The Role of the Teacher in Second Language Learning

Nasir Abbas Syed

Abstract

In a comprehensive study of perception and production of English consonants by students of three post-graduate colleges of Southern Pakistan, it was found out that the students were unable to differentiate between [v] and [w], and [ʒ] and [j]. They also could not differentiate between aspirated [p^h t^h] and unaspirated [p t] allophones of /p t/ phonemes of English. Unlike native speakers of English, they produce English /d/ as retroflex with pre-voicing. The alumni of the same colleges now living in and around London were studied in a second experiment. They were also found to face the same problems in their perception and production of these consonants of English. The L1 of the students has aspirated and unaspirated sounds, which means they are already familiar with these sounds and their inability to differentiate between aspirated and unaspirated allophones of English plosives cannot be ascribed to negative transfer from the L1. On the basis of these results, we developed a hypothesis that the difficulties of students in proper acquisition of English consonants are due to improper input at early stage of learning in Pakistan. To test this hypothesis, we repeated the same experiment with a group of ten English language teachers who were teaching in the similar post-graduate colleges from where the student participants were selected for the previous experiments. A computer-based acoustic analysis of productions of teacher participants shows that the teachers also perceptually assimilate English [v] with [w] and [ʒ] with [j]; they produce English /d/ as retroflex with pre-voicing and could not differentiate between aspirated and unaspirated allophones of /p t/. These results confirm that students produce what they receive from their teachers. They do not change their pronunciation even after obtaining input from native speakers of English. This confirms everlasting impact of teachers' pronunciation on students' learning. The findings of this study support teacher-centered approach of teaching and suggest that more attention on teachers' pronunciation may be paid if we have to improve pronunciation of students.

Keywords: *second language learning, aspirated, retroflex*

Introduction

The current paper reports on an experiment which was conducted in the perspective of the debate on teacher-centered and learner-centered

teaching (Edens, 2000; Hansen & Stephens, 2000; Villaume, 2000). Like Kain (2003), it supports a merger of the two approaches in pedagogy. The data presented in this paper points out everlasting impact of teacher's pronunciation on learners' pronunciation. It demonstrates that we cannot thoroughly negate the role of teacher in learning. The current paper addresses this issue with reference to acquisition of English consonants by Pakistani learners. Recordings of English speech of Pakistani learners were analyzed using latest computational techniques which highlight the nature of errors of students learning English as a foreign language in Pakistan and another group of Pakistani learners who were learning English as a second language in London. Both groups of learners were doing the same errors. Later on, the speech of their teachers were recorded and analyzed which shows that the teachers also had the same pronunciation errors in their speech. This confirms that learners develop phonetic categories for L2 sound on the basis of teacher input. Once they develop specific phonetic representations for L2 phonemes, it is very difficult to restructure these representations. Even correct input from native speakers of the L2 also cannot change those categories. This demonstrates the everlasting impact of teacher on students' learning. The findings support the view that even in learner-centered teaching, vital role of teacher as a model may not be thoroughly neglected.

L2 learners make errors of specific nature in acquisition of English consonants. For example, Spanish learners are reported to perceptually assimilate English [d] and [ð] (Archibald, 1998), Korean learners confuse [s] with [ʃ] (Eckman & Iverson, 2013) and Japanese learners perceptually assimilate [l] with [r] (Brown, 2000; Flege, Takagi, & Mann, 1996). Since Lado's seminal publication (1957) on contrastive analysis hypothesis, interference of L1 has been considered one of the major motivations for errors in L2 acquisition. The speakers of Indo-Aryan family of languages living in the subcontinent of Pakistan and India and the adjacent countries have been reported to perceptually assimilate English [v] with [w] (Gargesh, 2004; Ghenghesh, 2010; Iverson et al., 2008; Iverson, Wagner, Pinet, & Rosen, 2011; Mahboob & Ahmar, 2004). Researchers consider the L1 interference as a major resistance in acquisition of L2 sounds. The current study is conducted with a view that there may be other possible reasons for the difficulties of L2 learners. Syed (2012) studied acquisition of English [d], English alveo-palatal fricative [ʒ] (Syed, 2013c), English [v w] (Syed, 2013a), and voiceless plosives (Syed, 2014) by a group of Pakistani students who were doing MA English in post-graduate colleges of Southern Punjab, Pakistan. All these students speak Saraiki as L1. Later on, Syed (2013a) also studied a group of advanced learners of English living in and around London. The participants of this study had obtained early education from the post-graduate colleges of the Southern Punjab. They

also speak Saraiki as L1. The findings of all these studies are summarized below:

1. Unlike native speakers of English who produce [d] as [+anterior]and with post-burst voicing, the Pakistani learners produce English [d] as retroflex (i.e. [-anterior]) and with pre-voicing. In terms of feature geometry (Clements & Hume, 1995), a sound is [+anterior] if the passive place of articulation is alveolar ridge or teeth; if the place of articulation is behind the teeth-ridge, the sound is [-anterior] or posterior. And if the tongue is curled back in production of a sound, the sound produced is called retroflex.
2. They assimilate English alveo-palatal fricative [ʒ] with the approximant [j] in perception and production.
3. They also could not differentiate between [v] and [w].
4. The Pakistani learners also could not produce aspirated and unaspirated allophones of English /p t/ phonemes with accurate voice onset time (VOT). However, they could produce aspirated and unaspirated allophones of English velar stop /k/. Voice onset time or VOT is time interval between the burst of a stop and onset of the following vowel (Docherty, 1992). VOT is measured in milliseconds and it indicates the quantity of air accompanied with the burst of a stop. Aspirated stops e.g. [p^h t^h k^h] have bigger VOT ranges than unaspirated stops like [p t k] (Roca & Johnson, 2007). VOT is considered a major acoustic correlate of stops (Foulkes, Docherty, & Jones, 2010). If voicing of vocal folds starts before the burst, it is called pre-voicing. Pre-voicing is measured in negative values. If voicing of vocal folds starts after the burst, it is called short-lag VOT and is measured in positive values. Aspirated sounds have long-lag VOT which are also measured in positive values.

Most of these findings may be ascribed to the interference of L1. The participants in all the studies quoted above speak Saraiki. Corresponding to English [v w], there is a single labio-dental approximant in Saraiki (Shackle, 1976; Varma, 1936). That is why [v w] pair makes a single category type of sound for these learners (Best & Tyler, 2007). Single category type of sounds are normally very difficult to differentiate for L2 learners (Best, 1994). English palato-alveolar fricative [ʒ] does not exist in Saraiki (L1), so it may be perceptually assimilated with another closer sound. Corresponding to English [d], the L1 of the students has a pre-voiced retroflex stop. Therefore, according to the spirit of contrastive analysis tradition started by Lado (1957), it is natural for Pakistani students to produce English [d] with retroflexion and pre-voicing. However, the L1 of participants of the studies referred to above has aspirated and unaspirated phonemes in its consonant inventory (Shackle, 1976). According to the

contrastive analysis hypothesis (Lado, 1957), these learners can easily acquire aspirated and unaspirated plosives of English. The findings of this study put a question mark on the contrastive analysis hypothesis. Some other studies conducted with Pakistani students also found the same results (Rahman, 1991). These studies challenge the contrastive analysis hypothesis and L1 interference tradition and invite us to investigate some other possibilities with a more scientific approach.

The current study was conducted with a hypothesis that some factors other than L1 interference also cause learning difficulties for Pakistani learners of English. These factors resist in acquisition of consonants of English. One of these may be inadequacy of input. The research question was whether the difficulties faced by Pakistani learners in acquisition of English consonants were due to the L1 interference, or improper input. For this purpose, a perception and production experiment was conducted with ten English language teachers in the three post-graduate institutions of southern Punjab from where the participants of the studies quoted in the above discussion were selected. The selection of the participants was based on convenient sampling. The experiment was conducted with a view that if the teachers perceive and produce English consonants accurately, the inaccurate perception and production of English consonants by Pakistani students and London-based Pakistani learners may be ascribed to the L1 interference, but if the teachers themselves face the same difficulties in perception and production of English consonants as the students do, then the problems lie in the input that Pakistani learners receive at schools and colleges during their academic setting. The teachers speak the same L1 (Saraiki) as their students. The details of the participants and of the experiment are given in the next section.

Research Methodology

Questionnaire and a perception & production experiment were major tools used for data collection. The questionnaire was served to each of the participants to elicit information about the linguistic and academic background. The details of the participants and experiments are given in the following sections.

Participants

Ten English language teachers were selected on the basis of availability. Their ages ranged between 23 and 48 years with an average of 33.90 (standard deviation (s.d. = 8.71). According to their own statements, they speak English for approximately 2.20 (s.d. = 1.14) hours daily. Before perception and production tests, a questionnaire was served to the participants to elicit the information given in this section. According to the

information provided by the teachers/participants, they all speak Saraiki as their L1. Their length of teaching English ranges between 4 and 29 years with an average of 13.20 (s.d. = 8.43) years. Although a sample size of ten participants seems apparently too small for developing big generalizations, but there were six repetitions for each of the stimuli in production task and two perception tasks for each of the target sounds in perception test (with three repetitions for each of the stimuli in identification task). Thus, three different kinds of tasks and several repetitions in each task yielded a large number of tokens for each of the stimuli which are fairly sufficient for reaching a solid conclusion about perception and production of the participants.

Perception Tests

Two tasks were arranged to determine perception of the participants of English [v], [w] and [ʒ]. The first one was an identification task in which VCV syllables recorded in the voice of a female native speaker of English were presented to the participants. The V in the stimuli was a low vowel [a] on both sides of the C which was a target sound. Thus, the stimuli for the identification task were [aʒa], [ava], [awa], etc. The low vowel was used in the stimuli because it is neutral in its effects on the adjacent consonant (Syed, 2011).That is why researchers prefer to use the low vowel [a] in experiments as a carrier of the target consonants (Guion et al., 2000).Some other CVC syllables were also included in the list of stimuli so that the participants may not know the target consonants. The participants were asked to identify the consonant between two vowels and write it on a given piece of paper in English and Urdu in the relevant column (See answer sheet in Appendix I). Since the respondents were English language teachers, they were also familiar with the IPA symbols. They were apprised that they could also use an IPA symbol instead of a Latin letter if they feel convenient with it. In this task each of the target sounds was presented three times. In this way a total of 30 (3 repetitions* 10 participants=30) responses were obtained.

A 3 alternative forced choice (hereafter 3AFC) discrimination activity was another task in the perception test. In this task, the CVC syllable followed by two more syllables of the same structure, one of which carried the consonant which may be confused with the target consonant, was presented to the participants. They were asked to discriminate the sounds from each other. The following stimuli were presented to the participants:

[ava]	[awa] [aja]
[awa]	[ara] [ava]
[aʒa]	[aja] [aða]

Some other sets of syllables of similar VCV structure (e.g. aba, ada, etc.) recorded by the same female native speaker of English were also included in the above list so that the participants may not know the target sounds. The participants were asked to answer by ticking in the relevant column of a given answer sheet whether the consonant in the first syllable was the same as that in the second or third syllable, or the consonant in the first syllable was different from the consonants in the second and third syllables. The main purpose of this task was if the participants could discriminate [v] from [w] and [ʒ] from [j]. Normally an AX discrimination test with two consonants is used in such experiments. In the current study, a third apparently irrelevant syllable was added to make the task relatively more difficult so that the participants are attentive to the task and decide on careful listening. Besides, previous research shows that there is more probability of guesswork in a simple AX discrimination task (Best, McRoberts, & Goodell, 2001, p. 782). Therefore, a relatively more complicated discrimination task was developed to avert guesswork. This task did not have any repetitions.

Production Task

The purpose of this task was to determine if the participants could produce English consonants /ʒ v w d/ and the allophones of English /p t k/ accurately. For this task, a list of written words of English was provided to the participants, and they were asked to read in natural normal speed the given words of English. The list had two sections; in the first section the target words were embedded in a carrier sentence and in another the target words were written in isolation. The carrier sentence was “I say . . . again” with the target word between “say” and "again.” The readings of the participants were recorded and analyzed acoustically. The purpose of recording the target words in sentences and in isolation was to see if there is any difference in production of the participants in isolated words and continuous speech. Previous studies show that sometimes the performance of L2 learners is different in production of isolated words and continuous speech (Birdsong, 2007). The target words in this task were *deal, Venus, weed, measure, pleasure, treasure, peak, keep, teeth, speak, ski,* and *steal*. Some other distracters were also included in the list of stimuli so that the participants may not know the target sounds. These recordings were analyzed acoustically using Praat (Boersma & Weenink, 2012) software. The results are presented in the following section.

Results

As delineated earlier, perception and production tests were arranged for [v w ʒ] sounds; but for English [d] and voiceless aspirated [p^h t^h k^h] and unaspirated stops [p t k], only production tests were arranged.

Perception tests were not arranged for voiceless English plosives because aspiration contrast is allophonic in English. Due to allophonic variance, there is a complementary distribution between aspirated and unaspirated plosives (Davenport & Hannahs, 2010). Since the two allophones do not make minimal pairs, there is no probability of perceptual assimilation between aspirated and unaspirated allophones of English plosives at semantic level. Similarly, no solid perceptual difficulties regarding [d] have already been reported in Pakistani learners of English. However, difficulties in production of these stops have already been reported in Pakistani learners (Mahboob & Ahmar, 2004; Rahman, 1990, 1991). Therefore, only production tests were arranged for English plosives. The results are presented in the following sub-sections.

Perception and Production of [v w]

Perception and production of English [v w] consonant pair by the teacher participants was studied in identification, discrimination and word-reading tasks. In the identification task, there were three repetitions for each of the target sounds giving 30 responses in all. Twenty times the participants identified English [v] and 10 times English [w] accurately. It is important to point out here that the participants were asked to write their responses in English and Urdu (See answer sheets in Appendix I). These are the responses of the participants in English. In the third column of the answer sheet in which they were asked to write their responses in Urdu language, the participants wrote Urdu letter "و " for both English [v] and [w]. The Urdu letter "و " represents the approximant sound in Urdu and other Pakistani languages including the L1 of participants. In other words, in their L2 phonemic inventory, they had developed a single representation for [v] and [w]. That single representation is the same as the labio-dental approximant of the L1 of the participants. However, since English provides two letters for these sounds, they sometimes wrote 'v' and sometimes 'w' in their responses. But actually, as the word-reading task shows, they had a single representation in their L2 phonemic inventory for these two sounds of English. This was also confirmed in the 3 AFC discrimination test. In 80% of the responses, the participants could not discriminate [v] from [w]. These results show that the participants cannot perceive English [v] and [w] as two different sounds.

In the production test, the participants' production of the words, 'Venus' and 'weed' were analyzed acoustically. As the previous studies show, Pakistani speakers of English equate English [v] and [w] with the labio-dental approximant of their L1. In other words, unlike native speakers of English, Pakistanis produce English [w] without lip-rounding and [v] without frication. For understanding the real nature of the sounds produced by the participants, the repetitions of the English word 'Venus'

produced by the participants were analyzed using Praat software (Boersma & Weenink, 2012). A concentration of acoustic energy above 5000 Hz is an acoustic correlate of fricatives (Ladefoged & Maddieson, 1996). The spectrograms of the productions of 9 out of 10 participants show no concentration of acoustic energy above 5000 Hz which means nine out of ten teacher participants had produced [v] without frication. As an illustration, one of the spectrograms, obtained in the analyses of the productions is given in figure 1.

As the spectrogram and waveform in figure 1 show, there is no concentration of acoustic energy in the initial part which reflects the production of [v]. The final part of the spectrogram shows that [s] in the word 'Venus' was produced with frication which is reflected in form of concentration of dark energy in the upper half of the spectrogram and also in the final part of the waveform. No such acoustic correlates are visible in the initial part of the spectrogram.

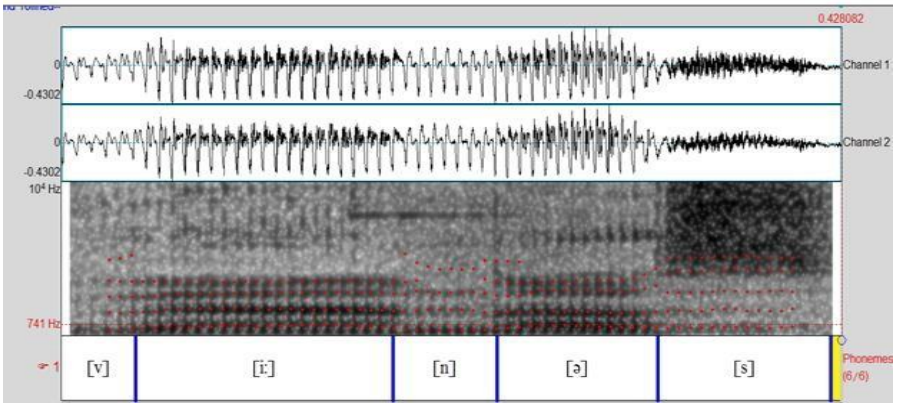

Figure 1: Spectrogram of the Word 'Venus' Produced by a Participant

For comparison, we produce below in figure 2, a spectrogram of the word ‘Venus’ produced by a female native speaker of English living in the suburbs of London.

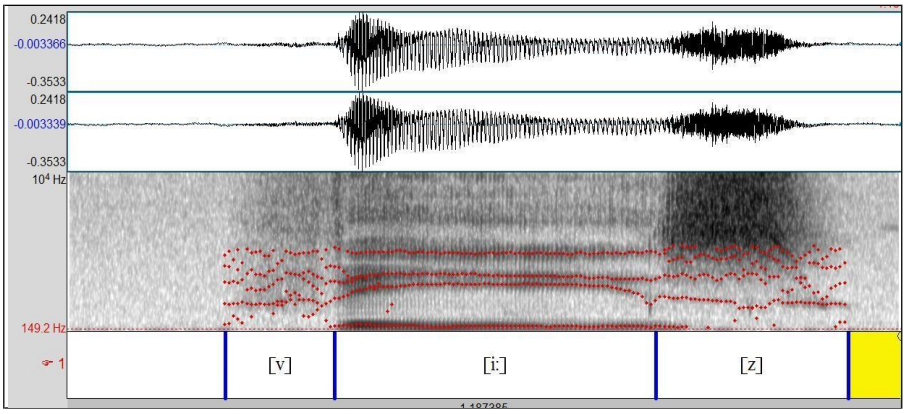

Figure 2: Spectrogram of the Word 'Venus' Produced by a Native Speaker of English

As the initial part of the spectrogram shows, the consonant [v] was produced with a fricative noise by the native speaker. The concentration of acoustic energy which indicates frication is there (though not so strong in the initial part of the spectrogram, which is lighter dark compared with the last part of the spectrogram because the last consonant in the word is a sibilant whereas the initial consonant [v] is a non-sibilant). In the spectrogram in figure 1, no such frication is apparent. The acoustic analysis reveals that [v] is produced by 9 out of 10 participants without frication. Only one of the participants produced [v] with frication but he also produced [w] with frication like [v].

It has already been observed that Pakistani learners produce English [w] without lip-rounding. One of the research questions, in the current study, was to determine whether Pakistani teachers also produce English [w] without lip-rounding? F3 lowering is an acoustic correlate of lip-rounding (Ladefoged, 2006). Human sound travels in waves making resonances at different frequencies each of which is called a formant (Spencer, 1996, p. 26). First formant is called F1, second is called F2 and the third one is called F3. For acoustic analysis, F3 in the productions of the word 'weed' by the participants was taken using Praat (Boersma & Weenink, 2012) (See Appendix I for details of F3).

There were six repetitions (3 repetitions in words and 3 in continuous sentences) of the target word 'weed' produced by each of the participants. Thus we obtained 60 tokens (6 repetitions*10 participants) of the word 'weed' for analysis. Consistency of participants in repetitions was 80% (Cronbach's alpha = .800) in [v] and 82% (Cronbach's alpha = .816) in [w]. Overall, there was neither a significant difference between the F3 of [v] and [w] of participants in exclusive words (Z = - .153, Sig. = .878) nor in

sentences (Z = -.968, Sig. = .333). The mean F3 of the participants' productions of [v] in exclusive words was not significantly different from that in continuous sentences (p>.1). However, the mean F3 was 80 Hz higher for [w] in words than that in continuous sentences. The difference was marginally significant (Z = -1.99, Sig. = .047). Both F3 values in words and sentences were combined and an average was taken. There was no significant difference between mean F3 of [v] and [w] obtained in six repetitions (Z = -.97, p>.3).

As pointed out earlier, F3 lowering is an indication of lip-rounding. These data show that the participants produce [w] without lip-rounding because there is no significant difference between the F3 of [v] and [w] produced by the participants. Had they produced [w] with lip-rounding, there average F3 would have been significantly lowered in [w] than in [v]. But no F3 lowering in the productions of the participants was observed. Thus, the results confirm that the participants produce [w] and [v] in the same way. Their [v]'s are without frication and [w]'s without lip-rounding.

Perception and Production of Alveo-Palatal [ʒ]

In the identification test, the participants identified English [ʒ] as [j] in 20 out of 30 trials. In 2 trials, they identified it as [z] and in the remaining 8 trials they identified it correctly. Therefore, the accuracy percentage in the identification of English [ʒ] is 26.67%. The participants identified it as approximant [j] in 66.67% of the trials. In the 3AFC discrimination test, they perceptually assimilated [ʒ] with [j] in 60% of the trials whereas in the remaining 40% of the trials they accurately discriminated it from [j].

In the production test, the participants produced English words 'measure, pleasure, treasure'. The spectrograms of the productions were analyzed acoustically. The spectrograms show that 7 out of the total 10 participants produced the target sound as approximant. There was no concentration of acoustic energy in the upper half of a spectrogram of 10^4 Hz resolution. This shows that 70% of the teachers cannot differentiate between [ʒ] and [j] in production and approximately 67% of them cannot differentiate between these consonants in perception. This conclusion is based on the identification test because normally Identification tests (not discrimination test) are accurate indicators of perception of L2 learners. It is because there is a probability of guesswork in discrimination test (Boersma & Hamann, 2009). That is why discrimination test results are normally better than identification test results (Archibald, 2005). Other studies with Pakistani learners also came up with the similar results (Syed, 2013b).

For an illustration we recorded the word 'measure' spoken by a female native speaker of English living in the suburbs of London. Figure 3 shows a comparative view of the same word produced by a native speaker and one of the teacher participants of this study.

A concentration of acoustic energy after the vowel in the left spectrogram shows that the consonant was spoken as the fricative [ʒ] by the native speaker whereas less clear formants and absence of acoustic energy after the vowel in the right spectrogram clearly indicates that the consonant after the vowel was produced as an approximant [j]. The right spectrogram is based on the production of one of the teacher participants and the left one is that of the female native speaker of English.

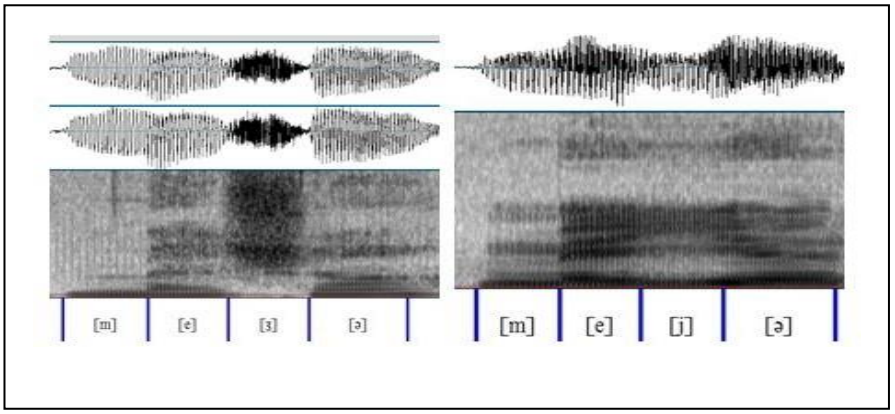

Figure 3: Spectrograms of the Word 'Measure'

Production of English [d]

The production of English [d] by the participants in the word 'deal' and 'weed' was recorded and analyzed acoustically. It is already observed that Pakistani learners produce English [d] with retroflexion and pre-voicing, whereas in native speech English [d] is produced without retroflexion with post-burst short-lag VOT. F3 lowering is an acoustic correlate of retroflexion (Hamann, 2005). For the study of retroflexion in productions of the participants, F3 values of the word 'weed' were taken using Praat. The F3 in the spectrograms of word 'weed' were taken at two points. F3 of midpoint of the vowel (when formants are stable) was compared with that of the same vowel in final phase of the same formant (F3) in which a vowel subsumes into [d] in the word 'weed.' If the word-final [d] is produced with retroflexion, the final phase of the F3 of the vowel must be lower than the F3 in midpoint of the same vowel, but if the final [d] is produced without retroflexion, F3 of the final part of the vowel should not be lower than that of midpoint. F3 of the word 'weed' (not

'deal') was studied for retroflexion because retroflexion is clearer on the spectrogram in vowel to consonant transition than in consonant to vowel transition (Hamann, 2005; Ladefoged & Maddieson, 1996). Each of the participants produced six repetitions of 'weed' (3 in words and 3 in sentences) in the word-reading task. The consistency among repetitions in F3 lowering was 86% (Cronbach's alpha = 0.861). The difference between F3 in words and sentences was non-significant (p>.1). The F3 obtained in words and sentences were averaged. The mean frequencies of F3 of the vowel in mid and in the final phase are given below.

Table 1: Frequencies of F3 of the Vowel in 'Weed' Produced by the Participants

Part of the F3 Studied	N	Minimum	Maximum	Mean	Std. Deviation
Final	10	2352.00	2783.17	2596.32	129.48
Mid	10	2551.33	3267.83	2822.20	246.28

A non-parametric test shows that difference between the means is significant (Z = -2.497, Sig. = .013). The means of F3 in table 1 show that F3 of the vowel in 'weed' was lowered in the final phase in which the vowel subsumes into the following consonant. This indicates that the participants produced [d] with retroflexion.

The following figure shows mean F3 of the vowel in mid and final phase. The trend line reflects the third formant. As figure 4 shows, third formant of the vowel is higher in the mid point when it is not influenced by the following consonant which is [d]. But it lowers when it subsumes into the following consonant in the final phase. This clearly indicates that the word final [d] was produced by the participants as a retroflex.

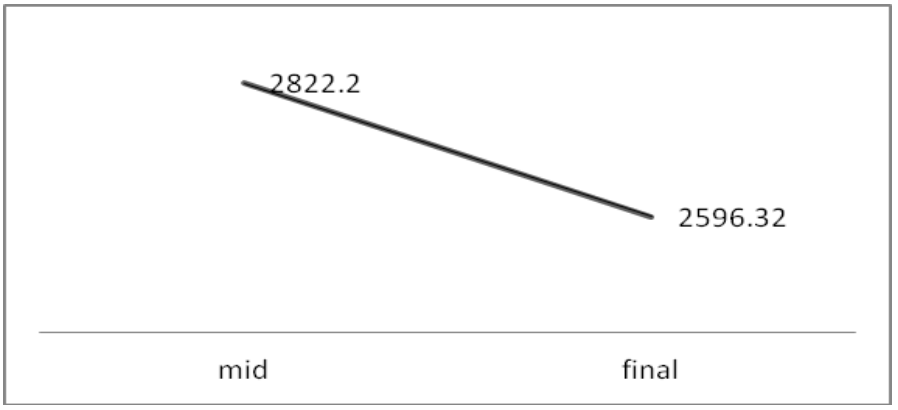

Figure 4: F3 of the Vowel in the Word 'Weed'

Another question related to this sound was whether the English language teachers produce [d] with or without pre-voicing. There were six repetitions of the target word carrying [d], three in exclusive words and three in sentences. The mean VOT of the participants' productions in words and sentences was -115.10 and -54.10 ms respectively. The consistency in the repetitions was 41% (Cronbach's alpha = 0.412) in words and 50% (Cronbach's alpha = 0.504) in sentences. There was a significant difference between the mean VOT of the participants' productions in words and sentences (Z = 2.599, Sig. = .009). These results show that the participants produce English [d] with pre-voicing although the duration of pre-voicing is longer in exclusive words than that in continuous sentences.

Production of English Voiceless Plosives

For determining acquisition of allophonic variance of English voiceless plosives, the participants were asked to produce words carrying English aspirated and non-aspirated plosives. Each of the stimuli had three repetitions as exclusive words and three repetitions in continuous sentences. A Cronbach's alpha reliability test was applied to determine consistency of the participants in repetitions. The results of the reliability test show that there was 75% to 90% consistency in the repetitions (See Appendix I for detailed results of the reliability test). The VOTs for plosives produced in words were not significantly different from those produced in continuous sentences (p>.1) (See Appendix I for detailed results). Therefore, the VOTs obtained in words and those obtained in sentences were merged and an average was taken. Table 2 shows the mean VOTs.

Table 2: Mean OTs

Sounds	N	Minimum	Maximum	Mean	Std. Deviation
[p^h]	10	7.67	41.17	17.33	10.72
[t^h]	10	8.67	43.17	19.34	11.30
[k^h]	10	29.17	101.17	47.63	20.72
[p]	10	5.67	51.67	16.90	14.99
[t]	10	12.67	56.50	26.63	16.92
[k]	10	19.00	65.83	34.48	16.70

A non-parametric test shows that the difference between VOT of aspirated and unaspirated labials is non-significant whereas that between the allophones of coronal and velar plosives is significant. The following table shows results of the non-parametric test.

Table 3: VOT of Aspirated and Unaspirated Stops Compared

	[p] - [p^h]	**[t] - [t^h]**	**[k] - [k^h]**
Z	-.714	-2.193	-2.599
Asymp. Sig. (2-tailed)	.475	.028	.009

This means that the teacher participants cannot differentiate between aspirated and unaspirated allophones of English /p/ and produce both allophones without aspiration. They produce the allophones of English /k/ with two separate ranges of VOT. Although the mean VOT for aspirated velar [k^h] produced by participants is smaller than mean VOT of native speakers of English, it is confirmed that the participants have two significantly different VOTs for aspirated and unaspirated allophones of English /k/ which indicates that the participants have two separate phonetic categories for these two sounds of English though the phonetic category for aspirated stop is a little deflected away from that of native speakers of English. It is interesting to find out that the participants produce English [t] with a mean VOT of 56.50 ms whereas they produce [t^h] with a VOT of 43.15 ms. The reason for this is that the participants produce English coronal /t/ on word-initial position without aspiration. A bigger VOT for [t] in the word 'steal' is because of some articulatory constraints which activate in such clusters (See Syed, 2013a) for detailed discussion about this issue). We finally conclude that the participants of this study have acquired aspiration contrast in English plosives at dorsal position only but they are unable to produce allophones of English labial and coronal stops with accurate VOT values.

Analysis, Discussion and Conclusion

The results presented in the previous section show that the participants of this study who are English language teachers in the same institutions from where student participants were taken in the studies quoted in section 1, cannot differentiate between [v] and [w] and [ʒ] and [j] of English. They also cannot acquire proper allophonic variance between aspirated and unaspirated allophones of English labial /p/ and coronal /t/ although they have acquired allophonic variance in aspirated dorsal [k^h] and unaspirated dorsal [k] allophones of English /k/. They also produce English [d] with retroflexion and pre-voicing. The same performance was observed in the studies conducted with the students of the same institutions from where the teacher participants of this study were selected (See results of studies conducted with students in (1) in section 1). These findings confirm that there is no difference between performance of the teachers and their pupils.

As pointed out earlier, both the student participants who participated in the studies quoted in section 1 and the teacher participants of this study are from the same area and speak the same L1. Following the classical tradition of contrastive analysis, we can ascribe some of learning difficulties faced by the teachers and students of English in Pakistan to interference of the L1. It may be justified in cases of perceptual assimilation between [v w] and [ʒ j] pairs of consonants. The difficulties faced by the student learners and their teachers in acquisition of [d] may also be ascribed to the L1 because voiced stops in the L1 of participants are produced with pre-voicing (Syed, 2013d). However, the results in production of voiceless stops /p t/ pose a big challenge to the proponents of contrastive analysis hypothesis. Saraiki, the L1 of the participants, has both aspirated and unaspirated stops (Syed, 2013d). Therefore, they are expected to perform better in acquisition of these sounds of English by positive transfer from the L1. The inability of the English language teachers and students of Pakistan in proper acquisition of English aspirated and unaspirated variants of English stops /p t/ shows that it is not only L1 interference which is root cause of difficulties in L2 acquisition. There are some other motivations for this. The student learners experience similar difficulty as experienced by their teachers in development of two separate categories of allophones of English labial and coronal plosives. This is not because of L1 interference. Rather, Pakistani learners could not acquire English aspiration contrast even at advanced stage of learning because they had received inaccurate input from their teachers at early stage of learning in Pakistan. The findings of this study confirm the hypothesis that the input which the students receive from their teachers also plays effective role in L2 acquisition not only while the learners are acquiring a language; it rather continues exerting influence in later life. The reason is that the phonetic categories of L2 consonants once acquired are extremely difficult to re-structure or unlearn. Other researchers also support this view (Flege, 2009).

We do not totally negate the role of L1 in L2 acquisition. L1 interference is of course one of the major reasons for inaccurate production of L2 sounds but it is not the only reason for that. The problem in Pakistan started after the departure of English rulers from the subcontinent. English was introduced in the subcontinent of India and Pakistan in sixteenth century C.E. (Baumgardner, 1990, p. 59). After departure of the British colonial rulers, there was no native speaker model in front of Pakistani learners. As a result, the first generation of English language teachers of Pakistan depended on written English only. Since English orthography does not differentiate between aspirated and unaspirated plosives, they could not differentiate between these sounds of English. After some generations, this continuous practice fossilized and

turned into what is now called Pakistani English. Pakistani English claims the status of a variety of English (Rahman, 1990). In this way, the initial deviation from native Standard English pronunciation started due to different factors including absence of native input and interference of the L1. Later on, it transferred to students cyclically and became a standard practice. The practice is so strong that even those Pakistani learners who live in London also cannot deviate from this practice. Thus, we conclude that it is the input which, along with other factors like L1 interference, plays effective role on L2 acquisition. Therefore, if we want improvement in our students, first we need to improve our teachers.

Appendix I

Mean F3 [v w]

Context	N	Minimum	Maximum	Mean	Std. Deviation
F3 of [v] in Words	10	2333.00	3136.00	2703.50	287.11
F3 of [w] in Words	10	2490.33	2968.00	2697.17	167.48
F3 of [v] in Sentences	10	2470.33	2937.00	2678.30	166.02
F3 of [w] in Sentences	10	2228.00	2831.33	2618.00	189.40
F3 of [v] Combined	10	2420.50	2947.50	2690.90	203.059
F3 of [w] Combined	10	2366.00	2862.50	2657.58	165.57

Reliability Test Applied on Six Repetitions

	[p^h]	[t^h]	[k^h]	[p]	[t]	[k]
Cronbach's Alpha	.762	.875	.931	.866	.906	.904
%age Consistency	76%	88%	93%	87%	91%	90%

A Comparison of the VOTs Produced in Words and Sentences

	[p^h]	[t^h]	[k^h]	[p]	[t]	[k]
Z	-.178	-.867	-1.326	-.919	-1.174	-.561
Sig. (2-tailed)	.858	.386	.185	.358	.241	.575

Appendix II

Identification Test Answer Sheet

The participants were asked to listen to the stimuli and respond by writing in the relevant column in Urdu and English, which consonant they heard between two a's. They were apprised that if they feel they did not know any letter of English and Urdu script which represents a sound in the stimuli, they might point out that in writing the last column.

Answer Sheet for Identification Test			
S. No.	English	Urdu	Remarks
1	a…….a		
2	a…….a		
3	a…….a		
4	a…….a		
5	a…….a		
6	a…….a		
7	a…….a		
8	a…….a		
9	a…….a		
10	a…….a		
11	a…….a		
12	a…….a		
13	a…….a		
14	a…….a		
15	a…….a		
16	a…….a		
17	a…….a		
18	a…….a		

References

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

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

Baumgardner, R. J. (1990). The indigenization of English in Pakistan. *English Today, 6*(1), 59-65.

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., 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., & 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: John 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.

Boersma, P., & Hamann, S. (2009). Loanword adaptation as first-language phonological perception. In A. Calabrese & W. L. Wetzels (Eds.), *Loanword phonology* (pp. 11-58). Amsterdam: John Benjamins.

Boersma, P., & Weenink, D. (2012). Praat: Doing phonetics by computer. Computer program. Retrieved September 20, 2012.

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.

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.

Davenport, M., & Hannahs, S. J. (2010). *Introducing Phonetics and Phonology* (3rd ed.). London: Hodder Education.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. Berlin: Foris Publications.

Eckman, F., & Iverson, G. K. (2013). The role of native language phonology in the production of L2 contrasts. *Second Language Acquisition, 35*(1), 67-92.

Edens, K. M. (2000). Preparing problem solvers for the twenty-first Century through problem-based learning. *College Teaching, 48*(2), 55-60.

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., 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.

Foulkes, P., Docherty G. J., & Jones, M. J. (2010). Analyzing Stop. In P. Foulkes, G. J. Docherty & M. J. Jones (Eds.), *Sociophonetics: A student's guide* (pp. 58-71). London: Routledge.

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.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107*(5), 2711-2724.

Hamann, S. (2005). The diachronic emergence of retroflex segments in three languages. *Link, 15*, 29-48.

Hansen, E. J., & Stephens, J. A. (2000). The ethics of learner-centered education: Dynamics that impede the process. *Change, 32*(5), 41-47.

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.

Iverson, P., Wagner, A., Pinet, M., & Rosen, S. (2011). Cross-language specialization in phonetic processing: English and Hindi perception of /w/-/v/ speech and nonspeech. *Journal of Acoustical Society of America, 130*(5), EL 297-EL 303.

Kain, D. J. (2003). Teacher-centered versus student-centered: Balancing constraint and theory in the composition classroom. *Pedagogy, 3*(1), 104-108.

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

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

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

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.

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.

Roca, I., & Johnson, W. (2007). *A course in phonology*. Oxford: Blackwell.

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

Spencer, A. (1996). *Phonology : Theory and description.* Oxford: Blackwell.

Syed, N. A. (2011). Perception and production of consonants of English by Pashto speakers. *The Journal of Humanities and Social Sciences, XIX*(1), 119-146.

Syed, N. A. (2012). Perception and production of English [d] by Pakistani L2 learners. *Essex Graduate Students Papers in Language and Linguistics, 13*, 134-157.

Syed, N. A. (2013a). *The acquisition of English consonants by Pakistani learners* (Unpublished Ph.D Dissertation). University of Essex, Colchester.

Syed, N. A. (2013b). Acquisition of alveo-palatal fricative [ʒ] by Pakistani learners of English. *Journal of Social Sciences and Interdisciplinary Research*, *2*(1), 39-50.

Syed, N. A. (2013c). *Role of equivalence classification in L2 acquisition.* Paper presented at the Language at the University of Essex Postgraduate Conference (LangUE), 2012. Colchester: University of Essex.

Syed, N. A. (2013d). Voice onset time for plosives in Saraiki: Implications for the acquisition of English aspiration contrast. *EFL Annual Research Journal, 14*, 1-15.

Syed, N. A. (2014). Influence of L1 laryngeal contrast in acquisition of allophonic variance in English plosives. *Balochistan Journal of Linguistics, 1*, 45-65.

Varma, S. (1936). The phonetics of Lahnda. *Journal of Royal Asiatic Society of Bengal, 2*, 47-118.

Villaume, S. K. (2000). The Necessity of Uncertainty: A Case Study of Language Arts Reform. *Journal of Teacher Education, 51*, 18-25.