Acquisition of New L2 Sounds without Separate Category Formation*

Nasir A. Syed

This study focuses on acquisition of English [b g] by Pakistani learners. Three groups of learners, two from Pakistan and one from England participated in this study. The findings are based on acoustic analyses of the target sounds produced in words and sentences. The findings demonstrate that Pakistani learners of English experience difficulty in the acquisition of native-like VOT for English voiced stops [b g]. They rather transfer L1 VOT values for English [b g] producing them as pre-voiced. The results are analyzed in light of the speech learning model (Flege 1995) of second language acquisition and element theory (Backley 2011). The findings of this study accord with those of the previous studies which claim that acquisition of voiced stops of aspiration languages like English is difficult for the speakers of voicing languages. The study also points out that the speech learning model lacks a solid scale to accurately calculate L2 learning.

Key words: *English, L2 Acquisition, pre-voicing, Saraiki, VOT*

English was introduced in the Subcontinent (India and Pakistan) in the 17th century (Baumgardner 1993). It has been the official language in this area since 1857. After Pakistan came into being in 1947, English speaking rulers returned to their homeland but their language remained the official language of the newly created state of Pakistan and still enjoys the same status in the country. It is taught as a compulsory subject of study at all levels of education starting from class 1 onto BA as a compulsory part of study even in the educational institutions which use Urdu, the national language of the country, as a medium of instruction. Therefore, anyone educated in Pakistan knows English. Saraiki is a Pakistani language of the Indo-Aryan family (Masica 1993, Varma 1936). The current study focuses on the acquisition of English voiced stops [b g] by adult Pakistani learners who speak Saraiki as L1.

The paper is divided into the following five sections. Section 1 provides a theoretical background relevant to the study and section 2 is based on the research methodology used. The results are presented in section 3 and analyzed and discussed in section 4. The paper ends with a conclusion in section 5.

1 Theoretical Background

The current paper studies the acquisition of English [b g] by Pakistani learners. The study focuses on the voice onset time of the target consonants. The current study is based on the postulates and paradigms of the speech learning model (Flege 1995) and element

* My special thanks are due to the organisers of CECIL'S 2 and Visegrad Foundation for providing financial sponsorship which enabled me to participate in the conference and the post-conference Summer School at Pázmány Péter Catholic University Piliscsaba (Hungary).

theory (Harris 1994, Backley 2010). In the following subsections, a brief introduction of VOT, element theory and the speech learning model is given.

1.1 Voice Onset Time

Voice onset time, commonly called VOT, is the time interval between the burst of a stop and onset of the following vowel (Docherty 1992). It is calculated in milliseconds. Lisker & Abramson (1964) divide sounds of world languages into three classes on account of VOT ranges; first, short-lag VOT, if the vocal folds of speakers start vibrating soon after the burst, such sounds are phonologically voiceless, unaspirated consonants. (However, in languages like English (Kager et al. 2007) and German (Hamann 2011), voiced stops are also produced with short-lag VOT). Second, if the vocal folds start vibrating long after the burst, this is called long-lag VOT and the sounds are called voiceless, aspirated consonants. Third, if the vocal folds start vibrating before the burst, this is called lead-voicing and the sounds are called pre-voiced or truly voiced consonants. The duration of lead-voicing is calculated in negative values. The examples of languages having pre-voiced stops are Japanese (Nasukawa 2010), Saraiki (Syed 2012a), Arabic (Flege & Porte 1981), Hungarian (Lisker & Abramson 1964), Dutch (Simon 2009, 2011), Spanish (Flege & Eefting 1988), Russian (Backley 2011), etc.

The languages of the world are divided into aspiration and voicing languages on account of the nature of laryngeal contrasts for consonants (Backley 2011, Iverson & Salmons 1995, Harris 1994). The voicing languages differentiate between phonologically voiced and voiceless unaspirated stops on the basis of the feature [voice] whereas the aspiration languages differentiate between these consonants on the basis of the feature [spread glottis]. Phonologically voiced stops are normally produced with pre-voicing (negative VOT) in the voicing languages but with short-lag VOT in the aspiration languages. The examples of aspiration languages are German (Hamann 2011), English (Honeybone 2005), Swedish, Korean, Icelandic (Backley 2011), etc. The L1 of the participants of this study (Saraiki) is a voicing language and the L2 (English) is an aspiration language.

There is a large body of literature on the acquisition of English stops by adult L2 learners of different linguistic backgrounds. It has been observed that L2 learners of English who speak voicing languages can acquire English aspirated stops but face a lot of difficulty in the acquisition of English voiced stops. The studies by Simon (2009) on Dutch learners, Syed (2012b) on Pakistani learners and Flege & Eefting (1988) on Spanish learners show that the speakers of Dutch, Saraiki and Spanish which are voicing languages, can acquire English voiceless aspirated stops. However, the studies by Flege & Porte (1981) on Arabic learners, Simon (2009, 2011) on Dutch learners and Shimizu (2011) on Korean, Thai and Japanese learners of English show that it is very difficult for the L2 learners who speak voicing languages to acquire voiced stops of English with short-lag VOT.

1.2 The Speech Learning Model (SLM)

Several models of second language acquisition emerged during the last five decades after Lado (1957) which attempted to find out the reasons for errors in the acquisition of L2 sounds in L1 grammar. The speech learning model (Flege 1995) is one of these models. According to the SLM, L2 learners develop equivalence classification between an L2 and the corresponding L1 sound if they do not perceive some phonetic difference between

the two sounds. Such equivalence classification results in the blockage of a new phonetic category for the new L2 sound in the L2 phonemic inventory of learners. In such a context, the learners have the same representation for both the new L2 and the corresponding L1 sound in their L2 phonemic inventory. In some cases of equivalence classification, L2 learners can perceive some difference between an L2 sound and the corresponding L1 sound but they do not perceive the difference big enough to enable them to develop a new phonetic category for the L2 sound. In such situations, learners develop a representation for the new L2 sound which is a merger of the phonetic characteristics of the new L2 and the closest L1 sound. This is called the 'merger hypothesis' (Flege 1987).

The SLM further claims that if L2 learners perceive phonetic distance between a new L2 sound and the corresponding L1 sound, they develop a separate phonetic category for the new sound. Thus, according to the SLM, there may be three possible learning outcomes; first, strong equivalence classification between L1 and L2 in which case learners simply transfer to the new L2 sound the characteristics of the closest L1 sound; second, weak equivalence classification in which case a little improvement is observed in learners' performance as a result of which they develop a representation for a new L2 sound which is a merger of the L2 and the corresponding L1 sound; third, a situation in which L2 learners perceive some phonetic distance between a new L2 and the closest L1 sound clearly and develop a separate phonetic representation for the L2 sound. However, the SLM further predicts that the phonetic representation developed by L2 learners for the new L2 sound may be deflected away from that of monolinguals of the L2 (Flege 1995, 239).

As mentioned before, it has been observed that for speakers of a voicing language it is very difficult to acquire accurate phonetic representations for L2 voiced stops of an aspiration language. In light of the existing literature, we develop a hypothesis that it is very difficult for Pakistani learners of English to acquire English voiced stops [b g][1] with accurate VOT ranges because the L1 of these learners is a voicing language (i.e. has stops with lead-voicing) whereas English is an aspiration language (i.e. has voiced stops which are commonly produced with short-lag VOT). The current study aims to test this hypothesis.

1.3 Element Theory (ET)

Feature geometry factors out sounds into smaller properties called features (Clement 1985, 225). The feature geometry uses place features like [labial], [coronal] and [dorsal] to

[1] The reason for not including [d] in this study is that the English voiced alveolar stop [d] does not exist in Saraiki as such. Corresponding to the English alveolar [d], Saraiki has a retroflex and a dental stop. In other words, Saraiki coronal stops are not only different from English [d] phonetically in terms of VOT, but also they are different phonologically in terms of coronal features in that the Saraiki dental stop is [+anterior, +distributed] and retroflex is [-anterior, -distributed] whereas English [d] is [+anterior, -distributed]. In the language of Element Theory, English alveolar [d] is ⎪A⎪ whereas the Saraiki dental stop is ⎪I⎪ and the Saraiki retroflex is ⎪A⎪. Owing to these differences, the English coronal stop [d] does not make part of this study. The plosives [b g] have been exclusively selected for this study also because, according to the ET classification, they make a single class by sharing U element (Backley 2011) (although some proponents of the Element Theory (e.g. John Harris & Geoff Lindsey) disagree with the idea of labials and velars sharing a single element.) Some models of feature geometry (e.g. Rice & Avery 1993) also consider labial and velar stops lying under the same peripheral node of place of articulation, thus classifying them into a single type of sound.

explain places of articulation of sounds. It uses features like [spread glottis], [constricted glottis], [voice], etc. for laryngeal specifications of sounds and the features [nasal], [continuant], etc. for the manner of articulation of sounds (Botma, Kula & Nasukawa 2011). But element theory, on the other hand, uses the elements U, I and A for different places of articulation.

According to Backley (2011, 5) "An element is the smallest unit of segmental structure to appear in phonological representations." The Element Theory (ET) does not assume separate class of elements for laryngeal settings and manner of articulation (ibid, 114). Rather, it uses the same elements H ʔ L for aspiration, manner and voicing. H represents frication as a primary position and aspiration as a secondary articulation. L reflects voicing and nasality whereas the element ʔ represents stops. The element-based representation of English allophones [p^h p b] and Saraiki pre-voiced [b] is given in the following figure adopted with some modifications from Backley (2011, 140, 151).

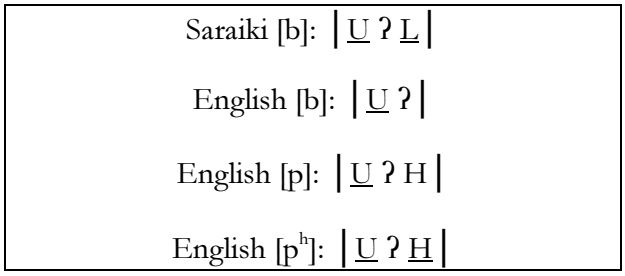

Figure 1: ET representation of plosives[2]

The above representation shows that in Saraiki, the phoneme /b/ is a bilabial (U) stop (ʔ) which has pre-voicing (L) as a salient acoustic feature, and place and pre-voicing are head elements in Saraiki /b/ whereas the absence of the element L from the representation of the phoneme /b/ in English indicates that it lacks pre-voicing. The presence of H as a head element in the phonological representation of the voiceless aspirated allophone [p^h] of the phoneme /p/ in English indicates the existence of strong aspiration (long-lag VOT) as a salient acoustic cue whereas its presence without headedness in [p] indicates weak aspiration (short-lag VOT). The head elements are underlined in element theoretic representation. In the succeeding sections we will recapitulate some relevant studies on the acquisition of English stops by learners of voicing languages like Dutch and Saraiki.

[2] According to an anonymous reviewer, element theoretic expressions may not have two heads. However, the current paper follows the Element Theory version presented by Backley (2011) which uses two heads for expressing some of the phonemes as reproduced above. It is also worth pointing out that L without having headed representation in the Element Theory stands for nasality. Since nasality is not topic of discussion in this paper, non-headed L does not make part of the discussion. It is also important to note the difference between H and L in the Element Theory. H stands for aspiration as a secondary articulation. If it is headed, it represents voiceless aspirated stops and if it is non-headed, it represents unaspirated voiceless stops because phonology differentiates between aspirated and unaspirated stops which are different from each other on the basis of gradient phonetic difference in the quantity of aspiration or VOT. But there is no such phonological difference between less pre-voiced and more pre-voiced stops in L languages so headed and non-headed representations of L do not reflect two different allophones of the pre-voiced stops.

2 Literature Review

Simon (2009) studied the acquisition of English voiced and voiceless aspirated stops by Dutch learners. In Dutch, voiced stops are pre-voiced but in English they are produced with short-lag voicing. Dutch does not have voiceless aspirated stops. It only has voiceless unaspirated stops. The aim of this study was to determine if Dutch learners of English transfer their L1 laryngeal contrasts by equating the short-lag and long-lag stops of English with the pre-voiced and short-lag stops of Dutch respectively, or they develop new representations for English [b p^h]. English voiceless stops were studied in two stressed positions, namely stops followed by vowels and those followed by sonorants (e.g. in words like 'play, pray,' etc.).

For the study, 16 Dutch advanced learners of English selected from a university, were asked to speak Dutch and English in 8 groups of dyads (two persons in each group) for 30-45 minutes. The participants spoke Dutch first and English later in a spontaneous conversation session which was recorded. The second task of the experiment was to read some words presented on a screen with 3 seconds inter-stimulus-intervals (ISI). Ten out of the sixteen Dutch subjects who had taken part in the spontaneous conversation task also participated in the word-reading task. Thirty-seven Dutch and thirty eight English words spoken by the Dutch informants were recorded in all. Out of 38 English words, 20 carried word initial voiced stops, 3 carried voiceless stops in word-initial stressed position, and 5 were word-initial clusters of stop+sonorants (e.g. pray, play etc).[3] The remaining words started with fricatives (which served as control sounds). Ten native speakers of English were also recorded as a control group for comparison. The target sounds were edited and transcribed using Praat.

The first research question in this study was related to the acquisition of the voiced stops of English by the Dutch learners. The reading list (stimuli) had 20 words of English and Dutch, each with voiced stops in initial position, which gave 200 tokens of the voiced stops in English and an equal number of tokens in Dutch spoken by the Dutch learners of English. An equal number of tokens in English were also recorded from the English control group. The results show that 93% of Dutch and 92.5% of English voiced stops were pre-voiced in the reading of the Dutch learners. On the other hand, 72.5% of the English voiced stops produced by the native speakers of English (the control group) were not pre-voiced. The results show that the Dutch learners of English had failed to suppress the transfer of the L1 pre-voicing to the L2 plosives.

The study of aspiration contrast was another objective of this research. For this purpose, voiceless word-initial stops in stressed syllables were selected where aspiration is quite clear. For voiceless aspirated stops, 525 tokens of English words spoken by the Dutch participants were analyzed, out of which 509 were used for further discussion. In the word-reading task, the Dutch speakers produced voiceless stops word-initially with 80 msec. VOT in the English words and 21 msec. VOT in the Dutch words. The average VOT of the native English speakers (the control group) for the voiceless stops of English in stressed word-initial position was 76 msec. However, the VOT of the Dutch speakers in the production of word-initial stressed English voiceless stops in spontaneous English speech (continuous conversation) was 48 msec. which is much less than that of the English native speakers (control group). These results show that the learners had acquired English aspirated stops with accurate VOT values (80 msec.) in

[3] The results relating to stop+sonorant are not discussed here because they are irrelevant for the present study.

word-reading task but not in the spontaneous conversation task, because their VOT is 48 msec. for aspirated stops in spontaneous conversation task whereas the VOT of the native speakers is 76 msec. for the same sounds.

In light of these results, the author concludes that the learners transfer pre-voicing to English stops from the L1. Overall, English aspirated stops with long-lag VOT have been acquired by the Dutch learners in exclusive word-reading task, but those with a short-lag VOT have not been acquired. In the opinion of the author, it is because of the acoustic salience of aspiration cues (Simon 2009, 401) that the learners perceive the English aspirated stops easily. It is also possible, in the opinion of the author, that the participants have shown good performance in the acquisition of aspirated sounds on account of the training that they had received from their university, since they were learning English and had already been taught to produce aspirated stops. However, they had not been taught to produce the voiced stops of English, because voiced stops are part of the phonemic inventory of their L1, whereas the aspirated phonemes are new for them. Regardless of the reason, the results of the study show that the Dutch learners acquired voiceless aspirated stops but they could not acquire voiced stops of English.

Syed (2012b) studied the acquisition of English voiceless aspirated and unaspirated stops by a group of 29 adult Pakistani learners of English who were doing MA in English in Pakistan. The study was focused on the acquisition of allophonic variance of English voiceless stops. The stimuli used in the experiment were English words 'peak, speak, teeth, steal, key, ski.' The participants were asked to read a list containing these words in continuous sentences and as isolated words. There were six repetitions for each of the words. The results were based on 174 repetitions (6 repetitions* 29 participants) of each of the allophones of the English plosives. The results show that Pakistani learners of English acquired aspiration contrast in velar stops but they did not acquire the same contrast for bilabial stops. They rather neutralized the aspiration contrast on bilabial position and produced both aspirated and unaspirated bilabial stops ([p] and [p^h]) as unaspirated [p]. The current study is based on the acquisition of the voiced stops [b g] of English by the same group of Pakistani learners along with four other groups of adult learners.

As pointed out earlier, the previous research on acquisition of English stops shows that it is relatively easier for L2 learners speaking voicing languages to acquire voiceless stops of English but more difficult for them to acquire voiced stops of English. The current study aims to investigate if Pakistani learners who speak a voicing language can acquire voiced stops of English.

3 Research Methodology

The current study focuses on the acquisition of English [b g] phonemes by adult Pakistani learners of English. The details of the participants and the methods used for data collection are explained in the following sub-sections.

3.1 Participants

A total of 105 subjects participated in this experiment. The participants were divided into five groups. Group one comprised 29 students of MA English in Pakistan. In this study they will be called '*Student*' learners. The second group of participants was of 30 teachers in Pakistan who were teaching different subjects of study in post-graduate colleges but

none of them was teaching English language. They will be referred to as '*Teacher*' group in this study. The participants of both these groups were selected from public sector colleges of central Pakistan. The third group comprised 22 UK-based Pakistani learners of English. They will be referred to as '*UK*' group in the following discussion. The UK-based learners migrated from those areas of Pakistan where the first two groups of Pakistan-based learners of English were living. The UK participants had been living in England in the County of Essex for an average of 65 months (standard deviation 77.45) at the time of the experiment. All the participants of these three groups speak the same L1 Saraiki which is predominantly spoken in central Pakistan (Shackle 1976).

Two control groups of participants were also selected to obtain the VOT values for the target sounds from monolingual speakers of English and Saraiki. For this purpose, ten Saraiki monolinguals were selected from the same area to where the three groups of learners belonged. They will be referred to as '*Saraiki*' monolinguals in the following discussion. A number of 14 native speakers of English, who were living in Essex in the same area from where the UK-based Pakistani learners of English were selected, also participated in the experiment. They will be referred to as '*native English*' (NE) group in the following discussion. The purpose of selecting the control groups and the study groups from the same areas was that the learners may be judged against the same standard of the L2/L1 which they were listening around them. The details of the participants are given in the following table.

Group	**No**	**Age (in years)**	**Speaking English (No. of hours/day)**	**Listening English (No. of hours/day)**
Student	29	22.28 (2.56)	2.03 (1.21)	2.79 (1.61)
Teacher	30	32.66 (7.8)	1.23 (1.5)	0.63 (0.96)
UK	22	32.91 (7.46)	5.73 (2.73)	5.36 (3.05)
Saraiki	10	29.6 (10.32)	--	--
NE	14	45.92 (22.09)	--	--

Table 1: Details of the participants

The above table shows the mean age of the participants in years and number of hours they speak and listen to English per day. The standard deviations are given in the parentheses. The above table shows that the UK-based learners speak and listen to English more than five hours daily whereas the 'Student' learners speak English for almost two hours and listen to it for 2.79 hours daily. The 'Teachers' listen to English less than an hour and speak for approximately 1.23 hours daily. It is important to point out that the two Pakistan-based groups of learners only listen to Pakistani English

whereas the 'UK' group of participants mostly listen to the English spoken by native speakers.

3.2 The Data Collection

All the participants were given a word-reading task. English words *beak* and *geese* written on a paper exclusively and in a carrier sentence were given to the participants. The carrier sentence was *I say beak/geese again*. The list of the stimuli also had some other words along with the target words so the participants did not know the target words. They were asked to read in a natural normal speed the list of sentences and words. Each word had six repetitions in the list, three times in the carrier sentence and three times as exclusive word. The purpose of recording the target words in two different contexts i.e. carrier sentence and exclusive words, was to find out if there was any difference in the VOT of the participants in continuous speech and exclusive words. Previous studies (Birdsong 2007) show that accuracy in word-production is a necessary but not sufficient condition of L2 acquisition because competence in continuous speech implies competence in exclusive word-production but not vice versa. M-audio track-II digital recorder was used for recording the productions and Praat (Boersma & Weenink 2012) was used for acoustic analyses of the target sounds.

The Saraiki monolinguals did not read English words. They were given another list of Saraiki words starting with [b g] sounds of their L1. They also produced words carrying the target sounds three times in a carrier sentence and three times as exclusive words. The words of Saraiki used as stimuli were *beebee* ('sister') and *geesee* ('slope') and the carrier sentence was *ay beebee/geesee hey*. ('This is a sister/slope').

4 Results

The mean VOT values obtained in the context of the continuous sentences and those obtained as exclusive words by all five groups of the participants were compared. There was no significant difference between the two VOTs ($p>.1$). Therefore the VOT values obtained in both contexts were merged together. The following table shows the mean VOT values of the target sounds produced by the participants in 6 repetitions (3 repetitions in words and 3 in sentences).

Group	[b]	[g]
NE	8.91 (2.40)	31.71 (4.73)
UK	-77.18 (55.45)	-41.77 (52.70)
Teacher	-107.41 (27.67)	-76.04 (28.99)
Student	-106.92 (22.85)	-70.89 (23.85)
Saraiki	-99.77 (14.98)	-72.81 (13.53)

Table 2: The average VOT values

In the above table, the average VOT values of the participants are given with standard deviations in the parentheses. The table shows that the L2 learners and the Saraiki monolinguals produced the target sounds with negative VOT values. The Saraiki monolinguals produced the target sounds in their L1 with lead-voicing which confirms that the L1 of the participants has pre-voiced stops.[4] However, the native speakers of English (NE group) produced these sounds with short-lag VOT.

A repeated measures ANOVA with place of articulation (labial and velar) as within group contrast and grouping as between group factor shows that the place of articulation has a highly significant ($F=98.544$, $p>.001$) effect on the VOT. Overall group variance is also strongly significant ($F_{4,100}=44.177$, $p>.001$). However, Bonferroni post hoc comparisons further confirm that the differences between the mean VOT values of the 'Teacher, 'Students' and 'Saraiki' monolinguals are non-significant ($p>.1$). The UK group of learners and NE group are significantly different from each other as well as from all other groups ($p<.005$). The interaction between grouping and place of articulation is non-significant ($p>.1$).

The individual results show that all Saraiki monolinguals produced [b g] with negative VOTs in all repetitions and all native speakers of English produced them with short-lag VOTs in all repetitions. Most of the learners produced [b g] with negative VOT in most of the repetitions but with positive VOT in some of the repetitions as the following table shows.

[4] This is the first detailed empirical study on the VOT of Saraiki, so there is no previous research which could confirm that voiced stops in Saraiki are pre-voiced.

Repetitions	No. of the participants who produced repetitions with positive VOT					
	[b]			**[g]**		
	UK R=132[5] N=22	Teacher R=180 N=30	Student R=174 N=29	UK R=132 N=22	Teacher R=180 N=30	Student R=174 N=29
0	10	23	28	9	21	15
1	2	4	1	1	3	13
2	2	2	0	1	2	0
3	1	1	0	3	2	1
4	3	0	0	2	2	0
5	2	0	0	4	0	0
6	2	0	0	2	0	0
Total positive VOT Repetitions	43	15	1	52	21	16
Percentage	32.58	8.33	0.57	39.39	11.67	9.20

Table 3: No. of time [b g] were produced with short-lag VOT

The above table shows that most of the learners did not produce even a single token of English [b g] with short-lag VOT and only two of the 'UK' participants produced both target sounds with short-lag VOT in all six repetitions consistently. The remaining L2 learners produced some of the repetitions with short-lag but most of them with lead-voicing. Among all three groups of learners, the performance of the 'Student' group in the production of [b] is most L1-like in that only one out of 29 participants of this group produced English [b] with short-lag VOT in one repetition only. As the total number of repetitions in different groups varies, the productions with short-lag VOT are reflected in the following figure in percentage to compare the group performance.

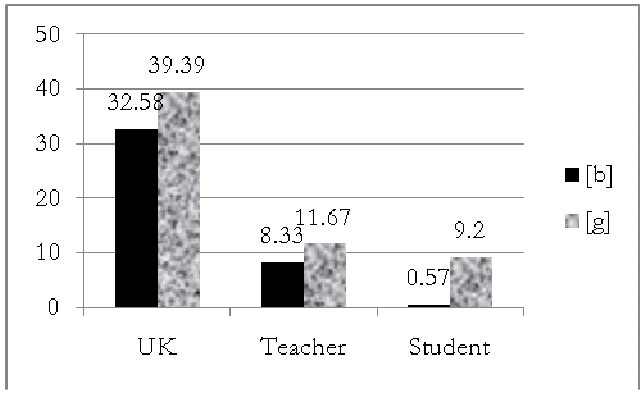

Figure 2: No. of time (in percentage) a sound was produced with short-lag VOT

[5] In this row R stands for total number of repetitions and N for total number of participants. Total numbers of repetitions are obtained by multiplying total number of participants with 6 because there were six repetitions for each sound by each of the participants (3 in words and 3 in sentences). However, in the columns only the number of participants is given who produced English [b g] with short-lag VOT.

The above figure and table show that the 'Student' group of learners are the poorest and the 'UK' group are the best in the production of English [b g] with short-lag VOT. The results also show that the performance of the participants is better on [g] than on [b]. Overall, none of the groups as a whole could acquire English [b g] with accurate VOT. We shall discuss and analyze these results in the following section.

5 Analysis and Discussion

The results presented in the preceding section show that the 'Student' group of participants performed worst of all as only one of the 'Student' participants produced one out of 174 repetitions (29 participants* 6 repetitions) of [b] with English-like short-lag VOT. The performance of the 'UK' group is the best of all. Although the UK-based participants could not acquire English-like VOT range for the target sound, but they performed significantly better than the other two groups. The 'Teacher' group of participants is in between the other two groups of learners in their performance. The performance of the participants on [b] and [g] is different but the pattern of learning is the same in that only a little improvement is observed in the 'UK' group on both sounds. No significant interaction between grouping and place of articulation confirms this trend. However, as an individual factor, the effect of place of articulation is significant in that the overall performance of all learners is relatively better on velar [g] than on labial [b].

The best performance of the 'UK' group is ascribed to their stay in the United Kingdom during which they have been receiving input from native speakers of English. As the results show, two participants of the 'UK' group consistently produced both of the target sounds with native-like VOT ranges in all six repetitions. In other words out of 22 'UK' learners, two were able to develop native-like accurate phonetic representations for English [b d] (i.e. with short-lag VOT). Although the 'Student' group of learners spend more time on speaking and listening to English than the 'Teachers' do (see table 1 above), their performance is not better than the 'Teacher' group. The results of repeated measures ANOVA show that the group variance between 'Student,' 'Teacher' and 'Saraiki' monolinguals groups is non-significant. This means the two Pakistan-based groups of learners simply transferred the L1 VOT values to the L2 stops. Following the SLM, we conclude that the learners of the 'Student' and 'Teacher' groups do not perceive the phonetic difference between the L1 and L2 [b g] sounds, so they have developed an equivalence classification between the L2 and the corresponding L1 sounds in their L2 phonemic inventory. This also confirms the hypothesis that it is very difficult for speakers of a voicing language to acquire voiced stops of an aspiration language.

The results of the 'UK' group need a separate analysis. The mean VOT of the 'UK' participants is significantly different from all other groups, namely those of the Saraiki monolinguals, English native speakers and the two Pakistan-based learners groups (i.e. 'Student' & 'Teacher'). The mean VOTs of the 'UK' learners in the production of English [b] and [g] are -77.18 and -41.77 msec., whereas the L1 (Saraiki) VOT values for these sounds are -99.77 and -72.81 msec., respectively. The differences between the L1 and L2 VOT values of the productions by the UK-based participants are significant ($p<.005$). This means the 'UK' learners tried to suppress L1 transfer of pre-voicing but they could only improve their productions to some extent (since the VOT values of their productions are significantly different from the L1 VOT values for the same sounds). However, they could not reach the native level (since their VOTs are also significantly different from those of the native speakers of English). Maximum improvement in this

regard is observed in the velar stop [g]. The mean VOT value of the 'UK' group for [g] is -41.77 msec. which is between the L2 VOT (short-lag VOT) and the L1 VOT (-72.81 msec.). This can be an example of a merger. The learners perceive some of the differences between the L2 and the L1 sound but not as much as may be helpful for them to acquire a native-like VOT. The improvement in the production of [b] by the 'UK' participants is not so much. Thus we conclude that the Pakistan-based learners of English failed to acquire accurate VOT for English [b g]. Only the 'UK' participants who have access to native speech show some improvement particularly in learning English [g]. The findings are according to the predictions of the SLM that equivalence classification between L2 and L1 sounds blocks the establishment of a new phonetic category for the L2 sound and in case of a weak equivalence classification, learners perceive some phonetic difference between L2 and the closest L1 sound but they improve little or only develop a merger of the two sounds. Now we try to analyze why the UK-based Pakistani learners of English who have been living in England for such a long time (approximately 65 months) and receiving input from native speakers of English, could not acquire English [b g] accurately.

The reason for the 'UK' learners' not acquiring English-like [b g] is that some of the native speakers of English also produce voiced stops with negative VOT (Simon 2009, Docherty 1992). Therefore, they mostly perceive a prevoiced stop of non-native speakers as native-like (Syed 2013). Even though some of them may not perceive it as native-like, they at least do not confuse it with any other sound. Thus there is no communication gap likely to occur between the non-native L2 speakers and the native English listeners. On the other hand, if L2 learners confuse English [p^h] with [p] and produce words like 'peak' without strong aspiration, there is a probability that the native speakers of English perceive such a production as 'beak' because there is no significant difference between unaspirated [p] and [b] in terms of VOT in most of the dialects of English. In other words, there is no lexical semantic load on the learners for acquiring voiced stops in a native-like manner but there *is* a strong lexical semantic load on them for acquisition of English aspiration contrast. That is why the L2 learners are more likely to acquire the aspiration contrast rather than the voicing contrast in English stops.

The reason for the difficulty in the acquisition of voiced stops of English faced by Pakistani learners is also of a phonological nature. We can interpret it using the terminology of the Element Theory (Backley 2011, Harris 1994). The Element Theory studies sounds on the basis of elements (Botma, Kula & Nasukawa 2011). According to the theory, L represents pre-voicing and H represents aspiration. In this way pre-voiced stops are L-headed and voiceless aspirated stops are H headed. But the stops with short-lag VOT are not specified for the elements H (aspiration) or L (voicing). Since beginning, Pakistani learners of English (like the speakers of other voicing languages), produce English voiced stops as pre-voiced on account of negative transfer from the L1 and under the influence of Pakistani English, which means their English voiced stops are actually L-headed. The following figure reflects this.

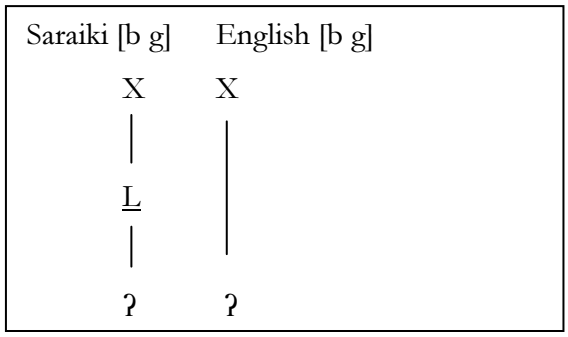

Figure 3: Voiced stops of Saraiki & English

For acquiring English voiced stops with accurate VOT ranges, Pakistani learners have to delete the element L which, being a head element in their grammar is very prominent. The head element on account of being the most salient element in a sound may be very difficult to delete. In simple language, it is less liable for the learners to neglect very salient acoustic cues but relatively easier to neglect less prominent cues. Therefore, the acquisition of English voiced stops is always very difficult for the Saraiki learners and voiceless aspirated stops are easier for them to acquire. It is because of this acoustic saliency of the aspiration cues which, Simon (2009) thinks, helps L2 learners to perceive and produce English voiceless plosives accurately.

On the other hand, these Pakistani learners can easily acquire voiceless aspirated stops of English. The English aspirated stops are produced unaspirated by Pakistani learners (Rahman 1990, 1991, Mahboob & Ahmar 2004). Therefore, for acquiring the aspirated stops, these learners have to simply add headedness to the already existing H element of the voiceless stops in their L2 phonemic inventory. This is not difficult for them because such H-headed stops already exist in most of the Pakistani languages including Saraiki since aspiration contrast is phonemic in most of the languages spoken in Pakistan. In this way, Pakistani learners of English can easily acquire English aspiration contrast on account of a positive transfer from the L1 and the specific nature of the elements involved in the target sounds. The following figure shows the difference between the initial stage production of Pakistani L2 learners (including Saraiki learners) and the target sounds of English as L2.

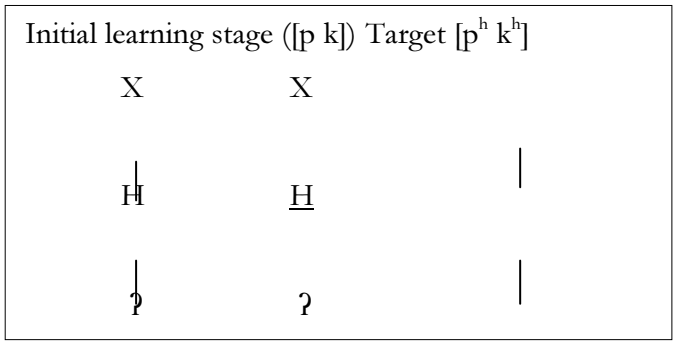

Figure 4: Stages of acquisition of voiceless stops of English by Pakistani learners

As the above figure shows, for acquiring voiceless aspirated stops, the L2 learners who have only voiceless unaspirated stops, or who equate both aspirated and unaspirated stops of English with only unaspirated stops of the L1 (as the adult Pakistani learners do), need to add headedness to the H element which already exists in voiceless unaspirated sounds in their L2 phonemic inventory. That is why the speakers of voicing languages who have unaspirated stops in their L1 find it easier to acquire English aspirated stops but difficult to acquire voiced stops of English (Pater 2003, Simon 2009, 2011).

Lastly, the findings of this study point out a possible gap in the SLM. The SLM predicts that a new phonetic category acquired by L2 learners may be deflected away from monolinguals of that language. But it does not provide any statistical scale to determine how 'deflected away' may be the new phonetic category of the L2 sound from the monolingual category. In the current study, we came across many cases of [g] which were produced with a negative VOT which was significantly different from the VOT for the L1 [g]. One interpretation of such productions is that the pre-voiced [g] of these learners, which is different from both the L1 and L2 [g], is an indication of a merger situation in the L2 phonemic inventory of these learners. However, we can also claim that the learners have acquired a new phonetic representation for English [g] which is a little deflected away towards pre-voicing from the native-like category of English [g]. This viewpoint is further strengthened by the fact that some native speakers of English also produce voiced stops with pre-voicing (Docherty 1992, Simon 2009). The SLM does not provide any statistical measure to determine the border-line between different stages of learning, like a little improvement without formation of a new phonetic category, which occurs in weak equivalence classification context, and a new category which is deflected away from that of monolinguals of an L2. It also does not provide any statistical matrix to calculate the perceptual distance between two sounds. These results point out the gap in the SLM which has already been pointed out in previous studies (e.g. Larson-Hall 2004, Harnsberger 2001, Schmidt 1996, Levy 2009, Wester et al. 2007, etc.) i.e. that the SLM lacks a solid statistical matrix to calculate learning and perceptual distance between sounds. In the absence of such a scale, a categorical boundary may not be clearly drawn between new phonetic categories and learning without establishment of separate phonetic categories for new L2 sounds.

6 Conclusion

The current study was based on the acquisition of English [b g] by adult Pakistani learners. The findings show that it is very difficult for Pakistani learners to produce voiced stops of English with short-lag VOT. They normally produce these plosives with pre-voicing. Most of the 'UK' participants in this study could not produce the target sounds with native-like VOT despite living among native speakers for more than five years. The study also points out a gap in the speech learning model, that it does not provide any statistical yardstick to calculate the improvement observed in L2 learners. This leads to confusion as for how to differentiate between different learning outcomes. Developing a possible yardstick to calculate the perceptual distance between L2 and L1 sounds may be an interesting research question for future research.

7 References

Backley, Phillip. 2011. *An Introduction to Element Theory*. Edinburgh: Edinburgh University Press.

Baumgardner, Robert. J. 1993. *The English Language in Pakistan*. Karachi: Oxford University Press.

Birdsong, D. 2007. Nativelike Pronunciation among Late Learners of French as a Second Language. In Ocke-Schwen Bohn and Murray J. Munro (eds.), *Language Experience in Second Language Speech Learning,* 99-116. Amsterdam: John Benjamins.

Boersma, Paul & David Weenink. 2012. Praat: Doing Phonetics by Computer. http://www.fon.hum.uva.nl/praat/ (accessed Jauary 2012).

Botma, Bert, Kula, Nancy & Nasukawa, Kuniya. 2011. Features. In Nancy C. Kula, Bert Botma & Kuniya Nasukawa (eds.), *The Continuum Companion to Phonology*, 33-63. London: Continuum International Publishing Group.

Clements, N. 1985. The Geometry of Phonological Features. *Phonology Yearbook* 2. 225-52.

Docherty, G. J. 1992. *The Timing of Voicing in British English Obstruents*. Berlin: Foris publications.

Flege, James Emile. 1987. The Production of 'New' and 'Similar' Phones in Foreign Language: Evidence for the Effect of Equivalence Classification. *Journal of Phonetics* 15. 47-65.

Flege, James E. 1995. Second Language Speech Learning: Theory, Findings, and Problems. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research,* 233-77. New York: Timonium, MD.

Flege, James E. & Eefting, Wieke. 1988. Imitation of a Vot Continuum by Native Speakers of English and Spanish: Evidence for Phonetic Category Formation. *Journal of the Acoustical Society of America* 83(2). 729-40.

Flege, James E. & Port, Robert. 1981. Cross-Language Phonetic Interference: Arabic to English. *Language & Speech* 24, (2). 125-46.

Hamann, Silke. 2011. Phonetics-Phonology Interface. In Nancy C. Kula, Bert Botma & Kuniya Nasukawa (eds.), *The Continuum Companion to Phonology*, 202-24. London: Continuum international publishing group.

Harnsberger, J. D. 2001. On the Relationship between Identification and Discrimination of Non-Native Nasal Consonants. *Journal of the Acoustical Society of America* 110(1). 489-503.

Harris, John. 1994. *English Sound Structure*. Oxford: Blackwell.

Harris, John & Geoff Lindsey. The Elements of Phonological Representation. In Jean Durand & Francis Katamba (eds.), *Frontiers of Phonology: Atoms, Structures, Derivations*, 34-79. Harlow: Longman, 1995.

Honeybone, P. 2005. Diachronic Evidence in Segmental Phonology: The Case of Laryngeal Specifications. In Marc Van Oostendorp & Jeroen Van Weijer (eds.), *The Internal Organization of Phonological Segments,* 317-356. Berlin: Mouton de Gruyter.

Iverson, Gregory & Joseph Salmons. Aspiration and Laryngeal Representation in Germanic. *Phonology* 12(3). 369-96.

Kager, Rene, Feest, Suzanne va der, Fikkert, Paula, Kerkhoff, Annemarie. &. Zamuner, S. Tania. 2006. Representations of [Voice]: Evidence from Acquisition. In Erik J. Van der Torre & Jeroen Van de Weijer, *Voicing in Dutch*, 41-79. Amsterdam: John Benjamins, 2006.

Lado, Robert. 1957. *Linguistics across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor: University of Michigan Press.

Larson-Hall, J. 2004. Predicting Perceptual Success with Segments: A Test of Japanese Speakers of Russian. *Second Language Research* 20(1). 33-76.

Levy, Erika. 2009. On the Assimilation-Discrimination Relationship in American English Adults' French Vowel Learning. *Journal of the Acoustical Society of America* 126(5). 2670-2682.

Lisker, L., & Abramson. A. S. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word* 20(3). 384-422.

Mahboob, A., & Ahmar, N. H. Pakistani English: Phonology. In E. W. Schneider (ed.), *A Handbook of Varieties of English: A Multimedia Reference Tool*, 1003-15. Berlin: Mouton de Gruyter, 2004.

Masica, Colin. 1993. *The Indo-Aryan Languages*. Cambridge University Press, 1993.

Nasukawa, Kuniya. 2010. Place-Dependent Vot in L2 Acquisition. In Mathew Prior (ed.), *Selected Proceedings of the 2008 Second Language Research Forum*, 198-210. Somerville: Cascadilla Proceedings Project.

Pater, J. The Perceptual Acquisition of Thai Phonology by English Speakers: Task and Stimuli Effect. *Second Language Research* 19. 209-23.
Rahman, Tariq. 1991. Pakistani English: Some Phonological and Phonetic Features. *World Englishes* 10(1). 83-95.
Rahman, Tariq. 1990. *Pakistani English: The Linguistic Description of a Non-Native Variety of English.* Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University.
Rice, Kerene, & Peter Avery. 1993. Segmental Complexity and the Structure of Inventories. *Toronto Working Papers in Linguistics* 12(2). 131-53.
Schmidt, A. Marie. 1996. Cross-Language Identification of Consonants. Part 1. Korean Perception of English. *Journal of the Acoustical Society of America* 99(5). 3201-11.
Shackle, Cristopher. 1976. *The Siraiki Language of Central Pakistan: A Reference Grammar.* London: University of London School of Oriental and African Studies.
Shimizu, K. 2011. Study on Vot of Initial Stops in English Produced by Korean, Thai and Chinese Speakers as L2 Learners. Paper presented at the. International Congress of Phonetic Sciences XVII, Hong Kong, 17-21 August.
Simon, E. 2009. Acquiring a New Second Language Contrast: An Analysis of the English Laryngeal System of First Language Dutch Speakers. *Second Language Research* 25(3). 377-408.
Simon, E. 2011. Laryngeal Stop Systems in Contact: Connecting Present-Day Acquisition Findings and Historical Contact Hypotheses. *Diachronica* 28(2). 225-54.
Syed, Nasir A. 2013. Perception and Production of Consonants of English by Pakistani Learners. University of Essex, PhD dissertation.
Syed, Nasir A. 2012. Perception and Production of English [d] by Pakistani L2 Learners. *Essex Graduate Students Papers in Language and Linguistics, University of Essex* 13. 134-57.
Syed, Nasir A. 2012. *Why an L1 Contrast Does not Prime in L2?* Saarbrucken, Germany: Lap Lambert academic publishing.
Varma, S. 1936. The Phonetics of Lahnda. *Journal of Royal Asiatic Society of Bengal. Letters* 2. 47-118.
Wester, Femke, Gilbers, Dicky & Lowie, Wander. 2007. Substitution of dental fricatives in English by Dutch L2 speakers. *Language Science* 29. 477-491.