Pakistan Journal of Language Studies
Vol. 4, Issue 1, 2020
ISSN (Online) 2664-1461
ISSN (Print) 2664-1453

Pakistani Speakers' Difficulties in Learning English [v]-[w] Consonants

Nasir Abbas Syed[1], Firdos Atta[2]

[1]Associate Professor of English, Lasbela University of Agriculture, Water and Marine Sciences Uthal Balochistan, Pakistan.
Correspondence: nasirabbassla@gmail.com

[2]Assistant Professor of English, Lasbela University of Agriculture, Water and Marine Sciences Uthal Balochistan, Pakistan.
Email: firdosmalghani@gmail.com

Abstract

Pakistani English (PakE) does not differentiate between [v] and [w]. This study focuses on pronunciation and comprehension of such adult Pakistani speakers of English who, after learning PakE in Pakistan, immersed with native British English (BrE) speakers in England. The study aims to determine if native BrE input has any positive effect on pronunciation of the adult speakers who have already acquired English in a non-native environment like Pakistan. Comprehension and production experiments were arranged with 30 adult Pakistani speakers of English who, after acquiring PakE migrated to England and at the time of experiment, were receiving input from native speakers of English for 5-6 hours daily. The results show that they still confuse [v] and [w]. These results demonstrate that if adult L2 speakers first learn an L2 dialect, which does not differentiate between two consonants, even long immersion with native speakers in later life cannot bring any improvement in their pronunciation of such problematic consonants.

Keywords: Comprehension, consonant, English, L2, learning, production

1. Introduction

There is a large body of literature on the problems of learning of English plosives (Nasukawa, 2010), dental fricatives (Hatten, 2009; Syed, 2013; Wester, Gilbers, & Lowie, 2007) and liquids (Flege, Takagi, & Mann, 1996) by non-native speakers. However, the research on learning English [v]-[w] is limited to a very small number of studies. Previous research reveals that Pakistani-English (PakE, henceforth) (Khalid, 2011; Mahboob & Ahmar, 2004; Rahman, 1990, 1991) and Indian-English (IE) (Gargesh, 2004; Sailaja, 2009), do not differentiate between [v] and [w]. This is because English is learnt in academic institutions in Pakistan; therefore, the learners are compelled to receive non-native input. They only listen to their non-native teachers who themselves are unable to differentiate between these two consonants (Syed, 2015). This is one of the prominent features of PakE and IE. It would be very interesting to know whether such speakers can learn to differentiate between [v] and [w] if an accurate input

is provided to them in a natural environment later on. This study fills this gap in the literature.

The current experiment is conducted with a group of adults that had learnt PakE in Pakistan and moved to England, later on. They speak Saraiki as their mother tongue. Saraiki is an Indo-Aryan language spoken in Pakistan (Atta, 2019; Kula & Syed, 2020). Though the Saraiki speakers are scattered throughout the country but majority of the native Saraiki speakers reside in southern Punjab. At the time of experiment, they were living in the UK and used to listen to ample native English. The aim of this study is to know if immersion in a native speaker community of an L2 can bring any improvement in such phonological features as are absent in the first acquired dialect of the L2 (PakE) but present in the second dialect (British English).

2. Previous studies on acquisition of English [v]-[w]

Iverson, Ekanayake, Hamann, Sennema and Evans (2008), studied the comprehension and production of English /v/-/w/ by Sinhala, German, and Dutch speakers. In its consonant inventory, Sinhala has only a labio-dental approximant [ʋ] and German has a voiced fricative /v/ corresponding to English /v/-/w/; Dutch has a labio-dental approximant /ʋ/ and a labial voiced fricative [v] corresponding to English [w] and [v], respectively. Highly experienced Sinhala and less experienced German and Dutch speakers of English were selected as participants of this study. A group of native speakers of English was also included as a control group.

Overall, the comprehension and pronunciation of Sinhala speakers were very poor; Germans were relatively good and Dutch speakers were best of all. Age was found to have no effect but experience had a positive effect on pronunciation and comprehension of only German speakers. Sinhala and Dutch speakers had shown no positive effect of duration of immersion with native speakers, since Sinhala speakers could not gain accuracy in English [v]-[w] despite living with native speakers of English for almost 28 years, whereas, Dutch speakers developed command over these consonants with a little experience.

As far as the role of L1 is concerned, the German and Sinhala speakers were similar as both these languages have a single phoneme corresponding to English [v]-[w] pair. So, they develop a single phonetic representation for these two consonants of English. But the German speakers were better than Sinhala speakers in their performance. One possible reason for this may be allophonic variance of /v/ in German. German /v/ has two allophones; it is produced with frication word-initially but fricative noise in the rest of the positions. These two variants of the same phoneme may help German learners in developing separate phonetic categories for English [v]-[w], equating one variant with

English [v] and the other with [w] in their perceptual map. But no such context is available to Sinhala speakers of English. Another reason for their weak performance is that Sinhala learners start learning English at a very early age in their native country where the input they receive also does not differentiate between these two consonants. After they have learnt English in their country and the single category representation for two English consonants was quite mature in their L2 phonemic inventory, they moved to the English-speaking country and started receiving British English input. At this stage, the already developed representations were so much hardened and fossilized that the Sinhala learners were too insensitive to the existing difference between [v] and [w] in native English. Therefore, long immersion with the native speakers did not accrue any benefit to them.

In terms of the L1 interference, the performance of Sinhala speakers is most relevant for the present study because Sinhalese is similar to Saraiki, in that, both Indo-Aryan languages have one labio-dental approximant corresponding to English [v] and [w]. The Pakistani speakers of English in the present study also underwent the same circumstances as Sinhala speakers did. Initially, they acquired English in Pakistan receiving input from non-native speaker teachers who do not differentiate between [v] and [w] (Rahman, 1991).

From this perspective, the present study aims to know if Pakistani speakers of English can maintain a difference between [v] and [w] after receiving input from native speakers of English. English [v]-[w] pair is a huge problem for non-native speakers of English in Pakistan. Indian learners of English have also been reported to perceptually assimilate [v] and [w] (Gargesh, 2004; Ghenghesh, 2010; Sailaja, 2009). This contrast has almost vanished in India in the pronunciation of even educated speakers (Fuchs, 2019). The same scenario persists in Pakistan.

3. Research methodology

A group of 30 Pakistani adults residing in Essex (England) was selected as a sample for this study. They used to listen to native English regularly in the Essex County. It is important to note that before arriving at the UK they had already received input from non-native teachers in Pakistan during their academic lives. Their ages ranged from 21-59 years (Mean=33.26, Standard Deviation (SD) =7.21) and the average duration of their stay in England was 70.8 months (Range=4-360, SD=80). Best and Tyler (2007) based on previous studies (Flege & Liu, 2001) suggest that 6-12 months experience is required for learning new sounds through immersion. Three participants of this study had lived in the UK for 6-10 months. The remaining 27 had lived there for more than a year. They had started learning PakE at the average age of 16.3 years (range=5-33, SD=8.15) and

through immersion with native speakers of English after their arrival in the UK at the average ages of 19-36 years (mean=26.26, SD=4.58). All participants of this study speak the central Saraiki dialect (Shackle, 1976) as their L1. They self-reported that they listen to native BrE for an average of 5.50 hours daily (range=1-14, SD=3.59), and speak it (with native speakers during their jobs) for an average of 5.90 hours daily (range=2-12, SD=2.84). Henceforth, we use them as 'target participants' in our discussion.

Besides the target participants, ten native Saraiki speakers were also included in this study as a control group. They participated only in the comprehension test. These monolinguals also speak central Saraiki. Their ages were 18-45 years (Mean=29.6 SD=10.3). They were selected on the basis of convenience sampling. We shall call them the control group in further discussion.

3.1 Experiments

The experiments were conducted in various places of convenience for participants. All participants were doing jobs in super stores so they had ample opportunity to talk to their customers during duty hours. The nature of experiments was thoroughly explained to them concealing the specific purpose. Written permission was obtained from them for recording their productions for research purpose. They were also asked to complete a questionnaire which elicited information about their linguistic and academic background. Ethical approval for this study was granted by the University of Essex.

3.1.1 Stimuli

A comprehension and a production test were planned for this study. The comprehension test was conducted twice by shuffling the stimuli. A set of stimuli of VCV structure (where V stands for a low vowel /a/ in all contexts of the stimuli as it has relatively minimum effect on the adjacent consonant and C stands for consonant) for the comprehension test were recorded in the voice of a female native speaker of English (aged 27) in a laboratory. The used stimuli were nonce words [ava]-[awa]. Some other consonants namely [ɟ] [d] [r] [l] [j] were also recorded in the same structure for use as distracters. Before using the stimuli in the experiment, five native speakers of Essex English (aged 63, 59, 64, 29 and 50), who, according to their statements, had no listening deficiency, verified the stimuli as correct English consonants. The accuracy in their identification was between 87% and 100%.

The stimuli for the pronunciation test were English words 'Venus, thieve, Weed'. The target participants were asked to pronounce. The experiment with target participants was conducted in England but that with the control group was conducted in Pakistan. The

comprehension test had two components namely, identification and discrimination while the production test was conducted through reading. A detail of the experiments is given below.

3.1.2 The Identification Task

In the identification task, VCV (vowel-consonant-vowel) recordings (i.e. [ava] & [awa]) were played and the target participants were asked to identify the consonant of English between the two [a]'s in each stimulus. The control group also listened to the same stimuli (English [v]-[w]) but they were asked to identify if there was any consonant of their mother tongue between two vowels in these productions. The participants of both native Saraiki and L2 speaker groups were asked to write their responses in Urdu and English letters on a provided answer sheet.[1] The control group were also asked to point out if any or some of these consonants did not exist in their L1. The target stimuli were presented along with the distracters listed above. Each stimulus with the target consonant had three repetitions presented in a mixed randomized sequence. One mark was awarded to each correct response.

3.1.1.2 The Discrimination Task

The discrimination task was performed after the identification task. In the discrimination task, recordings of three different syllables of VCV type were played and the respondents of both groups were asked to determine which sound matched with which one. This carried tokens of ABX type for the target sound pair. Each of the sets of stimuli had two syllables containing target consonants and one syllable with another consonant between two low vowels. The following sequence of stimuli (along with similar sets of distracters containing other consonants of English) was used for this test.

ava awa aya

awa ara aya

On hearing these stimuli, the respondents were asked to determine whether the first syllable/nonce-word was the same as the second or the third, or if it was different from both, or all three were identical. Some other sets of syllables/nonce words of similar structure carrying different consonants of English were also included as distracters. There was no repetition in this task. The respondents noted their responses on a printed sheet of paper (provided to them by the author) by ticking one of the multiple answers. A total of

[1]The L2 learner participants also knew Urdu very well. However, since the monolinguals know Saraiki and Urdu. The control group participants transcribed their responses in Urdu only.

twenty-nine tokens were recorded in this experiment. One of the participants could not complete this task.

3.2 The Production Test

The final task in this experiment was a production test conducted with only target participants. They were asked to read two lists, one consisting of isolated words having the target consonants and another with the same words embedded in a carrier sentence. The carrier sentence was 'I say..... again.' These productions were recorded by the author. The stimuli carrying the target consonants were three words 'Venus' 'thieve' and 'weed'. The purpose was to test the participants' pronunciation of [v]-[w] in word-initial and final positions. The primary focus was on those stimuli which had target consonants in word initial position ('weed' and 'Venus') because it is a strong position and is easier to learn (Archibald, 1998). Some other words were also included in the list as distracters. The target words (weed, Venus) start with [w] and [v] respectively, followed by a high tense vowel. Thus, any effect of the adjacent vowel on pronunciation of the target consonants was controlled or equalled. The participants were asked to read the stimuli at normal speed which is a standard practice recommended for such studies (e.g. see Strange, 2007, p. 41). Many researchers for example, Shea and Curtin (2011), Hanzawa (2018), etc. have used this technique in their experiments[2].

3.3. Evaluation of the Recordings

The production test recordings were evaluated by four native speakers of English. They evaluated the productions against their own pronunciation. That is why monolingual English native speakers were not included in this study for comparison. This purpose was served by the judges who were asked to consider their own pronunciation as a standard while evaluating the recordings. All evaluators (judges) were from Essex where the target participants were living, so the effect of dialectal variation between the judges' pronunciation and the input that the participants were receiving was also controlled. The judges were asked to mark the target sounds on a Likert scale of 1-5 where 5 meant 'native-like' (where native means the judges' own pronunciation), 4 meant 'near native-like', 3 meant 'different from natives but understandable', 2 meant 'hardly understandable' and 1 meant 'unintelligible'. The judges were asked to evaluate only the pronunciation of [v] and [w] neglecting the vowels in the productions. They evaluated the recordings in the researcher's office in the latter's presence. The evaluations took place in separate sessions. No two judges were ever allowed to sit

[2]There were fifty participants in the study by Hanzawa (2018) and fifteen participants in Shea and Curtin (2011). The participants in the latter study were divided into three groups i.e., five participants in each group.

together for marking. They even did not know each other. Keeping in view the potential effects of tiredness and boredom, the judges were allowed to evaluate the productions in only one-hour session in a day. The judges were paid for their time.

All recordings were originally in two types: one consisting of isolated words and the others in sentences. Before analysis, the target words were extracted from the carrier sentences for evaluating. The judges were not informed of the details of the recordings and editing process.

The reliability of given score is settled on by a Cronbach's alpha test. Normally 60% reliability (or a Cronbach's alpha value of 0.60) is considered a threshold for acceptance (Scholfield, 1995). The evaluation of [v] in coda position showed less reliability than the required level as shown in table 1. Given the small number of judges, these agreement coefficients can be considered strong indicators of reliability, because a bigger agreement coefficient can be obtained with a large number of judges (Larson-Hall, 2016).

Table 1

Reliability of the production test results

Context	Position	Sound	Reliability coefficient
Word	Onset	[w]	0.779
	Onset	[v]	0.660
	Coda	[v]	0.610
Sentence	Onset	[w]	0.659
	Onset	[v]	0.757
	Coda	[v]	0.491

The following analyses will be based on those evaluations which meet the reliability standard. As mentioned above, the main concern of this study is to evaluate the target participants' pronunciations on word-initial positions. And the reliability coefficient in the onset position for both target consonants is 0.66-0.78 which implies 66%-78% reliability, which is considered an acceptable range by researchers. The results are presented in the following section. In the above six contexts, only in one context, i.e., on [v] in the coda position in sentences, the reliability coefficient is below the standard reliability. However, in the remaining five contexts, it is above the standard cut off point.

4. Results

In this study, data were collected from two groups of participants; one group comprised of ten Saraiki monolinguals in Pakistan and the other comprised of 30 adult L1-Saraiki speakers of English living in England. The results obtained from the native group will be used to determine possible perceptual overlap of the target consonants with the corresponding L1 consonant.

4.1 The Control Group

As mentioned earlier, only the comprehension test including discrimination and identification tasks was conducted with the control group. One (10%) out of ten native Saraiki speakers accurately discriminated English [v] from [w]. In other words, 90% of them perceptually confused [v] with [w].

In the identification task, they listened to each of the stimuli three times and identified [w] as their L1 labio-dental approximant in all 30 trials (10 participants* 3 repetitions) but [v] was identified as the L1 labio-dental approximant in 28 out of 30 (93.33%) trials. It means, Saraiki native speakers before learning English have 100% overlap between English [w] and the corresponding L1 approximant, and 93.33% overlap between English [v] and the labio-dental approximant. Based on these results, we expect that English [v]-[w] consonants may be equated with the Saraiki labio-dental approximant by Saraiki learners of English. Since, they can realize a minimal fraction of difference between English [v] and the closest L1 labio-dental approximant, slightly better performance on comprehension and production of [v] is expected from them.

4.2 The Target Group

There were 30 Pakistani respondents in this group. One of them did not participate in the discrimination task. Only two out of the remaining 29 participants could accurately discriminate [v] from [w] yielding an accuracy of only 6.9% in this task.

Another task in the comprehension experiment was identification of English [v] and [w]. The following table provides means of identification scores out of a total of 3. There were three repetitions for each target consonant. One mark was awarded for each correct response and zero for an incorrect response.

Table 2

The identification task result

Stimulus	Minimum	Maximum	Mean	Std. Deviation
[w]	.00	3.00	1.97	1.30
[v]	.00	3.00	1.20	1.37

The results show that the target participants are extremely poor in identification of English [v]-[w] consonants. It is important to point out that in all there were 30 participants and each identified each target sound three times. In this way, a participant could score a maximum of three points. In the above table, only mean of the group out of the three marks are given. The following table provides the same results in frequencies showing how many of them identified which of the target sounds accurately in how many trials.

Table 3

The identification of [v]-[w] by the target participants

	[w]		**[v]**	
Correct Response[3]	No. of participants	Percentage	No. of participants	Percentage
0	7	23.3	15	50
1	4	13.3	4	13.3
2	2	6.7	1	3.3
3	17	56.7	10	33.3

There were 90 responses (3 repetitions x 30 participants) for each of the target consonants in this task. The results show that in 59 (65.56%) trials [w], and in 36 (40%) trials [v] were identified accurately. Participants' performance is better in this task than in discrimination. Table 3 also shows that most of the participants consistently identified the stimuli correctly or incorrectly in all three repetitions although the stimuli were presented in a randomized order along with some distracters. In very few cases, the participants identified a stimulus correctly in one or two repetition(s) and incorrectly in the remaining one(s). In all three trials, 24 (80%) identified [w] either accurately (N=17) or inaccurately (N=7) and in the identification of [v], 25 (83.33%) were either thoroughly accurate (N=10) or thoroughly inaccurate (N=15). Such high consistency in responses enhances

[3]In this column '0' indicates that in none of the three trials was the target consonant identified correctly. '1' means, one correct response, '2' means, it was identified correctly twice and '3' means, in all three trials the target consonant was identified correctly.

reliability of these results. Briefly speaking, the results suggest that the participants were better at identifying [w] than they were at [v]. We shall return to these results in discussion section.

The results of the production test are given in Table 4. Four native speakers evaluated/marked [v] and [w] produced by the participants on a five-point scale. The following table shows mean scores of the whole group out of a total of five.

Table 4

Production test results

Context	Sound	Position	Minimum	Maximum	Mean	Std. Deviation
Words	[w]	Onset	1.50	4.00	2.91	.72
	[v]	Coda	2.00	4.33	3.24	.57
	[v]	Onset	1.50	4.25	3.22	.81
Sentences	[w]	Onset	1.25	3.75	2.40	.53
	[v]	Coda	2.33	4.67	3.57	.67
	[v]	Onset	1.75	4.25	3.43	.67

Table 4 reveals that not a single participant could obtain a score of 5 in this test. (Recall that '5' denotes native-like pronunciation on the scale used for evaluation). The results show that even a mean score of 4 which denotes a 'near native-like' pronunciation also could not be obtained by the participants. It means, on average, there is below near native-like pronunciation. They are significantly better in pronunciation of [v] than of [w] in isolated words (t= -6.412, p <.001). But in pronunciation of the same consonants in sentences, the difference between [v] and [w] is statistically non-significant (p>.1), it means the participants are very poor in pronunciation of [w] in this context.

A 2*2 repeated measure ANOVA with position (onset and coda) and context (words and sentences) as repeated conditions applied on the production test scores of [v] shows that only the effect of context is significant ($F_{(1,29)}$=4.586, p<.05), but that of position is non-significant (p>0.5). The interaction between position and context is also non-significant (p>0.5). The individual pair-wise comparisons show that the target participants pronounce [v] better (t= -2.053, p<.05) in sentences than in words on the coda position only. They produced [w] better in words than sentences (t=3,617, p<.001). We summarise these results in the following points before proceeding to further analysis.

- Overall, Pakistani speakers of English living in England are extremely poor in the discrimination of [v] from [w].
- They can identify [v] in 40% and [w] in 65.56% of listening trials
- Their pronunciation of English [v]-[w] is below near native-like level.

- Their pronunciation of [v] is slightly better than of [w].
- Overall, there is no significant difference in their pronunciation of [v] and [w].
- The role of context is significant in a varying way, in that, they pronounce [w] better in words, but in production of [v] on the coda position, their performance is better in sentences. Therefore, based on these results, no generalization about the role of context can be developed in this study.

Overall, the performance of the target participants is not better in these experiments. Because of their immersion with the native speakers of English, they were expected to perform better, since input is normally expected to play a very positive role in L2 learning (Flege, 2009, 2019; Flege & Wayland, 2019; Tyler, 2019, etc.). But the above results show that they cannot accurately perceive a difference between English [v] and [w]. They also cannot produce these consonants correctly. These issues are further discussed in the following section.

5. Analysis and discussion

In this section, first we discuss the anomaly regarding the unexpectedly better performance of Pakistani speakers of English in the identification task compared with their poor performance in discrimination. Later on, we determine the level of Pakistani speakers in pronunciation of [v]-[w]. Next, we take up the issue of slightly better performance of the participants on [v] than [w]. Lastly, we comment on why Pakistani speakers cannot improve their pronunciation of English [v]-[w] despite long time immersion in an environment of native speakers of English.

5.1 Comparison of the Target Participants' Performance in Comprehension Tasks

According to Boersma and Hamann (2009) learners' real comprehension of an L2 phoneme can be determined by performance in identification tasks. This is because in discrimination tasks, listeners are asked to decide whether they can discriminate among different sounds. They can either guess, or realize the surface phonetic difference between two phonemes in a stimulus on the basis of acoustic cues even without having developed separate representations for those phonemes in their L2 phonemic inventory. But in an identification task, they have to retrieve from their existing L2 phonemic inventory a phoneme matching with the one they hear in a stimulus. That is why identification tasks are considered more difficult as well as more trusted than discrimination tasks. Keeping this in mind, the primary purpose of conducting a discrimination task after identification was to double check the results obtained in the identification task. The target participants showed 40% and 65.56% accuracy in identification of [v] and [w], respectively, but that in the discrimination task is as low as

6.9%. It is an unexpected result. Speakers of an L2 are normally expected to show better performance in discrimination than in identification tasks (Archibald, 2005).

Overall results also confirm that Pakistani speakers produce and perceive English [v] and [w] alike, since they cannot discriminate between these consonants. But they can discriminate these consonants from other sounds of English. Therefore, on hearing a stimulus carrying either [v] or [w], they equate it with the single category of consonant existing for both these consonants in their mind. For this single category, incidentally, they have two separate letters available in English orthography. Therefore, in the identification task, when they heard the stimuli carrying [v] and [w] consonants, they responded by writing either English letter 'v' or 'w' for both target sounds. Naturally, one of the two responses was correct. As a result, they obtained good scores in one but poor in identifying the other. But their good scores do not imply their accurate comprehension because their performance on discrimination of the same pair is extremely poor. Therefore, we can only conclude that although they do not perceptually assimilate [v] and [w] with other phonemes of English, Pakistani speakers cannot hear a difference between these two consonants. The same is also further confirmed in a correlation test applied on the target participants' scores for [w] and [v]. The test shows a strong negative correlation (*Spearman's rho*= -.717, $p<.001$) between the identification scores for [v] and [w]. It means, if their performance is better on [v], it is poorer on [w] and vice versa. This is further confirmed if we have a look at their responses written in Urdu. (Let us recall that the target participants were asked to write their responses in English and Urdu). Urdu language has one letter 'و' for the single approximant in its phonemic inventory which corresponds to two English consonants, [v] and [w]. On listening to both these sounds, they wrote in Urdu that they identified 'و' sound (a letter representing the corresponding approximant in Saraiki and Urdu)[4]. Although they were also given an option in the answer sheet to point out if they realize that any consonant in the set of stimuli did not have a corresponding letter in either Urdu or English, none of them pointed out any such mismatch. This means they confused English [v] with [w] considering these two consonants as one. Thus, the asymmetric results in the two comprehension tasks can only be ascribed to methodological issues or orthography.[5]

[4]It is relevant to point out that Saraiki also uses Urdu orthography with some minor modifications (Bashir & Conners, 2019).

[5]An anonymous reviewer is of the view that it could be that participants are better at [w] because that in fact is a better exemplar of their native L1 labio-dental approximant than [v] (a labio-dental fricative). This would then in fact correlate with why they are better at producing [v] – it has less interference with an L1 sounds that [w] does which they produce less well. The discrimination results are different as this is quite a different task. The issue is open for further research.

5.2 How Pakistanis Pronounce [v-w]?

Now we discuss the production test results. The target group's pronunciation of [v] was evaluated on word-initial and final positions but [w] was studied only in the onset of syllables. The literature on L2 acquisition demonstrates that accuracy in production of consonants in isolated words alone does not imply accuracy in global (continuous) speech. Birdsong (2007) calls segmental and global level accuracies as '*necessary*' and '*sufficient*' conditions of learning respectively. The relationship between accuracy in L2 consonants in exclusive words and continuous speech is unidirectional i.e., better performance in continuous speech implies that in isolated words, but not vice versa. In the current study, average scores obtained by the target participants in pronunciation of [v] on the onset and coda positions in the isolated words are not significantly different. Although their scores are different on these two positions in sentences, we shall use for comparison scores obtained on word-initial position in isolated words because we have recorded the productions of [w] in the onset position only. Another reason for comparing results on word-initial position is that reliability coefficients for these data on coda position were lower than on word-initial position. Besides, the relevant literature shows that adult L2 learners first acquire the target sounds in isolated words in the onset of a syllable (Archibald, 1998).

The scale used for evaluation of pronunciation of the target participants had five points. Each point represents one stage. If we count the number of participants in each stage on account of their pronunciation, the following picture emerges.

Table 5

The production test results in frequencies

Category	**Category defined**	**[v]**	**[w]**
5	Native-like	0	0
4	Near native-like	4	3
3	Different from natives but understandable	17	12
2	Hard to understand	5	14
1	Unintelligible	4	1
	Total participants	30	30

Table 5 shows that none of the target participants is native-like in pronunciation of [w-v] and only 3-4 participants are 'near native-like'. Keeping in view the previous literature, we assume that 4 is the highest possible score on this scale that can normally be achieved by average adult L2 learners, since native-like pronunciation is improbable (though not quite impossible) in adult language learning. Previous research shows that,

despite learners' successful acquisition of an L2, normally, their pronunciation of some consonants of the L2 is slightly deflected away from native speakers' pronunciation (Flege, 1995). As the results of the present study show, with the exception of only 3-4 participants, Pakistanis could not learn to pronounce [v]-[w] consonants in a near native-like manner.

5.3 A Feature-based Analysis

Now we come to know why the scores of the target participants are slightly better for [v] than [w]? For this, we need to understand the nature of the problem in the present context. Actually, there is a two-way confusion in the mind of Pakistani speakers of English. In Saraiki, there is one labio-dental approximant [ʋ] corresponding to English [v]-[w]. Therefore, on the one hand, Pakistanis confuse English [w] with English [v], and on the other, they equate these two consonants of English with the corresponding L1 approximant [ʋ][6]. This indicates a perceptual assimilation of three consonants, namely [v]-[w] of L2-English and [ʋ] of the L1. There are examples of such perceptual assimilation in the L2 literature. For example, Mooney (2017) identified similar assimilation of Occitan-French vowels in the speech of bilinguals. However, an anonymous reviewer has a different point of view that the participants have different perception so they clearly treat these two consonants differently. Our point of view is that the difference in their performance on [v] and [w] is an artifact of orthography. The question needs further research.

This scenario may be better explained in terms of phonological features as defined by Clements and Hume (1995). Phonologically, Saraiki [ʋ] is [+sonorant, -round], whereas English [v] is [-sonorant, -round] and [w] is [+round, +sonorant]. In this way, [ʋ] shares one feature ([+sonorant]) with [w] and one feature ([-round]) with [v]. Similarly, it is different from each of these two consonants in respect of one feature; i.e., it lacks feature [+round] of [w] and [-sonorant] of [v]. In other words, the target participants of this study produced English [v] as [+sonorant] (without fricative noise) and [w] as [-round] (without lip-rounding). One reason for their relatively better performance on [v] is that perhaps the judges noticed the absence of feature [+round] more clearly and penalized them in the evaluation but did not notice the absence of feature [-sonorant] in [v]. We assume this because lip rounding results in lowering of the second and third formants which is a very solid acoustic cue. The absence of such a robust signal in production can be easily identified by the native speakers. Thus, they awarded poorer marks for [w]. In the literature, it has been noticed that although Japanese learners perceptually assimilate English [l] and [r] with each other and equate these consonants

[6]All major Pakistani languages have one approximant corresponding to English [v-w].

with a single consonant of their L1 but they performed better on one than the other consonant. Similarly, in the study by Iverson et al. (2008), some German participants performed better on [v] than [w]. Even the Sinhala speakers, who otherwise performed poorly, also showed better performance on [v] in the study by Iverson et al. (2008).

Brown has analyzed learning difficulties with reference to phonological features. According to her (Brown, 1998, 2000), if two new L2 sounds are different from each other with reference to a phonological feature which is active in the L1 feature geometry of learners, they may acquire the new L2 sounds with some effort. But, if the crucial feature which differentiates between such sounds is not active in the L1, they cannot acquire the new sound contrast. English [v] and [w] are different from each other on account of the features [round] and [sonorant]. These features are not active in Saraiki. The reason for this is that although we have sonorant sounds but they are differentiated from non-sononrant sounds on the basis of other features also. No two sounds of Saraiki are solely differentiated by either of these features. So, we cannot safely claim that this feature is active in L1 Saraiki. Therefore, English [v]-[w] pair is expected to be very difficult for L1-Saraiki speakers of English. Along with, this sound pair also presents single category type of assimilation for Pakistani learners according to the Perceptual Assimilation Model (Best, 1995), so they are expected to be extremely difficult for adult learners of Pakistan.

5.4 Why Pakistanis Fail to Differentiate [w] from [v]?

A question then arises here, why long immersion with native speakers could not enable Pakistanis to improve their pronunciation? Input definitely has a very effective role in learning (Chu, Yang, & Liu, 2019). But in this study, we notice that Pakistanis, despite having a very long interaction with native speakers of English, could not improve their pronunciation and comprehension of [v]-[w]. Previous studies also mention cases where L2 learners could not improve despite living among native speakers of the target L2 for a long time (e.g. Sinhala speakers in Iverson et al. (2008). On the other hand, in some studies, extended exposure to native input has been found to have a positive effect. For example, the German learners of English in Iverson et al. (2008), Japanese learners in Flege et al. (1996) and those in Best and Strange (1992) seem to have benefited from their interaction with native speakers. In the current study, the target participants seem to have not benefited from their interaction with native speakers of English.

The reasons for German and Dutch learners' better acquisition of English [v]-[w] is that these learners have two consonants in their L1 corresponding to English [v]-[w] which help them discriminating these English consonants by positive transfer. According to researchers, such consonant pairs of L2 which have one equivalent in the L1 are easier

than those pairs which have two equivalents in the L1 of learners (Best, 1995). Discrimination between the latter L2 phonemes is found to be extremely difficult (see e.g. Best (1994), Best, McRoberts, and Goodell (2001), Wiltshire (2005), Wiltshire and Harnsberger (2006), Antoniou, Best, and Tyler (2013) Fenwick, Best, Davis, and Tyler (2017), etc.

Another important factor is frequency of occurrence of the target L2 consonants. Whereas, in the other studies quoted above, target sounds (e.g. [l]-[r], etc.) occur frequently in English vocabulary making a large number of minimal pairs, [v]-[w] pair do not provide a large number of words which contrast in English. Since, Pakistani speakers of English do not experience any communication problem because of this perceptual assimilation due to a very small number of minimal pairs of English starting with [v] and [w], they do not have a strong motivation to struggle for accurate pronunciation of these consonants. The role of vocabulary size on acquisition of L2 is established in the literature (Bundgaard-Nielsen, Best, Kroos, & Tyler, 2012; Bundgaard-Nielsen, Best, & Tyler, 2011a, 2011b).

Based on results of this experiment, we conclude that Pakistani learners first develop a single consonant category for English [v] and [w] in Pakistan[7], and later on, when they move to England as adults, it becomes very difficult for them to re-structure their L2 phonemic inventory and re-develop two separate representations for these consonants. That is why long exposure to native English input does not accrue any benefit to them. Flege and Liu (2001) have also pointed out to this fact that after that initial period of immersion very little benefit seems to accrue from additional listening experience if the L2 speakers' already learnt representations are fossilized. Mooney (2017), along with others, also predicts that L2 categories cannot be modified later with long exposure to the L2, if learners have established a single category representation for two L2 phonemes. The same occurs with Pakistani speakers of English who learn English in Pakistan and move to English speaking countries. Thus, we conclude that Pakistanis cannot differentiate English [v] from [w] because of a specific L1-L2 perceptual map which results into one-to-two correspondence of consonants, low frequency of minimal pairs of [w] and [v] in word-initial position and initial input that Pakistani learners receive from their PakE teachers who also do not differentiate between these two consonants.

[7] It is already known in the previous research that Pakistani English speakers, who receive input solely from non-native speakers, confuse [v] with [w].

6. Summary and Conclusion

This study was based on an analysis of comprehension and production of English [v]-[w] by adult Pakistanis who first received education from Pakistan and after having learnt PakE, migrated to England. On arrival in England, they started receiving ample input of native English speech. Recordings of production and comprehension of these Pakistanis were evaluated when they had spent 5-6 years in England. The results show that overall, such Pakistanis are very poor in comprehension and production of these consonants. They equate both English [v]-[w] consonants with a labio-dental approximant of their L1 which blocks further learning. That is why they cannot discriminate [v] from [w] in pronunciation despite receiving native input for a long time. The study concludes that in such a context, where first learning of adult speakers of an L2 does not differentiate between two consonants of a foreign or second language, later immersion with native speakers and ample input of native speech even for whatever period of time does not accrue benefit towards improvement of pronunciation of such L2-speakers. This issue further enhances if the L1 grammar is also non-facilitative for the L2 learning and the frequency of words having the target phenomenon is very low in the L2. The best solution to such learning difficulties is that at the very initial stage of learning, proper input must be provided to learners in schools. Specifically, for Pakistani learners of English, the major difficulty is that they miss lip-rounding in production of [w] and fricative noise in that of [v]. These articulatory problems can be rectified by proper training of the teachers in Pakistan.

Before we conclude, it is significant to point out that the generalizations developed in this study about L1-Saraiki speakers of Pakistan may be equally valid for Pakistanis who have similar social circumstances (migrate to the UK or any native English speaking country after acquiring English in Pakistan) but speak other Pakistani languages like Urdu, Punjabi, Sindhi, Balochi, Pashto, Kashmiri, etc. because (a) all these major Pakistani languages also have only one approximant corresponding to English [v] and [w] (see Elfenbein (1997) for Balochi, Koul (2007) for Kashmiri, Khubchandani (2007) for Sindhi, (Schmidt, 2007) for Urdu, Shackle (2007) for Punjabi and David (2014) for Pashto phonemic inventory); (b) all Pakistani learners receive similar PakE input at earlier stage of learning which does not differentiate between [v] and [w] (Rahman, 1991; Syed, 2015). Therefore, although the current experiment was conducted with only L1-Saraiki speakers of Pakistan, the findings of this study may be generalized to almost all adult Pakistani learners/speakers of English.

References

Antoniou, M., Best, C. T., & Tyler, M. D. (2013). Focusing the lense of language experience: Perception of Ma'di stops by Greek and English bilinguals and monolinguals. *Journal of the Acoustical Society of America, 133*(4), 2397-2411.

Archibald, J. (1998). *Second language phonology*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Archibald, J. (2005). Second language phonology as redeployment of phonological features. *Canadian Journal of Linguistics, 50* (1-4), 1000-1030.

Atta, F. (2019). *Phonetics and Phonology of the Saraiki Language: A descriptive exploration and an anayslis from the perspective of Optimality Theory.* (PhD), Shanghai International Studies University.

Bashir, E., & Conners, T. J. (2019). *A Descriptive Grammar of Hindko, Panjabi, and Saraiki*. Berlin: De Gruyter Mouton.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In C. Goodman & H. Nasbaum (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. *The Journal of the Acoustical Society of America, 109*(2), 775-794.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*(3), 305-330.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emile Flege.* (pp. 13-34). Amsterdam: J. Benjamins.

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 99-116). Amsterdam: John Benjamins.

Brown, C. A. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research, 14*(2), 136-193.

Brown, C. A. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 4-63). Malden Mass: Blackwell.

Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C., & Tyler, M. D. (2012). Second language learners ' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics, 33*, 643-664.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011a). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition, 33*, 433-461.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011b). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics, 32,* 51-67.

Chu, J., Yang, C., & Liu, G. (2019). Analysis of Second Language Acquisition (SLA) Speech Perception Model & the Perception of Second Language Prosody. *Revista De Cercetare Si Interventie Sociala, 64*, 334-351.

Clements, G. N., & Hume, E. V. (1995). The internal organization of speech sounds. In J. Goldsmith (Ed.), *A handbook of phonological theory* (pp. 245-306). Oxford: Blackwell.

David, A. B. (2014). *Descriptive grammar of Pashto and its dialects* (Vol. 1). Berlin: Walter de Gruyter.

Elfenbein, J. (1997). Balochi Phonology. In S. A. Kaye (Ed.), *Phonologies of Asia and Africa* (Vol. 2, pp. 761-776).

Fenwick, S. E., Best, C. T., Davis, C., & Tyler, M. D. (2017). The influence of auditory-visual speech and clear speech on cross-language perceptual assimilation. *Speech Communication, 92*, 114-124.

Flege, J. E. (2009). Give input a chance. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175-190). Bristel: Multilingual Matters.

Flege, J. E. (2019). A non-critical period for second-language learning. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. Hjortshøj (Eds.), A sound approach to language matters: In honor of Ocke-Schwen Bohn, Aarhus University (pp. 1-23). Aurhus University Library: Aurhus University.

Flege, J. E., & Liu, S. (2001). The effect of experience on adult's acquistion of a second language. *Studies in Second Language Acquisition, 23*, 527-552.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical Familiarity and English-language Experience Affect Japanese Adults' Perception of [ɹ] and [l]. *Journal of Acoustical Society of America, 99*(2), 1161-1173.

Flege, J. E., & Wayland, R. (2019). The role of input in native Spanish Late learners' production and perception of English phonetic segments. *Journal of Second Language Studies, 2*(1), 1-33.

Fuchs, R. (2019). *Almost [w]anishing: The elusive /v/-w/ contrast in educated Indian English.* Paper presented at the International Congress of Phonetic Sciences (ICPhS), Melbourne.

Gargesh, R. (2004). Indian English: Phonology. In B. Kortmann (Ed.), *A Handbook of Varieties of English* (pp. 815-829). Oxford: Blackwell.

Ghenghesh, P. (2010). The Motivation of L2 Learners: Does it decrease with age? *English Language Teaching, 3*(1), 128-141.

Hanzawa, K. (2018). The Development of Voice Onset Time (VOT) in a Content-Based Instruction University Program by Japanese Learners of English: A Longitudinal Study. *Canadian Modern Language Review, 74*(4), 502-522.

Hatten, D. J. (2009). Substitution patterns for English inter-dental fricatives by L1 Latin American Spanish speakers. Retrieved from Rutgirs Optimality Archive website: http://andromeda.rutgers.edu/~hattem/Academic%20Research/Sample%20Research%20Paper%20Substitution_Patterns_for_EnglishInter-Dental_Fricatives_(David_J_Hattem)_FINAL_ PAPER.pdf

Iverson, P., Ekanayake, D., Hamann, S., Sennema, A., & Evans, B. G. (2008). Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers.

Journal of experimental psychology: human perception and performance, 34(5), 1305-1316.

Khalid, N. (2011). Existence of [v] and [w] in Urdu language. Retrieved from http://www.crulp.org/Publication/Crulp_report/CR02_08E.pdf.

Khubchandani, L. M. (2007). Sindhi. In G. Cordona & D. Jain (Eds.), *The Indo Aryan Languages* (pp. 683-721). New York: Routledge.

Koul, N. O. (2007). Kashmiri. In G. Cordona & D. Jain (Eds.), *The Indo Aryan Language* (pp. 991-1051). New York: Routledge.

Kula, N. C., & Syed, N. A. (2020). *Non-myopic Nasal Spreading in Saraiki*. Radical: A Journal of Phonology.

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

Mahboob, A., & Ahmar, N. H. (2004). Pakistani English: Phonology. In E. W. Schneider (Ed.), *A Handbook of Varieties of English: A Multimedia Reference Tool* (pp. 1003-1015). Berlin: Mouton de Gruyter.

Mooney, D. (2017). Phonetic transfer in language contact: Evidence for equivalence classification in the mid-vowels of Occitan-French bilinguals. *Journal of the International Phonetic Association*, 1-33.

Nasukawa, K. (2010). *Place -Dependent VOT in L2 Acquisition*. Paper presented at the Proceeding of the 2008 Second Language Research Forum, Tohoku Gakuin University.

Rahman, T. (1990). *Pakistani English: The linguistic description of a non-native variety of English*. Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University.

Rahman, T. (1991). Pakistani English: some phonological and phonetic features. *World Englishes*, *10*(1), 83-95.

Sailaja, P. (2009). *Dialects of English: Indian English*. Edinburgh: Edinburgh University Press.

Schmidt, R. L. (2007). Urdu. In G. Cordona & D. Jain (Eds.), *The Indo Aryan Languages* (pp. 286-350). New York: Routledge.

Scholfield, P. (1995). *Quantifying language: A researcher's and teacher's guide to gathering language data and reducing it to figures*. Clevendon/Philadelphia: Multilingual Matters Ltd.

Shackle, C. (1976). *The Siraiki language of central Pakistan: A reference grammar*. London: University of London School of Oriental and African Studies.

Shackle, C. (2007). Punjabi. In G. Cordona & D. Jain (Eds.), *The Indo Aryan Language* (pp. 581-621). New York: Routledge.

Shea, C. E., & Curtin, S. (2011). Experience, representations and the production of second language allophones. *Second Language Research, 27*(2), 229-250.

Strange, W. (2007). Cross-language phonetic similarity of vowels: Theoretical and methodological issues. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 35-55). Amsterdam/Philadelphia: John Benjamins.

Syed, N. A. (2013). Acquisition of dental fricatives by adult Pakistani learners. *Newcastle Working Papers in Linguistics 19, 2*, 64-73.

Syed, N. A. (2015). The Role of Teacher in Second Language Learning. *The NUML journal of Critical Inquiry, 13*(1), 71-91.

Tyler, M. D. (2019). PAM-L2 and Phonological Category Acquisition in the Foreign Language Classroom. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, M. Hjortshøj, & Sørensen (Eds.), *A Sound Approach to Language Matters - In Honor of Ocke-Schwen Bohn* (pp. 607-630). University of Arhus: Department of English School of Communication and Culture.

Wester, F., Gilbers, D., & Lowie, W. (2007). Substitution of dental fricatives in English by Dutch L2 speakers. *Language Sciences, 29*(2-3), 477-491.

Wiltshire, C. R. (2005). The Indian English of Tibeto-Burman language speakers. *English world-wide, 26*(3), 275-300.

Wiltshire, C. R., & Harnsberger, J. D. (2006). The influence of Gujarati and Tamil L1s on Indian English: A preliminary study. *World Englishes, 25*(1), 91-104.