Author: Perwitasari, A.

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

Issue Date: 2019-02-19

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.

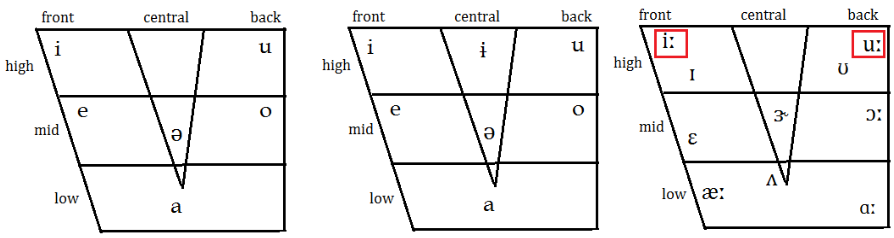


Figure 3.1 The monophthongs of Javanese (left panel), Sundanese (central panel) and English (right panel). The vowels in the squares are *similar* vowels. The remaining English vowels are *new* vowels.

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.

<table>
<tr><td rowspan="2"></td><td colspan="6">Place of Constriction</td></tr>
<tr><td colspan="2">Front</td><td colspan="2">Central</td><td colspan="2">Back</td></tr>
<tr><td colspan="7">Native Language: Javanese</td></tr>
<tr><td>high</td><td colspan="2">i</td><td colspan="2"></td><td colspan="2">u</td></tr>
<tr><td>high-mid</td><td colspan="2">e</td><td colspan="2">ə</td><td colspan="2">o</td></tr>
<tr><td>low-mid</td><td colspan="2"></td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td>low</td><td colspan="2"></td><td colspan="2">a</td><td colspan="2"></td></tr>
<tr><td colspan="7">Native Language: Sundanese</td></tr>
<tr><td>high</td><td colspan="2">i</td><td colspan="2">ɨ</td><td colspan="2">u</td></tr>
<tr><td>high-mid</td><td colspan="2">e</td><td colspan="2">ə</td><td colspan="2">o</td></tr>
<tr><td>low-mid</td><td colspan="2"></td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td>low</td><td colspan="2"></td><td colspan="2">a</td><td colspan="2"></td></tr>
<tr><td colspan="7">Target Language: English</td></tr>
<tr><td></td><td>Tense</td><td>Lax</td><td>Tense</td><td>Lax</td><td>Tense</td><td>Lax</td></tr>
<tr><td>high</td><td>i:</td><td></td><td></td><td></td><td>uː</td><td></td></tr>
<tr><td>high mid</td><td></td><td>ɪ</td><td>ɝ</td><td></td><td></td><td>ʊ</td></tr>
<tr><td>low mid</td><td></td><td>ɛ</td><td></td><td></td><td>ɔː</td><td></td></tr>
<tr><td>low</td><td>æː</td><td></td><td></td><td>ʌ</td><td>ɑː</td><td></td></tr>
</table>

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

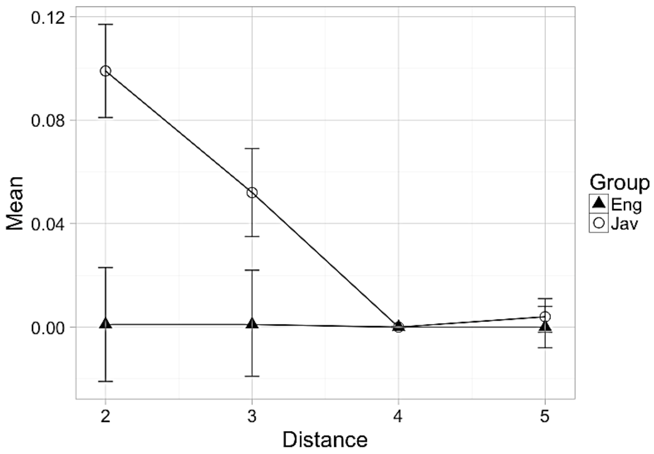
There was an interaction of Distance × Group [$F(1.44, 69.32) = 14.55$, $p < .001$, $\eta_p^2 = .23$], and Distance × Category × Group [$F(1.58, 75.67) = 3.89$, $p < .05$, $\eta_p^2 = .08$]. Last, there was a between-subject effect for Group [$F(1, 48) = 23.98$, $p < .001$, $\eta_p^2 = .33$]. Pairwise 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: Javanese (N = 30) $\bar{x}$	Javanese (N = 30) SD	English (N = 20) $\bar{x}$	English (N = 20) SD	*p*
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

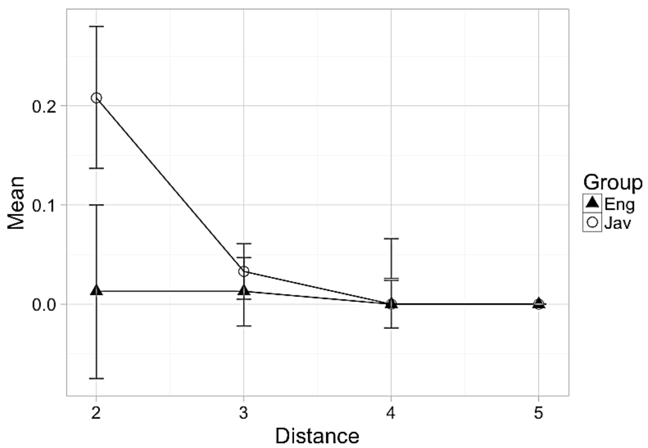


Figure 3.2a-b Mean Error rates (proportion) of new (panel a) and similar (panel b) vowel category for each level of phonetic distance performed by Javanese listeners, 95% confidence intervals for Distance levels 2-5.

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

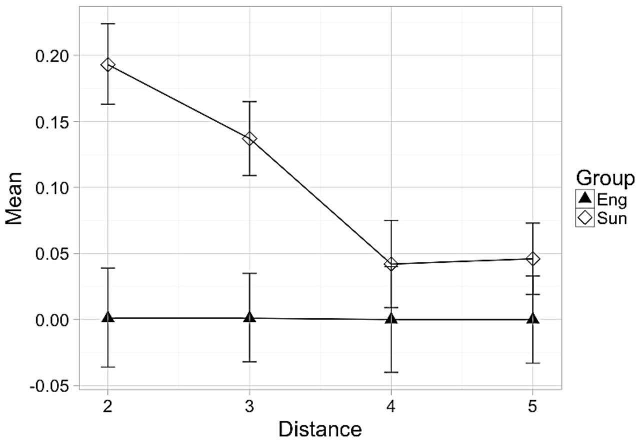
There was an interaction of Distance × Group [$F(2.04, 98.19) = 40.06$, $p < .001$, $\eta_p^2 = .46$], Category × Group [$F(1, 48) = 7.51$, $p < .05$, $\eta_p^2 = .14$], Distance × Category × Group [$F(1.20, 95.77) = 13.26$, $p < .001$, $\eta_p^2 = .22$]. There was a between-subject effect for Group [$F(1, 48) = 55.32$, $p < .001$, $\eta_p^2 = .53$]. 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 Sundanese (N = 30)		English (N = 20)		*p*
		$\bar{x}$	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

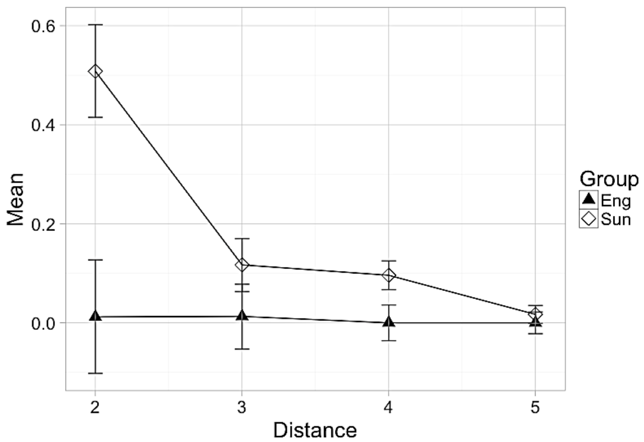


Figure 3.3a-b Mean Error rates (proportion) of new (panel a) and similar (panel b) vowel category for each level of phonetic distance performed by Javanese listeners, 95% confidence intervals for distance levels 2-5.

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

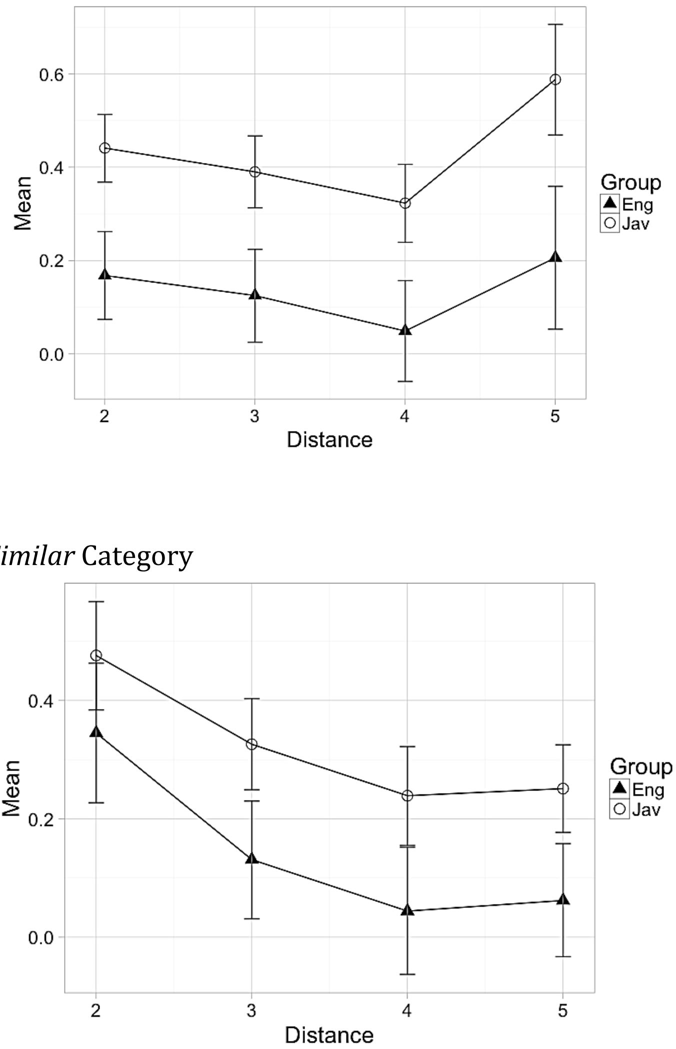


Figure 3.4a-b Area Under the Curve (mean AUC) of new (panel a) and similar (panel b) vowel category for each level of phonetic distance obtained by Javanese and American listeners, 95% confidence intervals for Distance levels 2-5.

3.5.2.2 Sundanese vs. English Listeners

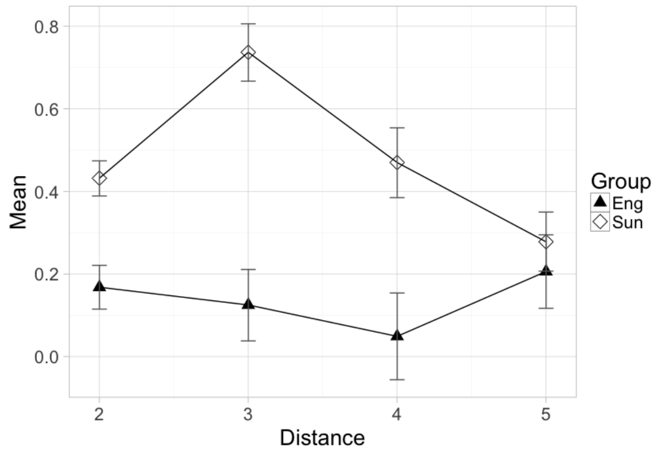
We excluded 2 outputs from the English group and 2 outputs from the Sundanese group in this analysis. There was an interaction of Distance × Group [$F(2.60, 114.27) = 8.57$, $p < .001$, $\eta_p^2 = .16$], Distance × Category × Group [$F(2.34, 103.05) = 24.80$, $p < .001$, $\eta_p^2 = .36$]. There was 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)				p
		Sundanese (N = 30)		English (N = 18)		
		$\bar{x}$	SD	$\bar{x}$	SD	
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

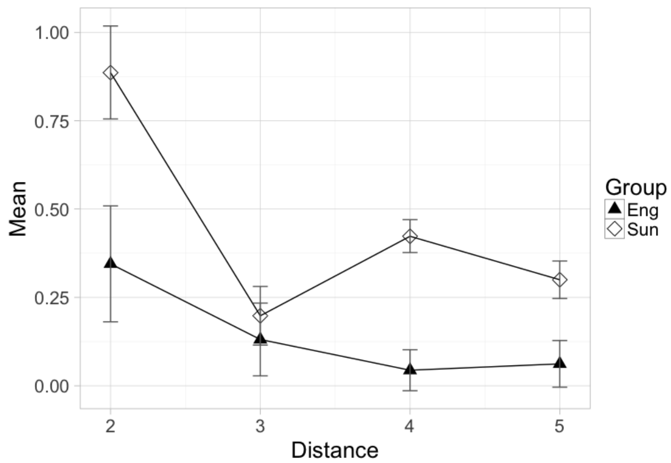


Figure 3.5a-b Area Under the Curve (mean AUC) of new (panel a) and similar (panel b) vowel category for each level of phonetic distance performed by Sundanese and American listeners, 95% confidence intervals for Distance levels 2-5.

3.5.3 Initiation Time

3.5.3.1 Javanese vs. English Listeners

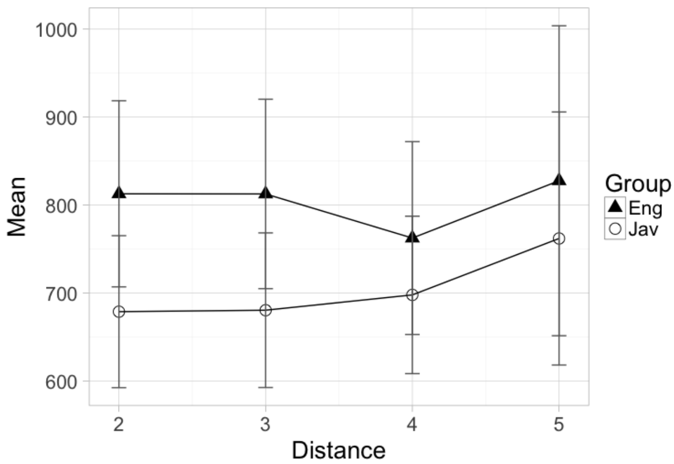
There was an interaction of Distance × Group [$F(2.13, 102.22) = 9.48$, $p < .001$, $\eta_p^2 = .17$], Distance × Category × Group [$F(2.52, 121.38) = 14.7$, $p < .001$, $\eta_p^2 = .23$]. There were no other main effects or interactions. Pairwise comparisons are presented in Table 3.8.

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

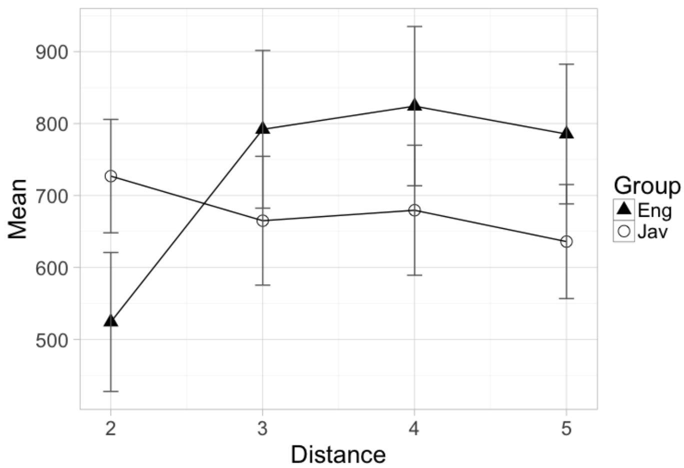


Figure 3.6a-b Mean Initiation Time (ms) obtained for new (panel a) and similar (panel b) vowel category for each level of phonetic distance from Javanese L2 and American L1 listeners, 95% confidence interval for Distance levels 2-5.

3.5.3.2 Sundanese vs. English listeners

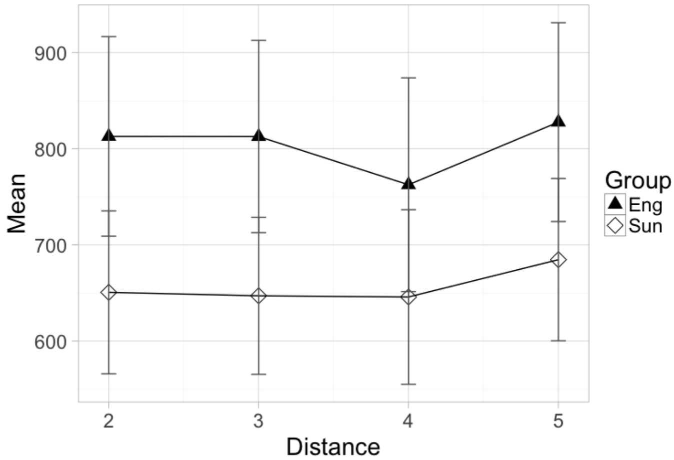
We excluded two listeners from the American group and two listeners from the Sundanese group in this analysis. There was an interaction of Distance × Group [$F(2.16, 103.74) = 12.21$, $p < .001$, $\eta_p^2 = .20$], Category × Group [$F(1, 48) = 7.12$, $p < .05$, $\eta_p^2 = .13$], Distance × Category × 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

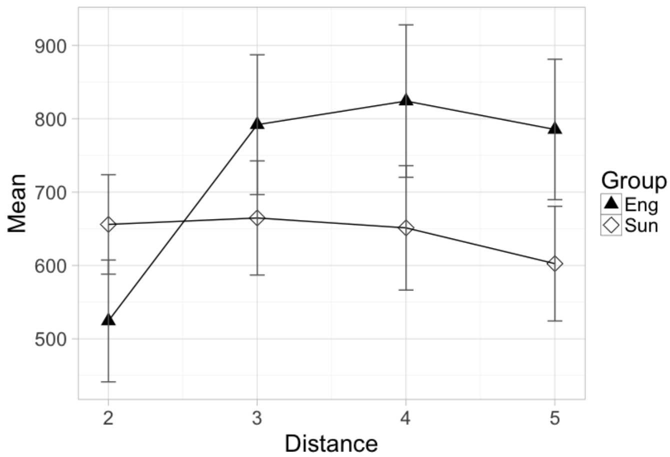


Figure 3.7a-b Mean Initiation Times (ms) of new (panel a) and similar (panel b) vowel category for each level of phonetic distance obtained from Sundanese and American listeners, 95% confidence interval for Distance levels 2-5.

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.

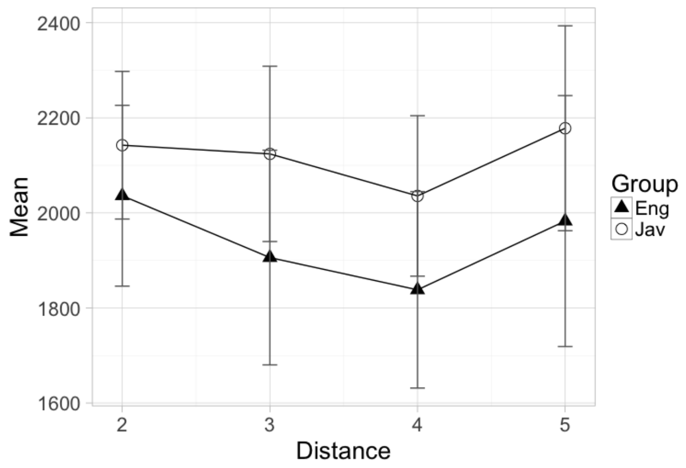
There was an interaction of Distance × Category × Group [$F(1.99, 95.95) = 3.89$, $p < .05$, $\eta_p^2 = .08$]. There were no other main effects or interactions. Mean and standard deviations are presented in Table 3.10.

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

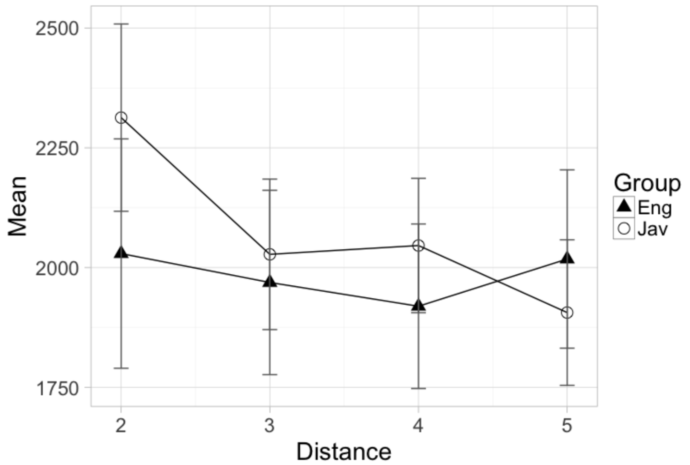


Figure 3.8a-b Mean Reaction Times (ms) of new (panel a) and similar (panel b) vowel category for each level of phonetic distance performed by Javanese and American listeners, 95% confidence interval for Distance levels 2-5.

3.5.4.2 Sundanese vs. English Listeners

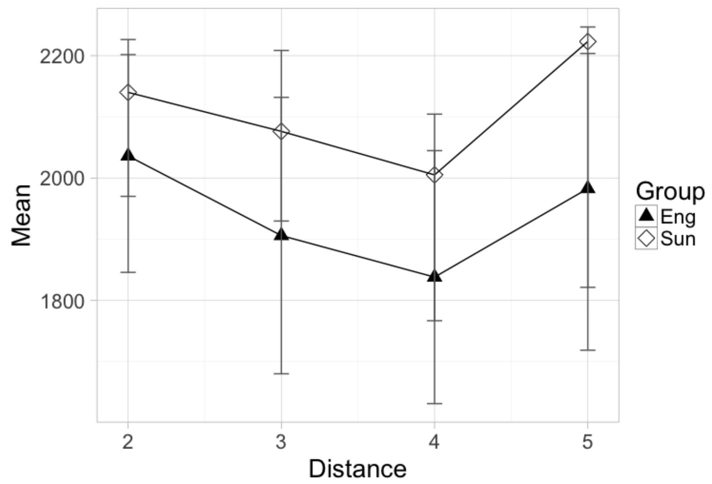
There were no other main effects or interactions. Mean and standard deviations are presented in Table 3.11.

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

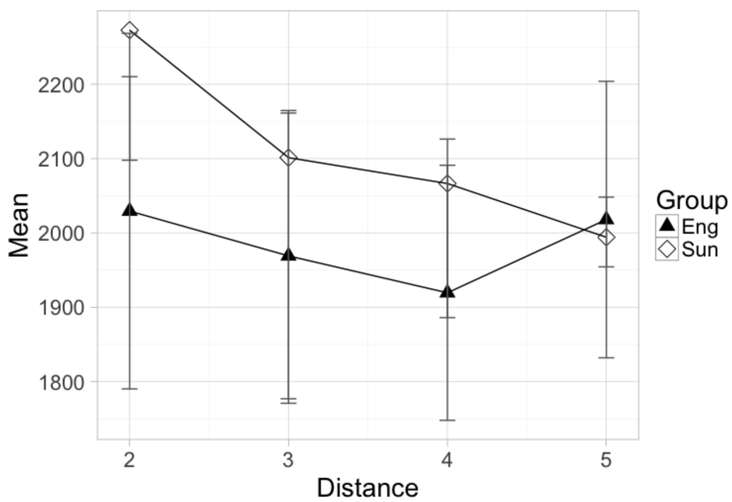


Figure 3.9a-b Reaction Times (mean) of similar category for each level of phonetic distance performed by Sundanese listeners, 95% confidence interval for Distance levels 2-5.

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.

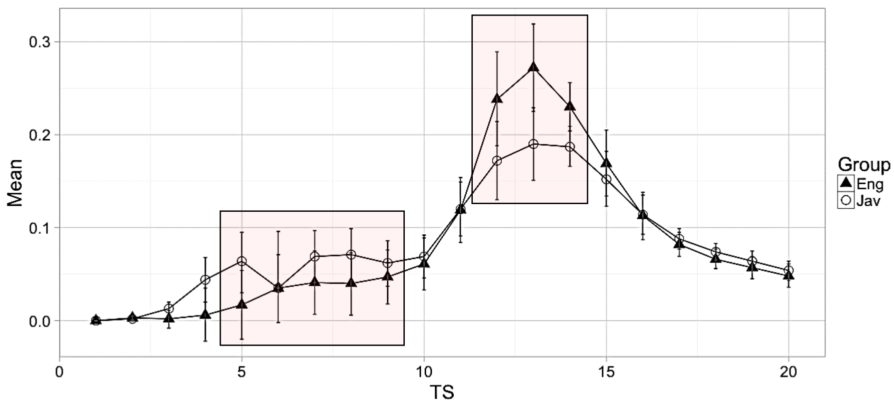


Figure 3.10 Mean hand movement velocity (in pixels/ms) of Javanese and American listeners perceiving English vowels over time (0-1500 ms), divided into 20 bins of 75 ms, TS = time step.

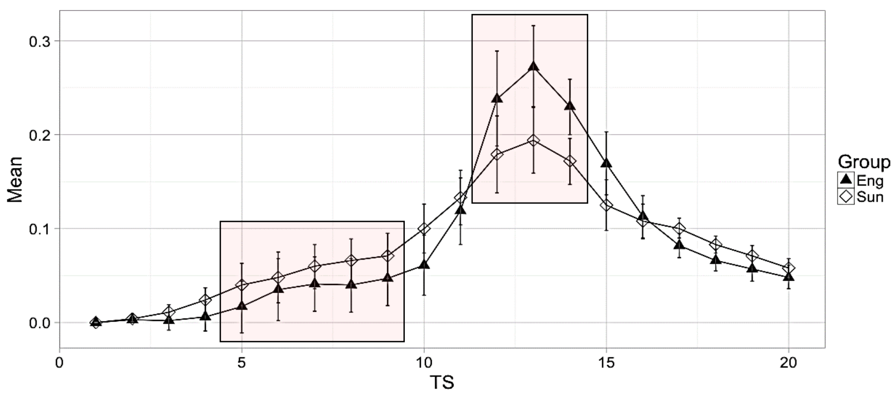


Figure 3.11 As Figure 3.10 but for Sundanese and American listeners.

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, x̄ = .1643, SD = .044) moved more slowly than the American listeners (N= 20, x̄ = .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, x̄ = 0.1605, SD = .038) show slower velocity than the American listeners (N= 20, x̄ = .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
ɝ - ʊ	bird/heard - buddhist/hood	−2.474	−2.312	-	-
ɝ - 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
ɝ - 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