Author: Perwitasari, A.

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

Issue Date: 2019-02-19

The Acquisition of English Vowels by Javanese and Sundanese Native Speakers

Second language (L2) learners often face difficulties while learning L2 sounds. Evidence suggests that difficulties in learning L2 sounds are affected by the first language (L1). Given the prominent status of English as a foreign language in the multilingual context of Indonesia, it is important to investigate whether Javanese and Sundanese learners of English show systematic problems in learning English. The results could not only serve as a test of L2 speech learning models, but also improve English education in Indonesia. Therefore, speech production and mouse-tracking experiments were carried out to investigate both the L1 vowel systems- and the pattern of L2 acquisition problems among Javanese and Sundanese learners of English.

This book gives an overview of the native vowels of Javanese and Sundanese as well as the perception and production of English vowels by these speakers. The first empirical chapter offers a presentation of the theoretical background and insight into the L1 vowel systems of Javanese and Sundanese speakers. Subsequent data chapters examine perceptual and articulatory diffuclties these L2 learners have in the acquisition of English.

The general conclusion is that English vowel perception and production is difficult for Javanese and Sundanese learners of English. The results of the present thesis showed that the L2 speakers do not accurately perceive the *new* L2 vowels /ɑː, ʌ, æː, ɛ, ɪ, ʊ/ and the *similar* L2 vowels /iː, uː/. In terms of pronouncing English vowels, the L2 learners overshorten both long and short English vowels.

This book offers approaches to improve the perception and production of English vowels among the Javanese and Sundanese speakers. We recommend that teachers of English design vowel identification tasks to improve their sound perception of the English vowels. We also suggest that Javanese and Sundanese learners of English should be trained to pronounce vowels /ɑː/, /ɪ/, /æː/, and /iː/ correctly with more openness and a frontal tongue position. The phonetic training should also focus on lengthening short and long English vowels.

This book is of interest to scholars, educators, practitioners and students of EFL.

ISBN 978-94-6093-317-2

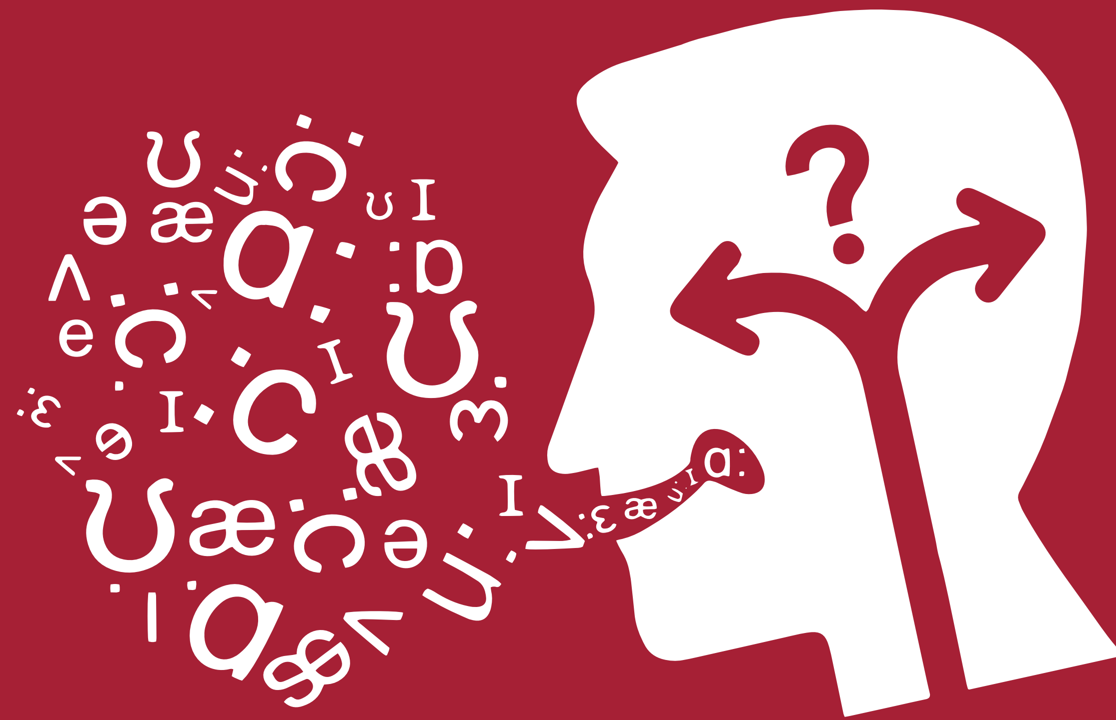
LOT
Netherlands Graduate School of Linguistics
Landelijke Onderzoekschool Taalwetenschap

LOT 532

Arum Perwitasari

The Acquisition of English Vowels by Javanese and Sundanese Native Speakers

LUCL



Arum Perwitasari

The Acquisition of English Vowels by Javanese and Sundanese Native Speakers

LOT
Netherlands Graduate School of Linguistics

LUCL

The Acquisition of English Vowels by Javanese and Sundanese Native Speakers

Published by
LOT
Kloveniersburgwal 48
1012 CX Amsterdam
The Netherlands
e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration:
Speech Confusion by M. Badruzzaman@SmartDesign

ISBN: 978-94-6093-317-2
NUR 616

The Acquisition of English Vowels by Javanese and Sundanese Native Speakers

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 19 februari 2019
klokke 13.45 uur

door

ARUM PERWITASARI

geboren te Dabo Singkep, Indonesia
in 1985

Promotores : Prof. dr. Niels O. Schiller
Prof. dr. Marian Klamer

Co-promotor : Dr. Jurriaan Witteman

Promotiecommissie:
Prof. dr. Claartje C. Levelt
Prof. dr. Vincent J.J.P. van Heuven (Pannonia University)
Prof. dr. Abigail C. Cohn (Cornell University)
Dr. Nivja de Jong

ACKNOWLEDGEMENTS

I would like to express my deep sense of gratitude to a number of people who shared their experience, knowledge and support. Without their guidance and persistent help this dissertation would not have been possible.

Firstly, I would like to thank my supervisors and promotors, prof. dr. Niels O. Schiller, prof. dr. Marian Klamer and dr. Jurriaan Witteman, who offered their consistent guidance, generous time and valuable advice, without which this study would not have been a success. I am very grateful to have been able to carry out this work under their supervision and I owe them the greatest degree of appreciation.

Secondly, I am genuinely appreciative of prof. dr. L.J. Slikkerveer and Ibu Mady Slikkerveer for their valuable scientific assistance during my study. I really feel honored to have had such a great opportunity to work as a guest researcher in the Leiden Ethnosystem and Development Programme (LEAD). My thankfulness also goes to the members of the LEAD Programme for their great support and their kind care.

Thirdly, my appreciation extends to my colleagues in Educational Testing Service (ETS) in Princeton, Amsterdam and all over the continents. For my dearest colleagues at ETS, Eileen, Mavi, Miriam, Stella and Wouter, thank you for all the supports. I also extend my gratitude to the members of the Language and Cognition Group (LACG) and the Leiden Institute for Brain and Cognition (LIBC) for their constructive comments and suggestions during my study.

My appreciation extends to my amazing friends at Leiden University Centre for Linguistics (LUCL) who have been with me during every stage of my PhD research. I am very grateful to all the Javanese-, Sundanese- and English-speaking participants who have been willing to contribute to my research. I would also like to express my thanks to the Indonesian Student Association in Leiden (PPI Leiden) that have supported me during my study in the Netherlands.

Above all, I am indebted to my family for their love and support over the years. Last but not least, my gratitude goes to all of my friends who directly and indirectly helped me to complete this dissertation. Thank you for helping me to stay sane throughout this process. Any omission from these brief acknowledgements does not mean lack of gratitude.

For my reason to BREATH,
Hafidz Ibnu Azzam

Table of Contents

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 oficially 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; Dardjowijojo, 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' 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-Eurpean 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-lamguage 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 of 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], [iː], [ʌ], [ɑː], [ʊ] and [uː]. 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.

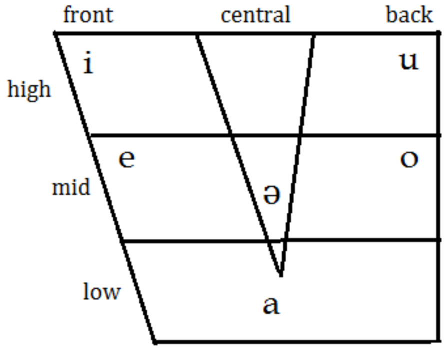


Figure 1.1 Javanese vowels as described by Wedhawati et al. (2006).

Sundanese also has a small vowel inventory, which consists of the seven vowels /i/, /a/, /ə/, /ɨ/, /e/, /u/, and /o/ (Crothers, 1978; Sudaryat, 2007) (see Figure 1.2). These seven vowels are classified as the high front /i/, high central /ɨ/, high back /u/, mid front /e-ɛ/, 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.

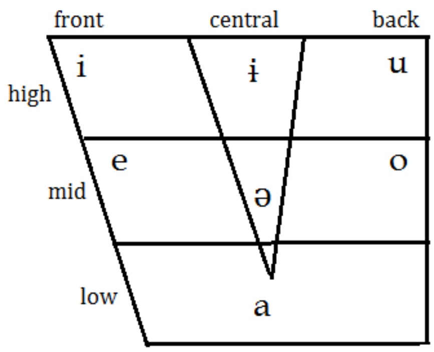


Figure 1.2 Sundanese vowels as described by Crothers (1978) and Sudaryat (2007).

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 /iː, ɪ, ɛ, æ/, central /ʌ, ɝ/, and back /ɑː, ɔː, ʊ, uː/ (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 /iː, ɪ, ʊ, uː/ are considered high vowels, /ɛ, ɝ, ʌ, ɔː/ mid vowels and /æː, ɑː/ 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. /iː, uː, ɝ, ɔː, ɑː/ 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).

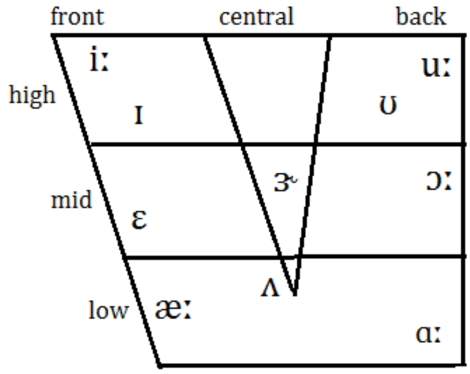


Figure 1.3 American English vowels as described by Ladefoged (2001, 2006).

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 /ɑː/, /ɝ/, /ɔː/, /ʌ/, /æː/, /ɛ/, /ɪ/, /ʊ/. S*imilar* 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 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 /iː/ and /uː/ have the same IPA symbols but make a difference in their length marking. Thus, they can be considered similar L2 vowels. The English vowels /ɑː/, /ɝ/, /ɔː/, /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 /iː/ and /uː/ and easily perceive and produce the new L2 vowels /ɑː/, /ɝ/, /ɔː/, /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 /ɑː/, /ɝ/, /ɔː/, /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 /iː/ and /uː/.

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 (Escurero, 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 Javanese speakers would have in perceiving English vowels. First, we expect Javanese speakers will easily identify and produce English sounds that are similar to the Javanese sounds, such as /iː/ and /uː/. However, the Javanese learners are predicted to have more difficulty in acquiring English sounds which are new to them, such as /ɑː/, /ɝ/, /ɔː/, /ʌ/, /æ/, /ɛ/, /ɪ/, and /ʊ/.

The predictions of the L2LP model regarding Sundanese speakers are identical: the new L2 sounds /ɑː/, /ɝ/, /ɔː/, /ʌ/, /æ/, /ɛ/, /ɪ/, and /ʊ/ will be difficult to acquire, and the similar L2 sounds /iː/ and /uː/ 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 /iː - ɪ/ and /uː - 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 - ɝ/. 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 /iː, uː, ɝ, ɔː, æː, ɑː/ 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 acquision 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 desensitivization. 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 /iː, ɝ, ɑː, ɔː, uː/ 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, ʊː, ɝ, ɛ, ʌ, ɔː, æ, ɑː/ will be successfully produced and the *similar* L2 vowels /iː/ and /uː/ 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 data[1].

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.

2

Quality of Javanese and Sundanese Vowels

A preliminary version of this chapter appeared as:
Perwitasari, A., Klamer, M., and Schiller, N. O. (2017). Quality of Javanese and Sundanese vowels. *Journal of the Southeast Asian Linguistics Society*, *10*(2), 1-9. DOI: http://hdl.handle.net/10524/52406.

Abstract

The aim of this study is to describe the vowel systems of Javanese and Sundanese. The acoustic properties of vowels in Javanese and Sundanese vowels have not been instrumentally examined. The current study seeks to investigate to what extent the vowels produced by Javanese and Sundanese speakers match the impressionistic description of the Javanese vowels found in Wedhawati et al. (2006) and Sundanese vowels described by Crothers (1978). We recorded the vowel production of four Javanese and four Sundanese native speakers and measured the formant frequencies F_1 and F_2. The results confirm that the Javanese schwa is considerably higher than its Sundanese counterpart. Javanese schwa was also found to be higher than Javanese /e/ and /o/. Sundanese male speakers (but strangely not the female speakers) produced a closed central vowel /ɨ/. Overall, the results fit Crothers' description mentioning that Sundanese has one closed and one mid central vowel in a 7-vowel sound system. Overall, the findings of the formant frequencies of the Javanese and Sundanese vowels are consistent with the description of the vowels in the earlier studies by Wedhawati et al. (2006) and Crothers (1978). In addition, the durations of Javanese and Sundanese are phonetically short, between 60 and 100 ms for all vowels.

Keywords: *acoustic analysis, formants, vowel quality, phonation, Javanese, Sundanese*

2.1 Introduction

Javanese has the most first-language speakers of any Austronesian language (Ogloblin, 2005; Oakes, 2009). Javanese is spoken by about 65 million people and considered the thirteenth most widely spoken language in the world (Comrie, 2003). Nothofer (1982) classifies Javanese as a member of the Malayo-Polynesian subgroup, which includes Malay, Madurese, Sundanese, and Lampung. Javanese is spoken primarily in the central and eastern region of Java Island (Oakes, 2009). There are three dialects of Javanese, which are mutually intelligible: Solo-Yogyakarta, East Javanese, and West Javanese (NVTC, 2007; Cole, Hara, & Yap, 2008). The Solo-Yogyakarta dialect is spoken in the center of Java and is considered the standard form of Javanese. The East Java dialect is spoken in Surabaya, Malang, and Pasuruan, and the West Javadialect is spoken in Banten, Cirebon, and Tegal (Gordon, 2005). The present study examines the vowel quality of the Solo-Yogyakarta dialect.

Sundanese is spoken by approximately 34 million people in Indonesia, making it the second most widely spoken first language in Indonesia after Javanese (Lewis, 2009). Sundanese is spoken in the western half of the island of Java (Hardjadibrata, 1985) (see Figure 2.1). Sundanese has four dialects: Banten, Bogor - Karawang, Priangan, and Cirebon (Nothofer, 1977; Muslim et al., 2010). The Banten dialect is spoken in Karesidenan Banten, the Bogor-Karawang dialect is spoken in Tangerang, Bogor, Purwakarta, Krawang, and Subang, the Priangan dialect is spoken in Karesidenan Priangan, and the Cirebon dialect is spoken in Karesidenan Cirebon, Brebes, and Cilacap. The subjects in the current study spoke the Bogor-Karawang dialect of Sundanese.

Figure 2.1 Map of traditional languages spoken on the island of Java (Source: Simons & Fennig, 2018).

2.2 Javanese and Sundanese Vowels

The current study aims to describe the vowel system of Javanese and Sundanese. The study will further examine whether the Javanese and Sundanese-accented vowels in Standard Indonesian are identical to their Javanese and Sundanese counterparts. Javanese comprises six vowel phonemes: /a, ə, i, u, e, o/ (Uhlenbeck, 1963; Horne, 1961; Clynes & Rudyanto, 1995). In the view of most scholars, it has four allophonic pairs: [i] - [ɪ], [u] - [ʊ], [e] - [ɛ], and [o] - [ɔ] (Dudas, 1976; Wedhawati et al., 2006; Nothofer, 2009). According to Wedhawati et al. (2006), the Javanese vowels are classified as high front /i - ɪ/, high back /u - ʊ/, mid front /e - ɛ/, mid central /ə/, mid back /o - ɔ/, and low central /a/ (Wedhawati et al., 2006).

Sundanese comprises seven vowels (Crothers, 1978; Van Zanten & Van Heuven, 1984; Sudaryat et al., 2007). According to Crothers (1978) and Sudaryat et al. (2007), Sundanese vowels are classified as high front /i/, high central /ɨ/, high back /u/, mid front /e - ɛ/, mid central /ə/, mid back /o - ɔ/, and central low /a/. Conforming to the standardized orthography showed by Tamsyah (1996), Hardjadibrata (2003), and Danadibrata (2006), Kurniawan (2013) mentions that /ɨ/ represents a central unrounded vowel and is produced in a higher position than the schwa /ə/.

Van Zanten & Van Heuven (1984) showed that /i/ and /u/ are closer to one another in Javanese accented Bahasa Indonesia than in Sundanese accented Bahasa Indonesia. Moreover, they also found that the position of schwa in Javanese is much higher than in Sundanese. The Javanese-accented schwa was in fact found to be higher than that of the /e/ and /o/ counterparts, both in sound production and in the perceptual representation of the vowel system. The Javanese and Sundanese vowel are illustrated in Figure 2.2.

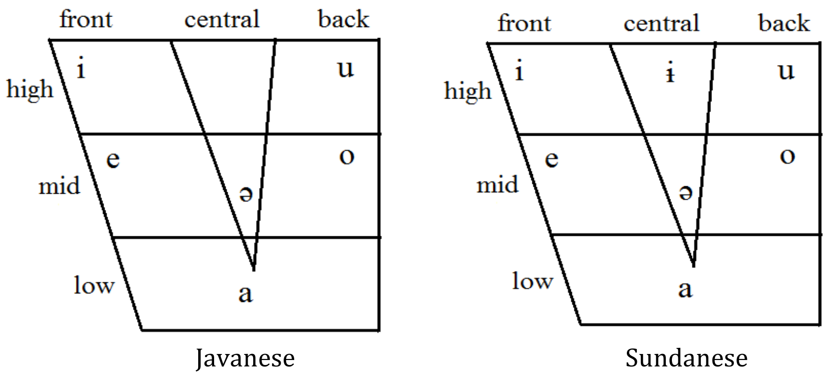


Figure 2.2 The positions of the vowels of Javanese vowel inventory (left) (Dudas, 1976) and Sundanese (right) (Crothers, 1978).

2.2.1 Vowel quality in Javanese and Sundanese

Vowel quality in Javanese and Sundanese is claimed to be influenced by phonation differences, which are correlated with the stops in the preceding syllables (Fagan, 1988; Hayward, 1993, 1995; Thurgood, 2004). Fagan (1988) analyzed acoustic differences between the slack-voiced stops /b/, /d̪/, /ɖ/, and /g/ and the stiff-voiced stops /p/, /t̪/, /ʈ/, and /k/ followed by /a/. His study shows that Javanese slack-voiced stops are characterized by a lower F_0, a lower F_1, and a higher F_2. Fagan (1988) claims that vowels following stiff-voiced stops are pronounced with a clear voice, while vowels following slack-voiced stops are pronounced with breathy voice.

A later investigation by Hayward (1993) yielded similar results: Javanese vowels are pronounced with lower F_1 and F_0 after the slack-voiced /b/. She also observed a high F_2 after voiced stops. Hayward (1995) extended her study by comparing the Voice Onset Time (VOT) of the stiff-voiced /p/ and the slack-voiced /b/, as well as the vowels /i/, /u/, /ɔ/, and /a/ following the stops. Her study shows that Javanese slack-voiced stop /b/ is characterized by having a lower F_0 at the vowel onset and was pronounced with negative VOT. Hayward (1993) summarized that the slack-voiced stop /b/ characterizes breathiness in Javanese, and the breathiness is manifested in the vowels, not in the consonants themselves.

In her acoustic study of Javanese vowels /u/, /ɔ/, and /a/, Thurgood (2004) found that vowels after the stiff-voiced stops /p/ and /k/ or the slack-voiced stops /b/ and /g/ have different formant frequencies. The vowel /a/ is the only vowel which is characterized as having a lower F_0 after the voiced stops. The vowel /ɔ/ was articulated with a lower F_0 after a velar stop, higher F_0 after a bilabial stop. The vowel /u/ was articulated with a higher F_0 after bilabial and velar stops. The raising of F_2 was found after slack-voiced stops in all vowels. Thurgood (2004) concluded that the phonetic realization of vowels after slack-voiced and stiff-voiced stops in Javanese includes the distinct breathy voice used for emphasis.

A recent study by Gordon et al. (2012) reported F_1 values of Javanese vowels by two speakers and found that F_1 distinguished between four heights where /i/ - /u/ (high) and /e/ - /o/ (mid) show roughly the same F_1 within the pairs. Schwa was found to be in a height category by itself, in between high and mid. Gordon et al. (2012) did focus on F_1, maximum intensity, acoustic energy, and perceptual energy of the vowels; the study did not measure F_2.

The present chapter is limited to the investigation of vowel quality independent of consonantal context - hence, although two consonantal contexts were used, we will only discuss vowel quality of the mean formant values averaged over the consonantal contexts.

Van Zanten & Van Heuven (1984) specifically reported that the central vowel /ə/ realization in Standard Indonesian appears to be more back and closed for the Javanese speakers than for their Sundanese and Toba Batak groups. The central vowel as pronounced by the Javanese speakers is in a mid-position, almost exactly half-way between /e/ and /o/, which contradicts the more recent findings by Gordon et al. (2012). The central vowel of the Sundanese speakers is considerably more closed. The Sundanese speakers have relatively closed realizations of /i/ and especially /u/. Toba Batak, on the other hand, has no central vowel in its phoneme system.

There is little recent published data on acoustic measurements of Javanese and Sundanese vowels which can be compared to the present study. For instance, Wedhawati et al. (2006) described the Javanese vowel system but did not acoustically examine the data. Thus, it is hard to reach conclusions regarding the accuracy of the tongue position of the speakers when they pronounced Javanese vowels.

The current study investigates to what extent the Javanese and Sundanese vowels produced by the Javanese and Sundanese speakers is identical to, respectively, the Javanese vowel system as described by Wedhawati et al. (2006) and to the Sundanese vowel system described by Crothers (1978). We predict that the Javanese speakers produce the Javanese central vowel /ə/ as found by Wedhawati et al. (2006) and Van Zanten and Van Heuven (1984). We also expect to find that the position of schwa pronounced by Javanese speakers is considerably higher than its Sundanese counterpart. Following Crother's (1978) description, Sundanese has one closed /ɨ/ and one mid /ə/ central vowel in a 7-vowel sound system. Overall, the present study seeks to characterize the vowel quality of Javanese and Sundanese vowels produced by Javanese and Sundanese speakers, respectively.

2.3 Materials and Methods

2.3.1 Participants

We collected speech data from 4 L1 Javanese speakers (2 male, 2 female, M_{age} = 34.75, SD = 6.9) and 4 L1 Sundanese speakers (2 male, 2 female, M_{age} = 35, SD = 4.7). The participants use the Javanese or the Sundanese language for daily interactions. The Javanese participants were considered to speak the Solo and Yogyakarta dialect while the Sundanese participants were considered to speak the Priangan dialect. All the participants demonstrated normal speech and hearing abilities.

2.3.2 Stimuli

Javanese vowels /i/, /e/, /a/, /ə/, /u/, and /o/ and Sundanese vowels /i/, /a/, /ə/, /ɨ/, /e/, /u/, and /o/ were inserted in /b/...$C and /h/...$C where $ refers to a syllable boundary. The target vowels and the syllables were embedded in a carrier phrase, *Kula ngendika ... malih* "I say … again" for Javanese speakers, and *Abdi nyarios … deui* "I say … again" for Sundanese speakers. The participants read the lists three times in random order. In total, the Javanese dataset comprises 3 repetitions × 6 vowels × 4 speakers = 72 items, and the Sundanese dataset comprises 3 repetitions × 7 vowels × 4 speakers = 84 items. The list of stimulus words is shown in Table 2.1.

Table 2.1 Javanese and Sundanese vowels in /b/...$C and /h/...$C sequences.

/b/...$C sequences			
Javanese	Target Word	Transcription	English Gloss
/a/	***ba**dhe*	/'**ba**dhe/	will (be)
/ə/	***be**cik*	/'**bə**cɪk/	main
/i/	***bi**narung*	/'**bi**naruŋ/	in a row
/o/	***bo**dho*	/'**bo**do/	stupid
/u/	***bu**deg*	/'**bu**'dəg/	deaf
/e/	***be**lekan*	/'**be**leʔan/	sore eyes
Sundanese	Target Word	Transcription	English Gloss
/a/	***ba**tur*	/'**ba**tur/	colleague
/ə/	***be**legbeg*	/'**bə**ləgbəg/	murky
/i/	***bi**tu*	/'**bi**tu/	explode
/o/	***bo**lotot*	/'**bo**lotot/	goggle
/u/	***bu**ni*	/'**bu**ni/	sealed
/e/	***be**ntes*	/'**be**ntes/	clear
/ɨ/	***beu**reum*	/'**bɨ**rɨm/	red
/h/...$C sequences			
Javanese	Target Word	Transcription	English Gloss
/a/	***ha**kekat*	/'**ha**kekat/	truth
/ə/	***he**mpas*	/'**hə**mpas/	smash
/i/	*wa**hi**ng*	/wa'**hi**ŋ/	sneeze
/o/	***ho**bi*	/'**ho**bi/	hobby
/u/	*dhu**hu**r*	/du'**hu**r/	high
/e/	***he**bat*	/'**he**bat/	great
Sundanese	Target Word	Transcription	English Gloss
/a/	***ha**ndap*	/'**ha**ndap/	under
/ə/	***he**nteu*	/'**hə**ntɨ/	no
/i/	***hi**deung*	/'**hi**dɨŋ/	black
/o/	***ho**ream*	/'**ho**ream/	lazy
/u/	***hu**rung*	/'**hu**ruŋ/	sparkle
/e/	***he**rang*	/'**he**raŋ/	shine
/ɨ/	***heu**rin*	/'**hɨ**rin/	narrow

2.3.3 Procedure

Participants were audio recorded one by one in a sound attenuated room. Before the recording started, participants filled in a demographic questionnaire and signed a consent form. Participants were then familiarized with the equipment, stimuli, and procedures for the production experiment. The stimuli (in the carrier phrase) were shown on a computer screen in random order. Immediately after a sentence appeared, participants read it aloud in a natural tone. The display of the monitor is set to 10 seconds. Speakers cut the sentences into three parts: *ngendik/nyarios xxx malih/deui* so that the speech becomes more intelligible. Recordings were made on a digital audio recorder (H4N Zoom, 44.1 kHz, 16 bit) using an adjustable microphone headset (Sennheiser PC 141). The microphone was placed 3 cm away from the right-hand corner of the participant's mouth.

2.3.4 Analysis

Using Praat (Boersma & Weenink, 2013), the beginnings and end points of the target vowels were located in the spectrogram. The first formant (F_1) and second formant (F_2) were estimated using the Burg Linear Predictive Coding (LPC) algorithm. Formant tracks were overlaid on the wideband spectrogram. Whenever a visual mismatch occurred between the tracks and the formant bands in the spectrogram, the model order of the LPC analysis was changed by trial and error until a satisfactory match was obtained. The values of F_1, F_2 and duration were then stored for offline statistical analysis.

After the formant frequency values were estimated, we took the mean of the /b/...$C items and /h/...$C items because the purpose of the current study is not to examine the effect of these contexts, but rather to describe the vowels of the Javanese and Sundanese independently of context.

Because the current data set was very small (n = 4 per group, k = 3 per vowel per context), we decided to present the data using descriptive statistics only. Formant frequencies will be plotted in vowel plots for descriptive purposes.

For plotting purposes, formant frequency measurements in hertz were converted to psychophysically more realistic Bark units using the formula suggested by Traunmüller (1990). Since the vocal tracts of female speakers are some 15 percent smaller than those of male speakers, the values of F_1 and F_2 values for the same vowel are different across speakers of different gender. In order to compare vowel formants across different speakers, vowel normalization was applied to the Bark-scaled measurements. We used z-normalization of F_1 and F_2 frequencies (Lobanov, 1971). To get the z-normalized scores, the speaker's mean formant frequency (for either F_1 or F_2) is subtracted from each token-individual formant frequency, and the difference is then divided by the speaker's standard deviation. Z-normalized F_1 values below 0 refer to relatively close/high vowels, whilst values larger than 1 correspond to open vowels. Positive z-values for F_2 correspond to front vowels, whilst negative F_2 values refer to back vowels.

2.4 Results

2.4.1 Formant Frequencies

Table 2.2 presents the means ($\bar{x}$) and standard deviations (SD) of the measured F_1 and F_2 values of the six Javanese vowels and seven Sundanese vowels, produced by the four speakers for each language. All values in the table are in hertz (Hz).

Table 2.2 Mean ($\bar{x}$) and Standard Deviation (SD) in Hz of the six Javanese and seven Sundanese vowels produced in a carrier sentence. F_1 and F_2 values are broken down by regional language and by gender, *N* = 12 per cell.

Vowel	Gender	Javanese				Sundanese			
		F_1		F_2		F_1		F_2	
		$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
/i/	female	381	67	2200	119	405	26	2236	46
	male	345	30	2179	142	307	32	2349	191
/e/	female	579	46	2073	159	569	20	2199	164
	male	486	27	1954	44	504	61	1863	85
/a/	female	788	88	1723	111	812	52	1702	115
	male	625	105	1650	272	599	49	1413	72
/o/	female	513	70	1108	50	648	86	1238	90
	male	473	82	1160	170	412	18	1030	82
/u/	female	451	53	1272	114	475	16	1151	53
	male	473	82	1160	170	378	35	979	39
/ə/	female	489	31	1614	214	571	46	1647	138
	male	405	51	1682	198	488	15	1356	250
/ɨ/	female	514	45	1709	95	-	-	-	-
	male	347	18	1479	17	-	-	-	-

Figures 2.3a-d present F_1 and F_2 mean values (in Bark units, after within-speaker z-transformation) of the six Javanese vowels produced by Javanese speakers and the seven Sundanese vowels produced by Sundanese speakers. The large phonetic symbols in the plots are placed at the centroids of the vowels, i.e. at the intersection of the mean F_1 and mean F_2 coordinate values. The individual vowel tokens are indicated by smaller-sized symbols. Spreading ellipses were drawn at ±1 SD along the two principal components optimally characterizing the scatter of the individual tokens around the centroid, theoretically capturing the most typical 46% of the distribution.[2]

[2] Figure 2.3 was produced using the Visible Vowels on-line facility (Heeringa & Van de Velde, 2017).

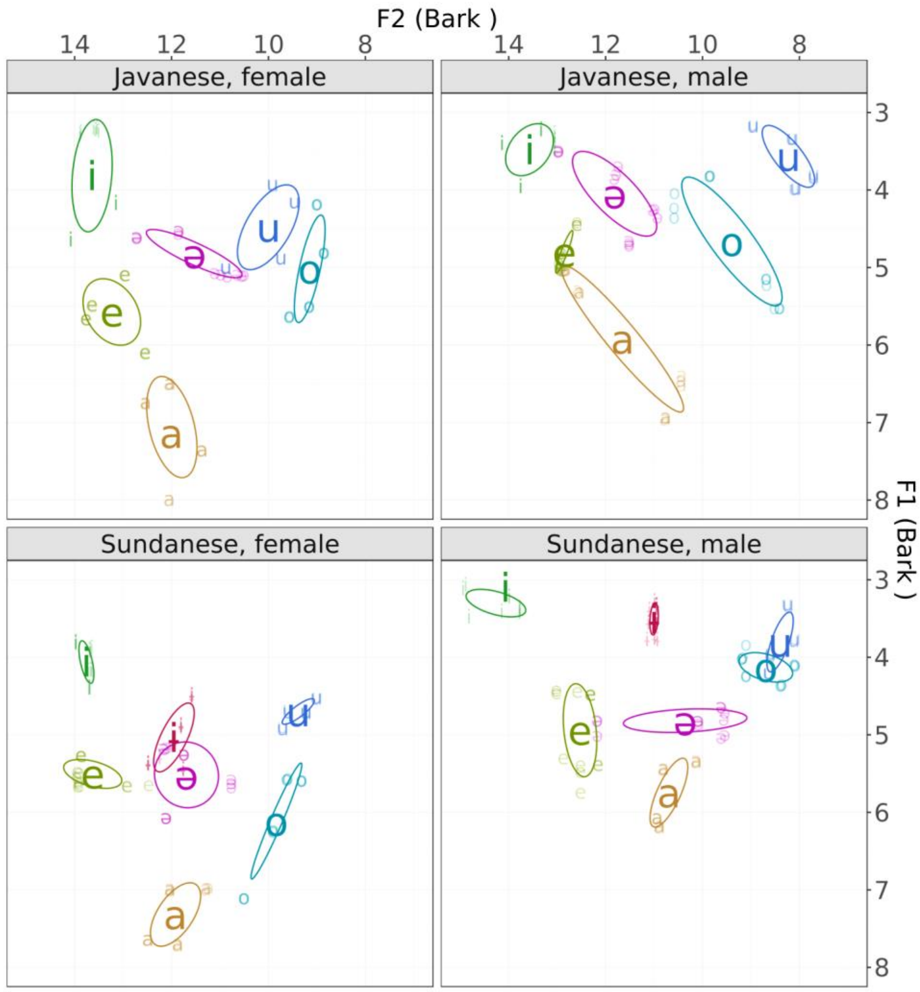


Figure 2.3a-d Javanese and Sundanese vowels plotted in an F_1 by F_2 plane (Barks).

In Figures 2.3a and b, Javanese has one high front vowel /i/, one mid front vowel /e/, one low vowel /a/, one mid central vowel /ə/, one high back vowel /u/, and one mid back vowel /o/. The vowel space area of the Sundanese male speakers is smaller than that of the females. But the relative distances between /i/, /e/ and /a/ are approximately the same, for both genders.

In Figures 2.3c-d, there are seven vowels: front /i/, mid front /e/, one low /a/, high central /ɨ/, mid central /ə/, high back /u/, and mid back o/. Interestingly, /e/ and /ə/ for Sundanese males are more open than for Sundanese female speakers and /o/ for Sundanese males is more closed than for the female speakers. Interestingly, the /ɨ/ ~ /ə/ contrast is almost absent for the female speakers but still clear for the male speakers. Conversely, /u/ ~ /o/ are virtually the same for the males but kept clearly distinct by the females.

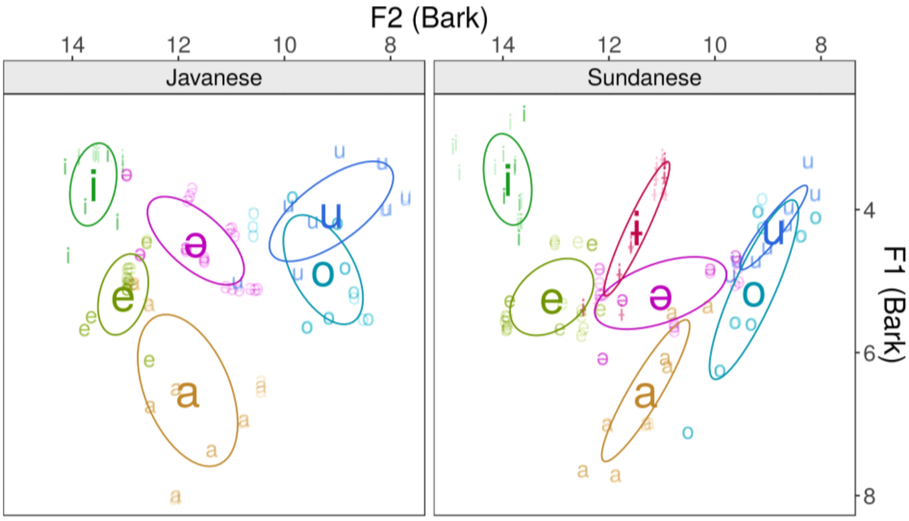


Figure 2.4a-b Javanese and Sundanese vowels plotted in an F_1 by F_2 plane across gender (Barks).

In Figure 2.4a-b, we present the plot of z-normalized F_1 and F_2 across male and female speakers. From the plot, it is shown that the Javanese /ə/ is higher than the Sundanese group. Sundanese speakers produce vowel /ɨ/ in a mid-high central vowel.

2.4.2 Duration

Table 2.3 presents the means (x̄) and standard deviations (SD) of the duration of six Javanese vowels and seven Sundanese vowels, produced by the four speakers for each language. All values in the table are in miliseconds (ms).

Table 2.3 Mean ($\bar{x}$) and Standard Deviation (SD) in ms of the six Javanese and seven Sundanese vowels produced in a carrier sentence. Duration values are broken down by regional language and by gender. *N* = 12 per cell.

Vowel	Gender	Javanese		Sundanese	
		x̄	SD	x̄	SD
/i/	female	93	34	87	18
	male	80	23	60	10
/e/	female	56	6	101	20
	male	91	22	57	30
/a/	female	78	31	93	3
	male	76	18	78	20
/o/	female	74	15	81	18
	male	79	9	99	14
/u/	female	95	26	130	16
	male	84	43	92	5
/ə/	female	67	27	98	15
	male	62	36	97	23
/ɨ/	female	-	-	185	40
	male	-	-	110	19

The results showed that the Javanese and Sundanese vowel durations are quite short between 60 and 100 ms for all vowel types, with the exception of /u/ (110 ms) and /ɨ/ (150 ms) for the Sundanese speakers.

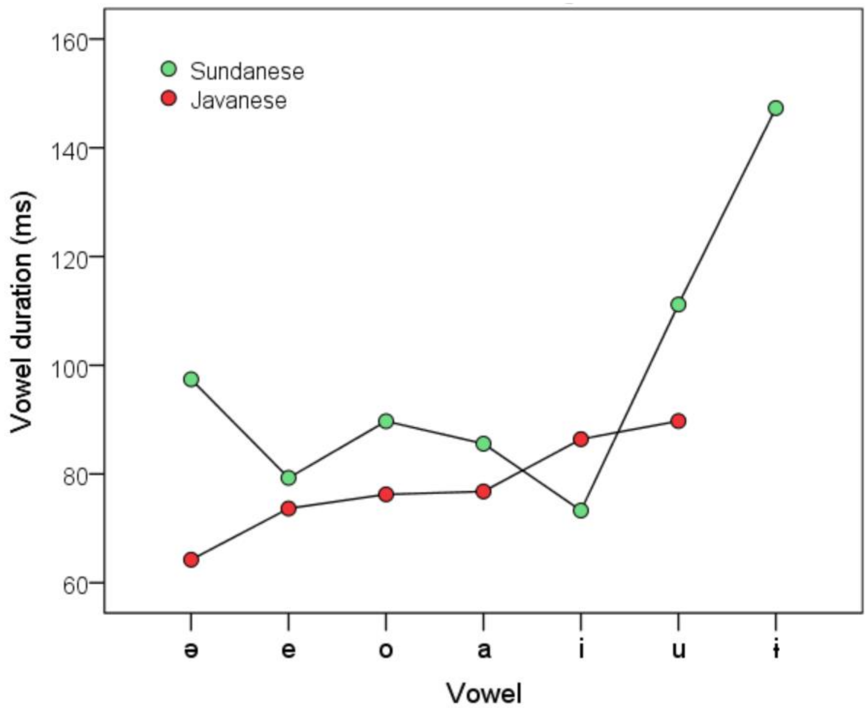


Figure 2.5 Javanese and Sundanese vowel durations across gender.

In Figure 2.5, we present the duration of vowels of Javanese and Sundanese across male and female speakers. From the graph, it is apparent that the duration of Javanese vowels is between 60 and 90 ms. For Sundanese, the durations are between 75 to 150 ms.

2.5 Discussion

In the present study, we examined the vowel quality of Javenese and Sundanese vowels by measuring the frequencies of the first two formants. The means per language group present a clear picture of general tendencies of difference in the first two-formant frequencies.

First, similar to Wedhawati et al. (2006), visual inspection of the mean values revealed that the Javanese female and male speakers produce a high front /i/, a mid front /e/, a low /a/, a mid central /ə/, a high back /u/, and a mid-back /o/.

Second, as found in Crothers (1978), Sundanese has seven vowels: front /i/, mid front /e/, low /a/, high central /ɨ/, mid central /ə/, high back /u/, and mid-back /o/. Earlier Sundanese vowel inventory studies claim that Sundanese has a high central vowel /ɨ/ (Crothers, 1978; Sudaryat et al., 2007). Our results demonstrate that the Sundanese male speakers still produce a high-central vowel /ɨ/, but the Sundanese female speakers do not seem to produce the vowel at all. The Sundanese female speakers' vowel /ɨ/ is visually indistinct from schwa.

The overall finding is inconsistent with Crothers (1978) and Kurniawan (2013) mentioning that vowel /ɨ/ in Sundanese is pronounced with a high central unrounded position of the tongue. Our results showed that vowel /ɨ/ is a high-mid central vowel, whereas it is a high central vowel in Crothers (1978).

This study shows that the Sundanese male speakers pronounce the vowel /ɨ/ differently than the females. The male pronunciation is compatible with Crothers (1978). It is possible that the Sundanese female speakers have lost the contrast between the high and the mid central vowels. This phenomenon where the Sundanese female speakers are losing the contrast between the two central vowels could be caused by language change occurring under the influence of Indonesian, which does not have this contrast. It is important to note that vowel /ə/ by Javanese speakers is higher than that of the Sundanese speakers. This result agrees with Van Zanten & Van Heuven (1984) who found that the Javanese schwa is higher than the Sundanese schwa. The Javanese schwa /ə/ is indeed shown to be higher than /e/ and /o/ in the Javanese sound system.

The durations of Javanese and Sundanese are phonetically short (between 60 and 100 ms) for all vowels. The exception is Sundanese /u/ (110 ms) and /ɨ/ (150 ms). The measurements of durations in the present study were taken from the Javanese and Sundanese in unstressed open CV syllables. The open CCV syllables were at the beginning of a two or three-syllable word inserted in a carrier sentence. Note that vowels in open syllables are longer than in closed syllables. However, at the same time, in unstressed syllables, vowels are likely to be shortened.

2.6 Conclusion

Previous studies by Van Zanten and Van Heuven (1984) and Van Zanten (1986) have presented the Standard Indonesian vowel system produced by the Javanese and Sundanese speakers. The current study aimed at extending the previous studies by investigating to what extent the Standard Indonesian vowels spoken with a Javanese or Sundanese accent as found in Van Zanten and Van Heuven (1984) and Van Zanten (1986) are similar to the Javanese and Sundanese vowels produced by the Javanese and Sundanese speakers.

The current study found that the Javanese schwa /ə/ is considerably higher than that of the Sundanese speakers. The Javanese schwa is higher than /e/ and /o/ produced by the Javanese speakers. For the Sundanese speakers, it is found that /ɨ/ is produced in closed central position, only for the male speakers. The results of the formant frequencies of the Javanese and Sundanese vowels confirm the description of the vowel system by Wedhawati et al. (2006), Van Zanten and Van Heuven (1984), Crothers (1978), and Kurniawan (2013).

It is important to note that Javanese schwa /ə/, which is remarkably closed and front, could possibly lead to pronounciation problem in L2. Some English words such as *cup, butter* and *but* would be problematic for the Javanese speakers. Both Javanese schwa /ə and vowel /a/ appear to be pronounced in the front part of the mouth and would therefore not be a good substitute for English /a/. English vowel /a/ is open and back position and thus Javanese speakers are expected to show a pronunciation problem with the vowel. Therefore, in chapter 4 of this thesis, potential pronunciation problems with the English vowel /a/ and schwa /ə/ among the Javanese and Sundanese speakers will be explored.

This study is limited to the vowel sound production of the Javanese and Sundanese speakers. It would be interesting for future research to extend the present study by exploring the effects of the onset consonant on the vowel production of these groups of speakers. Also, the present study has a relatively small sample size. Future studies should repeat the current experiment with a larger sample size in order to reach firmer conclusions.

Furthermore, the scope of the present study does not focus on the gender-related difference especially as the data set is too small to make the inferential analysis worthwhile. However, it seems reasonable to expect significant differences in the vowel production between male and female groups. The present results in the visual plots of Javanese and Sundanese speakers showed that the differences in vowel production between genders may well be significant.

References

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer* [Computer program]. Version 5.3.51, retrieved 01 September 2014 from http://www.praat.org.

Clynes, A., & Rudyanto, C. (1995). Javanese. In D. T. Tryon (Ed.), *Comparative Austronesian dictionary: An introduction to Austronesian studies*. Berlin: Mouton de Gruyter.

Cole, P., Hara, Y., & Yap, N. T. (2008). Auxiliary fronting in Peranakan Javanese. *Linguistics, 44*, 1-43. DOI: 10.1017/S002222670700494X.

Comrie, B., Matthews, S., & Polinsky, M. (2003). *The atlas of languages: The origin and development of languages throughout the world*. Singapore: Star Standard.

Crothers, J. (1978). Typology and universals of vowel systems. In J. H. Greenberg (Ed.), *Universals of human language*. Vol. 2: Phonology (pp. 94-152). Stanford, CA: Stanford University Press.

Danadibrata, R. A. (2006). *Kamus Basa Sunda*. Bandung: PT. Kiblat Buku Utama.

Dudas, K. (1976). *The phonology and morphology of modern Javanese*. Ann Arbor, MI: University Microfilms.

Fagan, J. L. (1988). Javanese intervocalic stop phonemes: The light/heavy distinction. In R. McGinn (Ed.), *Studies in Austronesian Linguistics 76* (pp. 173-202). Athens, OH: Ohio University.

Gordon, R. G. Jr. (2005). *Ethnologue: Languages of the World* (15th Ed.). Dallas, TX: SIL International.

Gordon, M. (2006). *Syllable weight: Phonetics, phonology, typology.* London: Routledge.

Gordon, M. Ghushcyan, E., McDonnell, B., Rosenblum, D., & Shaw. P. (2012). Sonority and central vowels: A cross-linguistic phonetic study. In S. Parker (Ed.), *The sonority controversy* (pp. 219-256). Berlin: Mouton de Gruyter.

Hardjadibrata, R. R. (1985). *A typological study of Sundanese* (Doctoral dissertation). Melbourne: La Trobe University.

Hardjadibrata, R. R. (2003). *Sundanese-English dictionary. Based on Soendanees-Nederlands woordenboek by F.S. Eringa.* Bandung: PT. Kiblat Buku Utama.

Hayward, K. (1993). /p/ vs. /b/ in Javanese: Some preliminary data. *SOAS Working Papers in Linguistics & Phonetics, 3*, 1-99.

Hayward, K. (1995). /p/ vs. /b/ in Javanese: The role of the vocal folds. *SOAS Working papers in Linguistics & Phonetics, 5*, 1-11.

Heeringa, W., & Van de Velde, H. (2017). Visible Vowels: A tool for the visualization of vowel variation. *Proceedings of Interspeech 2017*, Stockholm, 4034-4035. https://www.visiblevowels.org

Horne, E. C. (1961). *Beginning Javanese* (Vol. 3). Yale Linguistics Series. New Haven, London: Yale University Press.

Kurniawan, E. (2013). *Sundanese complementation* (Doctoral dissertation). University of Iowa. Retrieved from http://ir.uiowa.edu/etd/2554.

Labov, W. (1990). The intersection of sex and social class in the course of language change. *Language Variation and Change, 2*, 205-254.

Labov, W. (2001). *Principles of linguistic change. Social factors*. Oxford: Blackwell.

Lewis, M. P. (2009). *Ethnologue: Languages of the world* (16th ed.). Dallas, TX: SIL International, Online version.

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

Muslim, D., Haerani, E., Motohiko, S., & Hiroshi, Y. (2010). Language mapping based on geomorphology in Western part of Java, Indonesia. *Memoirs of Osaka Kyoiku University Ser. III, 58*, 11-18.

Nothofer, B. (1977). Dialektgeographische Untersuchung des Sundanesischen und des entlang der Sundanesischen Sprachgrenze gesprochenen Javanischen und Jakarta-Malaiischen. Ersten Teil. Köln: Philosophischen Fakultät der Universität zu Köln.

Nothofer, B. (1982). Central Javanese dialects. *Pacific Linguistics, Vol. 3/C-76*, 287-309.

Nothofer, B. (2009). Javanese. In *Concise Encyclopedia of Languages of the World* (pp. 560-561). Oxford: Elsevier.

NVTC (National Virtual Translation Center). (2007). *Javanese.* Languages of the World. Retrieved from http://www.nvtc.gov/lotw/months/june/Javanese.html

Oakes, M. (2009). Javanese. In *The world's major languages*. New York, NY: Routledge.

Ogloblin, A. K. (2005). *The Austronesian languages of Asia and Madagascar.* New York, NY: Routledge.

Simons, G. F. and Charles D. F. (2018). *Ethnologue: Languages of the World,* Twenty-first edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Sudaryat, Y, Prawirasumantri, H. A., & Yudibrata, H. K. (2007). *Tata Bahasa Sunda Kiwari.* Bandung: Yrama Widya.

Tamsyah, B. R. (1996). *Kamus Lengkap Sunda-Indonesia, Indonesia-Sunda, Sunda-Sunda*. Bandung: Pustaka Setia.

Thurgood, E. (2004). Phonation Types in Javanese. *Oceanic Linguistics, 43*, 277-295.

Traunmüller, H. (1990). Perceptual dimensions of openness in vowels. *Journal of the Acoustical Society of America, 69*, 1465 - 1475.

Uhlenbeck, E. M. (1963). Review of Beginning Javanese, by Eleanor C. Van Zanten Home. *Lingua, 12*, 69-86.

Van Zanten, E., & Van Heuven, V. J. (1984) The Indonesian vowels as pronounced and perceived by Toba Batak, Sundanese and Javanese speakers. *Bijdragen tot de Taal-, Land- en Volkenkunde, 140*, 497-521.

Van Zanten, E. (1986). Allophonic variation in the production of Indonesian vowels. *Bijdragen tot de Taal-, Land- en Volkenkunde, 142*, 427-446.

Wedhawati, Erni W., Nurlina, S., Setiyanto, E., Marsono, Sukesi, R., & Baryadi, I. P. (2006). *Tata Bahasa Jawa Mutakhir [Grammar of Modern Javanese]*. Yogyakarta: Penerbit Kanisius

3

Perception of English Vowels by Javanese and Sundanese Speakers: A Mouse-Tracking Study

Abstract

Second language (L2) learners often encounter difficulties due to the interference of their native language (L1) with the target language. The present study is concerned with L2 English learners with non-Western first languages—Javanese and Sundanese. The aim of the study is to investigate (1) whether the sound category (*new* vs *similar* vowels) affects the L2 sound perception based on L2 learning models, (2) whether the phonetic distance between target and distractor sounds influences the L2 sound perception. Thirty Javanese, 30 Sundanese, and 20 English native speakers participated in a mouse-tracking experiment. Participants were required to identify English vowels corresponding to an auditory token by clicking on one of two word strings presented on a computer screen. The results showed that phonetic distance between target and distractor plays a more important role in the sound perception of the Javanese and Sundanese listeners than the sound category of the target itself. The findings partially support the L2LP model indicating that new vowels are more problematic to be perceived by the L2 learners than similar vowels.

Keywords: *L2 perception*, *L2 learners*, *vowel perception*, *mouse tracking*

3.1 Introduction

Perception of foreign speech (L2) sounds is affected by the beginning age of L2 acquisition (Flege, Munro, & McKay, 1995; Baker, Trofimovich, Mack, & Flege, 2002), the amount of exposure to the L2 (Flege, 1987; Flege & Hillenbrand, 1984), as well as the L1 vowel and consonant system, the syllable structure system and the prosody (Bradlow, 1995; Fox, Flege, & Munro, 1995; Wang, 2007; Iverson & Evans, 2009; Elvin, Escudero, & Vasiliev, 2014; Van Heuven & Gooskens, 2017). Studies of English vowel perception have been done in some languages including Spanish (Flege, Munro, & MacKay, 1995; Escudero, 2000; Escudero & Chládková, 2010; Morrison, 2008, 2009), Catalan (Cebrian, 2006, 2007). Spanish-speaking learners of English struggled to perceive English contrasts, which are not present in their L1 (Escudero & Chládkova, 2010; Sisinni, Escudero, & Grimaldi, 2014). Previous studies examined the cross-language perception pattern in specific L2 sounds and contrasts (Escudero & Boersma, 2004; Morisson, 2009) and investigated complete vowel systems across languages (for Arabic: Ali, 2011; for Mandarin: Wang, 2007; Wang & Van Heuven, 2003, 2004, 2006; for Korean: Yoon, 2013; for Japanese: Kamiyama, 2011). Unlike these previous studies that tested L2 vowel sounds perceived by learners with western L1 backgrounds, the present study focuses on all of the English vowels perceived by learners from two non-western languages - the Indonesian languages of Javanese and Sundanese.

Previous studies have reported that L2 learners who have a smaller number of L1 vowels experience difficulty perceiving an L2 with a larger number of vowels (e.g. Flege, Bohn, & Jang, 1997; Iverson & Evans, 2007, 2009; Elvin et al., 2014). The reason is because that L2 contrasts are not found in the L1. Since the L1 has a small inventory, there will be fewer L2 contrasts available. Flege, Bohn, and Jang (1997) studied the interaction of L1 and L2 vowel systems of native German, Spanish, Mandarin, and Korean speakers. The study found that the nature of the L1 vowel system and its perceived relation to vowels affect the L2 vowel production and perception. Likewise, Iverson and Evans (2007, 2009) found that German and Norwegian speakers, who have larger L1 vowel systems, identified English vowels with more accuracy than Spanish and French speakers, who have smaller vowel systems. Specifically, the study demonstrated that despite the differences in their vowel systems, learners' perceptions are accurately predicted by detailed acoustic comparison.

Second language learning models have offered explanations of whether acoustic similarities/dissimilarities between L1 and L2 sounds play a role in cross-language speech perception. According to the Contrastive Analysis Hypothesis or CAH (Lado, 1957), L2 sound systems that are most dissimilar to the L1, will be most difficult to acquire. CAH claims that similar sounds across L1 and L2 will not create problems for L2 learners. This rationale of these models, however, is based on an analysis of phonemes and systematic allophones, not on fine-grained acoustic-phonetic comparison.

Another L2 learning model offered to explain the L2 learners' difficulties with L2 sound is the Speech Learning Model or SLM (Flege, 1987, 1995, 2002, 2003). SLM predicts that the difficulty of L2 learning is determined by the perceived phonetic similarities between the L2 sound and the closest L1 sound. The model posits that the larger the perceived phonetic difference is between two sounds, the more easily they are distinguished, and the more likely that the sound contrast will be acquired. In contrast to CAH, this model predicts that L2 sounds, which are similar to L1 sounds, yield more difficulties than new sounds. It is important to note that this model was originally focusing on *experienced* learners. At the early stage of L2 learning, two different L2 sounds, which are similar to a single category in the L1, will be assimilated to that single L1 category causing an acquisition problem (that is not explained by SLM, but see below). However, given enough time, the L2 sound inventory will be expanded to include all categories in L1 and the *new* categories in the L2 (Flege, 2003). At this later stage, SLM predicts that L2 learners will set up new categories for dissimilar L2 sounds.

The Perceptual Assimilation Model (PAM) (Best, 1995; Best and Tyler, 2007) predicts that the difficulty of L2 learning in its initial stages is determined by the degree to which of non-native contrasts are assimilated to native categories. While SLM targets experienced L2 learners, PAM seems more applicable to difficulties of beginners. PAM posits that a non-native contrast that is perceptually assimilated to two L1 categories (Two-Category Assimilation or TC) will not create difficulties in learning. However, a non-native contrast two different sounds which are assimilated to one L1 category (Single-Category Assimilation or SC) will be difficult to discriminate (Flege & McKay, 2004; Yoon, 2007).

Most recently, the Second Language Linguistic Perception model or L2LP (Escudero, 2005, 2006, 2009) was proposed. L2LP combines predictions by both SLM and PAM. L2LP attempts to explain the difficulties of L2 learning across all learning stages. In the early stage of learning, L2 sounds will be copied to the L1 system (Full Copying Hypothesis). The model posits that the shape of the full copying system is influenced by the particular properties of L1, such as accentual features. L2LP claims that beginners with the same L1 system differ in the way they map L2 categories onto L1. Unlike SLM and PAM, which do not consider inter-individual differences, L2LP considers individual variation as an aspect that will affect non-native sound perception. As L2 develops, each L2 learners have different learning problems and, at the later stage, have the ability to overcome their particular difficulties.

According to the L2LP model, L2 sounds that are similar to L1 sounds will be easier to acquire than new sounds. L2LP posits that in the similar sound scenario, or TC assimilation pattern in PAM, the learners will simply shift the existing L1 boundary between the two categories to match the L2 categories. On the other hand, in the new sound scenario, which involves either the SC or the CG assimilation patterns in PAM (see Chapter 1, section 1.3.3), the learners should either split up the single category to which both members of the non-native contrast were assigned or create a new L2 category to accommodate the tokens that are atypical exemplars of the L1 category. These category creation and split processes will impose difficulties on L2 sound learning that outlast the boundary shift problems associated with the similar sounds scenario. Empirical results, which match with this prediction, have been found among Dutch learners of Spanish (Escudero & Boersma, 2002) and Canadian English learners of Canadian French (Escudero, 2005, 2009).

3.2 Present Study

The present study examines the difficulties of L2 sound learning faced by Javanese and Sundanese learners of English. The study compares the acoustic differences between the Javanese and Sundanese vowel system as their native language and the American English (AE) vowel system as their L2 or target language. Specifically, in this study, we expected some sounds to be more difficult than others. To predict the difficulty that Javanese and Sundanese listeners will have with particular English vowel contrasts, the study aims at examining the effect of the sound category (similar vs new sounds) and the effect of phonetic distance between the non-native and native sounds:

a. The effect of the sound category (similar vs new sounds).
   In general, SLM posits that the greater the acoustic difference

between L1 and L2, the more likely it is that learners will categorize L2 sounds in native like fashion. Considering the IPA symbols to represent sounds of L1 and L2, SLM employs three criteria of classification: new, identical and similar sounds. SLM defines new sounds as the sounds, which have no same symbols between L1 and L2 and similar sounds as the sounds with same symbols, but different diacritics between L1 and L2. According to SLM hypothesis, new sounds can be acquired easily while similar sounds are very difficult almost impossible to learn at a native-like level. As a result, the new sounds create less errors, quicker reaction and initiation times, and smaller AUC than the similar sounds. In contrast, according to PAM hypothesis, which is incorporated within L2LP model, it should be easier to perceive similar sounds than new sounds than new sounds, resulting in smaller error rates, quicker reaction and initiation times for similar than new sounds.

b. The effect of phonetic distance between L2 sounds as measured by Levenshtein Distance.
In the present mouse tracking study, L2 learners will have to match a L2 stimulus sound against two L2 response sounds, one of which is the target (matches the stimulus sound) and the other the distractor. Hence, not only whether the stimulus sound falls into the similar or new category as described above can be expected to influence the results, but also the phonetic similarity of the two target sounds. This phonetic difference between the target sounds will be quantified by the phonetic distance between target and distractor vowel using a count of the distinctive phonetic features needed to differentiate two vowel sounds in English.

The aims of this study are therefore to test to what extent the sound category (similar vs new sounds) of an L2 sound and the phonetic distance between two-response alternatives affect the sound perception of the Javanese and Sundanese listeners. Two hypotheses were tested:

(H1) Sound category (new vs similar sounds) affects the L2 sound perception of Javanese and Sundanese listeners.
SLM (Flege, 1995) makes the prediction that Javanese and Sundanese listeners will perform worse than native speakers, but with relatively better performance for the new vowels /ɑː, ɝ, ɔː, ʌ, æː, ɛ, ɪ/ ʊ/ than for the similar vowels /iː, uː/. L2LP predicts

the opposite pattern, with (relatively) better performance for similar vowels than for new vowels.

(H2) Phonetic distance between L2 target and distractor affects the sound perception of the Javanese and Sundanese listeners. Phonetic distance was operationalized in the present study as the feature distance (FD, see methods section) between target and distractor vowel. Specifically, it is predicted that perception will be more difficult as FD is smaller.

3.3 Javanese, Sundanese and English Vowel System

Javanese and Sundanese, two of the most widely spoken Indonesian local languages (Lauder & Ayatrohaedi, 2006; Nothofer, 2009), and American English have different vowel systems. As discussed in Chapter 1 and 2, Javanese vowels are grouped into six phonemes, /i, u, e, ə, o, a/ and that Sundanese has seven vowels /i, ɨ, u, e, ə, o, a/. American English has a more complex vowel system with ten monophthongs, /iː, ɪ, ʊ, uː, ɛ, ɝ, ɔː, æː, ʌ, ɑː/, and (Ladefoged, 2001, 2006). It is important to note that in AE, the phonologically lax vowel /æ/ is phonetically tense and long (cf. Strange et al. 2004; Wang & Van Heuven 2006). Javanese and Sundanese do not distinguish vowels based on duration (Van Zanten & Van Heuven, 1997), whereas American English distinguishes vowels based on durational cues (Peterson & Lehiste, 1960; Hillenbrand, Clark, & Houde, 2000). The English, Javanese, and Sundanese vowels are illustrated in Figure 3.1.

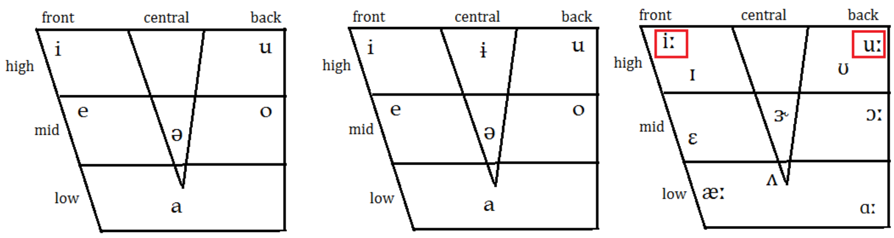


Figure 3.1 The monophthongs of Javanese (left panel), Sundanese (central panel) and English (right panel). The vowels in the squares are *similar* vowels. The remaining English vowels are *new* vowels.

Using the principle of Flege's SLM, English vowels /iː/ and /uː/ have an identifiable counterpart in Javanese /i/ and /u/ but differ in their length marking. Therefore, the vowels /iː/ and /uː/ are considered similar vowels for the Javanese and Sundanese native speakers. The other English vowels /ɑː, ɝ, ɔː, ʌ, æː, ɛ, ɪ, ʊ/ are represented by IPA symbols which are not used in any Javanese and Sundanese sounds, and thus are considered new L2 vowels. No identical vowels were found in the contrastive sounds of the English, Javanese, and Sundanese L1 vowel systems.

Taking into account the non-native vowel perception of Javanese and Sundanese learners of English, very few studies are available. A recent perception study involving speakers of Indonesian local languages is Perwitasari (2013), who investigated discrimination of American English vowels by Indonesian learners of English including Javanese and Sundanese speakers. A lexical discrimination task was conducted to measure the accuracy of the learners. The results showed that the Indonesian learners of English fail to perceive a word correctly because of the similarity of the vowels. The Indonesians were less accurate than the English listeners on four English vowel contrasts, /ɪ - iː/, /ɔː - ɑː/, /ʌ - ɑː/, and /uː - ʊ/. The confusion in the perception of English sounds occurred most often within the similar pairs of vowels that were not frequently heard. However, the study did not clearly determine how the L1 vowel system influences the perceptual difficulties.

To examine how the Javanese and Sundanese native speakers perceive English vowels, we used the *MouseTracker* software (Freeman & Ambady, 2010). Typically, two response alternatives are presented visually (i.e. in written form) in different corners of a computer screen while the participant hears a single spoken stimulus. The participant then has to move the mouse, starting from some neutral position at the bottom of the screen, as quickly as is reasonably possible, to the one of the two alternatives presented on screen that corresponds to the auditory stimulus. The technique can be used to estimate the participants' confusion in responding to a forced binary decision, in this case, between two different vowels (Spivey, Grosjean, & Knoblich, 2005; Dale, Kehoe, & Spivey, 2007; Farmer, Anderson, & Spivey, 2007). If respondents do not move straight to the target, it means that there is a processing difficulty due to conflicting information (Bruhn, Huette, & Spivey, 2013). Mouse-tracking data reflect real-time mental processing that appears as the result of a complex chain of thought (Freeman et al., 2011).

One benefit of adopting the MouseTracker technique is that one can obtain several online measures from the participant's hand movements. Not only can classical performance measures be measured such as error rate and response time, but in addition the technology offers new metrics and measures, some of which are even dynamic in time. For example, initiation times (the time it takes for participants to start moving the mouse after the visual target stimulus and distractor has been presented) might reflect the participants' impulsive behavior in making a response (Barca & Pezzulo, 2012). Another measure is the Area Under the Curve (AUC). AUC measures the area between the observed mouse trajectory and an idealized straight-line trajectory drawn from the START button to the correct response (Freeman et al., 2011). AUC indicates how much the hand movements are attracted toward the incorrect response and indexes the degree of uncertainty in decision-making process (Spivey et al., 2005; Farmer et al., 2007; Barca & Pezullo., 2012; Flumini et. al., 2015). MouseTracker can also generate dynamic velocity profiles (i.e., the speed at which the mouse is moved at evert time point) indicating the degree competition between the response alternatives at different time points.

3.4 Method

3.4.1 Participants

Thirty Javanese-speaking English learners (JEL) (15 female, 15 male, mean age = 22, SD = 1.4), 30 Sundanese-speaking English learners (SEL) (15 female, 15 male, mean age = 21, SD = 0.74), and 20 American English (AE) speakers (10 female, 10 male, mean age = 26.35, SD = 2.8) participated in the experiment. The Javanese and Sundanese participants used Javanese and Sundanese as their first language. They also spoke Indonesian as an L2 in formal situations. The JEL, SEL, and AE participants were tested at Universitas Gadjah Mada and Universitas Padjajaran, both in Indonesia. The AE participants reported having little knowledge of languages other than AE and considered themselves monolinguals.

The JEL and SEL speakers mainly used their L1 in daily conversation. The ages at which the JEL and SEL speakers started learning English were similar [JEL: M = 9.2, SD = 1.16; SEL: M = 8.9, SD = 2.02; $t(28) = 0.67$, $p = .5$]. At the time of the study, the JEL had more years of exposure (M = 11.23, SD = 2.17) to English than the SEL (M = 10.1, SD = 2.21; $t(28) = 2.04$, $p = .04$). To provide estimates of language proficiency, the participants completed a written English vocabulary test by Meara (2010). They had to write Y for the English words that they knew and marked N for the words that they did not know or were not sure about. Based on the vocabulary test result, their English level could be classified as intermediate. All participants provided written informed consent about the study indicating that they had made a rational and voluntary decision to participate.

3.4.2 Stimuli

Auditory stimuli comprised of ten American English vowels, /iː/, /ɝ/, /ɑː/, /ɔː/, /uː/, /ɪ/, /ɛ/, /æ/, /ʌ/, and /ʊ/, in two consonantal contexts, /bVd/ such as *bead* and *bid* and /hVd/ such as *heed* and *hid*, were produced in the carrier sentence: *Click bead, please* or *Click heed, please*. Each stimulus was recorded using Praat (version 5.5.04) at 44.1 kHz and produced by a single male native speaker of English originating from New York. He was asked to read out loud auditory stimulus items, which include English vowels. The target stimuli were composed of 35 minimal pairs (see Table 3.3). Pairs of stimuli were presented in a random order without repetitions. The visual target stimuli were letter strings presented using MouseTracker. One letter string appeared in the top left corner and the other in the top-right corner. The mouse movements were recorded at a sampling rate of 60 Hz. The display resolution was set to 1024 x 768 pixels.

3.4.2.1 Category Division

To predict the difficulty that Javanese and Sundanese listeners will have with particular English vowel contrasts, the study defines the sound category between the non-native and native sounds. Categories of speech sounds in the target language were defined on the basis of SLM. The first category is identical sounds. Sounds are categorized as identical if they have the same narrow IPA symbol in the native and the target language. The second category is similar sounds. These are sounds that are transcribed with the same IPA symbol in the native and target language but show differences in the diacritics marks only. The third category is new sounds. New sounds are target language sounds, which are not phonetically close to the native language sounds; they are phonetically transcribed with a different base symbol in L1 and L2. No identical sounds are indicated in the contrastive analysis of Javanese, Sundanese and English sounds. The table below indicates the similar sounds in grey cells and the new sounds in white cells.

Table 3.1 Contrastive vowel analysis of Javanese, Sundanese (Crothers, 1978) and English (Ladefoged, 2006). Grey and white cells in the target language represent similar and new sounds, respectively.

	Place of Constriction					
	Front		Central		Back	
Native Language: Javanese						
high	i				u	
high-mid	e		ə		o	
low-mid						
low			a			
Native Language: Sundanese						
high	i		ɨ		u	
high-mid	e		ə		o	
low-mid						
low			a			
Target Language: English						
	Tense	Lax	Tense	Lax	Tense	Lax
high	iː				uː	
high mid		ɪ	ɝ			ʊ
low mid		ɛ			ɔː	
low	æː			ʌ	ɑː	

It is unclear how the predictions of SLM apply. The vowels in the Javanese and Sundanese inventory have no length or tense-lax contrast. According to SLM, English vowels /iː/ and /uː/ have the same IPA base symbols as the Javanese and Sundanese counterparts and only make a difference in their length marking. According to SLM, the vowels /iː/ and /uː/ are therefore considered similar vowels for the Javanese and Sundanese native speakers. The other English vowels /ɑː, ɝ, ɔː, ʌ, æː, ɛ, ɪ, ʊ/ are represented by IPA base symbols which are not used in any Javanese and Sundanese sounds, and thus are considered new vowels.

3.4.2.2 Distance Level

We employ a simple distinctive feature distance (FD) to categorize the phonetic distances between each English vowel. FD is equal to the city-block distance between any two English vowels in the bottom part of Table 3.1. Because constriction place and tenseness are projected onto a single (horizontal) dimension in the table, the distance in the spectral space is determined first, using 4 steps (with a maximum difference of

3) along the vertical (vowel height) dimension and 3 steps (with a maximum difference of 2) horizontally (place of constriction). The largest possible spectral distance would be 3 + 2 = 5, which is observed for the /iː - ɑː/ pair. If there is a difference in tenseness between the members of a pair, the distance is incremented by 1.

Table 3.2 Feature distance between each pair of English vowels.

#	English words		iː	ɪ	ɛ	æː	ɑː	ɔː	ʊ	uː	ɝ	ʌ
1.	bead/ heed	iː	0									
2.	bid/ hid	ɪ	2	0								
3.	bed/ head	ɛ	3	1	0							
4.	bad/ had	æː	3	3	2	0						
5.	body/ hod	ɑː	5	5	4	2	0					
6.	bawd/ hawed	ɔː	4	4	3	3	1	0				
7.	buddhist/ hood	ʊ	4	2	3	5	3	2	0			
8.	booed/ who'd	uː	2	4	5	5	3	2	2	0		
9.	bird/ heard	ɝ	3	2	3	3	3	2	2	2	0	
10.	bud/ hudd	ʌ	4	2	1	3	3	2	2	4	2	0

From Table 3.2, it can be seen that there are five distances of the vowel pairs and word pairs in the present study. For instance, an FD of 1 would indicate that the vowels (and word pairs) are spectrally close and hence more difficult to distinguish, while the members of a vowel pair with an FD of 5 would be easier to distinguish since the members are spectrally far removed from each other and possibly also differ in tenseness. Table 3.3 provides a list of all 35 pairs of vowels used as stimuli in the mouse tracking experiment, irrespective of the order of the members within a pair. The pairs are listed in ascending order of feature distance. We excluded the long-long vowel pairs (except for æː) because the present study was initially focusing on examining the long versus short effect for the L2 learners.

Repeated measures ANOVAs were run for the different dependent variables: error rates, Area Under the Curve (AUC), reaction times, initiation times and velocity profiles.

Repeated measures ANOVAs were run for the different dependent variables: error rates, Area Under the Curve (AUC), reaction times, initiation times and velocity profiles. First, errors are calculated from the numbers of the feedback messages (with red X cross) appeared after the participants made incorrect responses.

Second, Area Under the Curve (AUC) is the geometric area between the actual mouse trajectory and the idealized trajectory (a straight diagonal line from the starting position to the target stimulus. Ideally, when a participant is attracted to only a final correct response, he will reach the response button straight away. Thus, their trajectory would a straight line from the /START/ button to the response button. However, when the participant is attracted to the incorrect response alternative, mouse trajectories would be less smooth and complex due to the fluctuating of directions. Hence there would be a deviation of the mouse trajectory from the diagonal, which can be quantified as the area under the curve between the ideal (diagonal) line and the observe mouse trajectory.

Third, once two visual stimuli (the target and the distractor) appeared on the upper left and right side of the screen, the mouse becomes active. The time from when the mouse is active and the time the participant first moved it, is called the 'initiation time'. Fourth, the 'reaction time' measure the time from when participants pressed /START/ until they reached and clicked on the response button.

Last, velocity profiles measure the participants' response speed in time. Depending on the efficiency of stimulus processing in time, fluctuations in mouse movement speed can be expected that may reflect processing difficulty. Hence, by analysing the velocity of mouse movement in time and comparing the velocity profile to that of native speakers, information can be gained about the decision process among L2 learners. To reflect the location of the mouse in raw times, we decided to retain trajectories without time-normalization. We opted to divide the raw time into 20 time bins to create between 0 ms and 1500 ms (as a cutoff).

The dependent variables were analyzed with the factor Group (Javanese/Sundanese, English) as a between-subjects factor, and Distance between target and distractor (1-5) and Category (new, similar) as within-subjects factors. Main and interaction effects of the consonantal context will be presented in appendix A only. To reduce the amount of results and focus on the hypotheses above, only main effects of, and interactions with, Group are reported (since only group main effects and interactions indicate deviance from L1 performance, and hence L2 acquisition difficulty). However, all results can be found in Appendix A. If the sphericity assumption was rejected (i.e. Mauchly's test was significant), Greenhouse-Geisser corrected p-values are reported. If effects were significant, ANOVAs were followed up by

Bonferroni-corrected post-hoc tests to test at what level of Distance and/or Category there were significant Group differences. Data were analyzed using SPSS version 22.0 (IBM, 2013).

3.5 Results

As explained in the introduction, we analyzed whether mouse-tracking responses were influenced by the phonetic distance between the two response vowels and whether the stimulus vowel was new or similar. Because there were only new vowels for Distance 1, this level of distance was excluded from the analyses. Hence, we performed an RM-ANOVA with Distance (level 2-5), Category (new, similar) and Group (native, non-native) as independents for every dependent variable (error rate, Area Under the Curve (AUC), initiation time, reaction time and velocity profiles).

Responses with a reaction time exceeding 2000 ms (6.23% of the total responses) were excluded from the analysis. Two participants' responses were discarded from subsequent analysis of the Area Under the Curve (AUC) because their responses deviated more than 3 SDs from the grand mean.

Table 3.3 Stimuli of the experiment. The word *buddhist* was presented as *budd*. All participants were well informed of the entire list of stimuli prior to the experiment, FD = Feature Distance.

Distance	Vowel Pairs			Orthographic word pairs		
FD = 1	ɛ	-	ɪ	bed / head	-	bid / hid
	ɛ	-	ʌ	bed / head	-	bud / hudd
FD = 2	ɔː	-	ʊ	bawd / hawed	-	buddhist /
	ɔː	-	ʌ	bawd / hawed	-	bud / hudd
	ɝ	-	ɪ	bird / heard	-	bid / hid
	ɝ	-	ʊ	bird / heard	-	buddhist /
	ɝ	-	ʌ	bird / heard	-	bud / hudd
	iː	-	ɛ	bead / heed	-	bed / head
	iː	-	ɪ	bead / heed	-	bid / hid
	ɑː	-	æː	body / hod	-	bad / had
	uː	-	ʊ	booed / who'd	-	buddhist /
	æː	-	ɛ	bad / had	-	bed / head
	ɪ	-	ʊ	bid / hid	-	buddhist /
	ɪ	-	ʌ	bid / hid	-	bud / hudd
	ʊ	-	ʌ	buddhist /	-	bud / hudd
FD = 3	ɔː	-	ɛ	bawd / hawed	-	bed / head
	ɔː	-	æː	bawd / hawed	-	bad / had
	ɝ	-	ɛ	bird / heard	-	bed / head
	ɝ	-	æː	bird / heard	-	bad / had
	iː	-	æː	bead / heed	-	bad / had
	ɑː	-	ʊ	body / hod	-	buddhist /
	ɑː	-	ʌ	body / hod	-	bud / hudd
	æː	-	ɪ	bad / had	-	bid / hid
	æː	-	ʌ	bad / had	-	bud / hudd
	ɛ	-	ʊ	bed / head	-	buddhist /
FD = 4	ɔː	-	ɪ	bawd / hawed	-	bid / hid
	iː	-	ʊ	bead / heed	-	buddhist /
	iː	-	ʌ	bead / heed	-	bud / hudd
	ɑː	-	ɛ	body / hod	-	bed / head
	uː	-	ɪ	booed / who'd	-	bid / hid
	uː	-	ʌ	booed / who'd	-	bud / hudd
FD = 5	ɑː	-	ɪ	body / hod	-	bid / hid
	uː	-	ɛ	booed / who'd	-	bed / head
	uː	-	æː	booed / who'd	-	bed / head
	æː	-	ʊ	bad / had	-	buddhist /

3.5.1 Error rates

3.5.1.1 Javanese vs. English Listeners

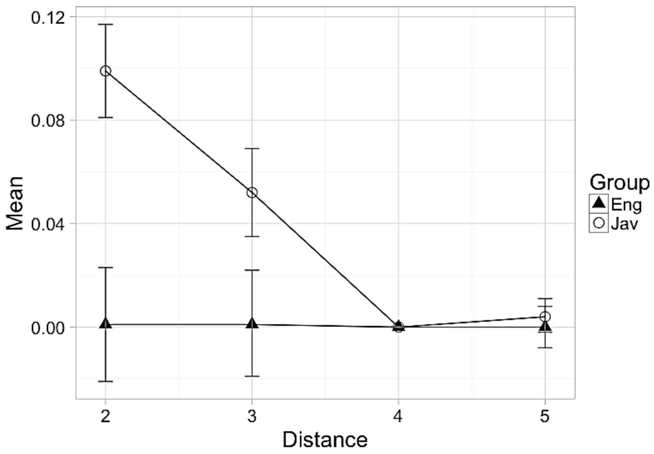
There was an interaction of Distance × Group [$F(1.44, 69.32) = 14.55$, $p < .001$, $\eta_p^2 = .23$], and Distance × Category × Group [$F(1.58, 75.67) = 3.89$, $p < .05$, $\eta_p^2 = .08$]. Last, there was a between-subject effect for Group [$F(1, 48) = 23.98$, $p < .001$, $\eta_p^2 = .33$]. Pairwise comparisons are presented in Table 3.4.

Table 3.4 Pairwise comparisons regarding proportion of errors for each English vowel within different 4 different distances and 2 categories in the Javanese and American English listeners. $\bar{x}$ = mean, SD = standard deviation, * = p < .05 (Bonferroni corrected significance threshold).

Distance	Category	Error Rates				*p*
		Javanese (N = 30)		English (N = 20)		
		$\bar{x}$	SD	$\bar{x}$	SD	
2	New	0.099	0.06	0.001	0.00	.000 *
	Similar	0.208	0.25	0.013	0.06	.001 *
3	New	0.052	0.06	0.001	0.00	.000 *
	Similar	0.033	0.09	0.013	0.06	.346
4	New	0.000	0.00	0.000	0.00	.000 *
	Similar	0.046	0.07	0.000	0.00	.005 *
5	New	0.004	0.02	0.000	0.00	.420
	Similar	0.000	0.00	0.000	0.00	.000 *

Figure 3.2a-b graphically shows the error rates (as proportions) obtained for Javanese versus American listeners, broken down by phonetic distance. Panel (a) does this for the new vowel category and panel (b) for the similar type.

a. *New* Category



b. *Similar* Category

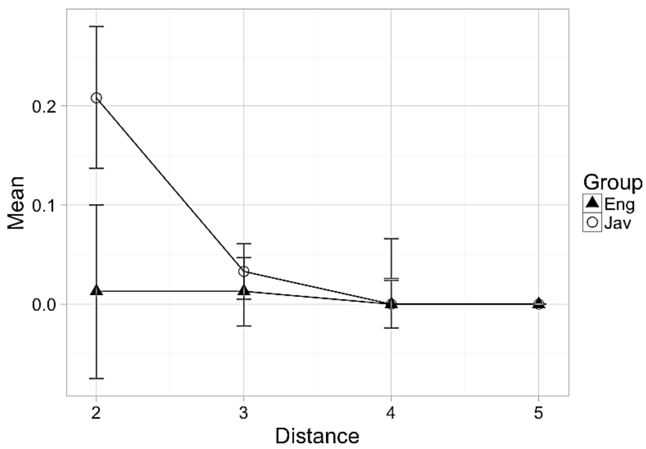


Figure 3.2a-b Mean Error rates (proportion) of new (panel a) and similar (panel b) vowel category for each level of phonetic distance performed by Javanese listeners, 95% confidence intervals for Distance levels 2-5.

The results show that at Distance 2, Javanese listeners show higher error rates than the English listeners for both new and similar vowels. At Distance 3, the Javanese listeners incurred higher error rates for new vowels only. Interestingly, the results show that the Javanese made more errors for similar vowels (but 0 errors for new) at Distance 4. At Distance 5, no effects remain (but 0 errors for similar vowels).

3.5.1.2 Sundanese vs. English Listeners

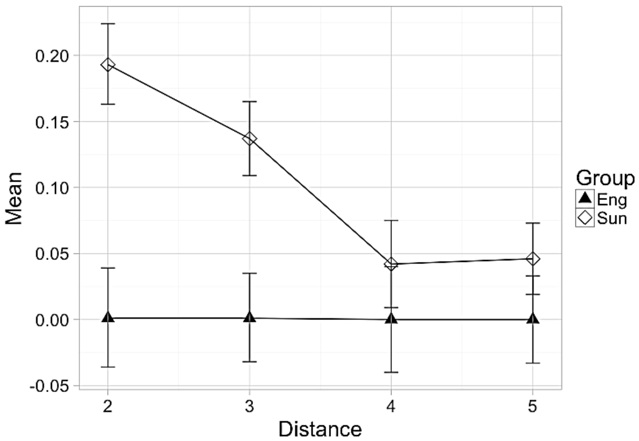
There was an interaction of Distance × Group [$F(2.04, 98.19) = 40.06$, $p < .001$, $\eta_p^2 = .46$], Category × Group [$F(1, 48) = 7.51$, $p < .05$, $\eta_p^2 = .14$], Distance × Category × Group [$F(1.20, 95.77) = 13.26$, $p < .001$, $\eta_p^2 = .22$]. There was a between-subject effect for Group [$F(1, 48) = 55.32$, $p < .001$, $\eta_p^2 = .53$]. Pairwise comparisons are presented in Table 3.5.

Table 3.5 Pairwise comparisons regarding error rates of each English vowel within different 4 different distances and 2 categories in the Sundanese and American English listeners. $\bar{x}$ = mean, SD = standard deviation, * = p < .05 (Bonferroni corrected significance threshold).

Distance	Category	Error Rates				p
		Sundanese (N = 30)		English (N = 20)		
		$\bar{x}$	SD	$\bar{x}$	SD	
2	New	0.193	0.11	0.001	0.00	.000 *
	Similar	0.508	0.32	0.013	0.06	.000 *
3	New	0.137	0.09	0.001	0.00	.000 *
	Similar	0.117	0.18	0.013	0.06	.017 *
4	New	0.042	0.11	0.000	0.00	.114
	Similar	0.096	0.10	0.000	0.00	.000 *
5	New	0.046	0.09	0.000	0.00	.038 *
	Similar	0.017	0.06	0.000	0.00	.247

Figure 3.3a-b graphically shows the error rates (as proportions) obtained for Sundanese versus American listeners, broken down by phonetic distance. Panel (a) does this for the new vowel category and panel (b) for the similar type.

a. *New* Category



b. *Similar* Category

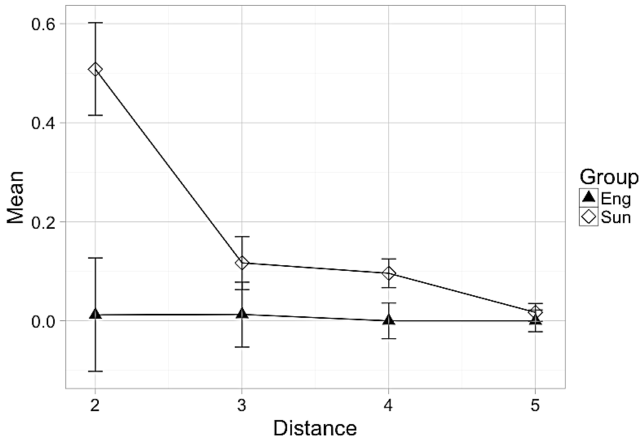


Figure 3.3a-b Mean Error rates (proportion) of new (panel a) and similar (panel b) vowel category for each level of phonetic distance performed by Javanese listeners, 95% confidence intervals for distance levels 2-5.

The results show that the Sundanese produce higher error rates than the American listeners for new and similar vowels at Distance 2 and Distance 3. At Distance 4, the Sundanese incurred higher error rates than the American listeners only for similar vowels. At Distance 5, the error rates by the Sundanese listeners are higher than the native listeners only for new vowels.

3.5.2 Area Under the Curve (AUC)

3.5.2.1 Javanese vs. English Listeners

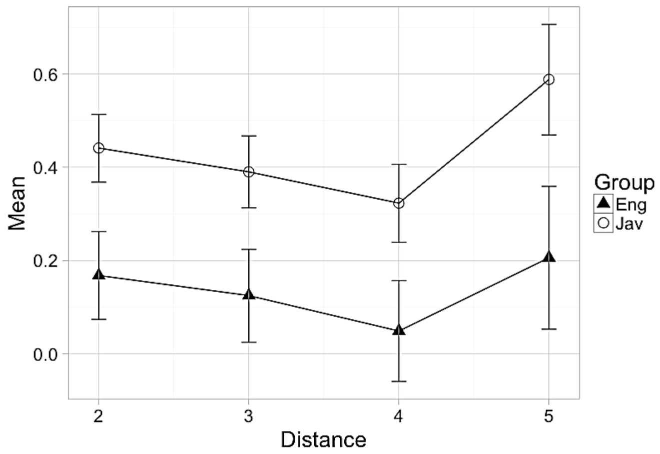
For the analysis of the Area Under the Curve (AUC) we excluded 2 subjects from the English group, because their mean scores were more than 3 SDs separated from the grand mean of the English group. There was an interaction of Category × Group [$F(1, 46) = 6.89$, $p < .05$, $\eta_p^2 = .13$]. There was a between-subject effect for Group [$F(1, 46) = 19.61$, $p < .001$, $\eta_p^2 = .30$. There were no other main effects or interactions. Pairwise comparisons are presented in Table 3.6.

Table 3.6 Pairwise comparisons regarding AUC of each English vowel within different 4 different distances and 2 categories in the Javanese and American English listeners. $\bar{x}$ = mean, SD = standard deviation, * = p < .05 (Bonferroni corrected significance threshold).

Category	Area Under the Curve (AUC)				p
	Javanese (N = 30)		English (N = 18)		
	$\bar{x}$	SD	$\bar{x}$	SD	
New	0.436	0.29	0.137	0.1	.000 *
Similar	0.323	0.25	0.146	0.14	.000 *

For both new and similar categories, the Javanese listeners performed with larger AUC values than the American English listeners. Moreover, the effect is larger for new vowels. This can be seen more clearly in Figure 3.4a-b, which shows the AUC values as a function of phonetic distance between target and distractor vowel for the two listener groups, separately for new vowels (panel a) and for similar vowels (panel b).

a. *New* Category



b. *Similar* Category

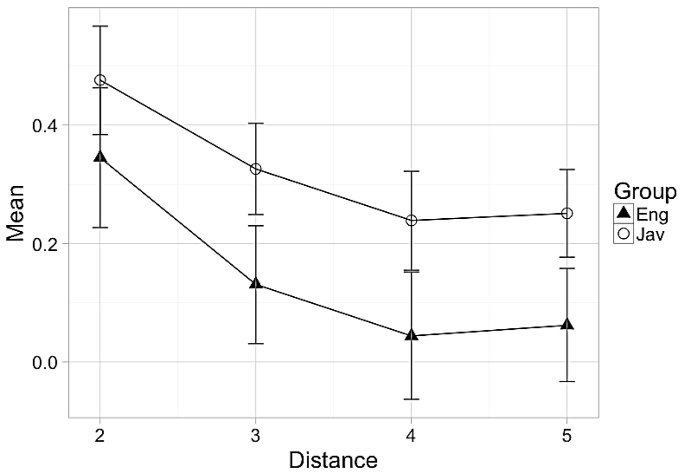


Figure 3.4a-b Area Under the Curve (mean AUC) of new (panel a) and similar (panel b) vowel category for each level of phonetic distance obtained by Javanese and American listeners, 95% confidence intervals for Distance levels 2-5.

3.5.2.2 Sundanese vs. English Listeners

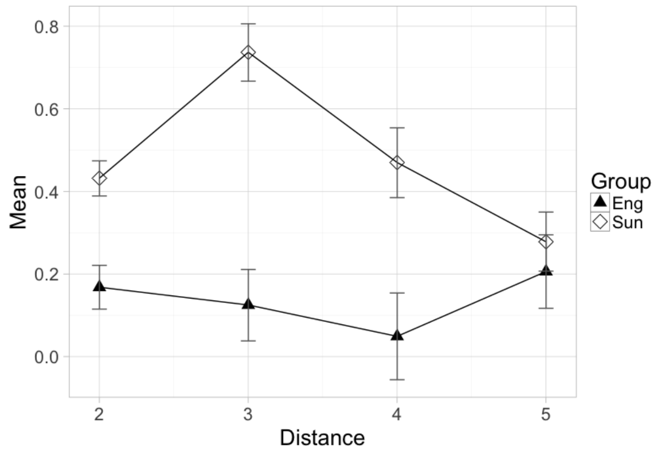
We excluded 2 outputs from the English group and 2 outputs from the Sundanese group in this analysis. There was an interaction of Distance × Group [$F(2.60, 114.27) = 8.57$, $p < .001$, $\eta_p^2 = .16$], Distance × Category × Group [$F(2.34, 103.05) = 24.80$, $p < .001$, $\eta_p^2 = .36$]. There was a between-subject effect for Group [$F(1, 44) = 94.34$, $p < .001$, $\eta_p^2 = .68$]. Pairwise comparisons are presented in Table 3.7.

Table 3.7 Pairwise comparisons regarding AUC of each English vowel within different 4 different distances and 2 categories in the Sundanese and American English listeners. $\bar{x}$ = mean, SD = standard deviation, * = p < .05 (Bonferroni corrected significance threshold).

Distance	Category	Area Under the Curve (AUC) Sundanese (N = 30) $\bar{x}$	Sundanese (N = 30) SD	English (N = 18) $\bar{x}$	English (N = 18) SD	*p*
2	New	0.432	0.12	0.168	0.09	.000 *
	Similar	0.886	0.39	0.345	0.25	.000 *
3	New	0.737	0.23	0.125	0.05	.000 *
	Similar	0.198	0.27	0.131	0.09	.313
4	New	0.47	0.27	0.049	0.07	.000 *
	Similar	0.423	0.14	0.044	0.08	.000 *
5	New	0.278	0.19	0.206	0.19	.206
	Similar	0.300	0.15	0.062	0.13	.000 *

The results are presented graphically in Figure 3.5a-b, which shows the AUC values as a function of Phonetic distance between target and distractor vowel for the two listener groups, separately for new vowels (panel a) and for similar vowels (panel b). Sundanese listeners showed larger AUC than English listeners for new and similar vowels at Distance 2 and 4. The L2 learners produced larger AUC values than the native listeners for new vowels at Distance 3, and for similar vowels at Distance 5.

a. *New* Category



b. *Similar* Category

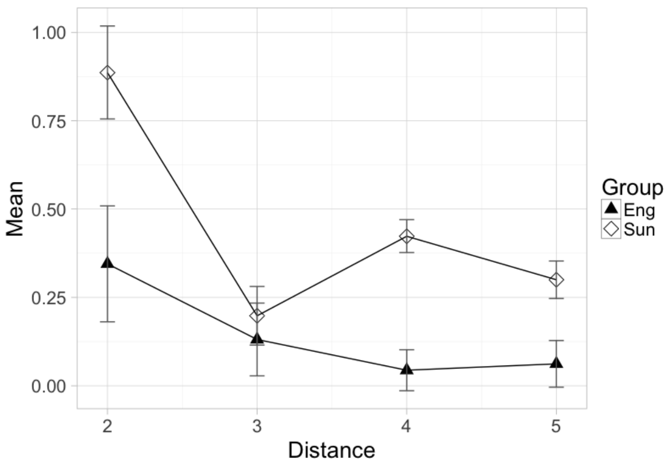


Figure 3.5a-b Area Under the Curve (mean AUC) of new (panel a) and similar (panel b) vowel category for each level of phonetic distance performed by Sundanese and American listeners, 95% confidence intervals for Distance levels 2-5.

3.5.3 Initiation Time

3.5.3.1 Javanese vs. English Listeners

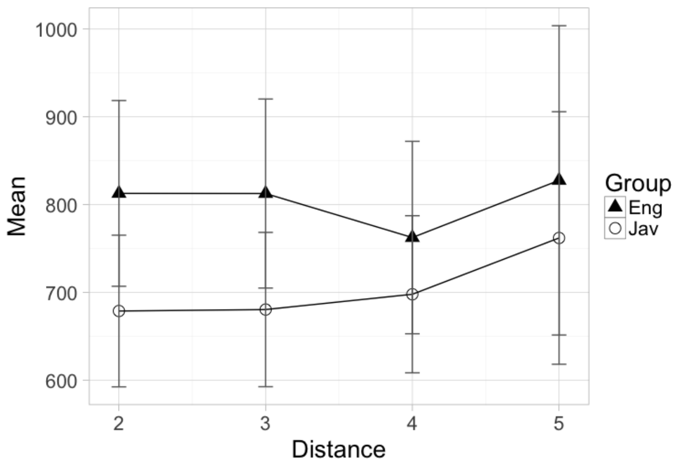
There was an interaction of Distance × Group [$F(2.13, 102.22) = 9.48$, $p < .001$, $\eta_p^2 = .17$], Distance × Category × Group [$F(2.52, 121.38) = 14.7$, $p < .001$, $\eta_p^2 = .23$]. There were no other main effects or interactions. Pairwise comparisons are presented in Table 3.8.

Table 3.8 Pairwise comparisons regarding Initiation Times (in ms) of each English vowel within different 4 different distances and 2 categories in the Javanese and American English listeners. $\bar{x}$ = mean, SD = standard deviation, * = p < .05 (Bonferroni corrected significance threshold).

		Initiation Time				
		Javanese (N = 30)		English (N = 20)		
Distance	Category	$\bar{x}$	SD	$\bar{x}$	SD	*p*
2	New	678.94	269.73	812.87	169.42	.054
	Similar	726.97	258.52	524.35	119.85	.002 *
3	New	680.61	272.36	812.72	177.33	.062
	Similar	665.03	282.99	792.01	166.68	.077
4	New	697.98	274.92	762.59	186.20	.363
	Similar	679.58	274.69	824.20	194.34	.047 *
5	New	762.10	475.46	827.72	206.12	.564
	Similar	636.04	242.84	785.45	166.23	.020 *

The results for Initiation Time are shown graphically in Figure 3.6a-b, which is organized in the same way as the earlier Figures 3.2-5. Javanese L2 listeners started to move the mouse later than the American L1 listeners for new and similar vowels at Distance 2 and 3. They initiated mouse movements after a longer delay for similar vowels at Distance 4 and 5.

a. *New* Category



b. *Similar* Category

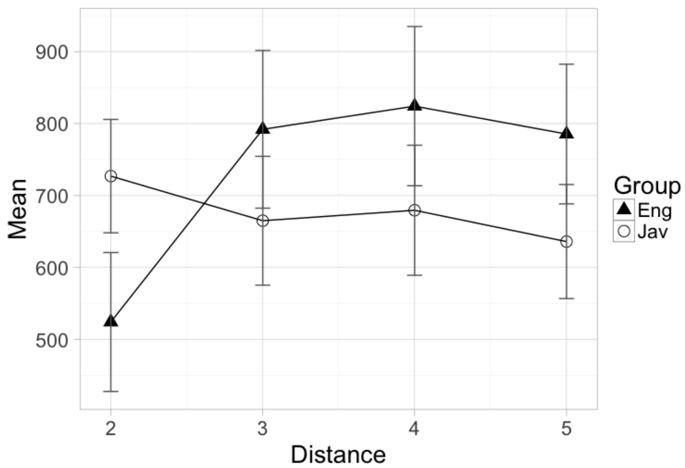


Figure 3.6a-b Mean Initiation Time (ms) obtained for new (panel a) and similar (panel b) vowel category for each level of phonetic distance from Javanese L2 and American L1 listeners, 95% confidence interval for Distance levels 2-5.

3.5.3.2 Sundanese vs. English listeners

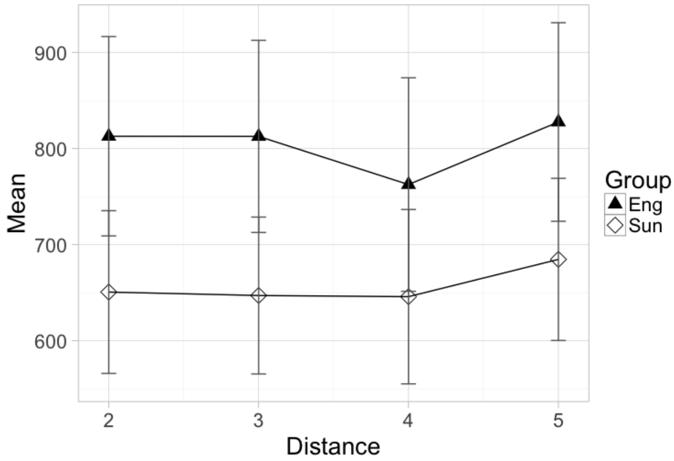
We excluded two listeners from the American group and two listeners from the Sundanese group in this analysis. There was an interaction of Distance × Group [$F(2.16, 103.74) = 12.21$, $p < .001$, $\eta_p^2 = .20$], Category × Group [$F(1, 48) = 7.12$, $p < .05$, $\eta_p^2 = .13$], Distance × Category × Group [$F(2.41, 115.59) = 16.15$, $p < .001$, $\eta_p^2 = .25$]. There was a between-subject effect for Group [$F(1, 48) = 4.06$, $p < .05$, $\eta_p^2 = .08$]. Pairwise comparisons are presented in Table 3.9.

Table 3.9 Pairwise comparisons regarding Initiation Times (in ms) of each English vowel within different 4 different distances and 2 categories in the Sundanese and American English listeners. $\bar{x}$ = mean, SD = standard deviation, * = p < .05 (Bonferroni corrected significance threshold).

Distance	Category	Initiation Time				*p*
		Sundanese (N = 30)		English (N = 20)		
		$\bar{x}$	SD	$\bar{x}$	SD	
2	New	650.65	278.21	812.87	169.42	.019 *
	Similar	656.00	262.32	524.35	119.85	.017 *
3	New	647.06	274.45	812.72	177.33	.013 *
	Similar	664.78	294.54	792.01	166.68	.043 *
4	New	645.85	286.98	762.59	186.20	.108
	Similar	651.27	298.45	824.20	194.34	.013 *
5	New	684.65	480.65	827.72	206.12	.036 *
	Similar	602.50	254.76	785.45	166.23	.005 *

The results are shown graphically in Figure 3.7a-b, which is constructed along the same lines as Figure 3.6, but now compares Sundanese and American listeners. Sundanese listeners initiated mouse movement after a longer delay than the American listeners did for new and similar vowels at Distance 2 and 3, and for only similar vowels at Distance 4.

a. *New* Category



b. *Similar* Category

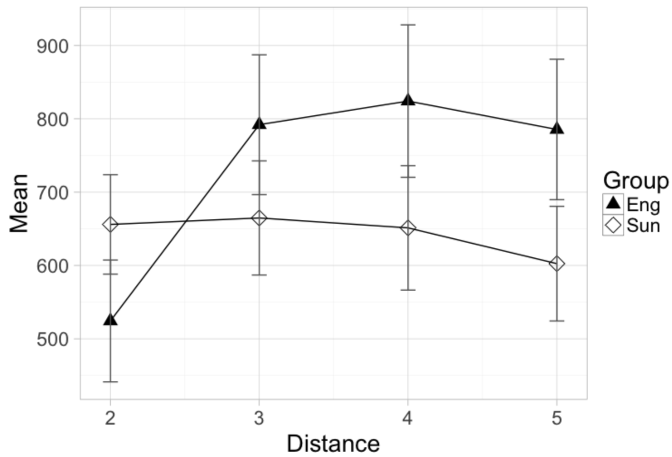


Figure 3.7a-b Mean Initiation Times (ms) of new (panel a) and similar (panel b) vowel category for each level of phonetic distance obtained from Sundanese and American listeners, 95% confidence interval for Distance levels 2-5.

3.5.4 Reaction Time

3.5.4.1 Javanese vs. English Listeners

Next, we analyzed reaction time**,** i.e. the time interval (in milliseconds) between the mouse click on /START/ button until the click on the target or distractor button.

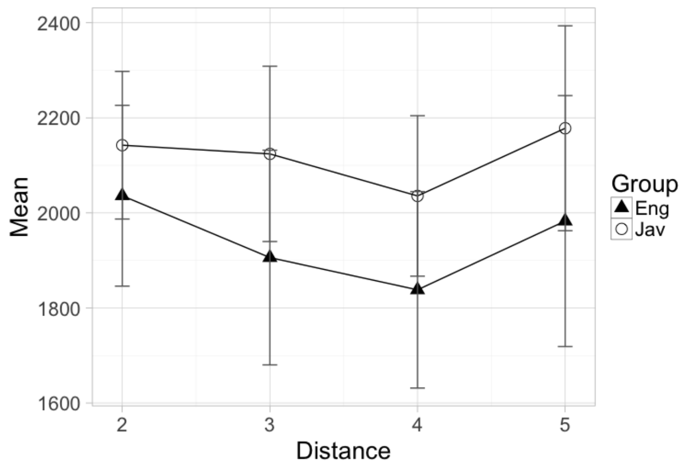
There was an interaction of Distance × Category × Group [$F(1.99, 95.95) = 3.89$, $p < .05$, $\eta_p^2 = .08$]. There were no other main effects or interactions. Mean and standard deviations are presented in Table 3.10.

Table 3.10 Mean and Standard deviation of Reaction Times (in ms) of each English vowel within different 4 different distances and 2 categories in the Javanese and American English listeners. $\bar{x}$ = mean, SD = standard deviation.

Distance	Category	Reaction Time			
		Javanese (N = 30)		English (N = 20)	
		$\bar{x}$	SD	$\bar{x}$	SD
2	New	2142.35	465.36	2036.07	348.58
	Similar	2312.95	623.23	2029.36	349.73
3	New	2124.09	599.10	1906.01	299.08
	Similar	2027.75	435.14	1969.01	415.60
4	New	2035.67	546.50	1838.15	279.20
	Similar	2046.14	416.71	1919.46	320.93
5	New	2178.01	675.91	1982.79	416.70
	Similar	1906.28	354.07	2017.99	491.08

The results are illustrated graphically in Figure 3.8a-b, on the analogy of the earlier figures. It took the Javanese listeners longer to move their mouse cursor to the location of the correct response alternative shown on screen than the English listeners but only for similar vowels at Distance 2.

a. *New* Category



b. *Similar* Category

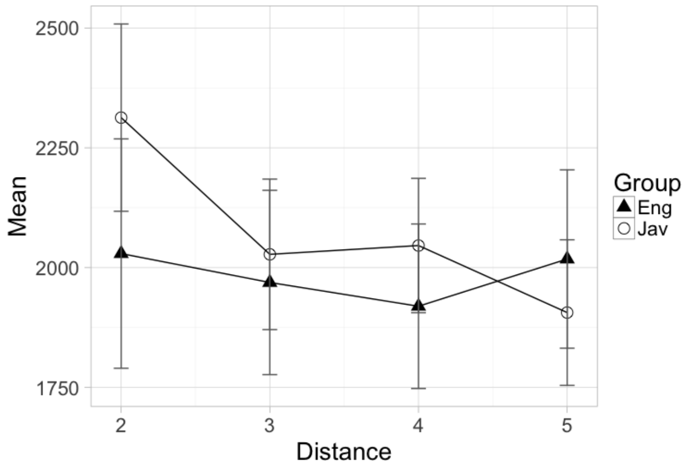


Figure 3.8a-b Mean Reaction Times (ms) of new (panel a) and similar (panel b) vowel category for each level of phonetic distance performed by Javanese and American listeners, 95% confidence interval for Distance levels 2-5.

3.5.4.2 Sundanese vs. English Listeners

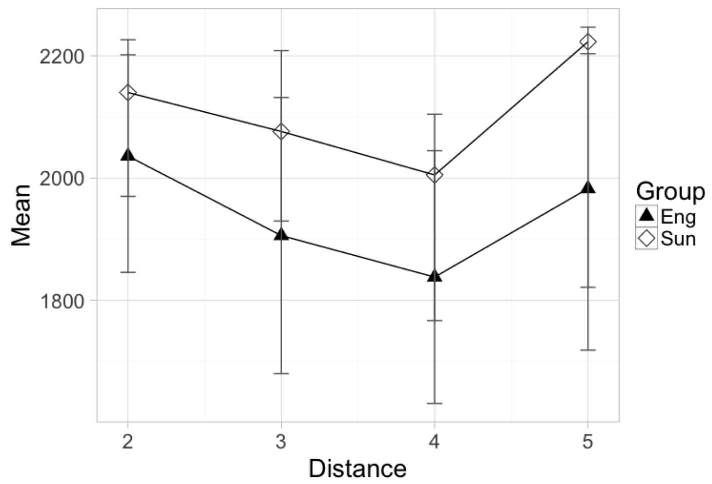
There were no other main effects or interactions. Mean and standard deviations are presented in Table 3.11.

Table 3.11 Mean and Standard deviations of Reaction Times (in ms) of each English vowel within different 4 different distances and 2 categories in the Sundanese and American English listeners. $\bar{x}$ = mean, SD = standard deviation.

Distance	Category	Reaction Time			
		Sundanese (N = 30)		English (N = 20)	
		$\bar{x}$	SD	$\bar{x}$	SD
2	New	2139.97	395.86	2036.07	348.58
	Similar	2272.87	440.04	2029.36	349.73
3	New	2076.28	403.32	1906.01	299.08
	Similar	2101.27	485.52	1969.01	415.60
4	New	2005.14	461.51	1838.15	279.20
	Similar	2066.49	401.02	1919.46	320.93
5	New	2223.02	638.63	1982.79	416.70
	Similar	1994.18	357.77	2017.99	491.08

A graphic representation of the results is provided in Figure 3.9a-b, on the analogy of Figure 3.8. The results show that no matter what combination of phonetic distance and vowel type is examined, it took the Sundanese listeners longer to move the mouse to the correct target word than the American listeners, except at the longest phonetic distance (Distance = 5) in the similar vowel category, where no difference between the Groups was found.

a. *New* Category



b. *Similar* Category

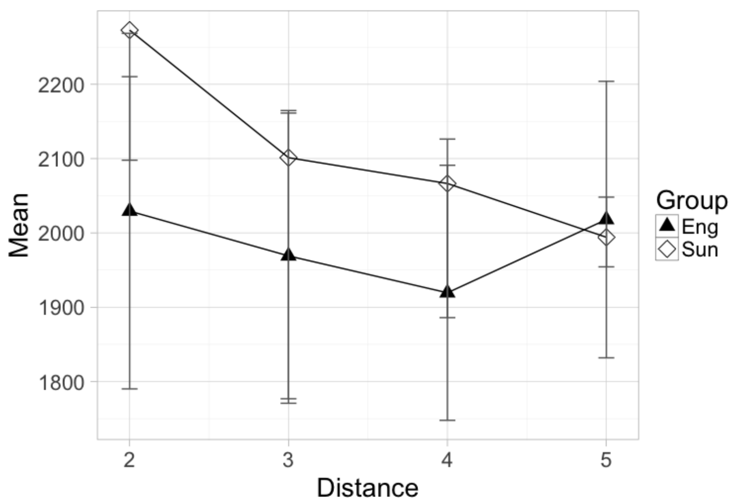


Figure 3.9a-b Reaction Times (mean) of similar category for each level of phonetic distance performed by Sundanese listeners, 95% confidence interval for Distance levels 2-5.

3.5.5 Velocity Profiles

Figures 3.10-11 show mean velocity (in pixels/ms) of the Javanese, Sundanese and American listeners in the time window from 0-1500 ms across all distances and vowel categories. Note that in an initial time window (steps 3-9) the non-native groups move the mouse faster than the native speakers, while in a window around the velocity peak (steps 12-14), the L1 listeners move faster than the L2 learners. As explained in the methods sections, these two windows were selected to determine whether velocity varied systematically between the windows as a function of the independent variables described above.

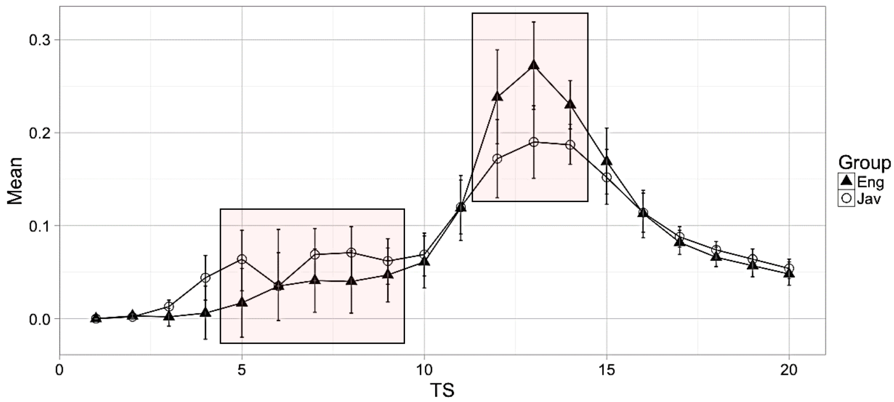


Figure 3.10 Mean hand movement velocity (in pixels/ms) of Javanese and American listeners perceiving English vowels over time (0-1500 ms), divided into 20 bins of 75 ms, TS = time step.

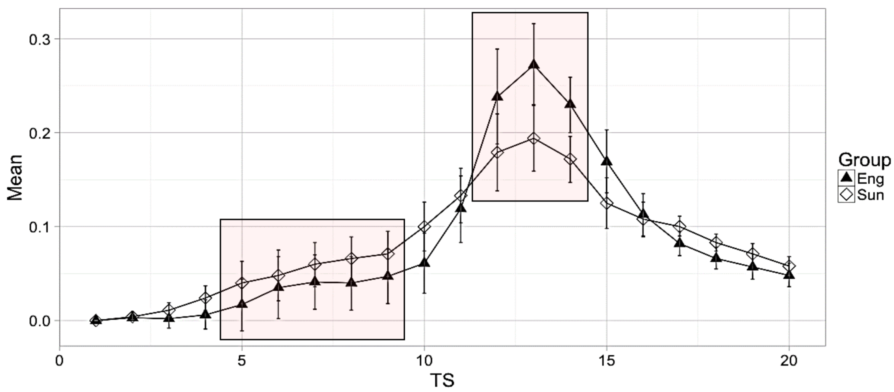


Figure 3.11 As Figure 3.10 but for Sundanese and American listeners.

Statistical analysis reveals no significant main effects or interactions between any of the independent variables for the velocity profiles of the hand movements during the Initial Time Window, i.e. the time steps 3-9 in the 20-point equidistant time steps of the trajectories between 0 and 1500 ms post-stimulus-onset (between 226 and 750 ms). The velocity profiles therefore did not differ significantly among the three listener groups during the first stage of the response process.

In the Later Time Window, i.e. the time steps 12-14 (between 751 and 1125 ms on the normalized time scale between 0 and 1500 ms), we find a between-subjects effect for Group [$F(1, 47) = 10.24$, $p < .05$, $\eta_p^2 = .18$]. This shows that the Javanese listeners (N = 30, $\bar{x} = .1643$, SD = .044) moved more slowly than the American listeners (N= 20, $\bar{x} = .2057$, SD = .051). Almost the same effect was found for the comparison of the Sundanese and American listeners, [$F(1, 48) = 13.60$, $p < .001$, $\eta_p^2 = .22$]. The Sundanese listeners (N = 30, $\bar{x} = 0.1605$, SD = .038) show slower velocity than the American listeners (N= 20, $\bar{x} = .2057$, SD = .051). No other main effects or interactions were found.

3.6 Explorative Analysis of Difficult Vowel Pairs

Since the category of new vowels, especially at the smallest distance between target and distractor (FD = 2), posed perception problems, as evidenced by greater AUC and higher error rates, we followed up these analyses to find out which specific vowel pairs were difficult for the L2 listeners (Table 3.12).

Table 3.12 Ten difficult vowel pairs for Javanese and Sundanese listeners.

Pairs		*t* for AUC		*t* for Error Rate	
Vowel	Word	Javanese	Sundanese	Javanese	Sundanese
ɔː - ʊ	bawd/hawed - buddhist/hood	−2.266	−1.988	−1.461	−2.717
ɔː - ʌ	bawd/hawed - bud/hudd	−2.781	−2.493	−2.191	−2.191
ɜ˞ - ʊ	bird/heard - buddhist/hood	−2.474	−2.312	-	-
ɜ˞ - ʌ	bird/heard - bud/hudd	−2.751	−2.323	-	-
ɑː - æː	body/hod - bad/had	−3.427	−4.218	−2.191	−2.869
æː - ɛ	bad/had - bed/head	−1.101	.5270	−9.026	−11.713
ɪ - ʊ	bid/hid - buddhist/hood	−3.183	−7.016	−1.719	−3.768
ɪ - ʌ	bid/hid - bud/hudd	−2.143	−4.238	−1.369	−2.499
ʊ - ʌ	buddhist/hood- bud/ hudd	−2.957	−7.256	−4.000	−6.707
ɜ˞ - ɪ	bird/heard - bid/hid	−2.075	.0390	-	-

As can be observed in Table 3.12, the correct identification of the target in the English vowel pair /æː - ɛ/ was the most difficult (relative to the Error Rate obtained by the American native listeners), as evidenced by the very large t-value for the Error Rate for both groups of Indonesian listeners. This contrast is phonetically rather small (FD = 2) and involves two new sounds.

3.7 Discussion

Lado (1957) suggest that L2 learners are almost entirely dependent on their mother tongue (L1) in the process of learning an L2. Second language learners often encounter difficulties due to interference from L1. The purpose of the present study was to examine whether the sound category (similar vs. new as compared to the L1 vowel inventory) affects L2 sound perception among Javanese and Sundanese listeners and whether the phonetic distance between L2 target and distractor influence the L2 learners' perception. To this end, a mouse tracking experiment measuring Error rate, Area Under the Curve (AUC), reaction time, initiation time and velocity profile was carried out.

The first aim of the study was to investigate whether familiarity with the L2 sounds (new vs similar sounds with respect to the L1 vowel inventory) can influence the L2 learning. The results of the experiment indicate that there is indeed an effect of sound category (new vs similar sound) on the English sound perception by Javanese and Sundanese listeners. Altogether, results suggest that the L2 learners experience more difficulty perceiving new vowels (relative to the L1 vowel inventory) than perceiving similar vowels. This lends partial support to the L2LP model (Escudero, 2005).

Indeed, the most sensitive dependent measures (which showed a clear deviation from native speakers and interacted with vowel category), i.e., AUC and Error Rate, showed that the L2 listeners generally had more problems with perceiving new than similar vowels, particularly when the target L2 vowel was spectrally close to the L2 competitor vowel.

However, results of the velocity profiles did not provide clear evidence in favor of any of the models of L2 speech learning with regards to the effect of the sound category. In the early time window (226-750 ms post-stimulus), Javanese listeners show a trend of higher velocity profiles than the American L1 listeners regardless the sound category. The Sundanese listeners, on the other hand, showed larger velocity than the English listeners for new sounds at Distance 3. In the later time window (751-1125 ms post-stimulus), the velocity of Javanese and Sundanese were smaller than those of the American L1 listeners for both new and similar vowels. Hence, these relative new behavioral measures do not seem informative in identifying specific perception difficulties among the L2 learners.

The second aim of the current study was to examine whether phonetic distance affects the L2 sound learning, assuming more perception difficulty as the phonetic distance is closer. As noted above, the two most sensitive measures of the mouse tracking experiment (AUC and particularly Error Rate), indeed showed an interaction with phonetic distance between the L2 target vowel and its competitor. The results of the error rates indeed showed that the L2 learners incurred relatively high error rates when the acoustic distance between target vowel and competitor was small. The results confirm that when the spectral distance between the target and competitor vowel is small the L2 listeners have difficulty with discriminating the L2 sounds.

Two unexpected findings need to be mentioned. First, the initiation times showed that the L2 listeners initiate the mouse movement faster than native speakers. Similarly, in the early time window, L2 speakers moved the mouse at a higher velocity than the native speakers. We therefore suggest that these results reflect general strategy differences between the groups of L2 and L1 speakers, which hence do not reflect linguistic processing differences. One possibility might be that the word 'directly' in the instruction to 'move the mouse directly to the target of their choice, within 2000 ms' was interpreted differently by the American and Indonesian listeners. The word directly has two meanings (i) in a straight line, and (ii) immediately. American native speakers may feel a contradiction in the instruction to move the mouse immediately but no later than 2 seconds after the initial mouse click. They will therefore be likely to conclude that their task is to wait for some time, and then move towards the target of their choice in a straight line (as fast as possible, because time is running out). The Indonesian listeners, on the other hand, may not have been aware of the semantic conflict in the instruction, and have opted for the interpretation that they should start moving the mouse immediately – in a neutral (vertical) trajectory making up their minds while moving the mouse. The latter strategy will lead to shorter trajectories to the target, once the listener has decided which one to choose. This account of what may have happened is perfectly compatible with the observed difference in behavior between the Indonesian and American participants in the experiment.

Prima facie our results lend some (at least partial) support to the L2LP model of L2 speech perception, since, if there were clear differences between similar and new vowels, perception differences between the native and non-native speakers were largest for new vowels. If we compare the effect sizes of interactions with the vowel category on the one hand, and the phonetic distance between the L2 target and competitor vowel on the other hand, however, it must be noted that the effect of distance is almost an order of magnitude larger than that of sound category. Hence, not L1-L2 interference but perceptual discrimination difficulty in distinguishing L2 sounds is driving the results of the present study.

On second thoughts, however, a different conclusion may also be appropriate. As stated in Chapter 1, the Speech Learning Model targets the L2 learner at the end of the acquisition process. The model does not say anything about the relative difficulty of similar versus new sound categories in the earlier stages of L2 acquisition. It is well documented in the literature that highly advanced learners of an L2 develop authentic, i.e. native-like, new sound categories, which do not interfere with the categories that already existed in their L1. At the same time these highly advanced learners continued to produce similar sounds in the L2 in a way that sounded odd to native listeners of the L2 (but not to the learners themselves). For similar sounds the learners expand an existing category in their L1 so as to include also the exemplars of what they tend to perceive as the same sound in the L2. This leads to a larger category, with a shifted prototype, which is deviant from the native norm for both the learner's L2 and his/her L1 (Flege 1987, 1995). Highly advanced learners of the L2, in Flege's examples, are American professors of French as an L2, with many years of experience with the L2 in France, or Dutch university lecturers and professors of English working in an English Department. It seems reasonable, in hindsight, that the Indonesian university students of English, who served as participants in the present study, should be considered intermediate (rather than advanced) learners of English, whose L2 acquisition is far from completed. In the earlier stages of L2 acquisition, a similar sound will be easier to identify as one of the sound categories in the L2, because there is a greater degree of overlap between the L1 and L2 categories. New sounds may well find themselves in between existing categories ('Uncategorized' in PAM terminology) and will certainly be atypical for any existing category in the learner's L1, which renders them a source of uncertainty (and hence delay the response) in the kind of perception task imposed on the participants in the present experiment. This is also the motivation for the L2LP model (and earlier transfer models of L2 perception) to consider similar sounds a lesser problem than new sounds in the beginning stages of the L2 acquisition.

A final point to consider in the present study is the way we operationalized the sound category, i.e. in terms of new versus similar, in the experimental task. A trial in the experiment involves a single auditory stimulus vowel in American English, which may be either new or similar sound to the Indonesian L2 learner of English. However, the participants have to identify the auditory sound with one of two written words printed in the corners of the screen in front of them. In 56 of the 70 trials the response alternatives are both new sounds. In the remaining 14 trials, however, the alternatives form a heterogeneous pair, i.e. one new and one similar sound. These pairs were used to represent the category of similar sounds, in the strength of the argument that the auditory stimulus was a similar sound. It is unclear at this time, to what extent the fact that the distractor vowel in the similar trial type was a new sound may have compromized the results. To control for this possibility a replication of the experiment should include also homogeneous pairs of similar sounds.

3.8 Conclusion

The purpose of the current study was to examine the effect of sound category (i.e., new or similar relative to the learner's L1 vowel inventory) and phonetic distance between target sounds in L2 sound perception among Javanese and Sundanese L2 listeners as compared to American native speakers of English as measured with mouse tracking. Results demonstrated that the perception of L2 sounds was affected partially by the sound category and affected clearly more strongly by the phonetic distance. The results indicated that L2 listeners had more difficulty perceiving new L2 sounds than similar sounds, particularly when having to discriminate between perceptually close L2 sounds.

There are three reasons for the large mean differences between the American-English listeners on the one hand, and the Javanese and Sundanese groups on the other. First, confusions of the members of English vowel pairs frequently occur because there are no similar vowel sounds in learners' first language. English vowel pairs /ɔː - ʊ/, /ɔː - ʌ/, /ɝ - ʊ/, /ɝ - ʌ/, /ɑː - æː/, /æː - ɛ/, /ɪ - ʊ/, /ɪ - ʌ/, /ʊ - ʌ/ are considered to involve exclusively new vowels for the Javanese and Sundanese listeners. In a related study, Dutch listeners made perception errors on American-English /ʊ - uː/, /ɪ - iː/, /æ - e/ vowel pairs because of the differences which exist in the listeners' L1 and L2 (Wang, 2007). For the Javanese and Sundanese listeners, the confusion in perceiving /æː - ɛ/ was due to a lack of a category boundary between /æ/ and /e/. It appears therefore, that the L2 listeners in the present study had more difficulty in discriminating L2 sounds than for example the Chinese, and especially the Dutch, L2 listeners in Wang (2007), possibly due to the lack of a representation of the English vowel /æ/ in the L1.

Second, the tense vs lax perception errors such as /ɔː - ʊ/, /ɔː - ʌ/, /ɝ - ʊ/, /ɝ - ʌ/, /æː - ɛ/ may have been caused by the duration difference between the English and the Javanese/Sundanese vowel systems. Hence, perceiving these vowel pairs can be attributed to interference from the Javanese and Sundanese speakers' L1 (Munro, 1993). The type of error among L2 speaking groups would then occur due to the total absence of contrastive length in the L1 vowel inventory.

Third, learning problems of English vowels pairs /ɔː - ʊ/, /ɔː - ʌ/, /ɝ - ʊ/, /ɝ - ʌ/, /ɑː - æː/, /æː - ɛ/, /ɪ - ʊ/, /ɪ - ʌ/, /ʊ - ʌ/ have been found to be the result of the closeness of such vowels in the vowel space. Indeed, this was confirmed by the large effect of phonetic distance on L2 perception performance in the current study.

This study was limited by the type of stimuli used. We specifically used vowel pairs that were only minimally different, zooming in on relatively difficult vowel pairs. Further research would be needed to show if cross-language perception shows similar patterns for all L2 vowel contrasts (Escudero & Boersma, 2004; Morisson, 2009).

References

Ali, E. M. T. (2011). *Speech intelligibility problems of Sudanese learners of English.* LOT Dissertation series nr. 270. Utrecht: LOT.

Baker, W., Trofimovich, P., Mack, M., & Flege, J. E. (2002). The effect of perceived phonetic similarity on L2 sound learning by children and adults. In A. Do, L. Dominguez, & A. Johansen (Eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press.

Barca, L., & Pezzulo, G. (2012). Unfolding visual lexical decision in time. *PLoS ONE*, *7*, e35932. DOI: 10.1371/journal.pone.0035932.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In: Strange W, editor. *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-*

language Speech Research (pp. 167-200). York: Timonium, MD.

Best C. T., & Tyler M. D. (2007). Nonnative and second-language speech perception: commonalities and complementarities, in Second *Language Speech Learning: The Role of Language Experience in Speech Perception and Production* (pp. 13-34). Amsterdam: John Benjamins.

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America, 97*, 1916-1925.

Bruhn, P., Huette, S., & Spivey, M. (2013). Degree of certainty modulates anticipatory processes in real time. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 525-538.

Cebrian, J. (2006). Experience and the use of non-L1 duration in L2 vowel categorization. *Journal of Phonetics, 34*, 372-387.

Cebrian, J. (2007). Old sounds in new contrasts: L2 production of the English tense-tax vowel distinction. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th ICPhS, Saarbrücken* (pp. 1637-1640).

Crothers, J. (1978). Typology and universals of vowel systems. In J. H. Greenberg (Ed.), *Universals of human language*. Vol. 2: Phonology (pp. 94-152). Stanford, CA: Stanford University Press.

Dale, R., Kehoe, C., & Spivey, M. (2007). Graded motor responses in the time course of categorizing exemplars. *Memory and Cognition, 35*, 15-28.

Elvin, J., Escudero, P., & Vasiliev, P. (2014). Spanish is better than English for discriminating Portuguese vowels: Acoustic similarity versus vowel system. *Frontiers in Psychology. Language Sciences, 5*, 1188.

Escudero, P. (2000). *Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels in Spanish speakers* (Unpublished M.Sc. thesis). University of Edinburgh.

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

Escudero, P. (2006). *Second language phonology: The role of perception*. In M. Pennington (Ed.), *Phonology in context* (pp. 109-134). Houndmills, Basingstoke, Hampshire, New York: Palgrave Macmillan.

Escudero, P. (2006). The phonological and phonetic development of new vowel contrasts in Spanish learners of English. In. B. O. Baptista & M. A. Watkins (Eds.), *English with a Latin beat: Studies in Portuguese/Spanish-English interphonology* (Studies in

Bilingualism 31, pp. 149-161). Amsterdam: John Benjamins.

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.

Escudero, P., & Boersma, P. (2002). The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish, *Proceedings of the 26th annual Boston University Conference on Language Development*, edited by B. Skarabela, S. Fish, and A. H. –J. Do (pp. 208-19). Somerville, MA: Cascadilla Press.

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., & Chládková, K. (2010). Spanish listeners' perception of American and Southern British English vowels. *Journal of the Acoustical Society of America, 128*, EL254-260.

Farmer, T. A., Anderson, S. E., & Spivey, M. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language, 57*: 570-595. DOI: 10.1016/j.jml.2007.04.003.

Flege, J. E. (1987). The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalent classification. *Journal of Phonetics, 15*, 47-65.

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 H. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development*, Papers in honor of Henning Wode (pp. 217-44). Trier: Wissenschaftlicher Verlag Trier.

Flege, J. E. (2003). Assesing constraints on second-language segmental production and perception. In A. Meyer & Schiller, N. (Eds.), Phonetics and phonology in language comprehension and production: Differences and similarities. Berlin: De Gruyter.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). The effect of experience on non-native subjects' production and perception of English vowels. *Journal of Phonetics, 25*, 437-470.

Flege, J. E., & Hillendbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America, 76*, 708-721.

Flege, J. E. & MacKay, I. (2004). Perceiving vowels in a second language.

Studies in Second Language Acquisition, 26, 1–34. DOI: 10.1017/S0272263104026117.

Flege, J. E., Munro, M., & MacKay, I. (1995). Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97*, 3125-3134.

Flumini, A., Barca, L., Borghi, A.M., Pezzulo, G. (2015). How do you hold your mouse? Tracking the compatibility effect between hand posture and stimulus size. *Psychological Research, 79*(6), 928-938. DOI: 10.1007/s00426-014-0622-0.

Fox, R., Flege, J. E., & Munro, M. (1995). The perception of English and Spanish vowels by L1 English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America, 97*, 2540-2551.

Freeman, J. B., & Ambady, N. (2010). MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*, 226-241. DOI: 10.3758/BRM.42.1.226.

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*, 59. DOI: 10.3389/fpsyg.2011.00059.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society, 108*, 3013-3022.

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.

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for L1 Spanish and German speakers. *Journal of the Acoustical Society of America, 126*, 866-877. DOI: 10.1121/1.3148196.

Kamiyama, T. (2011). Pronunciation of French vowels by Japanese speakers learning French as a foreign language: back and front rounded vowels /u y ø/. *Phonological Studies: Phonological Society of Japan*, 97-108.

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

Ladefoged, P. (2006). *A course in phonetics*. New York: Harcourt, Brace & Jovanovich.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers.* University of Michigan Press: Ann Arbor.

Lauder, M. R. M. T., & Ayatrohaedi (2006). The distribution of Austronesian and non-Austronesian languages in Indonesia: Evidence and issues. In T. Simanjuntak, I. H. E. Pojoh, & M. Hisyam (Eds.), *Austronesian diaspora and the ethnogeneses of people in Indonesian archipelago*. *Proceedings of the International Symposium* (pp. 361-391). Jakarta: LIPI Press.

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

Morrison, G. S. (2008). Perception of synthetic vowels by monolingual Canadian-English, Mexican-Spanish, and Peninsular-Spanish listeners. *Canadian Acoustics, 36*, 17-23.

Morrison, G. S. (2009). L1-Spanish speakers' acquisition of the English /i/ - /ɪ/ contrast II: Perception of vowel inherent spectral change. *Language and Speech, 52*, 437-462.

Nothofer, B. (2009). Javanese. In *Concise Encyclopedia of Languages of the World* (pp. 560-561). Oxford: Elsevier.

Perwitasari, A. (2013). Slips of the ears: Study on vowel perception in Indonesian learners of English. *Humaniora: Journal of Culture, Literature, and Linguistics, 25*, 103-110.

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.

Sisinni, B., Escudero, P., & Grimaldi, M. (2014). The perception of American English vowels by Salento Italian adult listeners: Longitudinal development in the classroom context. *Proceedings of the International Symposium on the Acquisition of Second Language Speech. Concordia Working Papers in Applied Linguistics*, *5*, 709-721.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the USA, 102*, 10393-10398. DOI: 10.1073/pnas.0503903102.

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*(4), 1791. DOI: 10.1121/1.1687832.

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.

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. (2003). Mutual intelligibility of Chinese, Dutch and American speakers of English. In P. Fikkert & L. Cornips (Eds.), *Linguistics in the Netherlands 2003* (pp. 213-224), Amsterdam/Philadephia: John Benjamins.

Wang, H., & Van Heuven, V. J. (2004). Cross-linguistic confusion of vowels produced and perceived by Chinese, Dutch and American speakers of English. In L. Cornips & J. Doetjes (Eds.), *Linguistics in the Netherlands 2004* (pp. 205-216). John Benjamins, Amsterdam/Philadelphia.

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.

Yoon, E.-K. (2013). The effects of perceptual training on speech production: Focusing on Korean vowels. *Studies in Foreign Language Education*, *27*(2), 1-27.

Appendix A. Summary of RM-ANOVA

Effects and interaction of Distance and Category on (i) Error Rate, (ii) AUC, (iii) Initiation Time, (iv) Reaction time, (v) Velocity in Initial Time Window, and (vi) Velocity in Later Time Window, comparing either Javanese or Sundanese L2 listeners with American L1 listeners of English.

Effect/interaction	Df1	Df2	F	p	η_p^2
Dependent = Error Rate; Javanese vs American					
Distance	1.44	69.32	17.27	< .001	.27
Category	1	48	7.05	< .05	.13
Distance × Category	1.58	75.67	4.35	< .05	.08
Dependent = Error Rate; Sundanese vs American					
Distance	2.04	98.19	43.59	< .001	.48
Category	1	48	9.94	< .05	.17
Distance × Category	1.20	95.77	14.34	< .001	.23
Dependent = Area Under the Curve (AUC); Javanese vs American					
Distance	2.73	125.4	23.58	< .001	.34
Category	1	46	5.03	< .05	.10
Distance × Category	2.64	121.62	24.74	< .001	.35
Dependent = Area Under the Curve (AUC); Sundanese vs American					
Distance	2.60	114.27	29.87	< .001	.40
Distance × Category	2.34	103.05	43.49	< .001	.50
Dependent = Initiation Time; Javanese vs American					
Distance	2.13	102.22	5.97	< .05	.11
Category	1	48	14.42	< .001	.23
Distance × Category	2.52	121.38	6.06	< .001	.11
Dependent = Initiation Time; Sundanese vs American					
Distance	2.16	103.74	10.78	< .001	.18
Category	1	48	15.06	< .001	.24
Distance × Category	2.41	115.59	14.51	< .001	.23
Dependent = Reaction Time; Javanese vs American					
Distance	2.09	100.49	8.22	< .001	.15
Dependent = Reaction Time; Sundanese vs American					
Distance	2.52	121.22	10.46	< .001	.18
Dependent = Velocity in Initial Time Window, Javanese vs American					
No effects or interaction					
Dependent = Velocity in Initial Time Window, Sundanese vs American					
Distance × Category	2.44	117.05	3.89	< .05	.07
Dependent = Velocity in Later Time Window, Javanese vs American					
Distance	2.72	127.75	10.18	< .001	.18
Category	1	47	26.72	< .001	.36
Dependent = Velocity in Later Time Window, Sundanese vs American					
Distance	2.67	127.48	7.47	< .001	.14
Category	1	48	14.28	< .001	.23

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, ɝ/. 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).

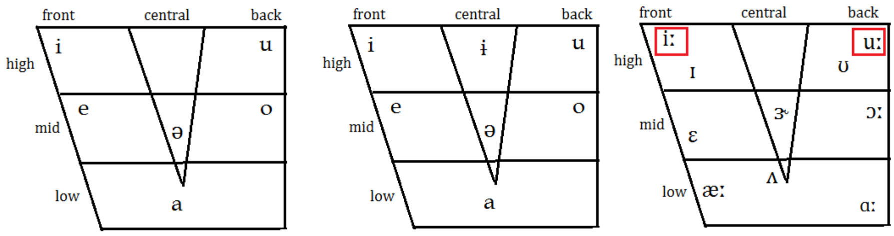


Figure 4.1 Articulatory plots for vowels of Javanese (left), Sundanese (middle) and English (right). The vowels in squares are considered similar L2 vowels for Indonesian learners of English.

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, ɝ/, 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, ɝ/ 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).

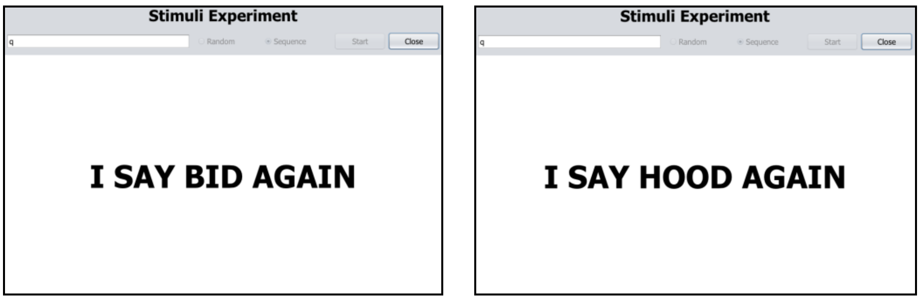


Figure 4.2 Examples of two stimulus strings as displayed on the computer screen.

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, ɝ, 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)					
	Sundanese		English		U	*p*
	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)					
	Sundanese		English		U	*p*
	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.

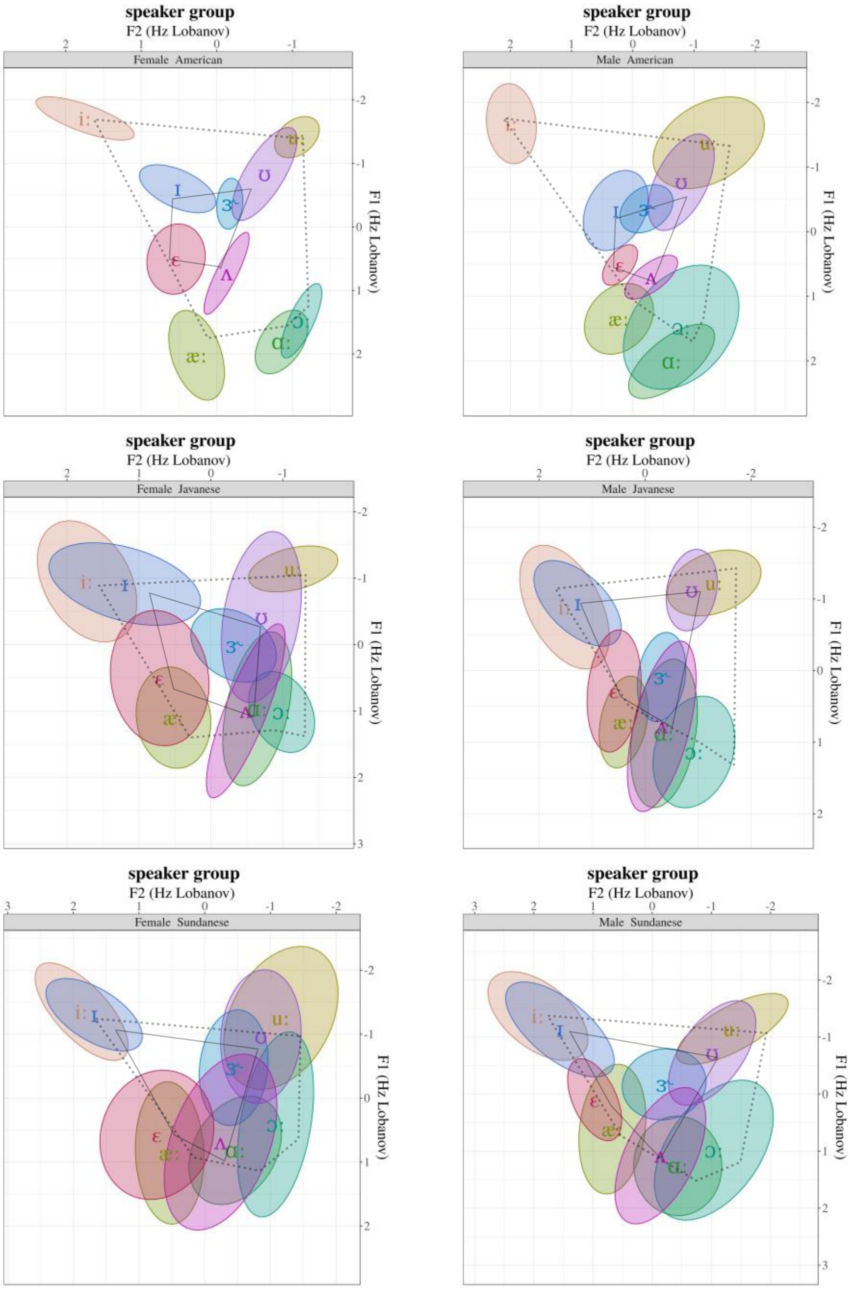


Figure 4.3 Centroids and ellipses of F_1 and F_2 values (z-normalized), for 10 English vowels produced by American English, Javanese and Sundanese female and male speakers. Dotted and solid polygons join tense and lax vowels, respectively.

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, ɝ, 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.

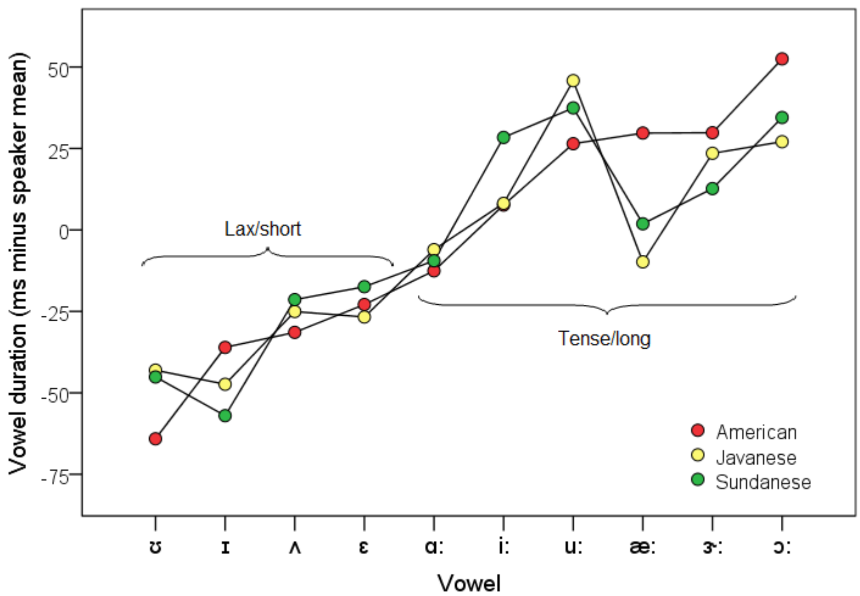


Figure 4.4 Adjusted vowel duration for the ten English monophthongs produced by American, Javanese and Sundanese speakers. Adjustment was done by subtracting the speaker-individual mean from the duration of each vowel token. Vowels are in ascending order of duration as established for the American speakers.

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.

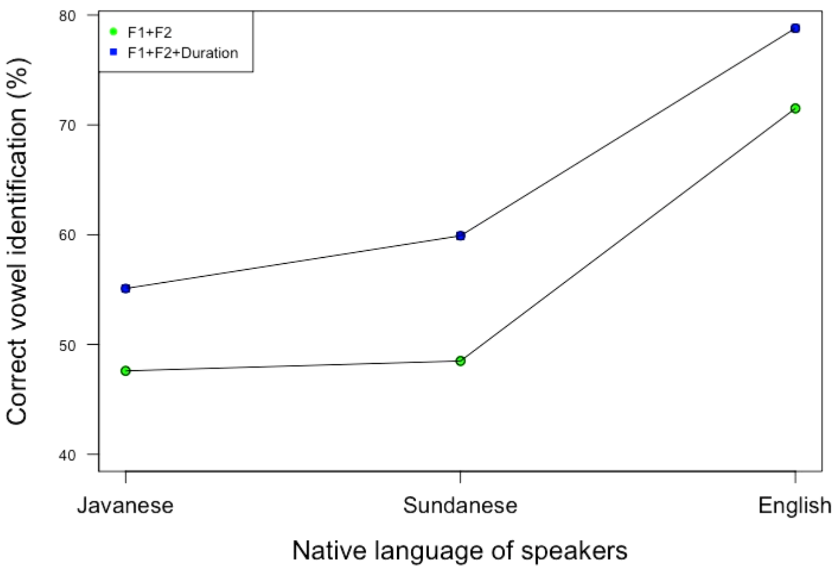


Figure 4.5 Mean correct English vowel identification (%) based on F_1 and F_2 (and duration, upper line) of Javanese, Sundanese and American speakers.

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	ɝ	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		
	ɝ		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	ɝ	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
	ɝ		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	ɝ	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		
	ɝ		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.

5
General Conclusion

5.1 Introduction

The purpose of this research was to investigate the perception and production of English vowels by Javanese and Sundanese speakers (henceforth L2 learners). This chapter reports the general conclusions with respect to L2 speech learning by the Javanese and Sundanese speakers and implications for the teaching of English as an L2 teaching in Indonesia. It also presents the limitations and provides recommendations related to L2 learning.

The use of two or more languages in a community of speakers, i.e. diglossia, is common in Indonesia because the country is home to more than 700 local languages spoken by a total of around 255 million people. Many studies have carried out cross-linguistic comparisons between English as a foreign language and western European languages such as Spanish, German, Dutch, Swedish, Danish, and Italian. However, few studies have investigated Austronesian languages, such as Javanese and Sundanese - two widely spoken local languages in Indonesia. Since English has officially become a foreign language to study in Indonesian schools, most Javanese and Sundanese learners of English face some problems learning English sounds. Given the status of English as a foreign language in the multilingual context of Indonesia it is important to study the production and perception of the Javanese and Sundanese L2 learners. The empirical results of this study can be used to improve strategies to learn English as an L2.

This thesis has addressed the issue of Javanese and Sundanese speakers' perception and production of English sounds. The aim of the current study was to examine how cross-linguistic differences in the vowel systems of Javanese and Sundanese affect the perception and production of English sounds.

An often-discussed model, which is rooted in the production of second language sounds, is Flege's (1995, 1999, 2002) Speech Learning Model (SLM). According to SLM, L2 learners can accurately produce L2 sounds if they have accurate understandings on phonetic distance between L1 and L2 sounds. As a heuristic, within SLM, similar sounds are L2 sounds which are phonetically transcribed with the same base symbol as a sound in the learner's L1, but with at least one difference in diacritics (denoting length, nasalisation, aspiration or some other subtle phonetic difference). New sounds are L2 sounds which have no counterpart in the learner's L1 that is written the same base symbol. Similar sounds are predicted to block the formation of a new phonetic category formation in L2 learning process. Instead, the existing L1 category is extended over the years to include also the L2 tokens, with a shift of the prototype roughly midway between the original prototypes of the L1 and L2, so that after completion of the acquisition process the category is wrong both for the L1 and the L2 – without the learner being aware of this. New L2 sounds, on the other hand, enable the learners over the years to develop new L2 categories, which are authentic in the L2 and do not interfere with the categories in the learner's L1 (Flege, 1995).

Another model on L2 sound learning is Escudero's (2005) Second Language Linguistic Perception model (L2LP). L2LP posits that the acquisition of L2 sounds should match the acoustic properties of L1 sounds (Escudero & Boersma, 2004; Escudero, 2005; Escudero & Williams, 2012). According to L2LP, the L2 sounds to be acquired are categorized into one of three scenarios, i.e. *new*, *subset* and *similar*. The *new* scenario occurs if the L1 perception grammar permits fewer perceptual categories than are required for native perception of the L2. As a result, the L2 environment produces phonological differences that do not exist in the L1 (Escudero, 2005).[4] The subset scenario occurs if the L1 perception grammar outputs more categories than the perception of the L2 requires. Thus, the L2 categories constitute a subset of L1 categories.[5] In the similar scenario, the L1 perception grammar outputs the same number of categories as the target sounds require, because the L1 and L2 categories are phonologically equivalent. This could be called the no (or neutral) transfer scenario.

In Chapter 1, we have introduced two L2 perception models, i.e. SLM and L2LP. SLM predicts that similar L2 sounds (with respect to L1) will always remain relatively difficult to produce and perceive. New L2 sounds, however, will ultimately cease to be perception and/or production problems because, given sufficient time/exposure, a new sound will trigger the formation of a new phonetic category in the L2 without any L1 interference. L2LP predicts that the acquisitions of the new L2 sounds are relatively difficult, and that new sounds are a greater source of difficulty than similar sounds, during the time that the L2 acquisition is in progress. The present thesis tested which of these two models was best supported by the results of the acquisition of the vowels by Javanese and Sundanese L2 learners of English.

We also tested the L2 Feature-dependent hypothesis in L2 speech production, which states that if duration is not used to signal

[4] This scenario is called underdifferentiation of the L1 relative to the L2 in Lado's (1957)'s transfer theory.

[5] This is equivalent to the overdifferentiation scenario in the transfer theory.

phonological contrasts in the learner's L1, it will be difficult to produce (McAllister et al., 2002). We compared this hypothesis with the Desensitization Hypothesis (Bohn, 1995). The Desensitization Hypothesis states that, even if the duration feature is not used contrastively in the L1, it will not be difficult to produce because duration cues are acoustically salient and easily acquired. The present thesis tested whether the Feature-dependent hypothesis vs. Desensitization Hypothesis was supported by our results.

5.2 L1 production of Javanese and Sundanese Speakers

The study in Chapter 2 set out to extend those of Van Zanten and Van Heuven (1984) and Van Zanten (1986), investigating to what extent the Standard Indonesian vowels spoken with a Javanese or Sundanese accent are similar to the Javanese and Sundanese vowels produced by the Javanese and Sundanese speakers.

The results showed that the Javanese female and male speakers produce a high front /i/, a mid front /e/, a low /a/, a mid central /ə/, a high back /u/, and a mid-back /o/. The Javanese schwa /ə/ is considerably higher than that of the Sundanese speakers. The Javanese produced schwa remarkably closed and front.

The results also confirmed that Sundanese female and male speakers produce a front /i/, mid front /e/, low /a/, high central /ɨ/, mid central /ə/, high back /u/, and mid-back /o/. Vowel /ɨ/ is produced in closed central position, only for the male speakers. The Sundanese female speakers produced /ɨ/ with an unclear distinction with the production of schwa.

Based on the descriptive analyses in Chapter 2, it was apparent that the Javanese schwa /ə/ is remarkably close and front. This might have resulted in a problem with pronouncing English words such as *cup, butter* and *but*. Additionally, /a/ appears to be produced in the front part of the mouth by the L2 learners, possibly because their L1 /a/ was relatively (as compared to English) front and open. This thesis discusses these potential pronunciation problems with schwa /ə/ and /a/ in chapter 4.

5.3 L2 Perception of Javanese and Sundanese Speakers

Chapter 3 of the thesis examined how Javanese and Sundanese speakers acquire perception of English (L2) vowels. One aspect analysed in this study was whether the sound category (similar vs. new as compared to the L1 vowel inventory) affects the L2 sound perception by Javanese and Sundanese listeners. Another aspect analysed was whether the phonetic distance between L2 target and distractor influence the L2 learners' task performance. To find out the answers, we carried out a mouse tracking experiment by specifically measuring the Error rate, Area Under the Curve (AUC), initiation time, reaction time and velocity profiles of the participants.

First, we examined whether the familiarity with the L2 sounds (new vs similar) influence the L2 learning. The results of Error rate and AUC showed that listeners had more problems when they perceived new than similar sounds. The L2 learners made more errors when the target L2 vowel was spectrally close to the competitor. Hence, the results support the L2LP model indicating that perceiving new sounds create more problems to L2 learners than similar sounds.

Interestingly, the results of velocity profiles did not produce any clear evidence for either of the L2 learning models regarding the effect of L2 sound familiarity. In the early time window (226 - 750 ms post-stimulus), the Javanese learners showed a trend of high velocity profiles regardless the sound category, while the Sundanese learners showed the higher velocity profiles than the American listeners for the new sounds. In the later time window (751 - 1125 ms post-stimulus), the L2 learners moved their mouse pointer more slowly than the American L1 listeners, for both new and similar target sounds. Altogether, the results of this new behavioral measure could not clearly identify specific perception problems by the L2 learners.

Second, we investigated whether phonetic distance affects the L2 sound leaning. We assume that the L2 learners will find it hard to perceive L2 sounds as the phonetic distance is spectrally close. The results of error rates and AUC showed that the L2 learners showed relatively high mistakes when the acoustic distance between target vowel and competitor is small. Thus, the results confirm that the smaller the acoustic distance is, the hard it is for the L2 learners to discriminate the L2 sounds.

To summarize, our data are in line with the L2LP model suggesting that L2 vowels which are acoustically new in the Javanese and Sundanese native languages may be harder to perceive than similar vowels as the former require learners to create new categories and perceptual mappings and to integrate them with the already categorized sounds.

5.4 L2 production of Javanese and Sundanese Speakers

In this section, we discuss the production of L2 English vowels by Javanese and Sundanese speakers. Models such as SLM and L2LP suggest that vowels which are difficult to perceive are also hard to produce. SLM postulates that similar sounds between the learner's native language L1 and the target language (L2) are more prone to trigger learning difficulties than new sounds because L2 learners will ultimately establish a new phonetic category which is different from any L1 sounds rather than considering it an equivalent of some L1 sound. SLM predicts that the L2 learners may continue to experience difficulties with the similar vowels /iː/ and /uː/ but would learn to be capable of successfully producing the new vowels /ɪ, ɛ, ʊ, æː, ɑː, ɔː, ʌ, and ɝ. On the other hand, L2LP holds that the production of new L2 sounds is more difficult than producing similar L2 sounds. Thus, the Javanese and Sundanese speakers may struggle more with producing the new L2 vowels than the similar L2 vowels.

The F_1 values of the new L2 vowels /ɑː/ and /ɪ/ were considerably lowered by the L2 learners as compared to native speakers. In addition, the F_2 values of the new L2 vowel /ɪ/ were produced significantly higher by the Sundanese speakers as compared to the American speakers. The lowered F_1 values indicate that the L2 learners produced vowels /ɑː/ and /ɪ/ with a more raised tongue. The high F_2 values of vowel /ɪ/ indicate that it is produced more frontally by the Sundanese speakers than by the American speakers. The Javanese speakers did not show any significant differences with the production of English new and similar vowels in F_2 values. Since the new vowels proved to be relatively difficult, these results, at least at first sight, would support L2LP rather than SLM. Again, however, the Indonesian learners of English who served as the participants in this dissertation may still be in the intermediate stage of the L2 acquisition process, so that the possibility that the new sound would ultimately, after the completion of the L2 acquisition, be set up as an authentic category in English (i.e., indistinguishable from the L1 counterpart even as judged by native listeners of the target language) cannot be ruled out.

5.5 Overall Conclusion Regarding L2 Learning Models

Second language learning models such as Speech Learning Model (SLM, Flege, 1987, 1995, 2002, 2003) and the Second Language Linguistic Perception model (L2LP, Escudero, 2005, 2006, 2009) explain L2 acquisition problems based on the similarity of L1 and L2 sounds. According to SLM, L2 sounds, which are similar to L1 sounds, will ultimately remain more difficult than new sounds. L2LP, on the other hand, predicts that new L2 sounds will be more difficult than similar sounds throughout the L2 acquisition process.

Chapter 3 explains that the L2 learners particularly show difficulties perceiving L2 sounds when the target vowel was spectrally close to its L2 competitor vowel. The results, however, additionally showed that perception differences between the native and non-native speakers were largest for new vowels. Chapter 4 indicated that L2 learners have difficulty in producing the similar vowels /iː/ (not for vowel /uː/), but also with new vowels (for Javanese: /ɪ, ɛ, æː, ɝ, ʌ, ɔː, ɑː, ʊ/; for Sundanese: /ɪ, ɛ, æː, ɝ, ʌ, ɑː, ɔː/). Taken together, the results of L2 perception give partial support to both models, but more to L2LP than SLM.

The next question would be if and how the L2 learning models can be supported simultaneously. One possibility is that the L2 learning models focus on different stages of learning. SLM focuses on *experienced* learners. L2LP, in contrast, assesses relatively unexperienced learners, which match with the speakers in the current study. Therefore, relative difficulty with *new* vowels, as observed in the present thesis, was to be expected.

We additionally tested two feature specific hypotheses, i.e. Feature-dependent Hypothesis (FH) by McAllister, Flege, and Piske (2002) and the Desensitization Hypothesis (DH) by Bohn (1995). FH predicts that L2 learners in the present work will have difficulties acquiring duration cues because these are not used contrastively in their L1. DH, in contrast, predicts that the L2 learners would have no dificulties acquiring the duration cues because duration remains relatively easy to access even if it is not used contrastively in the learners L1. In the present work, the L2 learners shortened all target vowels, both long/tense and short/lax. Specifically, they over-shortened short vowels. From this perspective, the results seem to support FH indicating that L2 learners have difficulties in producing duration in a native-like manner because the durational information is not part of their L1 phonology. On the other hand, the results showed that the Indonesian learners produced the relative durations in an almost native-like manner. The L2 durations correlated at better than r = .800 with the American L1 durations, and the long and short vowel categories were quite clearly separated. This would force us to accept DH at least in part. Moreover, there are alternative explanations for the overall (apparent) shortening of the L2 vowels by the Indonesian learners. The American speakers may have used longer (vowel) durations than is normal, since they may have been developed a strategy to speak more slowly to Indonesian listeners in order to be better understood. A slow rate of delivery is listed as a prominent characteristic of this so-called foreigner talk. Moreover, the Indonesian speakers will have pronounced the target vowels followed by a voiceless [t] (in stead of the L1 [d], and failed to apply vowel lengthening before the coda obstruent. The mere fact that the Indonesians adequately differentiated between short lax vowels and long(er) tense vowels is further, strong and positive evidence in favor of DH. There are (weak) indications that the Indonesians habitually speak with very short vowel durations, shorter even that the short vowels in English and Dutch. The vowel durations we measured for the six or seven vowels of Javanese are shorter that what is normally reported. Very short vowel durations were also found by Van Zanten and Van Heuven (1983) in their sample of ten speakers of Standard Indonesian, five of whom were of Javanese origin. On balance, then, our data lend support to both FH and DH, and - certainly in hindsight - it now seems that the two hypotheses are not mutually exclusive. Late L2 learners may remain more sensitive to duration than to other cues, and tune on to the distinctive length effects in the L2. But at the same they may have lost some (but not all) all sensitivity to duration when it is not part of their native-language phonology, and/or they may transfer the habit of producing very short vowel durations from their L1 to the L2.

5.6 Specific Relative Difficulties of the L2 Perception and Production of English Sounds

English vowel perception and production appear to be problematic for Javanese and Sundanese learners of English as the results of the experiments have shown.

The vowel pair /æː - ɛ/ is the most difficult L2 contrast to be perceived by our Indonesian L2 learners of English. Table 5.1 summarizes the production problems faced by our Javanese and Sundanese participants' as compared to American native speakers.

Table 5.1 The summary of the English sounds which are difficult to be produced by the Javanese and Sundanese speakers.

L1	Target vowels	Mispronounced as
Javanese	/iː/	/ɪ/
	/æː/	/ɛ/
	/ɑː/	/ʌ/
	/ɑː/	/ɔː/
Sundanese	/æː/	/ɛ/
	/ɑː/	/ʌ/
	/ɑː/	/ɔː/
	/ʌ/	/æː/
	/ɑː/	/æː/

The results of vowel perception in this study reveal the following relative difficulties of the L2 learners (Chapter 3):

a. Generally, the Javanese and Sundanese learners have relative difficulty perceiving new vowels (relative to the L1 vowel inventory)
b. Phonetic distance affects the L2 sound learning, showing that the Javanese and Sundanese have more perception difficulty as the phonetic distance is closer.
c. The effect sizes of interactions showed that the phonetic distance between the L2 target and competitor vowel is almost an order of magnitude larger than that of sound category (new vs similar sounds).

Acoustic measurements of the English vowels reveal that Javanese and Sundanese learners of English have problems in producing English vowels (Chapters 4). The results reveal the following relative difficulties of the L2 learners:

a. Javanese and Sundanese learners of English produced lower F_1 values than the English speakers.
b. The English vowel durations of Javanese and Sundanese speakers were shortened for both long and short vowels.
c. The Javanese and Sundanese speakers have difficulty contrasting English vowel pairs whose members are adjacent in the articulatory vowel space using spectral parameters while the use of duration is used contrastively, possibly with over-shortening.

5.7 Implications for L2 Teaching

The following recommendations can be made based on the current study with regard to the learning of English as a second language by Javanese and Sundanese university students, and possibly by Indonesian students in general.

The findings confirm that Javanese and Sundanese speakers do not accurately perceive the new L2 vowels /ɑː, ʌ, æː, ɛ, ɪ, ʊ/ and the similar L2 vowels /iː, uː/. It is therefore recommended that English teachers should design vowel identification tasks to help students perceive the differences between all adjacent pairs of the above-mentioned vowels. Laboratory discrimination training on English vowels using the MouseTracker software or other computer-based online training programs, such as the Perception of Spoken English (POSE) test <https://posetest.com/>, may exert a positive influence on speech perception.

As the findings in the current study show that the Javanese and Sundanese speakers have difficulty contrasting intended vowels using spectral parameters, we suggest that the English teaching and learning process for Javanese and Sundanese speakers focus on increasing the F_1 for the vowels /ɑː/, /ɪ/, /æː/, and /iː/. Javanese and Sundanese speakers should be trained to produce these vowels with more openness and a frontal tongue position. Secondly, as we found that the duration of Javanese and Sundanese speakers was shortened for both long and short English vowels, we recommend that Javanese and Sundanese speakers receive phonetic training focusing on lengthening short and long vowels, or perhaps talk more slowly in English than they are used to in their native language. One possibility to achieve these goals would be to use software to let L2 learners produce target vowels, record formant frequencies and durations, and then provide real time feedback to improve pronunciation. Using this approach, we expect that English pronunciation can be significantly improved among Javanese and Sundanese learners of English.

5.8 Limitations

We would like to outline the limitations of this study.

First, the current study focuses on vowels alone, neglecting other features such as consonants or consonant clusters which might compromise the intelligibility of Indonesian speakers of English as much or even more than incorrect pronunciation of the vowels. Thus, we suggest that further studies investigate how consonants or consonant clusters influence second language acquisition. However, it is important to note that the spectral and temporal properties of vowels play a role in speech intelligibility (Walker, 2001). Therefore, this study can be a useful starting point in addressing intelligibility problems in Javanese and Sundanese L2 speech learning.

The second limitation concerns the group of participants. The sample of the present study does not allow generalization to all native speakers of Javanese and Sundanese and all other Indonesian local languages because the characteristics of students from private universities, uneducated adults, and people from other regions may be different. L2 production and perception for other Indonesian local languages will also play a different role than it did in the case of Javanese and Sundanese. Ideally, we would have tested more groups from other Indonesian local languages. In the future, other researchers may include native speakers of other Indonesian local languages to illustrate a broader pattern of English vowel production and perception difficulties.

Third, the relatively small number of L1 participants tested, especially in the L1 production experiment, can be considered one of the limitations of this study. The number of Javanese (4) and Sundanese (4) native speakers in the L1 production test was very small. Testing more speakers of more languages will form a clearer description of the L1 vowel system of Indonesian local languages.

Finally, in this thesis we used a mouse-tracking technique to test L2 perception. It might have been harder for the non-native speakers of English to perform this task than for the native speakers of English. This might explain why the non-native speakers moved the mouse more slowly. It is interesting to see in the mouse-tracking data that the initial time window (226 - 750 ms post-stimulus) was not able to accurately show which specific English vowels are difficult to perceive. This may be due to the participants' lack of familiarity with the use of the computer mouse with which they interacted in the L2 perception experiment. This unfamiliarity may have influenced their mouse movements. We tried to control this issue by including practice trials in order to familiarize participants with the mouse-tracking software. Although we attempted to minimize familiarity effects in this way, we cannot exclude the possibility that the non-native speakers had a harder time to adjust to the task, resulting in non-linguistic differences between the native and non-native groups.

5.9 Future Research

The present dissertation and the studies reported in it cover only a small proportion of topics that may profitably be studied as a potential source of information on how the teaching of English pronunciation and of developing adequate listening comprehension in English could be improved for Indonesian learners of English as a foreign language.

The logical first step towards developing an insight into where the potential learning problems might reside would be to study in depth the perceptual assimilation patterns applied by Indonesian listeners when they are first confronted with the sounds of English. The Indonesian listeners' task would be to indicate for each sound of English (vowels and consonants) which sound in their native language they consider the foreign sound a token of, and how good a token of that native category it is. The results of this procedure will allow us to to assess, among other things, which English vowels map onto which Javanese or Sundanese vowels, and how well they fit the L1 vowel categories. This information, in turn, may be fruitfully used to rank-order the English vowels and vowel contrasts along a scale of difficulty for Indonesian learners of English. Examples of such studies are available in the literature and may serve as a blueprint for the kind of exercise required (e.g. Tsukada et al., 2005 for Korean learners of English; Sun & Van Heuven 2007 for Mandarin learners of English).

The present studies have not considered the intelligibility of Javanese and Sundanese accented English. The next step in the investigation should be to have groups of listeners, native as well as non-native listeners of varieties of English, identify the vowel sounds produced by our speakers. For lack of human native listeners, we have have taken recourse to computer-simulated (native) listeners using Linear Discriminant Analysis to generate a model that arguably approximated a human American native listener. This model can and should be verified using actual human listeners. It will be impossible to subject all 40 speakers (20 Javanese, 20 Sundanese) to such perceptual identification. Instead a smaller number of representative learners can be selected from the larger group, for instance using the individual scores obtained in the LDA as a criterion. Representative speakers would then be the persons closest to the centre of the distribution (see Wang, 2007; Wang & Van Heuven, 2014 as an example of how this can be done).

The perceptual awareness and production of the marked voiced – voiceless contrast in coda obstruents would be a topic that should be studied with high priority. The perception and production of coda consonants, which are largely absent from the phonologies of Indonesian languages are expected to be a major challenge to Indonesian learners of English. It is important to find out whether the Indonesian speakers in the present study were able to produce a properly voiced word-final [d]-sound, accompanied by the lengthening of the preceding vowel, as is usual in native English. It would be worthwhile, for instance, to repeat the study reported in Chapter 4, replacing the target words by counterparts ending in /t/: e.g. *heat*, *hit*, *bet*, *hat*, *Baht* ('Thai currency unit'), *bought*, *put*, *hoot*, *hurt*, *hut*. The results of such a follow-up study would allow us, among other things, to assess whether indeed the difference in vowel duration between the Americans and the Indonesians would be reduced.

A separate project would be to map out the perceptual representation of the vowel system of English in the mind of Indonesian learners. American English has ten monophthongs, which differ in vowel quality and duration. An artificial vowel space can be defined by the acoustic parameters F_1 (10 perceptually equidistant steps, capturing vowel height), F_2 (10 equidistant steps, capturing constriction place and lip rounding) and duration (5 steps). This yields a vowel set of (less than) 500 types, which can be embedded in a foces carrier, and offered to L1 and L2 listeners for perceptual identification and judgment of typicality, following the examples of e.g. Van Zanten and Van Heuven (1984) for Indonesian vowels, Van Heuven (1986) for Dutch vowels or Van Heuven (2017) for English vowels. The results will probably show that vowel quality and duration determine the vowel categories in roughly equal proportion for the L1 listeners of English, but the duration will outweigh the spectral cues for the L2 listeners, in much the same way that was found by Van Heuven (1986) for Dutch vowels as perceived by L1 Dutch and L2 Turkish learners.

As explained in Chapter 1, more than 700 languages are spoken in the Indonesian archipelago. Most of these, including Javanese and Sundanese, belong to the Austronesian language family, and therefore share many structural properties. The research reported in the present thesis has shown that the two learner groups pronounce the English vowels in more or less the same way. This should not come as a surprise considering that the vowel systems of Javanese and Sundanese are virtually the same, and differ only in the presence of a high central vowel in Sundanese, which is absent from the inventory of Javanese and of most other Austronesian vowel systems. A notable exception is Batak, an indigenous Austronesian language spoken by over two million speakers on the isle of Sumatra. Batak has a five-member vowel inventory without any central vowel (Van Zanten & Van Heuven, 1984), and has contrastive stress (Van Zanten & Van Heuven, 1997) instead of either fixed prefinal, variable or perhaps even no word stress at all, as is often claimed for other Austronesian languages (Van Zanten & Van Heuven, 1998, 2004; Goedemans, & Van Zanten 2007; Van Heuven, Roosman, & Van Zanten, 2008; Maskikit-Esed & Gussenhoven, 2016).

The presence of a second (high) central vowel in Sundanese would seem to offer no advantage to a learner of English, since English has no vowel in that part of the vowel-quality space that moght benefit from such a category. The absence of any central vowel, however, may be a source of learning difficulty for Batak learners of English. More generally, the question can be raised whether the teaching of English to speakers of different Austronesian languages should be different depending on the learner's specific L1, or whether a one-size-fits-all approach would be equally effective. As a first approximation to this issue, a study can be done to assess how well speakers of different Austronesian languages are able to determine whether an Indonesian speaker belongs to their own regional language community or not, and in the latter case, if they are able to identify the specific L1 – based on both the way they speak Standard Indonesian (which is a second language to most Indonesians) and how they pronounce English, along the lines sketched by Cui and Van Heuven (2011) for related languages spoken in China, or by Van Heuven and Gooskens (2017) for Scandinavian speakers of English. If the Indonesians are not able to differentiate their own pronunciation of the Standard language or of English from that of other regional groups of Indonesian speakers, there would be no point in developing English teaching materials for different regional learner groups in Indonesia.

The studies proposed here may serve to determine the causes of incorrect production and /or perception of the English vowels by Indonesian learners. They do not address the issue how the problems can be overcome. The pedagogy of the teaching of oral foreign-language skills is still very much in its infancy. As I said elsewhere in this thesis, computer-assisted teaching offers enormous advantages. Perceptual skills at all relevant linguistic levels (sound discrimination, word recognition, global listening comprehension) in the foreign language can be trained with computer feedback, and practically all aspects of foreign-language pronunciation, including the correct use of speech melody, can be practiced with computer-assisted supervision. Future research is needed to determine if the oral skills can be acquired more quickly and effectively if the technological aids zoom in on specific perception and production problems that spring from interference phenomena between Indonesian and English.

References

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.

Cui, R., Heuven, V. J. van (2011). Mutual intelligibility of English vowels by Chinese dialect speakers. In W.-S. Lee, & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, City University of Hong Kong, Hong Kong (pp. 544-547).

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., & Williams, D. (2012). Native dialect influences second-language vowel perception: Peruvian versus Iberian Spanish learners of Dutch. *Journal of the Acoustical Society of America, 131*, EL406–EL412. DOI: 10.1121/1.3701708.

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

Escudero, P. (2006). *Second language phonology: The role of perception*. In M. Pennington (Ed.), *Phonology in context* (pp. 109-134). Houndmills, Basingstoke, Hampshire, New York: Palgrave Macmillan.

Escudero, P. (2006). The phonological and phonetic development of new vowel contrasts in Spanish learners of English. In. B. O. Baptista & M. A. Watkins (Eds.), *English with a Latin beat: Studies in Portuguese/Spanish-English interphonology* (Studies in Bilingualism 31, pp. 149-161). Amsterdam: John Benjamins.

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.

Flege, J. E. (1987). The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalent classification. *Journal of Phonetics, 15*, 47-65.

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. (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 H. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development*, Papers in honor of Henning Wode (pp. 217-44). Trier: Wissenschaftlicher Verlag Trier.

Flege, J. E. (2003). Assesing constraints on second-language segmental production and perception. In Schiller, N. & A. Meyer (Eds.), Phonetics and phonology in language comprehension and production: Differences and similarities. Berlin: De Gruyter.

Goedemans, R. & Van Zanten, E. (2007). Stress and accent in Indonesian. In V. J. van Heuven & E. van Zanten (Eds.), *Prosody in Indonesian languages*. Utrecht: LOT. 35-62.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.

Maskikit-Essed, R., & Gussenhoven, G. (2016). No stress, no pitch accent, no prosodic focus: The case of Ambonese Malay, *Phonology*, *33*, 353-389.

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.

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.

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.

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 – 845). Berkeley: The Dutch Studies Program, U.C. Berkeley.

Van Heuven, V. J. (2017). Perception of English and Dutch checked vowels by early and late bilinguals. Towards a new measure of language dominance. In: Simone E. Pfenninger & Judit Navracsics (eds.) *Future research directions for Applied Linguistics* (Second Language Acquisition 109). Bristol, Buffalo, Toronto: Multilingual Matters (pp. 73-98).

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. Festschrift in honour of John Nerbonne,* no.32 London: College Publications (pp. 137-147).

Van Heuven, V. J., & Van Zanten, E. (1997). Effects of substrate language on the localization and perceptual evaluation of pitch movements in Indonesian. In *Proceedings of the 7th International Conference on Austronesian Linguistics*, Rodopi, Amsterdam/Atlanta, (pp. 63-80).

Van Heuven, V. J., Roosman, L.M., & Van Zanten. (2008). Betawi Malay word prosody. *Lingua, an International Review of General Linguistics, 118*(9), 1271-1287.

Van Zanten, E. (1986). Allophonic variation in the production of Indonesian vowels. *Bijdragen tot de Taal-, Land- en Volkenkunde, 142*, 427-446.

Van Zanten, E., & Van Heuven, V. J. (2004). Word stress in Indonesian: fixed or free? *NUSA, Linguistic Studies on Indonesian and other Languages in Indonesia, 48*, 1-18.

Van Zanten, E., & Van Heuven, V. J. (1983). A phonetic analysis of the Indonesian vowel system, a preliminary acoustic study. *NUSA, Linguistic Studies of Indonesian and other Languages in Indonesia*, *15*, 70-80.

Van Zanten, E. & Van Heuven, V. J. (1984) The Indonesian vowels as pronounced and perceived by Toba Batak, Sundanese and Javanese speakers. *Bijdragen tot de Taal-, Land- en Volkenkunde, 140*, 497-521.

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

Walker, R. (2001). Pronunciation for international intelligibility. Karen's linguistics issues: Free resources for teacher and students of English. *English Teaching Professional Magazine, 2*.

Wang, H., & Van Heuven, V. J. (2014). Is a shared interlanguage beneficial? Mutual intelligibility of American, Dutch and Mandarin speakers of English. In: Doel R. van den, Rupp L. (Eds.), *Pronunciation Matters. Accents of English in The Netherlands and elsewhere*. Amsterdam: VU University Press (pp. 175-194).

Summary

Second language (L2) learners often face difficulties while learning L2 sounds. Evidence suggests that difficulties in learning L2 sounds are affected by the first language. The interference by the first language could have a substantial impact on the production and perception of sounds for L2 learners. Many studies in L2 production and perception have performed cross-linguistic comparisons between English as a second language and western European languages such as Spanish, German, Dutch, Swedish, Danish, and Italian. So far, few studies have investigated Austronesian languages, such as Javanese and Sundanese—two widely spoken languages in Indonesia. Given that English has become a foreign language that is officially studied in schools in Indonesia, it is worth investigating whether Javanese and Sundanese learners of English show systematic problems in learning English, and if so, if the pattern of acquisition problems can be explained by interference by their native language (L1). Therefore, the current study investigates the native vowels of Javanese and Sundanese as well as the perception and production of English vowels by Javanese and Sundanese speakers.

Chapter 2 presents a description of these languages' sounds by examining the vowel quality and vowel duration of Javanese and Sundanese as produced by native speakers. Four native speakers of Javanese produced six Javanese vowels and four Sundanese native speakers produced seven Sundanese vowels. This chapter showed that Javanese speakers produced schwa considerably higher than that of the Sundanese speakers. For the Javanese speakers, schwa is produced higher than /e/ and /o/. For the Sundanese speakers, /ɨ/ is produced in closed central position, only for the male speakers. The formant frequencies of the Javanese and Sundanese vowels agree with the description of the vowel system by Wedhawati et al. (2006), Van Zanten and Van Heuven (1984), Crothers (1978), and Kurniawan (2013). The chapter also showed that durations of Javanese and Sundanese are phonetically short (between 60 and 100 ms) for all vowels.

Chapter 3 sets out to investigate the perception of ten English vowels among the L2 learners. This chapter examines (1) whether the sound category (new vs similar vowels) affects the L2 sound perception based on L2 learning models and (2) whether the phonetic distance between target and distractor sounds influences the L2 sound perception. Thirty Javanese, thirty Sundanese, and twenty English native

speakers participated in a mouse-tracking experiment. Participants were required to identify English vowels corresponding to an auditory token by clicking on one of two word strings presented on a computer screen. This chapter showed that phonetic distance between target and distractor plays a more important role in the sound perception of the Javanese and Sundanese listeners than the sound category of the target itself. The findings partially support the L2LP model indicating that new vowels are more problematic to be perceived by the L2 learners than similar vowels.

In **Chapter 4**, the current thesis investigated how the Javanese and Sundanese speakers produce ten English vowels. This chapter examines the formant frequencies and the duration of English vowels produced by the L2 learners. Forty Javanese and Sundanese speakers and ten native English speakers participated in the experiment. Speech Learning Model (SLM) predicts that highly advanced Javanese and Sundanese speakers of English should 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, ɝ/. 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. This chapter shows that the Javanese and Sundanese speakers have more difficulty with the new than with the similar vowels in English, indicating that the L2 acquisition process has not been completed. In addition, the Javanese and Sundanese speakers poorly contrasted the members of English tense-lax vowel pairs using spectral parameters while the use of duration is relatively sufficient.

The results of the present thesis are not only theoretically relevant but also have implications for teaching English as a second language among Javanese and Sundanese learners. First, the L2 speakers do not accurately perceive the *new* L2 vowels /ɑː, ʌ, æː, ɛ, ɪ, ʊ/ and the *similar* L2 vowels /iː, uː/. Hence, it is recommended that teachers of English design vowel identification tasks to familiarize Sundanese and Javanese students with and improve their identification of the above-mentioned vowels. Second, as L2 speakers have difficulty contrasting intended vowels using spectral parameters, we suggest that Javanese and Sundanese learners of English should focus on increasing the F_1 for the vowels /ɑː/, /ɪ/, /æː/, and /iː/. The training should emphasize producing these vowels with more openness and a frontal tongue position. Moreover, as the duration of Javanese and Sundanese speakers was shortened for both long and short English vowels, we recommend that Javanese and Sundanese speakers receive phonetic training focusing on lengthening short and long vowels. These approaches are expected to improve the perception and production of English vowels among the Javanese and Sundanese speakers.

Samenvatting

Leerlingen die een tweede taal (L2) leren, ondervinden een aantal moeilijkheden bij het leren van klanken. Onderzoek wijst uit dat deze moeilijkheden te wijten zijn aan hun moedertaal (L1). De invloed van de moedertaal heeft invloed op het vermogen van het produceren en begrijpen van klanken van de tweede taal. Een aantal L2-studies welke gebaseerd zijn op het produceren en op het interpreteren van een tweede taal, gebruiken een cross-linguïstische vergelijking tussen het Engels als tweede taal en West-Europese talen zoals: Spaans, Duits, Nederlands, Zweeds, Deens en Italiaans. Tot nog toe zijn er weinig studies die onderzoek hebben gedaan naar Austronesische talen, zoals het Javaans en Sundanees; twee talen die veel gesproken worden in Indonesië. Aangezien het Engels een vreemde taal is die officieel gestudeerd wordt op scholen in Indonesië, is het de moeite waard om te onderzoeken of Javaanse en Sundanese leerlingen van het Engels systematisch moeite ondervinden bij het leren van het Engels en als dat het geval is of het patroon van leermoeilijkheden bij het verwerven van deze taal verklaard kan worden door de inmenging van hun moedertaal.

Hoofdstuk 2 introduceert een beschrijving van de klanken van deze talen door te kijken naar de kwaliteit van de klinkers en de lengte daarvan bij Javaanse en Sundanese moedertaal sprekers. Vier Javaanse moedertaal sprekers produceren zes Javaanse klinkers en vier Sundanese sprekers produceren zeven Sundanese klinkers. Dit hoofdstuk laat zien dat Javaanse sprekers sjwa aanzienlijk hoger uitspreken dan Sundanese sprekers. Bij de Javaanse sprekers klinkt sjwa hoger dan bij /e/ en /o/. Bij de Sundanese sprekers wordt /ɨ/ alleen door mannelijke sprekers kort en gesloten uitgesproken. De formanten van Javaanse en Sundanese klinkers komen overeen met de beschrijving van het klinkersysteem door Wedhawati et al. (2006), Van Zanten en Van Heuven (1984), Crothers (1978) en Kurniawan (2013). Dit hoofdstuk laat tevens ziens dat de lengte van alle Javaanse en Sundanese klinkers kort is (tussen 60 en 100ms).

Hoofdstuk 3 onderzoekt de interpretatie van tien Engelse klinkers bij Javaanse en Sundanese sprekers (L2-leerlingen). In dit hoofdstuk wordt ten eerste onderzocht of de klankcategorie, ‘*nieuwe*’- (klinkers die niet in de modertaal voorkomen) versus *‘soortgelijke’* (L2 klinkers die lijken op L1 klinkers) klinkers de klankperceptie van L2-leerlingen beïnvloedt gebaseerd op L2-leermodellen. Ten tweede wordt er onderzocht of de fonetische afstand tussen de doel- en afleider geluiden de L2-klankperceptie beïnvloeden. Dertig Javaanse, dertig Sundanese en twintig Engelse sprekers hebben meegedaan aan een onderzoek waar muisbewegingen worden bestudeerd. Deelnemers moesten Engelse klinkers identificeren die overeenkwamen met een doelgeluid door te klikken op een van de twee woord alternatieven die gepresenteerd werden op een computerscherm. Dit hoofdstuk liet zien dat de fonetische afstand tussen doel- en stoorgeluiden een grotere rol speelt in de geluidsperceptie bij Javaanse en Sundanese luisteraars dan bij andere categorieën. Deze bevindingen ondersteunen het *Second Language Linguistic Perception* (L2LP)-model gedeeltelijk en suggereren dat nieuwe klinkers moeilijker zijn aan te leren bij de L2-leerlingen dan gelijksoortige klinkers.

In **hoofdstuk 4** wordt onderzocht hoe de L2-leerlingen tien Engelse klinkers uitspreken. Dit werd onderzocht door de formanten en de lengte van de Engelse klinkers zoals die door L2-leerlingen worden uitgesproken te meten. Veertig Javaanse en Sundanese sprekers en tien Engelse sprekers deden mee aan het experiment. Het *Speech Learning Model* (SLM) voorspelt dat gevorderde Javaanse en Sundanese sprekers van het Engels moeite zouden moeten hebben om klinkers uit te spreken die lijken op /iː/ en /uː/ (‘*soortgelijke’* klinkers) maar dat zij minder moeite zouden moeten hebben met *nieuwe* L2-klinkers, zoals: /ɪ, ɛ, ʊ, æː, ɑː, ɔː, ʌ, ɝ/. Daarentegen voorspelt het L2LP model dat het uitspreken van nieuwe L2-klinkers moeilijker is dan dat voor gelijksoortige L2-klinkers is, zolang het proces van het verwerven van de L2-taal nog niet voltooid is. Dit hoofdstuk toont aan dat Javaanse en Sundanese sprekers meer moeite hebben met nieuwe dan met gelijksoortige klinkers in het Engels, wat suggereeert dat het verwervingsproces van de L2-taal nog niet voltooid is. Bovendien presteerden de L2-leerlingensprekers slecht bij het gebruik van het Engels met betrekking tot gesloten en open klinkers in paren bij spectrale parameters, terwijl de tijdsduur redelijk voldoende was.

De uitkomst van deze dissertatie is niet alleen theoretisch van belang, maar heeft ook implicaties voor het leren van het Engels als tweede taal bij Javaanse en Sundanese L2-leerlingen. Ten eerste, deze L2-sprekers zullen de nieuwe L2-klinkers zoals: /ɑː, ʌ, æː, ɛ, ɪ, ʊ/ niet accuraat waarnemen, net zoals de gelijksoortige L2-klinkers /iː, uː/. Daarom is het raadzaam dat docenten Engels oefeningen ontwerpen om klinkers te kunnen identificeren, zodat L2-leerlingen vertrouwd kunnen raken met deze klinkers en de identificatie van de bovengenoemde klinkers kan worden verbeterd. Ten tweede, aangezien de L2-leerlingen moeite hebben met het onderscheid van klinkers bij het gebruik van spectrale parameters raden wij aan dat L2-leerlingen leerlingen van de Engelse taal zich oefenen met het meer open en frontaal uitspreken van de klinkers /ɑː/, /ɪ/, /æː/, en /iː/. Bovendien, omdat de L2-leerlingen klinkers relatief kort uitspreken, raden wij aan dat L2-leerlingen fonetische training krijgen die erop gericht is om korte en lange klinkers allebei te verlengen. Verwacht wordt dat deze benaderingen de waarneming en het uitspreken van Engelse klinkers door Javaanse en Sundase sprekers zal verbeteren.

Curriculum Vitae

Arum Perwitasari was born on January 30, 1985 in Dabo Singkep, Indonesia. She received a Bachelor of Arts in English Literature and Teaching at the College of Foreign Languages at the Indonesian-American Association in Yogyakarta, Indonesia in 2006, and enrolled as a Master of Arts student at Gadjah Mada University, Indonesia, in 2010 with DIKTI Graduate Scholarship Programme, where she graduated in 2012. From 2008, she was employed as an Associate Lecturer of English at the College of Aviation Technology (Sekolah Tinggi Teknologi Kedirgantaraan) in Yogyakarta, Indonesia and worked as a part-time English teacher at Gadjah Mada University Language Learning Centre. In 2013, she received another DIKTI Graduate Scholarship Programme in order to pursue a Doctoral degree and started as a PhD student at the Leiden University Centre for Linguistics (LUCL). She was invited to be a guest researcher at the Leiden Ethnosystem and Development Programme (LEAD), Faculty of Science, in 2015. In parallel to her PhD study, she started working as an Academic Relations Assistant at Educational Testing Service (ETS) Global Amsterdam in May 2017. ETS Global is a wholly owned subsidiary of non-for-profit ETS which develops and administers international standardized tests including *TOEFL®*, *GRE®* and *TOEIC®*.

Appendices

Appendix 1. Letter of Consent- Speech Production Experiment

CONSENT FORM
SPEECH PRODUCTION EXPERIMENT

Title of Research project:
THE ACQUISITION OF ENGLISH VOWELS BYJAVANESE AND SUNDANESE NATIVE SPEAKERS

Name of Researcher: Arum Perwitasari, Leiden University
Supervisors/promotors:
Prof. dr. Niels Schiller, Leiden University
Prof. dr. Marian Klamer, Leiden University
Dr. Jurriaan Witteman, Leiden University

Type of research
This project comprises an attempt to a fundamental research in the production and perception of English vowels by the Javanese and Sundanese speakers.

Participation of informants
If you are a native speaker of Javanese or Sundanese, you will be asked to produce and record a number of data sets consisting of English words. You have to read a given sentence.

Data collection usage
The data collection of speech production of native speakers of Javanese and Sundanese learning English, will be broken down using acoustic analysis. They are to contribute at answering the question on how native speakers of Standard Javanese and Sundanese learn to produce English vowels and also to glimpse whether the production of L2 vowels exploits duration similar to L1 pronunciation or whether it shows slightly different vowel duration.

Risks
There are no known hazards or risks involved in participation in this experiment.

Confidentiality
Your information in this research will remain confidential. The data will be described, analyzed and presented in a complete anonymity.

Consent
You will receive compensation for your participation. Should you change your mind and wish to withdraw yourself from the research at any time, you would be pleased to contact the researcher. You are free to withdraw from participation at any time with no need for explanation.

If you have any questions you can call Arum Perwitasari at +62-815-687-0173 or a.perwitasari@hum.leidenuniv.nl

INFORMED CONSENT FORM
SPEECH PRODUCTION EXPERIMENT

I, ______________________, agree to take part in this research project titled The Acquisition of English Vowels by Javanese and Sundanese Native Speakers. I understand that the study annotates a tape-recorded phonation. I understand that my voice will be analysed by speech analysis software.

The researcher has told me that the purpose of the study is to examine the vowel accuracy of L2 long-short vowels production. I may not receive any direct benefit from taking part in this study, but the study may help to increase knowledge that may help others in the future.

The researcher has promised that all information and the names of all informants will be kept confidential to the extent permitted by law. I understand that I do not have to take part in this research and may also withdraw myself at any time.

I have read and understand the above information and agree to take part in this study. I would contact Arum Perwitasari at +62-815-687-0173 or a.perwitasari@hum.leidenuniv.nl once I have any questions I have about the research.

Date : _______________________________
Sign : _______________________________
Venue : _______________________________

Appendix 2. Letter of Consent- Speech Perception Experiment

CONSENT FORM
SPEECH PERCEPTION EXPERIMENT

Title of Research project:
THE ACQUISITION OF ENGLISH VOWELS BYJAVANESE AND SUNDANESE NATIVE SPEAKERS

Name of Researcher: Arum Perwitasari, Leiden University
Supervisors/promotors:
Prof. dr. Niels Schiller, Leiden University
Prof. dr. Marian Klamer, Leiden University
Dr. Jurriaan Witteman, Leiden University

Type of research
This project comprises an attempt to a fundamental research in the speech of Second Language Learners (SLL).

Participation of informants
If you are a native speaker of Javanese or Sundanese, you will be asked to identify these stimuli. In this experiment, you are asked to give you response as accurately and quickly as possible.

Data collection usage
The data collection of speech perception experiment of Javanese and Sudanese learners of English, will be used to identify how Second Language Learners perceive and recognize words in a given sentences. Your hand movement through computer mouse is recorder to get a clear data of your response time and trajectories. There is no known hazard involved in this experiment.

Risks
There are no known hazards or risks involved in participation in this experiment.

Confidentiality
Your information in this research will remain confidential. The data will be described, analyzed and presented in a complete anonymity.

Consent
You will receive compensation for your participation. Should you change your mind and wish to withdraw yourself from the research at any time, you would be pleased to contact the researcher. You are free to withdraw from participation at any time with no need for explanation.

If you have any questions you can call Arum Perwitasari at +62-815-687-0173 or a.perwitasari@hum.leidenuniv.nl.

INFORMED CONSENT FORM
SPEECH PERCEPTION EXPERIMENT

I, ______________________, agree to take part in this research project titled English Vowel Production and Perception of Javanese and Sundanese Learners. I understand that the study record my hand movement through computer mouse track. I understand that my response time will be analyzed by mouse tracking software.

The researcher has told me that the purpose of the study is to analyze the error rate, response time and trajectory movement of accuracy. I may not receive any direct benefit from taking part in this study, but the study may help to increase knowledge that may help others in the future.

The researcher has promised that all information and the names of all informants will be kept confidential to the extent permitted by law. I understand that I do not have to take part in this research and may also withdraw myself at any time.

I have read and understand the above information and agree to take part in this study. I would contact Arum Perwitasari at +62-815-687-0173 or a.perwitasari@hum.leidenuniv.nl once I have any questions I have about the research.

Date : ________________________________
Sign : ________________________________
Venue : ________________________________

Appendix 3. Participant Questionnaire

KUESIONER PESERTA

Terimakasih atas kesediaan Saudara untuk bergabung dalam penelitian tentang *English Vowel Production of Javanese and Sundanese*. Mohon mengisi formulir dibawah ini.

Instruksi:
Silahkan jawab pertanyaan di dalam formulir ini secara lengkap dan akurat. Seluruh informasi yang anda tulis di dalam kuesioner ini akan dijaga kerahasiaannya. Tidak ada pernyataan salah ataupun benar untuk jawaban yang anda berikan. Anda harus membaca dan mengisi form kesediaan sebelum menjawab pertanyaan pada kuesioner ini. Tanggal hari ini:___________________.

A. Informasi Demografi Peserta
1. Nama/ Inisial:
2. Tanggal lahir:
3. Kota dan provinsi tempat anda lahir:
4. Jenis kelamin:
5. Kebangsaan :
6. Kota tempat anda tinggal:
7. Kota tempat anda menghabiskan masa kecil anda:
8. Bahasa pertama atau Bahasa ibu anda beserta dialeknya:
9. Bahasa yang anda gunakan di rumah:
10. Bahasa asing yang anda kuasai:
11. Informasi tentang orang tua anda.

	Ibu	Ayah	Nenek	Kakek
Tanggal lahir				
Kota dan provinsi tempat lahir				
Kebangsaan				
Kota tempat mereka menghabiskan masa kecil mereka				
Bahasa pertama atau Bahasa ibu				
Bahasa lain yang dikuasai				

B. Frekuensi Penggunaan Bahasa Ibu

12. Bahasa apakah yang anda gunakan di rumah dimasa kecil anda?
13. Adakah Bahasa lain yang digunakan di rumah saat itu? Jika ada, mohon sebutkan.
    a.
    b.
    c.
    d.
14. Seberapa sering anda menggunakan Bahasa ibu anda dimasa kecil anda? Mohon lingkari.
    a. Kadang-kadang
    b. Sekali seminggu
    c. Banyak kali seminggu
    d. Sekali sehari
    e. Setiap saat
15. Seberapa sering anda menggunakan Bahasa ibu saat ini? Please circle.
    a. Kadang-kadang
    b. Sekali seminggu
    c. Banyak kali seminggu
    d. Sekali sehari
    e. Setiap saat
16. Tataran Bahasa* apa yang anda gunakan kepada:
    a. Ibu:
    b. Ayah:
    c. Kakek/nenek:
    d. Saudara:
    e. Teman:

(*Bahasa Jawa: *Krama*, *Madya*, *Ngoko*; Bahasa Sunda: *Basa Lemes*, *Basa Loma*, *Basa Kasar*)

C. Latar Belakang Bahasa Inggris

17. Bagaimana Anda mengklasifikasikan input bahasa Inggris Anda? Mohon lingkari.
    a. British English
    b. American English
    c. Lainnya, ____________________
18. Pada umur berapa anda mulai belajar Bahasa Inggris?
19. Sudah berapa tahun anda belajar bahasa Inggris di sekolah?

20. Berapa jam rata-rata anda mendapatkan Bahasa Inggris di sekolah?
21. Bahasa asing apa yang pernah anda pelajari di sekolah, selain Bahasa Inggris?
22. Pernahkah anda menghabiskan waktu anda di Negara yang berbahasa Inggris? Jika pernah, berapa lama dan dimana? Sebutkan pula waktu da negaranya.
23. Pernahkah anda menerjemahkan dokumen dari dan atau ke dalam Bahasa Inggris? Berapa dokumen? Sebutkan jumlah halamannya.
24. Pernakah anda mengikuti ujian TOEFL? Kapan? Berapa skor anda?
25. Apakah anda menguasai Bahasa asing lain, selain Bahasa Inggris? Sebutkan.
26. Apakah anda memiliki masalah pada pendengaran anda?
    a. Yes   b. No
27. Apa kecenderungan kecakapan penggunaan tangan anda?
    a. Kanan
    b. Kiri atau Kidal
    c. Keduanya

Appendix 4. Native English Questionnaire

NATIVE ENGLISH QUESTIONNAIRE

Thank you for participating in my survey about English vowels production of Javanese and Sundanese learners. You need to fill in the data to provide background information for the study.

Instructions:
Please fill the form as completely and accurately as possible. All information from this questionnaire will be kept strictly confidential. There will be no right or wrong answer. You need to read and sign consent form beforehand. Today's date ____________________.

1. Name/ Initial:
2. Date of birth:
3. City and province of birth:
4. Sex:
5. Nationality:
6. What is your native language? Dialect?
7. What are languages that you have studied?

Language studied	Years of studied	Formal/ Informal Education

8. Do you master other languages in a native-like manner? Or Are you a bilingual speaker? Of what language?
9. Where did you stay in your home country? How long?

Country	Place or City	Years of residence

10. Did spend your life in any other country? How long?

Country	Place or City	Years of residence

11. Do you have any speech or hearing disorders? Please circle.
    a. Yes
    b. No

Appendix 5. A script of post-experiment

POST EXPERIMENT QUESTIONS

I am interested in your reactions to the stimuli that you listened to during the experiment today. Please answer the following questions as best you can. Your responses to these questions will be very helpful to us in our research. There will be NO WRONG answers! Please respond in the space provided. Be sure to write clearly so that we can read your answers. Thank you.

1) What were the stimuli that you heard?
2) Were there any new words to you? If yes, what were they?

3) Did you have problem pronouncing/ perceiving these stimuli? If yes, what words create difficulty to you?

Thank you for your assistance in this research. I do appreciate your support!

Name : ______________________________
Date : ______________________________
Venue : ______________________________