 Vowel Duration

4.5.2.1 Javanese vs. English speakers

The duration was significantly affected by Vowel [$F(6.4, 179) = 27.7$, $p < .001$] and Group [$F(1, 28) = 11.2$, $p < 0.01$]. Furthermore, there was a significant Vowel × Group interaction [$F(6.4, 179) = 2.2$, $p < .05$]. The Vowel × Group interaction was followed up with independent t-tests to

compare vowel duration between the two groups for each vowel. As can be seen in Table 4.3, significant differences in vowel duration between Javanese and English speakers occur in the vowels /iː, ɝ, ɑː, ɔː, ɪ, ɛ, æː, ʌ/. However, only the vowels /ɛ, ɪ, æː/ survived the Bonferroni correction. Table 4.3 presents the mean duration and standard deviation of the ten vowels for Javanese and English speakers of English.

Table 4.3 Independent t-tests comparing Javanese and English L1 speakers on duration of ten English vowels, x̄ = mean, * = p < .05, ** = p < .005 (Bonferroni corrected significance threshold).

English Vowel	Duration (ms)						
	Javanese		English		*t*	df	*p*
	$\bar{x}$	SD	$\bar{x}$	SD			
/iː/	155.5	51.3	212	56	2.755	28	.010 *
/ɝ/	175	63.5	232	48.7	2.485	28	.019 *
/ɑː/	145.7	49.6	184.5	38.4	2.159	28	.040 *
/ɔː/	187	52.7	252	66.9	2.932	28	.007 *
/uː/	196.5	74	229	56.7	1.217	28	.234
/æː/	144	51.9	235	60.1	4.292	28	.000 **
/ɪ/	98	37.8	163.5	34	4.360	28	.000 **
/ɛ/	121.7	43.9	175	38.2	3.255	28	.003 **
/ʌ/	119.5	50	161	33.5	2.361	28	.025 *
/ʊ/	113.5	50.5	137	35.5	1.312	28	.200

Phonetically, four vowels /ɪ, ɛ, ʌ, ʊ/ are considered short (lax) while the other vowels /iː, ɝ, ɑː, ɔː, uː, æː / are long (tense). Table 4.3 shows that the Javanese speakers pronounced all the English vowels shorter than the native speakers did but that the relative duration differences, especially those between the lax/short and tense/long vowels are well preserved, in non-overlapping ranges. The Javanese speakers produced lax vowels with means between 100 – 125 ms and the tense vowels between 130 – 230 ms).

4.5.2.2 Sundanese vs. English Speakers

Vowel duration was significantly affected by Vowel [$F(4.9, 139) = 34.8$, p < .001] and Group [$F(1, 28) = 7.9$, $p < .05$]. There was a Vowel × Group interaction, [$F(4.9, 139) = 2.4$, p < .05]. To compare the duration between Sundanese and English speakers for each English vowel, we followed up the interaction with independent t-tests. As can be seen in

Table 4.4, the long vowels /ɝ, ɑː, ɔː, æː/ and the short vowels /ɪ, ɛ/ showed significant differences in the speech production between the Sundanese and English speakers. However, only the vowels /ɪ, æː/ survived the Bonferroni correction. Duration measurements of English vowels as produced by native Sundanese and English speakers are shown in Table 4.4.

Table 4.4 Independent t-tests comparing Sundanese and English speakers on duration of ten English vowels, x̄ = mean, * = p < .05, ** = p < .0025 (Bonferroni corrected significance threshold).

English Vowel	Duration (ms)						
	Sundanese		English		*t*	df	*p*
	$\bar{x}$	SD	$\bar{x}$	SD			
/iː/	182.7	51.2	212	56	1.430	28	.164
/ɝ/	171	59.9	232	48.7	2.784	28	.010 *
/ɑː/	152	38	184.5	38.4	2.194	28	.037 *
/ɔː/	194	59.4	252	66.9	2.437	28	.021 *
/uː/	199	63.6	229	56.7	1.248	28	.222
/æː/	160	59	235	60.1	3.246	28	.003 **
/ɪ/	100.5	28.7	163.5	34	5.318	28	.000 **
/ɛ/	142.5	40.7	175	38.2	2.098	28	.045 *
/ʌ/	133.7	45.7	161	33.5	1.667	28	.107
/ʊ/	111.7	34.8	137	35.5	1.861	28	.073

Table 4.4 shows that the Sundanese speakers' vowel durations are similar to the Javanese realisations. The vowel /iː/ produced by Sundanese speakers is clearly longer than that produced by the Javanese speakers and comes rather close to the duration found for the American English speakers. The lax vowels are pronouced shorter (with means between 130 and 175 ms) than the tense vowels (with means between 190 and 250 ms) by the American English speakers.

4.5.2.3 Vowel Duration Reanalyzed

The RM-ANOVA performed on the raw vowel duration measurements showed a main effect of L1 speaker group, of Vowel type, as well as incidental interactions between the two effects. The main effect of Group was due to the fact that the American L1 speakers produced longer vowel durations overall (198 ms) than the two groups of L2

speakers (141 ms for Javanese and 153 ms for Sundanese speakers). This difference is unexpected since native speakers are normally found to talk faster (with shorter segment durations) and with less effort than foreign learners of the language. At least two reasons come to mind why the L1 speakers in the present case should be slower, and have longer vowel durations than the non-natives.

The first reason might be that the Americans in this study were well aware of the fact that the Indonesian listeners might have problems understanding them, and had learned over the years to slow down their rate of delivery in order to boost their intelligiblity. This way of talking to non-native listeners is a habit especially of language instructors abroad, and is often referred to as foreigner talk - the equivalent of motherese, the way caretakers speak to infants and todlers (e.g. Kuhl & Iverson, 1995).

The second reason may the in the choice of stimulus materials. The speakers in the present experiment were instructed to produce tokens of the word/items /hVd/ and /bVd/, in which the postvocalic coda should be voiced. English is one of a minority of marked languages in which coda obstruents do not undergo final devoicing. Keeping a coda obstruent voiced necessitates the phonetic lengthening of the preceding vowel. An important, if not the most important, cue to the voiced-voiceless distinction in English coda obstruents is therefore vowel lengthening before voiced obstruents against vowel shortening before the voiceless counterparts (House, 1961; Raphael, 1972; Flege & Port, 1981; Elsendoorn, 1985). The Indonesian learners of English will undoubtedly have pronounced the final obstruents in the stimuli with a voiceless counterpart, i.e. /t/. As a result of this almost universal pronunciation error, the Indonesians fail to lengthen the vowel, so that their vowel duration ends up shorter than that of the L1 speakers. The same problem is found when Dutch and German speakers pronounce English coda obstruents. It was found, for instance, by Wang and Van Heuven (2006) and Wang (2007), who measured vowel duration in /hVd/ structures as pronounced by Chinese, Dutch and American speakers of English: the American speakers had mean vowel durations of 217 ms, the Dutch L2 speakers of 207 ms and the Chinese L2 speakers 196 ms; the difference is smaller than in the present study but the effect was highly significant.

The overall shorter pronunciation of the English vowels by the Indonesian learners, therefore, is not due to an incorrect vowel duration per se but rather to an incorrect rendering of the voiced-voiceless distinction in coda obstruents. In the latter case, the vowel durations

may still be quite accurate. In order to check this possibility, we decided to factor out the possible effect of the final consonant by normalizing the vowel durations within speakers. This was done by computing the mean vowel duration per speaker and then subtracting the individual mean from each vowel token, i.e. by performing the first part of a z-transformation. The result of this normalization is that the mean vowel duration of every speaker (whether native or non-native) is changed to 0 but other than that all differences among the vowels within the speaker are left unchanged. Figure 4.4 presents the adjusted vowel durations for the three speaker groups. In the Figure the vowels are arranged along the horizontal axis in ascending order of duration as measured for the American native speakers.

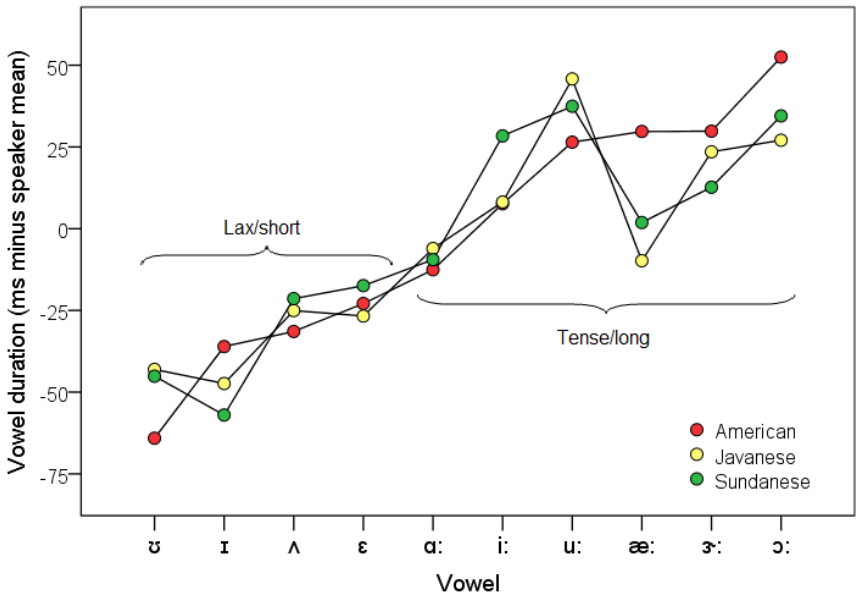

Figure 4.4 Adjusted vowel duration for the ten English monophthongs produced by American, Javanese and Sundanese speakers. Adjustment was done by subtracting the speaker-individual mean from the duration of each vowel token. Vowels are in ascending order of duration as established for the American speakers.

It is obvious from Figure 4.4 that the Indonesian speakers generally have an excelent conception of the duration of English vowels, since the three curves hardly differ from one another. For each speaker group the four lax vowels have the shortest duration, followed by the six tense vowels. The vowel durations produced by the three speaker groups are

very strongly correlated. Cronbach's alpha is .959. The correlation between Javanese and American vowel durations is r = .861 (N = 10, p = .001), between Sundanese and American speakers . 876 (N = 10, p < .001) and between the two Indonesian groups .948 (N = 10, p < .001).

Some discrepancies can be observed between the Indonesian and American speakers. The relative duration of the high vowels, especially /iː/ is somewhat longer than in the L1 data, whereas the relative duration of the more open vowels /ɝ, ɔː/ and especially /æː/, although clearly longer than the lax vowels, are shorter than in the L1 data.

In light of these findings, we should find that vowel duration will make a substantial contribution to the correct classification of the English vowels as spoken by Indonesian learners of English, probably as large a contribution as will be found for the American native speakers.

4.5.3 Linear Discriminant Analysis

A series of LDAs was performed to classify the English vowels produced by the American, Javanese and Sundanese speakers of English. Following Wang and Van Heuven (2006) the values of F_1 and F_2 were first converted to Barks, so as to do justice to the way the human hearing mechanism responds to differences in vowel quality. Because speakers differ in the size and shapes of their vocal tracts, especially between men and women (the latter have roughly 15% shorter vocal tracts), the Bark-transformed formants were then z-normalized within individual speakers, such that, for every speaker the mean F_1 and F_2 values were 0 and the Standard Deviations for F_1 and F_2 were 1. As a result, negative F_1 values correspond to relatively high vowels, and positive values characterize lower vowels. Similarly, negative F_2 values are obtained for back-rounded vowels, and positive values for front-spread vowels. The LDAs were trained on the vowel tokens collected for the American native speakers of English. The discriminant functions derived from this classification were then used to test the English vowels produced not only by the American speakers (with leave-one-out cross-validation so as to ensure that the tested vowel token was not in the training set) but also the vowel tokens produced by the Javanese and Sundanese speakers. The LDA was run twice. The first time only the formants F_1 and F_2 were included as predictors of the vowel type. The second time, vowel duration was added as a third predictor, so as to allow us to determine the specific added value of including vowel duration in the (simulated) recognition of American-English vowels.

Figure 4.5 shows the percentage of American English vowels correctly classified based on only spectral parameters and both spectral and durational information for the three speaker groups.

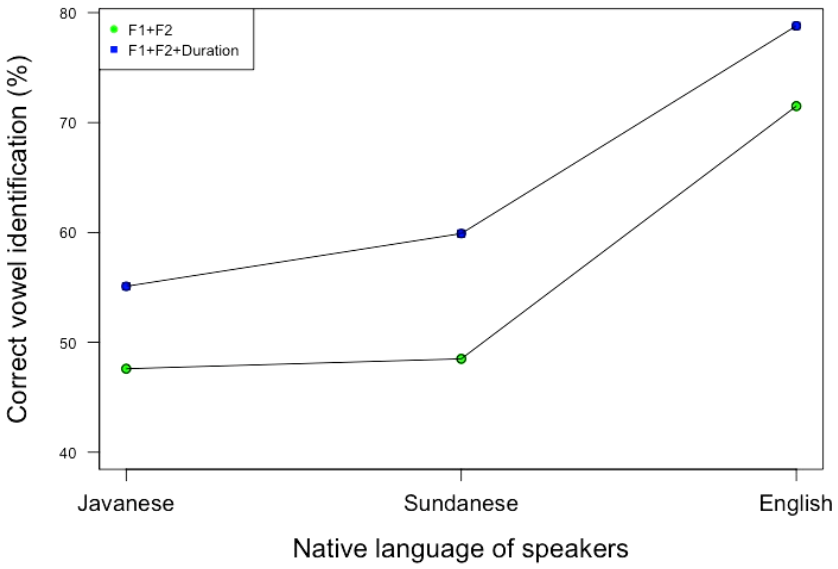

Figure 4.5 Mean correct English vowel identification (%) based on F_1 and F_2 (and duration, upper line) of Javanese, Sundanese and American speakers.

The results reveal that Javanese and Sundanese speakers have lower correct vowel identification than English native speakers. When the predictors consist of spectral parameters only, the LDA could identify the vowels spoken by the American L1 speakers at roughly 70%, which is seven times better than chance (= 10%), even when cross-validation was applied (see above). This performance is obviously better than of the scores obtained for the non-native speakers (48% correct vowel identification for both groups of Indonesian learners of English), indicating that these L2 learners had problems using F_1 and F_2 in contrasting the English monophthongs the way native L1 speakers do. When duration parameter is added to the set of predictors, the classification accuracy increases for all three speaker groups by about 10 points. This shows that Javanese and Sundanese learners exploit duration when producing English vowel contrasts as much as American L1 speakers, even while the duration is not a contrastive feature in the learners' L1.

Table 4.5 Confusion matrix of English vowels produced by Javanese, Sundanese and American L1 speakers as classified by LDA using F_1, F_2 and vowel duration as predictors. The LDA was trained on AE vowels. Correctly classified vowels are in the grey-shaded cells along the main diagonal. According to SLM, vowels /iː/ and /uː/ are similar vowels.

	Javanese learners of English: 60 % correct (N= 20)										
	Vowel classified as										
		iː	ɪ	ɛ	æː	ɜˑ	ʌ	ɑː	ɔː	ʊ	uː
Intended vowel	iː	**63**	**18**	13		6					
	ɪ	**49**	**36**	4		1	3			5	3
	ɛ	4	**17**	**44**	**21**	4	9				1
	æː		1	**39**	**47**	1	7	4	2		
	ɜˑ		11	2	7	**41**	**23**	3	10	4	
	ʌ		1		17	3	**25**	**34**	3	8	11
	ɑː			3	16	1	**28**	**25**	**20**	4	4
	ɔː			3	3		5	**26**	**63**		1
	ʊ		6		3	3	6	6		**54**	**23**
	uː	1	3			7				8	**83**
	Sundanese learners of English: 53 % correct (N= 20)										
		iː	ɪ	ɛ	æː	ɜˑ	ʌ	ɑː	ɔː	ʊ	uː
Intended vowel	iː	**85**	5	3		3	3				3
	ɪ	**50**	**43**				5			3	
	ɛ	3	**18**	**43**	**22**	5	8			3	
	æː		8	**17**	**41**	8	**18**	5	3		3
	ɜˑ		3	8		**59**	**25**		5	5	3
	ʌ			3	**18**	3	**20**	**43**		8	8
	ɑː				**18**		**25**	**38**	16	3	
	ɔː			3	5	13	8	**13**	**52**	3	5
	ʊ		3		3	3	7	5		**67**	14
	uː	0	1		3	7		3		10	**78**
	American native speakers: 80 % correct (N= 10)										
		iː	ɪ	ɛ	æː	ɜˑ	ʌ	ɑː	ɔː	ʊ	uː
Intended vowel	iː	**100**									
	ɪ		**75**	5		5				10	5
	ɛ			**80**		5	15				
	æː			10	**72**	3		10	5		
	ɜˑ		7			**88**	5				
	ʌ			5	10	5	**75**			5	
	ɑː				5			**70**	25		
	ɔː				15			10	**75**		
	ʊ		10			12	5			**68**	5
	uː		5			5				5	**85**

Table 4.5 gives us more precise information about *which* vowels precisely would be difficult to produce for the L2-speakers. The table shows for each speaker group the percentages of vowels that were classified correctly by the LDA algorithm trained on the vowel productions of the native speakers. Firstly, as can be seen in the bottom panel, the LDA algorithm trained on American English speakers performs relatively well (80% correct, cross validated) for productions of the same group as would be expected.

Table 4.5 shows relatively poor classification performance for intended English vowels /iː, ɪ, ɛ, æː, ɝ, ʌ, ɔː, ɑː, ʊ/ for the Javanese speakers. The Javanese speakers often mispronounced /iː/ as /ɪ/ and *vice versa*, /ɛ/ as either /ɪ/ or /æː/. They also mispronounced /æ/ as /ɛ/, /ɝ/ as /ʌ/, /ɑː/ as /ʌ/, both /ɔː/ and /ɔː/ as /ɑː/, and /ʊ/ as /uː/.

For the Sundanese speakers, the intended English vowels /ɪ, ɛ, æː, ɝ, ʌ, ɑː, ɔː/ are also often mispronounced, as shown in Table 4.5. Vowel /ɪ/ is mispronounced as /iː/, /ɛ/ as /ɪ/, /æː/ as either /ɛ/ or /ʌ/, / ɝ/ as /ʌ/, /ʌ/ as /æ/ and /ɑː/. The Sundanese speakers also mispronounced /ɑː/ as /æ/ or /ʌ/, /ɔː/ as /ɑː/, and /ʊ/ as /uː/.

4.6 Discussion

The primary goal of this chapter was to investigate what aspects of American English vowel production are particularly problematic for Javanese and Sundanese learners of English. Also, we examined whether or not the L2 learners have particular difficulty with producing the new L2 vowels /ɪ, ɛ, ʊ, æː, ɑː, ɔː, ʌ, ɝ/ vs the similar L2 vowels /iː, uː/. According to the SLM (Flege, 1995, 1999, 2002), the Javanese and Sundanese speakers should not exhibit differences in formant structure (relative to native American speakers) for the new L2 vowels when the L2 acquisition process is completed or fossilized. However, they will still show different formant frequencies for the similar L2 vowels. The L2LP model (Escudero, 2005) predicts that the Javanese and Sundanese speakers will produce non-native formant frequency values for the new L2 vowels and will produce rather more native-like frequencies for the similar L2 vowels at least in the earlier stages of the L2 acquisition process.

The results of F_1 analyses showed that F_1 frequencies of the new L2 vowels /ɑː/ and /ɪ/ were considerably lowered by the L2 learners as compared to native speakers (suggesting that the learners pronounced these vowels were with too elevated a tongue position). Overall, the results are in line with the prediction of the L2LP (Escudero, 2005) that

the L2 speakers will produce formant frequencies producing new L2 vowels and easily produce the similar L2 vowels of the English speakers, which suggests that the acquisition of the English vowels by the Indonesian participants was still in full swing.

The results of F_2 values showed that only the new L2 vowel /ɪ/ is produced significantly more to the front of the mouth (and is probably indistinct from /iː/ in terms of vowel quality) by the Sundanese speakers when compared to the English speakers. The Javanese speakers did not show any significant differences with the production of English new and similar vowels in F_2 values. In light of the results, it is confirmed that the production of new L2 vowels is difficult for the L2 speakers since they need to adjust the production of new L2 sounds with the L1 perceptual mapping and, at the same time, they need to create new L2 categories. Specifically, the L2 speakers seem to have problems with the correct openness of L2 vowels. Thus, as predicted by the L2LP (Escudero, 2005), the production of new L2 vowels would be challenging for the L2 speakers.

The study also taps into the durational properties of vowels produced by Javanese and Sundanese learners of English. The results showed that the long vowels /iː, ɝ, ɑː, ɔː/ were produced shorter by the Javanese than the English speakers. For the Sundanese, the long vowels /ɝ, ɑː, æː, ɔː/ were pronounced shorter than the English speakers. The data provide consistent support for the Feature-dependent Hypothesis (McAllister et al., 2002), which states that L2 learners have difficulties in producing duration in a native-like manner if the durational information is not found in their L1. The results can be explained by the fact that duration features are not prominently exploited in their first language. This finding confirms the prediction that contrasting categories in a second language would be difficult to acquire if the phonetic features do not exist in the first language.

Curiously enough, the shortening of duration also occurred in the production of the lax/short English vowels. The results showed that for the Javanese, short vowels /ɪ, ɛ, ʌ/ were produced significantly shorter than for the American speakers. For the Sundanese, the short vowels /ɪ, ɛ/ were shorter than those of the American speakers. The L2 learners have hence produced shorter target vowels, even for the English short vowels. Therefore, the Javanese and Sundanese speakers over-shortened short vowels compared to English speakers due to the absence of the duration cues in their L1, which is in line with the Feature-dependent Hypothesis (McAllister et al., 2002).

At second thoughts, however, it would seem to make little sense for speakers of a language without a vowel length contrast, such as Javanese and Sundanese, in which all the vowels are phonetically short, to shorten English short vowels even more than their own vowels. In Chapter 2, we presented the results for vowel duration of Javanese and Sundanese. We found that indeed the L1 vowel durations are quite short, i.e. between 60 and 100 ms for all vowel types, with the exception of /u/ (110 ms) and /ɨ/ (150 ms) for the Sundanese speakers. It should be born in mind, however, that these durations were measured for vowels in unstressed open CV syllables at the beginning of a two or three-syllable word in the middle of a carrier sentence. Cross-linguistically, vowels in open syllables are longer than in closed syllables (all else being equal), but at the same time, vowels are shortened in unstressed syllables. This makes it hazardous to compare the L1 vowel durations of the Indonesian speakers with their L2 English durations.

An alternative view of the production of the L2 vowel durations by the Javanese and Sundanese speakers was suggested in section 4.5.2. It was shown there that the English vowel durations produced by the Indonesian speakers, although shorter across the board, correlated very well with the vowel durations of the American speakers, with a clear contrast between short/lax vowels and long/tense vowels. The proper use of vowel durations by the Indonesian speakers was confirmed by the results of the LDA, which showed that adding vowel duration as a predictor yielded the same improvement in the automatic classification of the ten English monophthongs for each of the three speaker groups, whether American or Indonesian. Moreover, it was suggested that the shorter overall vowel durations produced by the L2 speakers was caused by the fact that they shortened the vowel durations relative to the American speakers because they mispronounced the coda consonant as voiceless [t], while the American native speakers correctly lengthened the vowels before the voiced coda obstruent [d]. If this account is accepted, then the Indonesian learners were shown to correctly produce the English vowel durations and to preserve the durational differences between tense and lax vowels. This, in turn, would be fully in line with Bohn's (1995) Desensitization hypothesis, which claims that adult language learners find it easy to tune in to duration differences even if they have lost their sensitivity to other phonetic parameters, such as differences in vowel quality.

The LDA results confirmed that Javanese and Sundanese speakers failed to spectrally distinguish the intended vowels. The addition of duration increased the correct vowel identification of the Javanese and

Sundanese speakers indicating that they successfully exploit the duration feature to contrast L2 vowels (while apparently over-shortening them). Therefore, L2 speakers seem to primarily have problems adequately using F_1 and F_2 (i.e. the correct degree of mouth opening and movement of the tongue) in pronouncing American English vowels.

Regarding difficulties with employing spectral features, the Javanese speakers reveal a symmetrical confusion for /iː - ɪ/, /æː - ɛ/, /ɑː - ʌ/, and /ɑː -ɔː/. Similarly, vowels /æː - ɛ/, /ɑː - ʌ/, /ɑː - ɔː/, /ʌ - æː/, and /ɑː - æː/ are symmetrically misclassified and seem relatively difficult for the Sundanese speakers. The Javanese speakers reveal asymmetrical confusion in /ɛ - ɪ/, /ɝ - ʌ/, and /ʊ - uː/ as indicated by relatively inaccurate classification performance by the algorithm. The Sundanese speakers also showed asymmetrical error patterns for vowel /ɪ - iː/, /ɛ - ɪ/, /ɝ - ʌ/, /ʊ - uː/, and /ɛ - ɪ, æː/. These results indicate that the Javanese and Sundanese speakers have difficulties with producing L2 vowels which are close to each other in the vowel space.

Regarding L2 production problems within the 'similar' vs 'new distinction, the LDA results indicate that the Indonesian L2 learners of English have difficulty in contrasting the similar vowels /iː/ but not /uː/. Both Indonesian groups also experience difficulties with the new vowels /ɪ, ɛ, æː, ɝ, ʌ, ɑː/. In addition, the Javanese speakers have a problem with the new vowel /ʊ/ and the Sundanese speakers with /ɔː/. Hence there seems to be mixed support for models, SLM and L2LP, and the question arizes how both models can simultaneously be supported. One possibility is that the two models contrasted in this thesis apply to different stages of the learning process. The SLM model applies to experienced learners who have formed new vowel categories for L2 and experience most interference from vowels that are similar to L1 (Flege, 2003). The L2LP model, on the other hand, can be seen as focusing on the relatively naive learners, which fits with the speakers in the current study. L2LP predicts that new sound categories (with respect to L1) are most difficult for learners because these categories still have to be formed (Escudero, 2005), while pronouncing similar vowels is relatively preserved because L2 learners can use the L1 equivalents as a basis to produce the respective L2 vowels. When look at the overall pattern of (mis)performance of the L2 learners in the current study then, most difficulties were observed with new vowels, as would be expected for these relatively inexperienced L2 learners.

4.7 Conclusion

The difficulties experienced by the Javanese and Sundanese speakers are mostly shown in the F_1 values, rather than in the F_2, indicating that they had different vowel height realizations, rather than differences in the degree of *backness*, when compared to the English speakers.

Javanese and Sundanese learners of English have no long and short (tense and lax) or vowel length attribute in their L1. They are predicted to produce vowels differently as compared to native English speakers due to the interference by their L1 with the L2 learning process. In the production of the English vowels, the Javanese and Sundanese speakers demonstrated a shorter duration for the long vowels /iː, ɝ, ɑː, æː, ɔː/ and for the short vowels /ɪ, ɛ, ʌ/. This result agrees with the Feature-dependent Hypothesis. However, the relative differences in durations between the English vowels (including the durational differences between tense and lax vowels) were quite well preserved in the L2 English vowels, which lend strong support to the Desensitization Hypothesis.

Javanese and Sundanese speakers did not differ much in terms of the ability in producing the English vowels. Both Javanese and Sundanese have difficulties producing contrasts between vowels that are close to one another in the spectral space, i.e. which differ in quality where such quality differences are not used in their L1. These are cases of Single Category (SC) or Category Goodness (CG) assimilation in terms of the Perceptual Assimilation Model of L2 perception. The L1 vowels systems of Javanese and Sundanese are very similar, yielding similar L1-L2 interference phenomena. Therefore, a similar pattern of L2 acquisition problems would be expected for the two groups of Indonesian learners of English. A practical implication of this result is that the same L2 teaching methods could be employed for both groups

To conclude, the results have shown that the Javanese and Sundanese speakers have difficulty contrasting intended vowels using spectral parameters while the use of duration is used contrastively, possibly with over-shortening. It is therefore recommended to train both speakers groups in adequately using the opening of the mouth and placement of the tongue to produce American English vowels. Specifically, the English teaching and teaching programs need to focus on training of the contrast between spectrally adjacent vowels the members of which are felt to be tokens of a single vowel category in the mother tongue, such as /iː - ɪ/, /ɛ – æː/, /ʌ - ɔː/, and /ʊ - uː/.

References

Bent T., Bradlow A. R., & Smith B. L. (2008). Production and perception of temporal patterns in native and non-native speech. *Phonetica, 65*, 131-147.

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer* [Computer Program]. Version 5.3.51, retrieved 01 August 2013 from http://www.praat.org/

Bohn, O. S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279-304). Baltimore, MD: York Press.

Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics, 11*, 303-328.

Bohn, O. S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition, 14*, 131-158.

Bradlow, A.R, Torretta, G.M, & Pisoni, D.B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication., 20*(3-4), 255–272. DOI:10.1016/S0167-6393(96)00063-5.

Chomsky, N., & Halle, M., (1968). *The sound pattern of English.* Boston: MIT Press.

Cruttenden, A. (2001). Gimson's Pronunciation of English: 6th Edn. of Gimson, A. C., *An Introduction to the Pronunciation of English*. London: Edward Arnold.

Elsendoorn, B. A. G. (1985). Production and perception of Dutch foreign vowel duration in English monosyllabic words. *Language and Speech, 28*, 231-254. DOI: 10.1177/002383098502800302.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26,* 551-585.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. LOT dissertation series 113. Utrecht: LOT.

Escudero, P. (2009). Linguistic perception of 'similar' L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in perception*. Berlin: Mouton de Gruyter.

Fant, G. (1973). Stops in CV syllables. In G. Fant (Ed.), *Speech sounds and features* (pp. 110 139). Cambridge, MA: MIT Press.

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

Flege, J. E. (1997). English vowel production by Dutch talkers: More evidence for the "new" vs "similar" distinction. In A. James & J. Leather (Eds.), *Second-language speech: Structure and process*. *Studies in Second Language Acquisition, 13* (pp. 11-52). Berlin: Mouton de Gruyter.

Flege, J. E. (1999). Age of learning and second-language speech. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 101-132). Hillsdale, NJ: Lawrence Erlbaum.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 217-244). Trier: Wissenschaftlicher Verlag Trier.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production, differences and similarities* (pp. 319-355). Berlin: Mouton de Gruyter.

Flege, J.E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech, 24*, 125-146. DOI: 10.1177/002383098102400202.

Gimson, A. C. (1980). *An Introduction to the Pronunciation of English* (3rd edn.). London: Edward Arnold.

Gimson, A. C., & Cruttenden, A. (1994). Gimson's Pronunciation of English (5th ed.), revised by A. Cruttenden. London: Edward Arnold.

Hazan, V., Markham, D., (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *Journal of the American Academy of Audiology, 116* (5), 3108–3118.

Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized /hvd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America, 406*, 3509-3523. DOI: 10.1121/1.424676.

House A. S. (1961). On vowel duration in English. *Journal of the Acoustical Society of America*, *33*, 1174–1178. DOI: 10.1121/1.1908941.

IBM Corp. (2013). *IBM SPSS Statistics for Macintosh*, Version 22.0. Armonk, NY: IBM Corp.

Jones, D. (1957). *Outline of English Phonetics*. Cambridge, UK: Cambride University Press.

Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movements and duration. *Journal of the Acoustical Society of America, 122*, 2842-2854. DOI: 10.1121/1.2783198.

Klecka, W. R. (1980). *Discriminant analysis*. Beverly Hills, CA: Sage Publications.

Krashen, S. D. (1981). *Second language acquisition and second language learning.* University of Southern California: Pergamon press, Inc.

Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of languages*. Malden: Blackwell.

Ladefoged, P. (2006). *A course in phonetics* (5th Ed.). Boston: Thomson Wadsworth.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lin, C. Y. (2013). Perception and production of English front vowels by Taiwanese EFL learners. *Theory and Practice in Language Studies, 3*(11), 1952-1958. DOI: 10.4304/tpls.3.11.1952-1958.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America, 49,* 606-608.

MacKay, I. R. A., Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica, 58*, 103- 125. DOI: 10.1159/000028490.

Makarova, A. (2010). Acquisition of Three Vowel Contrasts by Russian Speaker of American English (Doctoral Dissertation). Massachusetts: Harvard University. Retrieved from https://pqdtopen.proquest.com/doc/635753203.html?FMT=ABS.

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonia. *Journal of Phonetics, 30*, 229-258.

Meara, P. (2010). *EFL vocabulary tests* (2nd ed.). Swansea: Swansea University.

Meunier, C., Frenck-Mestre, C., Lelekov-Boissard, T., Le Besnerais, M. (2003). Production and perception of foreign vowels: does the density of the system play a role?, in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona.

Munro, M. J. (1993). Productions of English vowels native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech, 36*, 39-66.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175-184.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America*, *51*, 1296-1303. DOI: 10.1121/1.1912974.

Strange, W., Bohn, O. S., Trent S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America, 115*, 1791-80.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America, 88*, 97-100.

Van Heuven, V. J. & Gooskens, C. (2017). An acoustic analysis of English vowels produced by speakers of seven different native-language backgrounds. In: Wieling M., Kroon M., Noort G. van, Bouma G. (Eds.) *From semantics to dialectometry* (137-147). Festschrift in honour of John Nerbonne. no. 32 London: College Publications.

Wang, H. (2007). *English as a Lingua Franca. Mutual intelligibility of Chinese, Dutch and American speakers of English.* LOT Dissertation series 147. Utrecht: LOT.

Wang, H., & Van Heuven, V. J. (2006). Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers, In. J. M. van de Weijer & B. Los (Eds.), *Linguistics in the Netherlands 2006* (pp. 237-248). Amsterdam/Philadelphia: John Benjamins.

Weenink, D. (2006). *Speaker-adaptive vowel identification.* Doctoral dissertation, University of Amsterdam.