

Cover Page



Universiteit Leiden



The following handle holds various files of this Leiden University dissertation:

<http://hdl.handle.net/1887/68575>

Author: Perwitasari, A.

Title: The acquisition of English vowels by Javanese and Sundanese native speakers

Issue Date: 2019-02-19

1

General Introduction

1.1 Introduction

1.1.1 Bilingualism/Multilingualism in Indonesia

Indonesia is a nation with a great deal of linguistic diversity. Simons and Fennig (2017) report that Indonesia is home to more than 700 languages spoken by around 255 million people. Most of the local languages in Indonesia belong to the Austronesian language family and are thus related to the indigenous languages of Taiwan, the Philippines, Malaysia, and the Pacific islands. In eastern Indonesia, many of the local languages belong to non-Austronesian language families.

The history of *Bahasa Indonesia*, Indonesian, started when young Indonesian nationalists declared the Youth Pledge in October 28, 1928. On this day, the Indonesian youth congress proclaimed three ideals: one motherland, one nation and one language. In this declaration, they announced Bahasa Indonesia as the language of national unity (Ebing, 1997). Right after the time of independence, precisely in August 18, 1945 Indonesia decided to officially use the Indonesian language, *Bahasa Indonesia*, as its sole national language and principal lingua franca (Prentice, 1978; Steinhauer, 1980; Adelaar et al., 1996; Collins, 1998; Sneddon, 2003; Paauw, 2009). The consensus was incorporated into the 1945 Constitution of the Republic Indonesia. As an official language, Indonesian is used as the main means of communication in education, media, and government.

The Indonesian government's concern from the 1950s to the 1980s with diffusing the Indonesian language has led to an increase in multilingualism and the development of diglossia in Indonesia (Nababan, 1985; Sneddon, 2003; Cohn & Ravindranath, 2014). Indonesians whose first language (L1) is a local language, are also able to speak Indonesian (Nababan, 1985). The major influences on the Indonesian language are mainly from Javanese, followed by Sundanese and the colloquial variety of Indonesian spoken in Jakarta (Rubin, 1977). The influence of Javanese, Sundanese, and Jakartan Indonesian on standard Indonesian is visible in the lexicon as well as at the level of syntax, morphology, and phonology (Poedjosoedarmo, 1982).

1.1.2 Status of Indonesia in a Worldwide Ranking of English Proficiency

English has become part of the Indonesian education system since Indonesia's independence in 1945 (Dardjowidjojo, 2000; Sahiruddin, 2013). English is now taught for eight or nine years from Grade 4 or 5 of primary school through high school (Renandya, 2000). Previous studies on English teaching and learning found a lack of motivation and positive learning attitudes among students as they had not realized the importance of learning English (Sadtono, 1976; Dardjowidjojo, 1996; Panggabean, 2007; Marcellino, 2008; Mattarima and Hamdan, 2011).

A low level of English proficiency, and a lack of teaching preparation are common major problems in English language teaching in Indonesia (Dardjowidjojo, 1997, 2000; Nur, 2004; Marcellino, 2005). The English proficiency of teachers has not reached a mature level of language use, even at the university level (Dardjowidjojo, 2003). As a consequence, teachers and students communicate either in Bahasa Indonesia or in their local languages. In the end, they often have not acquired a sufficient English proficiency level after finishing the school program. Table 1.1 compares the level of English proficiency in Indonesia with that found in other countries.

Table 1.1 is based on the report of the Test and Score Data Summary for *TOEFL iBT®* Tests January 2017 – December 2017 Test Data. This report by the Educational Testing Services (ETS) summarizes the performance of *TOEFL iBT®* test takers. The test is designed to measure the level of English proficiency among adults across the world where English is typically a non-native language. The English test measured reading, listening, speaking and writing skills. The test and score data summary was calculated from test takers of 169 countries. The ranks were based on the total scale score classified by geographic region and native country, which ranged between 101 (#1, Ireland) and 59 (#169, Lao People's Democratic Republic). Table 1.1 shows that the level of English competence in Indonesia (85, #68) is below that of other Southeast Asian countries such as the Philippines (89, #42), and Malaysia (91, #31).

Table 1.1 Test and Score Data Summary for *TOEFL iBT®* Tests (ETS, 2017)

Rank	Country	Score	Rank	Country	Score	Rank	Country	Score
1	Ireland	101	58	Serbia	87	115	Uzbekistan	79
2	Austria	100	59	Aruba	86	116	Sudan	79
3	Netherlands	100	60	Jamaica	86	117	Nigeria	78
4	Belgium	99	61	Mexico	86	118	Thailand	78
5	Switzerland	99	62	Paraguay	86	119	Turkmenistan	78
6	Sou.Africa	98	63	Belarus	86	120	Armenia	78
7	Denmark	98	64	Cyprus	86	121	Turkey	78
8	Germany	98	65	Bahrain	86	122	Algeria	78
9	Singapore	97	66	Egypt	86	123	Libyan Arab Jam.	78
10	Trin.Tobago	96	67	Honduras	85	124	French Polynesia	78
11	Luxembourg	96	68	Indonesia	85	125	Uganda	77
12	UK	96	69	Sri Lanka	85	126	Guadeloupe	77
13	Mauritius	95	70	Georgia	85	127	Martinique	77
14	Canada	95	71	Moldova	85	128	Kyrgyzstan	77
15	Estonia	95	72	Monaco	85	129	Mongolia	77
16	Finland	95	73	Namibia	84	130	Palestine Ter.	77
17	NZ	95	74	El Salvador	84	131	Chad	76
18	Bahamas	94	75	Peru	84	132	Eritrea	76
19	India	94	76	Suriname	84	133	Sierra Leone	76
20	Iceland	94	77	Kazakhstan	84	134	Macao	76
21	Portugal	94	78	Ukraine	84	135	Kuwait	76
22	Costa Rica	93	79	Iran	84	136	Ethiopia	75
23	Greece	93	80	UAE	84	137	Reunion	75
24	Hungary	93	81	Zambia	83	138	Somalia	75
25	Norway	93	82	Bolivia	83	139	Rwanda	73
26	Slovenia	93	83	Ecuador	83	140	Yemen	73
27	Sweden	93	84	Panama	83	141	Mozambique	72
28	Pakistan	92	85	Puerto Rico	83	142	Tanzania	72
29	Australia	92	86	Venezuela	83	143	Cambodia	72
30	Uruguay	91	87	Azerbaijan	83	144	Iraq	72
31	Malaysia	91	88	Korea, Dem.	83	145	Afghanistan	71
32	Bulgaria	91	89	Korea, Rep.	83	146	Japan	71
33	Italy	91	90	Montenegro	83	147	Benin	70
34	Romania	91	91	Kenya	82	148	Cameroon	70
35	Israel	91	92	Swaziland	82	149	Gambia	70
36	Argentina	90	93	Tunisia	82	150	Niger	70
37	Croatia	90	94	Colombia	82	151	Gabon	69
38	Lithuania	90	95	Dom. Rep.	82	152	Saudi Arabia	69
39	Poland	90	96	Guatemala	82	153	Cape Verde	68
40	Slovakia	90	97	Taiwan	82	154	Burundi	67
41	US	89	98	Viet Nam	82	155	Liberia	67
42	Philippines	89	99	Kosovo	82	156	SSD	67
43	Czech Rep.	89	100	Oman	82	157	Togo	66
44	Spain	89	101	Qatar	82	158	Tajikistan	66
45	Lebanon	89	102	Ghana	81	159	Angola	65
46	Botswana	88	103	Madagascar	81	160	Mauritania	65
47	Zimbabwe	88	104	Chile	81	161	Burkina Faso	64
48	Nicaragua	88	105	Nepal	81	162	Senegal	64
49	Hong Kong	88	106	Albania	81	163	Cote D'Ivoire	63
50	Andorra	88	107	Morocco	81	164	Mali	63
51	Bosnia Her.	88	108	Syrian Arab	81	165	Haiti	63
52	France	88	109	Cuba	80	166	Guinea	61
53	Latvia	88	110	Myanmar	80	167	Congo	60
54	Brazil	87	111	Jordan	80	168	Djibouti	60
55	Bangladesh	87	112	Malawi	79	169	Lao Pep of Dem.	59
56	Macedonia	87	113	Neth.Ant.	79			
57	Russian F.	87	114	China	79			

It should be pointed out that the *TOEFL* iBT® scores are based on four tests, all of which target receptive English language skills only. The tests measure the participant's reading comprehension and listening comprehension through multiple choice questions. Low level listening skills such as phoneme identification are not tested nor are any productive language skills, such as speaking and writing. The *TOEFL* iBT® scores show that generally English learners whose L1 is a Germanic language have the best scores, followed by learners whose L1 belongs to one of the other Indo-European language families (i.e. Romance, Slavic, Finno-Ugric). The relatively high ranks for Singapore (#9), India (#19), Malaysia (#31), Phillippines (#42), and Hong Kong (#49) would seem to be related to the colonial past of these countries (e.g. India, Singapore, Malaysia, Hong Kong), or to more recently developed strong economic and political ties with the United States of America (e.g. Phillippines). Indonesia is in the upper half of the distribution, and its scores are better than those of many other South-East Asian countries, such as Sri Lanka (85, # 69), Thailand (78, #118), Myanmar (80, #110) and Cambodia (72, #143). We may conclude from this – highly provisional – survey that Indonesian speakers of English are in the middle of the range and hence that there is room for improvement.

1.1.3 History and Results of English Curriculum and Training in Indonesia

As part of the effort to introduce more English to the Indonesian education system, the Ministry of Education prepares curriculum guidelines containing basic outlines of English taught in Indonesian schools. The curriculum guidelines from 1945 to the present have undergone several innovations and changes. To improve the achievement level of learners and as the research on Second Language changed, changes in Indonesia's English language curriculum have occurred in at least six different periods with distinct areas of focus, as follows:

1. The 1945 period: Grammar translation-based curriculum
2. The 1958 period: Audiolingual-based curriculum
3. The 1975 period: Revised audiolingual-based curriculum
4. The 1984 period: Structure-based communicative curriculum
5. The 1994 period: Meaning-based communicative curriculum
6. The 2004 period: Competency-based curriculum

Indonesia's first English curriculum was introduced as a grammar translation-based curriculum. The curriculum in 1945, which focused on teaching English grammar, was implemented after the Dutch teachers of English left the country following Indonesian independence (Dardjowidjojo, 2000). Use of the grammar translation method was suitable for large classes since it was low-cost and only focused on the reading competence. The grammatical mastery was seen as just an instrument to help understand sentence structures and the translation component served as a test of grammatical mastery of English (Sahirudin, 2013).

In 1958, the government of Indonesia introduced its second English language curriculum, called the audiolingual-based curriculum. This curriculum was made based on the involvement the Ford Foundation of the United States, resulting in the creation of the Standard Training Course (STC) in Jakarta and Bukittinggi. The program was aimed at increasing the quality of teacher training. In selecting only fifty participants to join the program every year. All the teachers provided were native speakers, mostly American (Dardjowidjojo, 2000).

In 1975, the Indonesian government introduced a revised audiolingual-based curriculum. This curriculum was expected to provide systematic teaching guidelines including teaching approaches, objectives, materials, and assessments (Tjokrosujoso & Fachrurrazy, 1997). The curriculum, however, was not effective in raising the level of English competence due to the large class sizes and the absence of language laboratories (Wiramaya, 1991).

In 1984, a structure-based curriculum was introduced to develop communicative skills for Indonesian English language learners. The curriculum mainly focused on developing language skills and enriching vocabulary. The curriculum, however, was reported to be unsuccessful due to interference from the previous curriculum (Priyono, 2004). It was also inconsistent (Tjokrosujoso & Fachrurrozy, 1997), in that while the program itself was grammar-oriented, the implementation of the program focused on reading comprehension, the teaching approach aimed to be communicative, and the system of evaluation was again based on grammar skills.

In 1994, the curriculum for English was revised into a meaning-based communicative approach. This approach focused on reading, listening, speaking, and writing, and contained functional, situational, skills-based, and structure-oriented materials (Jazadi, 2000). In this curriculum, the teaching priority was still focused on reading (Priyono, 2004) and the tests in this curriculum emphasized reading

comprehension in multiple-choice questions. The communicative competence aspect was not fully measured (Sahirudin, 2013). The implementation of this curriculum showed disappointing results as Indonesian learners of English still appeared to be unable to comprehend, communicate, and write in English.

In response to the need for improving English instruction effectively, the 2004 competency-based curriculum was introduced. The concept of communicative competence pervades this curriculum. Listening, speaking, reading, and writing activities were integrated with language functions, language forms, and grammar. Yet, the implementation of this curriculum did not prove successful. Reasons included the lack of procedures for assessing English competence, the lack of teachers, resources, and facilities for supporting the implementation of a communicative approach to learning, and cultural barriers which prevent students from being interactive in the classroom (Masduqi, 2006).

The Indonesian government insisted on following the trends of implementing bilingual education by creating the English Medium of Instruction (EMI) program. In 2006, Indonesia launched the EMI as part of the National Education Law 20/3, 2003 by Directorate General for Primary and Secondary Education Management (DGPSEM). The EMI program is mainly being conducted in the international standard schools (SBI). National exam results indicated that students in the EMI program performed better in English rather than students in non-EMI programs (Sultan, Borland, & Eckersley, 2012).

The establishment of English in Indonesian universities and schools was partly due to the program initiated by America through the Ford Foundation (Candraningrum, 2008). In the 1958's the Ford Foundation was invited by the Indonesian government to improve the English teaching in Indonesia. All the teachers provided were Native American speakers. The Ford Foundation introduced the audio-lingual method, which was at the end successful to be implemented as many qualified teachers were produced.

In the 1960, FKIP Universitas Airlangga Malang set a project called the English Language Teacher Training Project (ELTTP). Again, the Ford Foundation supported the project by sending American professors to teach in the universities. Moreover, the Ford Foundation also provided them with financial and technical assistance. At the same time, through this program, some English language teachers in Indonesia were sent to US to get their master and doctoral degree (Sadtono, 1997).

The first seed disseminated through in-service training for English teachers and material developments were imported within the paradigm of the American English (Candraningrum, 2008). The Ford Foundation assisted the development of syllabi, instructional materials and manuals for secondary schools (Thomas, 1968). Also in the Suharto era, English teaching in Indonesia continued to be strongly supported by the US government scheme through the Ford Foundation (Candraningrum, 2008). The study chose American English as the control groups because American English is the first pronunciation norm in Indonesia after the initiation of Ford Foundation program and is mostly taught at Indonesian schools.

1.1.4 General Purpose of the Study

The goal of the present study is to make a contribution to the teaching of English as a foreign language in the context of the Indonesian school system. As was shown in the previous section, the level of attainment of foreign-language skills by Indonesian learners of English is in the middle of the global range, in spite of massive efforts made by the Indonesian and American governments to raise the level of English language skills in Indonesia. It would clearly be overly ambitious to attempt to address the teaching of all four skills (i.e. speaking, listening, reading, and writing) in one dissertation project. We therefore chose to address learning problems which specifically relate to the oral skills, i.e. speaking and listening. This choice follows from the theoretical insight that spoken language is the primary form of language use, whereas written language is considered a secondary form of language use, which is parasitic on spoken language. We also argue that speaking and listening are more demanding language skills in that they have to be performed in real time, in immediate exchanges of information between interactants, which exclude the use of language resources such as dictionaries, grammars, reference books and automatic translation software. Moreover, we endorse the view that within the area of spoken language communication the adequate use of pronunciation takes precedence over the correct use of syntax and (inflectional) morphology. Communication in spoken language is only possible if the listener recognizes (a sufficient number) of words per sentence. When no words are recognized, the question whether words are correctly inflected and/or in the correct syntactic order becomes irrelevant (e.g. Van Heuven, 1986, 2008; Hilton et al., 2013).

Spoken word recognition in English, as a prerequisite for understanding sentences, is largely determined by the quality of the vowels and the consonants. Moreover, when the segmental quality is poor or deviant, as is typically the case in foreign-accented English, the correct placement of the word stress (in polysyllabic words) is of the essence (Cutler, 2005; Van Heuven, 2008; Cutler & McQueen, 2014). Consonants, at least in the case of English and other Germanic languages, contribute more to the recognition of spoken words than vowels (Van Ooijen, 1994, 1996). In the context of foreign-language learning, differences between the sound structures of the learner's native language (L1) and those of the target language (L2) are often a source of difficulty. Typically, the sounds and sound structures (including melodies) of the target language are perceived in terms of the categories and sequential patterns of the learner's native language. To the extent that the structures of the L1 do not match those of the L2, sounds that belong to two different categories in the L2 may be perceived as tokens of just one category in the L1 (see section 1.3 below). A prominent source of difficulty is in the difference in the syllable structures of L1 and L2. English allows quite complex syllables with up to three consonants in the onset (i.e. preceding the vowel, as in the word *street*) and as many as four consonants in the coda (i.e. following the vowel, as in *thousandths* ending in /ndθs/). Many languages spoken in Indonesia do not allow such complex consonant clusters, so that these clusters in English present a challenge for speakers of those languages. This challenge is typically met by inadequate strategies on the part of the Indonesian learner of English, by simply omitting certain consonants, or inserting vowels between the consonants in a cluster (vowel epenthesis). Such inadequate strategies are highly detrimental to the intelligibility of the L2 speaker of English (e.g. Tajima et al, 1997).

In the present study, we concentrated on the perception and production of the vowels by Sundanese and Javanese learners of English as a foreign language. This choice was made for three reasons. First, the measurement of vowel duration and quality is relatively easy and involves only a small and uniform set of acoustic parameters (e.g. Dowd, Smith, & Wolfe, 1997). To simplify the analytic problem even further it was also decided to limit the present study to the monophthongs of English. During the production of a monophthong, the speaker needs to approximate just a single articulatory target, so that the spectral analysis can be limited to one single short time segment at or near the temporal midpoint of the vowel token.

Second, the analysis of diphthongs and consonants (and even more of consonants in sequence) would necessarily involve the tracing of dynamically changing acoustic parameter values. For instance, diphthongs involve a dynamic change from one articulatory target (near the beginning of the sound) to a second target (near the end of the sound). Moreover, the articulatory change between the two targets need not be linear, so that the proper description of diphthongs may, in fact, involve three or more sets of measurements and tracing their development over time. Consonants, by definition, are produced by a closing-and-opening gesture of the oral tract, with complex acoustic effects in which the degree of closure may or may not cause turbulence (i.e. acoustic noise) and in which the place of articulation is coded in quickly changing resonances with different frequency trajectories and/or different spectral composition of the noise bursts – which cannot be captured by measuring resonance frequencies.

Finally, consonants may or may not be produced with vocal fold vibration, which adds yet another set of acoustic parameters that has to be taken into account. Although the incorrect or deviant articulation of consonants will have a more negative effect on the intelligibility of a non-native speaker of English (see above), the effects of incorrect vowel pronunciation, in the case of Javanese and Sundanese learners of English, were expected to be large enough to substantially compromise the non-native speaker's intelligibility. Previous studies have shown that vowel quality is an important determinant of speech intelligibility (Flege, 1995) and the native vowel systems in the present research differ as much from that of English as the interfering vowel systems used by Flege (Italian, Spanish).

If indeed the results of the present study would indicate that certain target vowels are systematically mispronounced by Indonesian learners of English, they may be used to set up training programs in which the learner is explicitly instructed how to approximate the English vowels. Since articulation of monophthongal vowels can be easily visualized in a two-dimensional map (vowel space), it is technically quite feasible to use the measurements of vowel duration and resonance frequencies for visual feedback purposes as an aid to facilitate and improve the learning process (e.g. Povel & Wansink, 1986; Epps, Smith, & Wolfe, 1997), whereas visual feedback on the production of dynamically changing sounds such as diphthongs and consonants would be much more difficult to realize (or understand by the learner).

1.1.5 Research Questions

The overall research question of this thesis is:

How do Javanese and Sundanese learners of English acquire English vowels?

The sub-questions are:

1. Which English vowels are difficult to perceive and produce for Javanese and Sundanese speakers?
2. What are the causes of L2 speech learning difficulties for Javanese and Sundanese speakers?
 - a. Do the differences in vowel inventory cause English acquisition problems?
 - b. Does the absence of the vowel length feature cause English acquisition problems?

To answer the research questions, the thesis starts by describing the vowel systems of the Javanese and Sundanese and conducts an acoustic analysis of the languages. Second, the thesis tries to investigate how Javanese and Sundanese speakers perceive English vowels. Regarding L2 perception, the thesis tests the Speech Learning Model (SLM) (Flege, 1995, 2002) and the Second Language Linguistic Perception (L2LP) model (Escudero, 2005). Furthermore, the thesis tests the Feature Hypothesis (McAllister et al., 2002) and the Desensitization Hypothesis (Bohn, 1995) for vowel length acquisition.

1.1.6 Research Objectives

The current study aims to examine to what extent the properties of the vowel system in the first language interfere with the second language acquisition of American English for American English L2 learners who exhibit a smaller L1 vowel inventory than that of English, and to what extent they might have problems in producing L2 vowels and, therefore, are predicted to have difficulties with perceiving and producing English sounds (e.g. vowels). Previous studies have reported cross-linguistic comparisons between English as a second language and western European languages as a mother tongue (e.g. Spanish, German, Dutch, Swedish, Danish, and Italian). However, very little research has been done to reveal the problems Javanese and Sundanese learners have in acquiring English vowels. Overall, the goal of the current thesis is to examine the abilities and weaknesses in English sound perception and production of Javanese and Sundanese speakers.

The objectives of the present thesis are as follows:

1. To investigate the L2 learners' difficulties in English L2 sounds learning.
2. To identify the linguistic causes in the L2 sounds learning problems among Javanese and Sundanese speakers.

1.1.7 Significance of the Present Study

Recent studies have shown that Indonesian English language learners incorrectly pronounce English sounds. Riadi (2013) found that most of the students have problems in pronouncing tense and lax of English vowels. To be specific, the students are confused in distinguishing English vowels [ɪ], [i:], [ʌ], [ɑ:], [ʊ] and [u:]. This confusion is caused by the students being unfamiliar with the vowels or they did not practice pronouncing short and long English vowels.

Because non-native English teachers lack confidence in their own speaking and pronunciation (Mathew, 1997) there is considerable room for improvement in L2 production among Indonesian learners. To overcome this, comprehensive, direct and systematic investigation of the areas of difficulties in English pronunciation should be undertaken to inform pronunciation teaching. Therefore, the pronunciation problems identified in the present thesis could serve as a basis to improve English vowel pronunciation among Javanese and Sundanese L2 learners. Thus, the current thesis would improve L2 teaching in Indonesia.

1.2 Javanese, Sundanese, and English Vowels

Javanese, a spoken Austronesian language investigated in the current study, is a language which has a small inventory of vowel sounds. In the traditional analysis of its vowel inventory, Horne (1961) and Uhlenbeck (1963) report that Javanese has six vowel phonemes /a, ə, i, u, e, o/ (see Figure 1.1). Wedhawati et al. (2006) and Gordon (2006) state that Javanese has four allophonic pairs /ɪ-i/, /ɛ-e/, /a/, /ə/, /ʊ-u/, /ɔ-o/, with the members of each pair in complementary distribution: the low allophones only occur in closed syllables and high counterparts in open syllables.

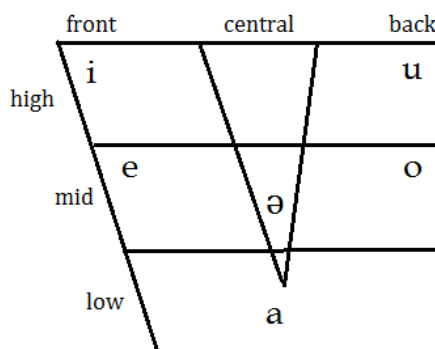


Figure 1.1 Javanese vowels as described by Wedhawati et al. (2006).

Sundanese also has a small vowel inventory, which consists of the seven vowels /i/, /a/, /ə/, /i/, /e/, /u/, and /o/ (Crothers, 1978; Sudaryat, 2007) (see Figure 1.2). These seven vowels are classified as the high front /i/, high central /i/, high back /u/, mid front /e-ε/, mid central /ə/, mid back /o-ɔ/, and central low /a/. Allophonic pairs are claimed here to exist for non-central mid vowels only, with the same complementary distributions as in Javanese.

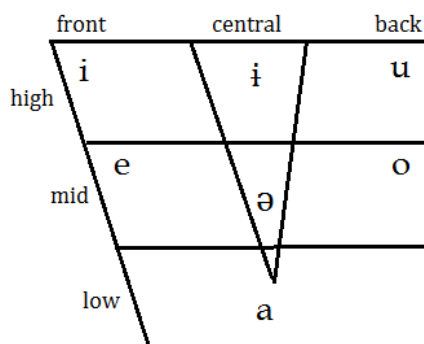


Figure 1.2 Sundanese vowels as described by Crothers (1978) and Sudaryat (2007).

The L2 investigated in this study is English, which has a complex vowel system. There are three articulatory dimensions that together distinguish the American English (AE) monophthongs from one another. The first dimension is the constriction place, along which three degrees are distinguished, as in the Indonesian languages: front /i:, ɪ, ε, æ/, central /ʌ, ɜ:/, and back /ɑ:, ɔ:, ʊ, u:/ (Ladefoged, 2001, 2006) (see Figure 1.3). The second dimension is vowel height, again with three degrees: high, mid and low, as is also applicable to the Indonesian languages. Here /i:, ɪ, ʊ, u:/ are considered high vowels, /ε, ɜ, ʌ, ɔ:/ mid

vowels and /æ:, ɑ:/ are low. English is different than the Indonesian languages in that it exploits a contrast between tense and lax vowels. Tense vowels are said to be produced with more articulatory effort so that the articulatory organs (tongue and lips) move further away from their neutral position than in the case of lax vowels. Since the articulators have to travel over a longer distance, tense vowels take more time to articulate than lax vowels, all else being equal. In English, the tenseness of a vowel causes it to be longer than its lax counterpart. Length is therefore parasitic on tenseness, and is considered a secondary rather than a primary feature in the English vowel system (Peterson & Lehiste, 1960; Hillenbrand, Clark, & Houde, 2000).

Other languages may exploit a pure or primary length contrast, such as German and Hungarian, in which vowels are contrasted in long-short pairs without any difference in quality. In English, the vowels are divided into tense/long and lax/short vowel (Chomsky & Halle, 1968; Bohn & Flege, 1990). Without exception all vowels that are marked as long in Figure 1.3, i.e. /i:, u:, ɜ:, ɔ:, ɑ:/ are tense. Phonologically, unlike tense vowels, lax vowels cannot occur at the end of a word; they have to be followed by a coda consonant. The low vowel /æ/ is phonologically lax, since it does not occur in word-final position. However, it is generally accepted that this vowel is phonetically tense in American English (but not in British English) both in terms of its extreme location in the vowel quality space and in terms of its duration (Strange et al., 2004; Wang & Van Heuven, 2006). Since the English pronunciation norm in the present study is American, we will consider /æ/ to be a tense and long vowel. Moreover, we excluded the tense vowels *ej* and *ow* which are diphthongal and the true diphthongs *aj*, *aw*, *oj*, and *ju* from in the present study.

Indonesian languages have no contrast in their phonology between tense versus lax vowels, nor between long versus short vowels. In Javanese and Sundanese, vowel duration does not contribute to any contrast in the vowel system. In Javanese, the vowel inventory lacks phonemic long vowels and diphthongs (Gordon, 2006). Javanese and Sundanese L1 speakers have no vowel length contrast (Van Zanten & Van Heuven, 1997). The correct realisation of the tense-lax (long-short) members of these contrasts in English may therefore pose a problem for Indonesian L2 speakers of English. Incorrect production of vowel length and tenseness compromises a speaker's intelligibility in English (Walker, 2001).

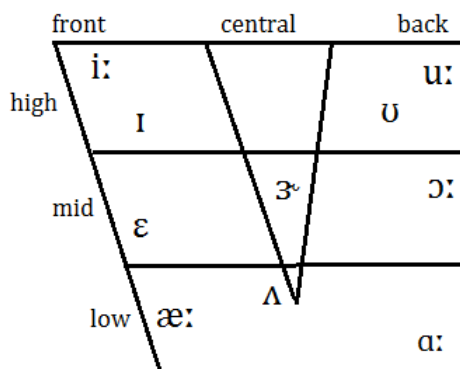


Figure 1.3 American English vowels as described by Ladefoged (2001, 2006).

The vowel systems of Javanese and Sundanese, then, differ from the English vowel system in terms of durational features and spectral characteristics. From the visual inspection of each vowel chart, it is clear that some English vowels are not found in Javanese and Sundanese—these vowels are represented by different IPA symbols. Such *new* L2 vowels (see section 1.3.1 for explanation) for the Javanese and Sundanese are /ɑ:/, /ɜ:/, /ɔ:/, /Λ/, /æ:/, /ε/, /ɪ/, /ʊ/. Similar L2 sounds among Javanese, Sundanese, and English are represented by the same IPA symbol and only differ in the diacritics. These vowels are Javanese/Sundanese /i/ and /u/, which have a length diacritic in English. Hence, the main difference between Javanese, Sundanese, and English lies in the number of vowels and the role of tenseness/duration.

1.3 Models of Second Language (L2) Speech Learning

This section reviews specific hypotheses of L2 speech learning which are relevant for the current thesis. It is not our main goal to introduce all the L2 learning models here. We will discuss three prominent models on learning L2 sounds, i.e. (i) the Speech Learning Model or SLM (Flege, 1995, 2002), (ii) the Second Language Linguistic Perception (L2LP) model (Escudero, 2005), and (iii) the Perceptual Assimilation Model (PAM) (Best et al., 1988, 2001). These models make contrasting predictions regarding L2 learning, allowing us to empirically test these models.

We also introduce in this section views on vowel length acquisition. We will focus on two hypotheses that make contrasting predictions regarding vowel length, allowing for an empirical test, namely the Feature Hypothesis and the Desensitization Hypothesis.

1.3.1 Speech Learning Model (SLM)

The Speech Learning Model (SLM) proposed by Flege (1995, 2002) posits that L2 features which are not used to signal phonological contrast in the L1 will be difficult to perceive and produce for L2 learners. This is because the L2 phonetic category formation may be blocked by the absence of the phonetic feature in the L1. The model postulates that an L2 learner, after an unspecified period of exposure to the L2, can establish a new phonetic category of an L2 sound, which is different phonetically from the closest L1 sound, as long as the learner recognizes some difference between the L1 and L2 sounds. It suggests that the perceived dissimilarity between an L2 sound and the closest L1 sound plays a role. If the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound is large, an L2 learner will be more likely to differentiate the phonetic differences between the sounds (Flege, 1995).

SLM classifies L2 sounds as either *identical*, *similar*, or *new*. To identify the difference between the three types of L2 sounds, SLM considers the IPA symbols which are used to represent the L1 and L2 sounds. IPA symbols can be used because cross-linguistic categorization occurs at a phonetic level rather than at a phonemic level (Flege, 1995). According to SLM, an *identical* L2 sound is represented by the same IPA symbol (including diacritics) used to transcribe a sound in the L1. A *similar* L2 sound is transcribed by the same base IPA symbol in the learner's L1 but differs in the diacritics. A *new* L2 sound is considered a segment, which is different acoustically and perceptually from the closest L1 sound and is represented by an IPA symbol which is not applicable to the L1 sound inventory (Flege, 1995).

Based on the classification of the L2 sounds, SLM makes specific predictions. *Identical* L2 sounds will be perceived and produced with ease because all the knowledge about the target sounds is available in the L1. Thus, in Lado's (1957) terminology, positive transfer will occur. *Similar* L2 sounds are less easily produced and perceived since the similar sounds in the L1 have perceptual equivalence and hence merge into the same category in the L2. Last, the model predicts that *new* L2 sounds will ultimately create no perception and production problems because these *new* sounds will trigger the formation of a new phonetic category in the L2 without any L1 interference. Thus, after prolonged exposure to the L2, learners start to form a new phonetic category for the L2 sounds.

Regarding Javanese, using the principle of Flege's SLM (1995), the English vowels /i:/ and /u:/ have the same IPA symbols but make a difference in their length marking. Thus, they can be considered similar L2 vowels. The English vowels /ɑ:/, /ɜ:/, /ɔ:/, /ʌ/, /æ/, /ɛ/, /ɪ/, and /ʊ/ are represented by IPA symbols which are not used in any Javanese sounds, and thus they are considered new L2 vowels. SLM predicts that the L2 speakers will have difficulty perceiving and producing similar L2 vowels /i:/ and /u:/ and easily perceive and produce the new L2 vowels /ɑ:/, /ɜ:/, /ɔ:/, /ʌ/, /æ/, /ɛ/, /ɪ/, and /ʊ/. However, given that SLM is not specific about the length and intensity of the exposure to the L1 needed to set up new sound categories, the expected difference between new and similar sounds may or may not be found. If not, the conclusion is not necessarily that SLM is wrong; it may also be an indication that the acquisition of the English vowel system by the Javanese (and Sundsnese) learners has not yet progressed to the formation of new categories. In this case both new and similar sounds will pose a problem.

Regarding Sundanese speakers, SLM makes the same predictions as for Javanese (with the same caveat): the new L2 vowels /ɑ:/, /ɜ:/, /ɔ:/, /ʌ/, /æ/, /ɛ/, /ɪ/, and /ʊ/ will be easily perceived and produced. In contrast, we would expect that the Sundanese speakers would have difficulty perceiving and producing similar L2 vowels /i:/ and /u:/.

1.3.2 Second Language Linguistic Perception (L2LP) Model

The Second Language Linguistic Perception (L2LP) model by Escudero (2005, 2009) is a relatively new model of second language learning. The L2LP model sets out five theoretical constructs: the optimal perception, the initial state, the learning task, the developmental state, and the end state. The optimal perception is the way in which an auditory speech signal is mapped onto the different phonological categories of the L2 learners depending on their phonological environment. In optimal perception, L2 learners categorize L2 phonetic signals into phonological vowels as intended by the speaker. The initial state in the model constitutes the perception of L2 sounds in which L2 learners have no prior knowledge of the target language. This phase involves a full copying of L1 perceptual mapping, i.e. all L2 sounds are mapped onto L1 phonological categories as if they were L1 sounds. Thus, L2 learners will initially perceive and produce L2 sounds by duplicating the L1 sound categories.

The learning task state in the model constitutes the different ways in which L2 learners try to reach optimal L2 perception by bridging mis-

matches between the L1 and target optimal perception (Escudero, 2005). The model distinguishes between two types of learning tasks in L2 sound perception - a perceptual task and a representational task. When the target L2 learn sounds with auditory dimensions that were not previously categorized in the L1 perception grammar, the L2 learner's perceptual task is to create new mappings to cope with the new production distributions. However, when L2 learners produce sounds with auditory dimensions that were already categorized in the L1, the L2 learner's perceptual task is to generate extra categories from the existing grammar through the redistribution or splitting of L1 perceptual mappings (Escudero, 2005). As for the representational task, the L2 learners will perceive the same vowel in the L2 by creating a new phonetic category and their L2 lexicon will contain the same lexical representation as the L1. To summarize, in the learning task state, L2 learners would be able to adjust their L2 initial perception, which is a copy of the L1 perception, and then shift it towards the optimal target L2 perception.

L2LP classifies L2 sounds as *subset*, *new*, and *similar* sounds. However, L2LP uses different definitions of what constitutes *new*, and *similar* L2 sounds than SLM. *Subset* L2 sounds occur when an L1 sound belongs to more than one category in the L2 (Escudero, 2005). *New* sounds are L2 sounds which are produced with at least one auditory dimension which has not been previously incorporated into the learner's L1 linguistic perception (Escudero, 2005). *Similar* L2 sounds are phonologically equivalent yet phonetically different from the sounds in the learner's L1 that are acoustically most similar (Escudero, 2009). L2LP assumes that the perception of similar L2 sounds is easier than of new L2 sounds because the L2 learners will preferably adjust their perceptual mapping rather than creating new L2 categories.

Based on the L2LP model, we made specific predictions of the difficulties Japanese speakers would have in perceiving English vowels. First, we expect Japanese speakers will easily identify and produce English sounds that are similar to the Japanese sounds, such as /i:/ and /u:/. However, the Japanese learners are predicted to have more difficulty in acquiring English sounds which are new to them, such as /ɑ:/, /ɜ:/, /ɔ:/, /ʌ/, /æ/, /ɛ/, /ɪ/, and /ʊ/.

The predictions of the L2LP model regarding Sundanese speakers are identical: the new L2 sounds /ɑ:/, /ɜ:/, /ɔ:/, /ʌ/, /æ/, /ɛ/, /ɪ/, and /ʊ/ will be difficult to acquire, and the similar L2 sounds /i:/ and /u:/ will be easy to acquire.

1.3.3 Perceptual Assimilation Model (PAM)

The Perceptual Assimilation Model (PAM, Best et al. 2001) tries to explain the acquisition of non-native (foreign-accented) speech sounds. PAM predicts perception and production difficulties mainly based on the similarity and difference between articulatory phonology across languages. In PAM's framework, listeners perceive information in speech through the articulatory properties (Best, 1995; Best & Tyler, 2007). Articulatory properties in the speech signal include articulatory organs (active articulator), constriction locations (place of articulation) and constriction manner (manner of articulation). Like all other models of L2 acquisition, PAM claims that the perception of non-native sounds is affected by the listeners' knowledge of native phonological classes.

According to PAM, listeners perceptually assimilate non-native speech sounds to native sounds based on detection of commonalities with their native articulatory properties. PAM posits three patterns of assimilation: (i) a non-native sound is *Categorized* as a native phoneme; (ii) a non-native sound is perceived as an *Uncategorized* sound. This occurs when the non-native sound falls between two (or more) native categories or is not consistently assigned to a single native category (Faris et al., 2016); (iii) a non-native sound is perceived as a *Non-assimilable* non-speech sound as it has no similarity with any native sound (Best, 1994; 1995). In short, listeners assimilate non-native sounds to their native phonological system by detecting similarities and discrepancies of articulatory properties of the non-native sounds and the phonological units of their native phonemes.

When a non-native sound is Categorized as a native phoneme (as in the first pattern), PAM predicts that listeners identify three assimilation types: Single-Category (SC), Category Goodness (CG) and Two-Category (TC) assimilation. SC assimilation occurs when two non-native sounds are assimilated to a single native category. This leads to a poor discrimination of the non-native contrast (Best, 1994). CG assimilation occurs when two non-native sounds are assimilated to the same native category. This assimilation leads to a moderate to very good discrimination of the non-native contrast depending on how much more one member of the contrast resembles the prototype of the L1 category than the other member does. TC assimilation occurs when two non-native sounds are assimilated to two different native sounds; this scenario yields good to excellent discrimination of the non-native contrast.

When English monophthongal vowel sounds have to be

assimilated to the native sound categories of either Javanese or Sundanese, there will be no English sounds that are outside the phonological space of the Indonesian languages, so that the Non-assimilable type will not apply. How the English monophthongs would be assimilated to the native Javanese and Sundanese vowel categories has not been established. This would require assimilation experiments as have been carried out by, e.g., Strange et al. (1998) and Bundgaard-Nielsen et al. (2011) for Japanese listeners of English, Tsukada et al. (2005) for Korean listeners and Sun & Van Heuven (2007) for Mandarin listeners. A reasonable expectation, however, would be that the /i: - ɪ/ and /u: - ʊ/ contrasts will conform to the SG or CG assimilation scenario's, predicting poor to moderate discrimination, and will therefore present a learning problem in the initial stages of the acquisition of English as an L2. The TC or CG scenarios will apply to other vowel contrasts in English as well, such as /ɛ - æ/ and /ʌ - ɜ:/. These are examples of contrasts between spectrally similar vowel qualities in parts of the vowel space where the Indonesian vowel systems have no subdivision, but where the unfamiliar length difference between the members of each pair may bring about a difference in category goodness (or 'typicality').

In the present thesis, the PAM model will not be formally tested since at this time we can only speculate which L2 sounds would be predicted to be relatively difficult for the L2 learners discussed in the present work. We will therefore specifically test the predictions of learning problems made by SLM and L2LP, and use PAM merely to provide an alternative interpretation of the results in those cases where the other models fail.

1.3.4 Feature-dependent Hypotheses

Two feature-dependent hypotheses that specifically address the acquisition of vowel length, the Feature Hypothesis by McAllister, Flege, and Piske (2002) and the Desensitization Hypothesis by Bohn (1995), are presented in this subsection.

1.3.4.1 The Feature Hypothesis

The Feature Hypothesis proposed by McAllister et al. (2002) states that L2 learners will have difficulties in acquiring specific features (e.g. duration) that are not used in the native language. According to the Feature Hypothesis the acquisition of new phonetic categories for

sounds in an L2 may be blocked by the absence of the contrastive use of a feature in the L1. Since the L2 learners in the present thesis do not exploit vowel length in their native language, whether phonologically or phonetically, they may have difficulty perceiving and producing vowel contrasts based on duration.

Several studies have provided evidence to support the Feature Hypothesis. McAllister et al. (2002) investigated the perception and production of Swedish involving twenty native speakers each group of Estonian, English, and Spanish L2 learners of Swedish. According to the L1 sound system, the role of duration is different in these three languages. Estonian (with a three-member contrast in vowel length) uses duration more than English (with a two-member secondary vowel length contrast), followed by Spanish (which does not exploit vowel length at all). The results of the production experiment showed that the Spanish speakers were less successful than the native Swedish speakers in producing Swedish quantity distinctions. Hence, the results support the Feature Hypothesis model stating that the duration feature, which has a less prominent role in the L1, may be difficult to perceive.

The absence of the duration cues in the Javanese L1 phonological system may cause learning difficulties for the Javanese speakers in producing English vowels. Following the Feature Hypothesis, the Javanese and Sundanese speakers, who do not exploit vowel length in their L1, would have difficulties with the production of English vowels that belong to different length categories. More particularly, they may be relatively unsuccessful in producing long/tense English vowels /i:, u:, ɜ:, ɔ:, æ:, ɑ:/ as compared to the short/lax English vowels /ɪ, ʊ, ɛ, ʌ/.

1.3.4.2 The Desensitization Hypothesis

New-born infants possess a genetically endowed sensitivity to any acoustic difference that can be found in human languages. In the first six months of their lives, however, infants quickly learn which acoustic contrasts matter in their ambient language and which differences between sounds are irrelevant (Kuhl & Iverson, 1995). The result of the acquisition of the mother tongue is that the child loses sensitivity to sound contrasts that are not part of the native phonology. Once the L1 acquisition process has been completed, it is difficult to undo this desensitization. However, some differences between sounds have been claimed to elude desensitization, possibly because they also fulfill useful roles elsewhere in the phonology or in perceiving non-linguistic sound properties, e.g. in music. The Desensitization Hypothesis proposed by Bohn (1995) postulates that L2 learners who do not use

duration contrastively in their first language will show no difficulty in acquiring vowels which are different in length because these cues are acoustically salient and easy to access even at a later stage in life when confronted with length differences in a foreign language. Thus, according to the Desensitization Hypothesis, the L2 learners will always rely heavily on duration cues even if duration is not used contrastively in their L1.

A number of studies in the investigation of vowel length production have supported Bohn's Desensitization Hypothesis. Cebrian (2006) investigated how L1 Catalan native speakers produced L2 Canadian English and found that despite not exploiting duration in their native L1 system, the L1 Catalan speakers showed a reliance on duration as the main cue to English vowel contrast. Similarly, Van Heuven (1986) tested the perception of the tense-lax contrast by Turkish learners of Dutch, which is simultaneously cued by vowel colour and duration. His results showed that the Turkish learners relied exclusively on the duration cue and ignored the quality difference, even though Turkish has no length contrast. Later, Nimz (2011) found that Turkish learners of German had no problems with the contrast between long and short vowels in the target language. These results support Bohn's Desensitization Hypothesis since the Turkish speakers relied on duration as the main cue to the German vowels despite not having experience with duration in their L1.

Following the Desensitization Hypothesis, Javanese and Sundanese speakers are predicted to have little difficulty in reproducing the correct long duration for the tense vowels /i:, ɛ:, ɑ:, ɔ:, u:/ and possibly /æ/ and the correct shorter duration of the lax vowels /ɪ, ʊ, ɛ, ʌ/ even though duration is not a contrastive feature in the learners' native languages. However, the quality differences between the tense-lax vowel pairs will remain a problem as the learners have become desensitized to the small quality differences that characterize the English tense-lax pairs. This result, if obtained, would replicate the findings for the production of the monophthongs by Mandarin Chinese learners of English as reported by Wang and Van Heuven (2006), where Mandarin, like the Indonesian languages, does not have any length contrasts in its phonology.

1.3.5 Models and Hypotheses of L2 Speech Learning: Summary

Current models of L2 speech learning, such as the Speech Learning Model (SLM) by Flege (1995) and the Second Language Linguistic Perception (L2LP) model by Escudero (2005), provide helpful heuristics in describing and interpreting phenomena observed in L2 perception and production. The models emphasize the need for comparing the sound systems of the L1 and L2 at a phonetic level. However, there are differences among these models.

Flege (1995) explicitly integrates the perception and production of L2 sounds in his SLM. Escudero (2005), however, discusses further studies which support the integration of perception data only. SLM is a theory of the ultimate attainment of the perception and production of the sounds in the L2. It is not explicit about the time course of the L2 acquisition process, and it makes no predictions of the length and intensity of exposure to L2 input needed for the learner to set up new categories or to increase the tolerance of existing categories in the L1 so as to include similar sounds in the L2. As long as the L2 acquisition process has not finished, similar sounds in the L2 may be more successfully approximated than new sounds. Only at the very end of the L2 acquisition process will the new sound categories be indistinguishable from those of native L1 speakers of the target language ('authentic'). Best's Perceptual Assimilation Model (PAM), in contrast to SLM, addresses the question of how listeners perceive foreign sounds when they listen to them for the first time in their lives, and tries to predict from the initial categorization of the foreign sounds how easy or difficult it will be for the L2 learner to learn to perceive and produce contrasts that matter in the L2. Escudero's L2LP model differs from the other models in that it tries to model, step by step, how the L2 learners adjust categories and category boundaries in their mental representation of the L2 sound system as the acquisition process progresses from the initial stage (in which the learners assume that the L2 categories are identical to the L1 categories) to the stage of final attainment of the L2, in which the definition of the categories is the same as those of native L1 listeners. L2LP may be used to account for the order in which the shifts of category boundaries take place but it makes no specific predictions when these shifts will be implemented.

The current study aims to investigate experimentally from both L2 perception and L2 production data which of these two models, SLM and L2LP, is best supported. Therefore, the present thesis will test the

contrastive predictions made by the two models. On the assumption that the learners L2 acquisition process has been fully completed, SLM predicts that the perception and production of *new* L2 sounds will be more successful, i.e. closer to the L1 norms, than of *similar* L2 sounds. L2LP, in contradistinction to this, predicts that creating *new* categories is more problematic than shifting the boundaries of L1 categories to accommodate *similar* sounds in the L2. Hence, for both Javanese and Sundanese SLM predicts that the *new* L2 vowels /ɪ, ʊ, ʌ, ɛ, ʌ, ɔ, æ, ɑ:/ will be successfully produced and the *similar* L2 vowels /i:/ and /u:/ will be more difficult to acquire, while on the other hand L2LP predicts the opposite pattern - that the perception of *new* L2 vowels will be difficult and the correct perception of *similar* L2 vowels will be less problematic to achieve.

Regarding L2 speech learning hypotheses that specifically focus on the acquisition of vowel length, the Feature Hypothesis (McAllister et al., 2002) and the Desensitization Hypothesis (Bohn, 1995), in principle, make opposite predictions. The Feature Hypothesis predicts that Javanese and Sundanese speakers, who do not exploit duration cues in the L1 vowel system, would have difficulties acquiring long vowels. In contrast, the Desensitization Hypothesis predicts that Javanese and Sundanese speakers should have little difficulty in making a distinction between long and short vowels. However, the two hypotheses need not be mutually exclusive. If Indonesian learners of English were to use vowel duration rather than quality differences to differentiate between the members of tense-lax vowel pairs in English, and at the same time use vowel duration less effectively than native English speakers do, both hypotheses are needed to account for the result.

1.4 Experimental design

This section provides a description of the methods of data collection and target groups of the experimental design.

1.4.1 Data collection

The current study adopts perception and production tests for data collection. For the perception task, the study used data obtained by studying real time responses of hand movements made by listeners when, after hearing one target vowel sound, they had to point to one of two targets shown on a computer screen, representing the target and a competitor. The production tests aim to provide insight into articulatory

and acoustic differences between the native L1 vowels of Javanese and Sundanese speakers, their approximations to the vowels of English, specifically in vowel duration (quantity) and formant frequencies (representing vowel quality), and those obtained from American native speakers of English. Data are archived in Data Archiving and Networked Services (DANS) Easy, an online archiving system for depositing research data¹.

1.4.2 Speaker groups

This research targets three groups of participants with different linguistic backgrounds. In order to be able to test the predictions of SLM, we tried to meet the assumption of completed acquisition of the L2 sound system by recruiting advanced learners of English, i.e. university students specializing in English language and literature, with at least nine years of English training (during primary and secondary school and one freshman year at the university). The Javanese university English L2 learners represent the first experimental group that participated in the production and perception experiments. The Sundanese university English L2 learners represent the second experimental group. Finally, native English speakers of General American English are involved in the experiments as a control group. The native speakers of English had just arrived in Indonesia by the time I invited them to participate in the experiment. The selection of participants varied in terms of nationalities, first and second language competence, frequency of native language use, and L1 and L2 experiences. The Javanese group was recruited from Gadjah Mada University in Yogyakarta, Indonesia, while the Sundanese group was recruited from Padjajaran University in West Java, Indonesia. The American native English speakers were recruited from a variety of locations in Indonesia.

1.5 Thesis outline

The structure of this dissertation is as follows.

Chapter 2 investigates the production of the vowels of Javanese and Sundanese by L1 speakers. It attempts to find out how Javanese and

¹ The dataset is now available to the public and can be cited as: Perwitasari, A (Leiden University) (2018): The Acquisition of English Vowels by Javanese and Sundanese Native Speakers. DANS. <https://doi.org/10.17026/dans-z3v-25xs>.

Sundanese speakers produce vowels in their native language. It provides an acoustic analysis of Javanese and Sundanese vowels spoken by Javanese and Sundanese speakers.

Chapter 3 identifies the perception problems of Javanese and Sundanese learners of English. It explores the extent to which these learners are able to identify speech sounds produced by American native English speakers using the so-called mouse tracking methodology.

Chapter 4 provides an acoustic analysis of English vowels spoken by Javanese and Sundanese speakers. It includes an analysis of the formant frequencies and vowel duration of Javanese and Sundanese speakers when they produce English vowels and compares these to the results obtained for the same materials as spoken by American native speakers.

Chapter 5 presents a general discussion and concluding remarks on the findings. The chapter formulates recommendations on how these findings could be applied in the field of second language learning, especially for Javanese and Sundanese learners of English.

References

- Adelaar, K. A., Prentice, D. J., Grijns, C. D., Steinhauer, H., & van Engelenhoven A. Th. P. G. (1996). Malay: Its history, role and spread. In S. A. Wurm, P. Mühlhäusler, & D. T. Tryon (Eds.), *Atlas of Languages of intercultural communications in the Pacific, Asia and the Americas* (pp. 673-693). Berlin: Mouton de Gruyter.
- Best, C. T., McRoberts, G. W., Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology Human Perception & Performance*, 4, 45-60.
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.
- Best, C. T. (1995). A direct-realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.

- Best, C. T., McRoberts, G. W. & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system, *Journal of the Acoustical Society of America*, 109, 775-794.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34). Amsterdam: John Benjamins.
- Bundgaard-Nielsen, R., Best, C. T., & Tyler, D. (2011). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, 32, 51-67. DOI:10.1017/S0142716410000287.
- 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.
- Candraningrum, D. (2008). *The challenge of teaching English in Indonesia's Muhammadiyah Universitites (1958-2005): Mainstreaming gender through postcolonial muslim women writers*. Berlin: Lit Verlag.
- Cebrian, J. (2006). Experience and the use of non-L1 duration in L2 vowel categorization. *Journal of Phonetics*, 34, 372-387. DOI: 10.1016/j.wocn.2005.08.003.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- Cohn, A. C., & Ravindranath, M. (2014). Local languages in Indonesia: Language maintenance or language shift? *Masyarakat Linguistik Indonesia*, 32(2), 131-148.
- Collins, J. T. (1998). *Malay, world language: A short history*. Jakarta: Dewan Bahasa dan Pustaka.
- Crothers, J. (1978). Typology and universals of vowel systems. In J. H. Greenberg (Ed.), *Universals of human language* (Vol. 2, pp. 94-152). Stanford, CA: Stanford University Press.
- Cutler, A. (2005). The lexical statistics of word recognition problems caused by L2 phonetic confusion. In *Proceedings of the 9th European Conference on Speech Communication and Technology* (pp. 413-416). Portugal: Interspeech 2005.

- Cutler, A., & McQueen, J. M. (2014). How prosody is both mandatory and optional. In J. Caspers, Y. Chen, W. Heeren, J. Pacilly, N. O. Schiller, & E. Van Zanten (Eds.), *Above and Beyond the Segments: Experimental linguistics and phonetics* (pp. 71-82). Amsterdam: Benjamins.
- Dardjowidjojo, S. (1996). The role of English in Indonesia: A dilemma. *Paper presented at the 44th TEFLIN Seminar, Surabaya, 7-10 October.*
- Dardjowidjojo, S. (1997). English policies and their classroom impact in some Asian countries. In G. Jacobs (Ed.), *Language classroom of tomorrow: Issues and responses*. Singapore: SEAMEO Regional Language Centre.
- Dardjowidjojo, S. (2000). English teaching in Indonesia. *English Australia Journal*, 18, 22-30.
- Dardjowidjojo, S. (2003). *Rampai Bahasa, Pendidikan, dan budaya*. Jakarta: Yayasan Obor Indonesia.
- Dowd, A., Smith, J. R. & Wolfe, J. (1997). Learning to pronounce vowel sounds of the vocal tract as feedback in real time. *Language and Speech*, 41, 1-20.
- Ebing, E. (1997). *Form and Function of Pitch Movements in Indonesian*. Leiden: Research School CNWS.
- Educational Testing Service. (2017). *Test and Score Data Summary for TOEFL iBT® Tests January 2017 – December 2017 Test Data*. Retrieved from <http://www.ets.org>.
- Epps, J., Smith, J. R., & Wolfe, J. (1997). A novel instrument to measure acoustic resonances of the vocal tract during speech, *Measurement Science and Technology*, 8, 1112-1121. DOI: 10.1088/0957-0233/8/10/012.
- 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* (pp. 151-190). Berlin: Mouton de Gruyter.
- Faris, M. M., Best, C. T., & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *Journal of the Acoustical Society of America*, 139(1), EL1-EL5. DOI: 10.1121/1.4939608.
- 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. (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.
- Gordon, M. (2006). *Syllable weight: Phonetics, phonology, typology*. London: Routledge.
- Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America*, 108, 3013-3022. DOI: 10.1121/1.1323463.
- Hilton, N. H., Gooskens, C., & Schüppert, A. (2013). The influence of non-native morphosyntax on the intelligibility of a closely related language. *Lingua*, 137, 1-18. DOI: 10.1016/j.lingua.2013.07.007
- Horne, E. C. (1961). *Beginning Javanese* (Vol. 3). Yale Linguistics Series. New Haven, London: Yale University Press.
- Jazadi, I. (2000). Constraints and resources for applying communicative approaches in Indonesia. *English Australia Journal*, 18, 31-40.
- Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect", in W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (121-154). Timonium, MD: York Press.
- 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.
- Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.
- Marcellino, M. (2005). Competency-based Language Instruction in Speaking Classes: Its Theory and Implementation in Indonesian Contexts. *A paper presented at the Third Annual International Conference on Education*, Hawaii, January 4-7, 2005.
- Marcellino, M. (2008). English language teaching in Indonesia: A continuous challenge in education and cultural diversity. *TEFLIN Journal*, 19(1), 57- 69.
- Masduqi, H. (2006). The competency-based curriculum of English subject for senior high school in Indonesia: A critical evaluation. *Jurnal Humanitas*, 1, 56-68.
- Mathew, R. (1997). *Report of the CBSE-ELT: Curriculum implementation study (1993-1997)*. India: Central Institute of English Languages.

- Mattarima, K., Hamdan, A.R. (2011). The Teaching Constrain of English as a Foreign Language in Indonesia: The Context of School Based Curriculum. *Sociohumanika*, 4(2), 287-300.
- McAllister, R., Flege, J., & 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. DOI: 10.1006/jpho.2002.0174.
- Nababan, P. W. J. (1985). Bilingualism in Indonesia: Ethnic language maintenance and the spread of the national language. *Southeast Asian Journal of Social Science*, 13, 1-18.
- Nimz, K. (2011). Vowel perception and production of late Turkish learners of L2 German. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1494-1497). Hong Kong: Department of Chinese, Translation, and Linguistics, University of Hong Kong.
- Nur, C. (2004). English language teaching in Indonesia: Changing policies and practices. In H. W. Kam & R. Y. L. Wong (Eds.), *English language teaching in East Asia today: Changing policies and practices* (2nd ed., pp. 178-186). Singapore: Eastern University Press.
- Paauw, S. (2009). One land, one nation, one language: An analysis of Indonesia's national language policy. In H. Lehnert-LeHouillier & A. B. Fine (Eds.), *University of Rochester Working Papers in the Language Sciences*, 5, 2.
- Panggabean, H. (2007). How to Motivate English Learners Faced with Psychological Burden. *Kata*, 9, 158-168. Surabaya: English Department, Faculty of Letters, Petra Christian University.
- Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32, 693-703. DOI: 10.1121/1.1908183.
- Poedjosoedarmo, S. (1982). Javanese influence on Indonesian. *Pacific Linguistics, Series D*(38). Canberra: ANU.
- Povel, D.-J., & Wansink, J. (1986). A computer-controlled vowel corrector for the hearing impaired. *Journal of Speech and Hearing Research*, 29, 99-105.
- Prentice, D. J. (1978). The best chosen language. *Hemisphere*, 22, 18-33.
- Priyono. (2004). Logical problems of teaching English as a foreign language in Indonesia. In B. Y. Cahyono & U. Widiati (Eds.), *The tapestry of English language teaching and learning in Indonesia* (pp. 17-28). Malang: State University of Malang Press.

- Renandya, W A. (2000). Indonesia. In Wah Kam Ho and Ruth Y. L. Wong (Eds). *Language Policies and Language Education: The Impact in East Asian Countries in the Next Decade* (pp. 113-137). Singapore: Times Academic Press.
- Riadi, A. (2013). *Students' problems in pronouncing short and long English vowels* (Unpublished article). Universitas Tanjung Pura, Pontianak.
- Rubin, J. (1977). Indonesian language planning and education. In J. Rubin, B. H. Jernudd, J. Das Gupta, J. A. Fishman, & C. A. Ferguson (Eds.), *Language planning process* (pp. 111-130). The Hague: Mouton.
- Sadtono, E. (1976). An interim report on the teaching of English at the five Centres of Excellence in Indonesia. In *Laporan Hasil-hasil Alumni RELC di perguruan tinggi* (pp. 95-107). Malang: IKIP Malang.
- Sadtono, E. (1997). *The Development of TEFL in Indonesia*. Malang: IKIP Malang
- Sahirudin. (2013). The implementation of the 2013 curriculum and the issues of English language teaching and learning in Indonesia. *Proceedings of The ASIAN Conference on Language Learning 2013*. Osaka, Japan.
- Simons, G. F., & Fennig, C. D. (2017). *Ethnologue: Languages of the world (20th ed.)*. Dallas, TX: SIL International.
- Sneddon, J. N. (2003). *The Indonesian language: Its history and role in modern society*. Sydney: UNSW Press.
- Steinhauer, H. (1980). On the history of Indonesian. In A. A. Barensten, B. M. Groen, & R. Sprenger (Eds.), *Studies in Slavic and General Linguistics* (Vol. 1, pp. 349-375). Utrecht: Rodopi.
- Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., & Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, 26, 311-344.
- Strange, W., Bohn, O.-S., Nishi, K., & Trent, S.A. (2004). Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, 118, 1751-1762.
- Sudaryat, Y. (2007). *Elmuning Sora Basa Sunda*. Bandung: Pustaka Luang Bandung.
- Sultan, Borland, H., & Eckersley, B. (2012). English medium of instruction in Indonesian public junior secondary school: Student's language use, attitude/motivation, and foreign language outcomes. *ACTA International TESOL Conference*. Cairns, Australia.

- Sun, L., & Van Heuven, V. J. (2007). Perceptual assimilation of English vowels by Chinese listeners. Can native-language interference be predicted?, in B. Los, & M. van Koppen, (Eds.), *Linguistics in the Netherlands 2007* (pp. 150-161). Amsterdam: John Benjamins.
- Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, 25(1), 1-24. DOI: 10.1006/jpho.1996.0031.
- Tjokrosujoso, H., & Fachrurrazy (1997). *Pengembangan materi bahasa Inggris dan kurikulum SMU (Curriculum and material development in senior high school English subject)*. Jakarta: Universitas Terbuka.
- Thomas, R.M. (1968). Indonesia: The English- language curriculum. In R.M.Thomas, L.B. Sands, & D.L.Brubaker (Eds.), *Strategies for curriculum change: Cases from 13 nations* (pp. 279-322). Scranton, ennsylvania: International Textbook Company.
- Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. E. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, 33, 263-290.
- Uhlenbeck, E. M. (1963). Review of Beginning Javanese, by Eleanor C. Home. *Lingua*, 12, 69-86.
- Van Heuven, V. J. (1986). Some acoustic characteristics and perceptual consequences of foreign accent in Dutch spoken by Turkish immigrant workers. In J. van Oosten, J. F. Snapper (eds) *Dutch Linguistics at Berkeley, papers presented at the Dutch Linguistics Colloquium held at the University of California, Berkeley on November 9th, 1985* (pp. 67-84). Berkeley: The Dutch Studies Program, U.C. Berkeley.
- Van Heuven, V. J. (2008). Making sense of strange sounds: (mutual) intelligibility of related language varieties. A review. *International Journal of Humanities and Arts Computing*, 2, 39-62.
- Van Ooijen, B. A. (1994). *The processing of vowels and consonants* (doctoral dissertation, Leiden University), pp. 110-117.
- Van Ooijen, B. A. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, 24, 573-583.
- Van Zanten, E., & Van Heuven, V. J. (1997). Effects of word length and substrate language on the temporal organisation of words in Indonesian. In C. Odé & W. A. L. Stokhof (Eds.), *Proceedings of the 7th International Conference on Austronesian Linguistics* (pp. 201-16). Amsterdam/Atlanta: Rodopi.

- Walker, R. (2001). Pronunciation for international intelligibility. Karen's linguistics issues: Free resources for teacher and students of English. *English Teaching Professional Magazine*, 22, 1-4.
- 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: John Benjamins.
- Wedhawati, E. W., Nurlina, S., Setiyanto, E., Marsono, Sukesi, R., & Baryadi, I. P. (2006). *Tata Bahasa Jawa Mutakhir [Advanced Grammar of Javanese]*. Yogyakarta: Penerbit Kanisius.
- Wiramaya, S. R. (1991). *Teacher Constraints* (Unpublished MA paper). Institute of Applied Linguistics, Heriot-Watt University.

