THE ACQUISITION OF ENGLISH STOPS BY SAUDI L2 LEARNERS

SAMI M. ALANAZI

A thesis submitted for the degree of Doctor of Philosophy (Ph.D.)

in Linguistics

Department of Language and Linguistics

University of Essex

JULY 2017

ABSTRACT

Researchers have studied voice onset time (VOT) in a number of languages but there is a scarcity of research on the acquisition of VOT of English, particularly by adult Saudi learners, and on the VOT of Saudi Arabic. The current study aims to fill these gaps. At the same time, we aimed to assess whether key claims of Flege's Speech Learning Model (SLM) were supported by this kind of data.

31 adult advanced Saudi learners of English and 60 monolinguals (30 native English and 30 Arabic monolinguals) participated in this study. The VOTs of the voiced and voiceless stops were measured followed by three different vowels, in both isolated word and word in sentence contexts.

The results show that the learners produced English voiceless stops with aspiration closer to Arabic than to the higher native English VOT values, and voiced stops with pre-voicing, similar to Arabic, rather than with native English short-lag VOT values. Context had an effect in English but vowel did not, while the reverse was true for the learners and Arabic native speakers. Overall, learners' acquisition was modest despite their level and exposure, in that they overwhelmingly resembled Arabic rather than English native speakers.

Several hypotheses based on SLM expectations were not confirmed in an unqualified way. However, support was found for learners' phonetic categories being ‘deflected’ away from both L1 and L2 categories.

All three groups produced longer positive VOT for aspirated than unaspirated or voiced plosives. All exhibited VOT increasing across places of articulation, front to back for the voiceless stops, but only English native speakers showed this clearly for the voiced stops. Length of residence in UK and daily use of English did not seem to affect nativelikeness of learner VOT.

ACKNOWLEDGEMENTS

Praise be to Allah, the Almighty, Creator and Sustainer of the universe. Prayers and blessings are sent on His Prophet Muhammad, the seal of all prophets and the leader of the righteous servants.

First, I would like to express my sincere gratitude to my academic supervisor, Dr. Wyn Johnson, who has been not only academically but also emotionally a great support. I would especially like to thank her for guiding me in this research and in the other courses that she taught. She has always been very supportive of whatever I do. My effort for the dissertation is too little in comparison to her encouragement and the support that she provided me with. A big thank you to you, Dr Johnson.

I would like to extend this gratitude to Dr Rebecca Clift, the chair of my supervisory board committee, for her extreme kindness and great support throughout the whole period of supervision and to all the committee members.

I would like to especially thank Prof. Nancy Kula, the head of the Language and Linguistics Dept. for all the models she taught me at the beginning of my PhD. I also would like to thank her for teaching me how to analyse data using Praat which was the main tool used in this study.

My words are beyond description and I cannot thank Phil Scholfield enough for his wit, statistical skills and generous feedback and for answering my countless questions regarding that, thank you very much, indeed.

My thanks also go Dr Muhammad Alzaidi, a professor at King Saud University, for coaching me on how to use the Prosody Pro script while he was doing his PhD here at the University of Essex and for all the fruitful discussions related to this topic and other various topics in Phonology.

My sincere gratefulness to all the members of our Research Group which was led by our supervisor, Sultan Aldaihani, Gary Taylor-Raebel, Nasir Abbas, Sara Altubuly, Eleni Galata, Huda Al Taisan, Demah Alqahtani, for all their fruitful comments and precious feedback while giving my presentations in the group.

Secondly, special thanks go to Northern Border University, which granted me this scholarship. I am also thankful to those who supported and helped me to obtain this scholarship and then to collect the data for this research. I am grateful to the subjects who participated in my study and to the Saudi Cultural Bureau in London, which continues to provide me with all the help and support that I need.

Thirdly, I am indebted to my beloved parents who brought me up till I reached this stage of my life and who are still at this moment praying for me and wishing me success.

Lastly, my deepest warm gratitude goes to all my friends in Saudi Arabia and here in Colchester— Fahad Hamed, Wael Alharbi, Usef Alanazi, Sami Alharthi, Abdulah Alanazi, Radi Alshammari, Bader Alrakan, Abdulaziz Aljohar, Asem Alseat, Hassan Alalmi, Mansour Almawidah, Mansour Alzahrani, Khalid Alshammari, Saud Alenezi —for their support and various types of assistance that they provided me with to help me to complete this degree.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1:

INTRODUCTION TO THE STUDY

1.1 Introduction

Although English is a foreign language in Saudi Arabia, there is a huge and growing interest in learning English at both governmental and individual levels. Indeed, English is the only foreign language that Saudi students learn in state schools. Hence research on all aspects of the acquisition of English by Saudis is of importance for the country, as well as of interest to the world of language acquisition research.

At the same time, there is a dearth of language research on many areas of the Saudi Arabic language in its own right, quite apart from its role as an L1 in foreign language learning. As chapter 2 will show, our chosen area of interest is one of those areas. Therefore, our findings will offer a contribution not only to existing research on bilingual phonological acquisition but also provide an up-to-date study of VOT patterns in Arabic as well as English. Indeed, the VOT patterns of our specific chosen dialect have never been investigated before (Northern Saudi Arabic, of the type spoken in and around Arar).

It is widely believed (e.g Scovel, 1988; Brown, 1998 etc.) that the acquisition of a second language is different from that of a first language, particularly as adult second language learners rarely attain the same native L2 proficiency that younger learners do when learning their first language. Furthermore, L2 learners usually experience greater difficulty learning their second language than their L1. This difference between second language and first language acquisition is perhaps most obvious with regard to the acquisition of second language sounds, and especially where the L1 and the L2 differ. It is widely observed (Scovel, 1988) that only learners who start the acquisition process of L2 at a very young age achieve a nativelike pronunciation. Our study is concerned with the more usual type of Saudi learner, who begins learning English only in school and later than that age.

There are numerous differences between English and Arabic which have been of interest to researchers for many years, and which often lead to challenges for Arab learners of English. Some of these differences are on the syntactical, orthographical, and phonological levels. The current study, however, focuses on sounds, and very specifically on the acquisition by Saudi L2 learners of English stop consonants, in particular on a phonetic feature of stops known as voice onset time (VOT).

Within this scope there are three main points on which English and Arabic are known to differ from existing literature, and which will therefore be key starting points for our attention throughout our study, both with respect to describing them in the two languages and considering their role in acquisition of English. These are: (1) Saudi Arabic lacks the stop /p/ which exists in English; (2) English possesses a clear allophonic distinction within voiceless stops which is not matched by Arabic (the aspirated pronunciation of /t/ in *till* [t^hɪɫ] versus the unaspirated in *still* [stɪɫ]); (3) both voiced and voiceless stops in general differ in their VOT values between English and Arabic, especially the former (voiced).

1.2 VOT and English - Arabic differences in stops

In previous studies, it has been quite common in the field of phonetics, phonology and second language acquisition of sounds to analyze the stop consonants in terms of their VOT. VOT is the time between the burst of a stop, when the closure phase opens, and the onset of voicing i.e vibration of the vocal cords, for the following vowel (Lisker and Abramson, 1964). If the voicing of the following vowel begins more than 30 msec after the burst, the duration is called ‘long-lag’, or aspirated, and if the duration between the burst of the stop and the onset of the following vowel is shorter than 30 msec it is called ‘short-lag’ (ibid), or unaspirated. In this way, the stops are divided into categories on the basis of short or long lag VOT. In the study by Lisker and Abramson (1964), a third type of stop was also identified, which is called ‘pre-voiced’ or 'voicing lead'. If the vocal folds start vibrating before the burst of the stop, such a stop is called pre-voiced and the duration is measured in negative VOT values. (see figure 2.1 for a schematic representation of the three VOT categories)

Thus, VOT is often seen as a primary phonetic cue for the voicing distinction, particularly in stops (Lisker & Abramson, 1964; Flege, 1980; Kent & Read, 1992). In numerous studies (e.g. Lisker & Abramson, 1971; Weismer, 1980; Yeni-Komshian, Caramazza & Preston, 1977; Flege & Port, 1981; Khattab, 2002; Keating, Linker, & Huffman, 1983; Cho & Ladefoged, 1999 etc) voice onset time (VOT) was used to measure the differences between similar sounds in a number of languages.

Some of the languages of the world have only stops with short and long lag VOT but there are many languages which have stops with negative VOT (pre-voicing). Arabic is a language which has pre-voiced stops in addition to short/long lag stops while English (Southern British English, specifically the RP variety) is a dialect which has only short and long-lag stops. On the basis of aspiration and voicing contrast, the languages of the world are therefore divided into ‘voicing’ and ‘aspiration’ languages (Simon, 2011). According to this classification, Arabic is a voicing language and English an aspiration language. This means that although both languages possess a distinction between stops which, in phonology, is labelled voiced versus voiceless, Arabic stops are discriminated phonetically by the feature [voice] but English stops are discriminated by the feature [spread glottis] (Kager et al., 2007, Honeybone, 2005).

As will be shown in chapter 2, a number of studies indicate that English aspirated voiceless stops are produced with long-lag and voiced stops are produced with short-lag or rarely pre-voiced (Lisker & Abramson, 1971; Weismer, 1980). Studies of Arabic vary particularly depending on what spoken variety is described, but the stop voicing contrast in Arabic is widely found to be different: positive VOT is for the voiceless stops and pre-voicing is for voiced stops (Flege & Port, 1981; Khattab, 2002). We may see, therefore, that one of the key differences between the phonological systems of English and Arabic is the Voice Onset Time (VOT) of their stops. As we will show in chapter 2, however, there remain gaps even in basic descriptive knowledge which need to be filled, such as the precise facts for a spoken dialect of Saudi Arabic such as ours, and the behaviour of stop VOT in detail before different vowels and in isolated word versus word-in-sentence contexts. To the best of my knowledge, so far only Flege & Port (1981) and Alghamdi (1990) have examined VOT patterns in Saudi Arabic monolinguals.

1.3 Acquisition of English stop VOT

We wish to investigate two aspects of acquisition. First, we wish to understand better whether Arabic learners of English who learnt in school in Saudi Arabia but have high enough general English proficiency to study in the UK, and who currently are living in a native English speaking environment, show evidence of having been able to acquire English stops with nativelike stop VOT. Second we wish to ascertain if data from such learners supports, or not, claims made by a particular theoretical framework, called the Speech Learning Model (SLM), about the acquisition of second language sounds.

Over the last few decades an extensive number of studies have been conducted on the acquisition of VOT in a second language in a variety of languages. However, only a few have examined Arabic learners and fewer still the VOT of Saudi learners of

English as L2. Indeed, to our knowledge only Flege & Port (1981) have studied the production of English stops by Saudi learners. Hence one of the major aims of the present study is to give a full account of the VOT patterns of learners whose L1 is a dialect of spoken Saudi Arabic.

There is a lot of literature in general on the acquisition of the VOT of aspiration languages by the learners of voicing languages. Factors like stress, the place of articulation of stops, the context in which a stop is produced and the nature of the vowel following the stop can influence the VOT of stops (Johnson and Babel, 2010). A review of the existing literature on the relevant topic (chapter 2) shows that there is a large body of literature on the acquisition of English stops by other language learners, but research on the acquisition of English stops by Saudi Arabic learners is not equally large. The current study is an attempt to fill this gap in the literature.

In order to thoroughly cover the ground we will not, as many studies in this area do, stop short at just considering learner production of voiceless aspirated and voiced stops in English, in one context. Rather, this study investigates both the aspirated [k^h] [t^h] [p^h], and the unaspirated [k] [t] [p] voiceless stops as well as the voiced stops [g] [b] [d], followed by each of three different vowels, and in two word/sentence contexts. In these ways, it is intended to be a comprehensive study of learner stop VOT production.

Aside from the literature on what VOT performance is found among learners, there is also a literature (chapter 2) on how the acquisition process occurs, which should explain differential performance by learners, where they seem to have acquired one sound better than another. This has led to much theorizing, from which, as we will show in chapter 2, one of the most influential theories that has emerged is Flege's SLM. Amongst other things, this takes a more careful view of differences between languages and their effect on acquisition success. For instance, it suggests that a very obvious difference between languages, such as the fact that Arabic lacks the /p/ found in English, may be easier for learners to overcome than a more subtle one, such as the fact that for Arabic voiced stops VOT is negative while for English it is positive. In the latter case, simple transfer of the L1 VOT habits would seem to be more likely.

Flege (1980) in fact argues that the VOT differences between languages may be one of the features that are especially difficult to acquire and so hinders the learner's ability to master second language pronunciation. Nevertheless, some VOT differences would seem easier for a learner to notice and imitate than others, such as the fact that English aspirated stops are more heavily aspirated than Arabic ones. Some recent studies, indeed, have established that the role of the L1 is not so direct in the phonological acquisition of the learner's second language (Flege 1995, Simon 2009).

The purpose of the study is then, in part, to investigate how well Saudi L2 learners of English acquire the VOT of English stops of different types and in different contexts etc., and to what extent those learners seem to simply transfer phonetic values for the newly acquired language from their native language. We are particularly interested to know where the Saudi learners transfer the L1 VOT and where they develop a separate category for L2 stops. If they develop a range of VOT for the English stops different from the L1, then what is the range of VOT of the newly established category for the target sounds? Is it like that of the native speakers of English or different from them?

Predictions of a theory such as the SLM concerning matters such as these need constant testing on new data so that theories may be supported, disconfirmed, or prompted to be refined. Hence we wish also to add to the small literature of this sort in the area of VOT acquisition research.

1.4 Purpose and objectives of the study

From the above, we can see that the current study has a number of specific objectives.

Firstly, there are descriptive aims, with respect to each of the three groups involved in our study separately. We wish to document in more detail some key descriptive facts concerning VOT for our chosen Saudi Arabic dialect (see 3.2.2), and for our kind of advanced learner of English (see 3.2.3), as well as to replicate such findings for native speakers of English (see 3.2.1). We therefore plan to measure the effect of the place of articulation (POA), of vowel context, of differences between stops in their voicing and aspiration category, and of word context, on the production of stop VOT by native speakers of English, native Saudi Arabic speakers, and Saudi advanced adult learners of English. Since, as chapter 2 will show, there are common, perhaps universal, findings concerning how differences in stop VOT change in relation to place of articulation in the mouth, increasing from front to back, we will also examine whether the data from all three of our groups are in line with this. In this thesis, when we speak of measuring the effect on VOT of voicing and aspiration in English, this wording is a shorthand for saying that we are measuring 'the reflection in VOT of differences between stops in terms of their phonemic or major allophonic categorisation by voice or aspiration'. These distinctions are often, but not necessarily, associated with VOT differences, since what phonologically are regarded as the English voiceless unaspirated stops and voiced stops have almost identical VOT. I.e. what we call a voice distinction may phonetically not be conveyed by VOT but by other features.

Secondly, we wish to find out the level of acquisition of stop VOT which our advanced learners have attained, and whether it shows any signs of successful acquisition (in the sense of their production being in some way nativelike in any area).

This involves comparing the VOT findings for the learners with those from the Arabic and English native speakers. Thus, we are concerned to know on which stops that the L2 learners will be significantly different in VOT from their Arabic L1 and on which stops the learners will be similar in VOT to their target English L2 (so able to be regarded as nativelike). Furthermore, regardless of being similar or different from ENS (English Native Speaker) or ANS (Arabic Native Speaker) in absolute VOT values, do the learners make some kind of VOT distinctions based on POA, voice/aspiration, following vowel and context parallel with Arabic or English native speakers? Also, relevant here is to ascertain whether there is any correlation between the learners' length of residence in the UK, or their amount of listening to and speaking English, and their accuracy in the acquisition of English stops.

A final aim is to pursue what we see as the most promising current theory of how L2 sounds are acquired, and test from our data some expectations which the hypotheses of that theory suggest (the Speech Learning Model (SLM) developed by Flege (1995)). Do learners perform in a more nativelike way in those areas of stop production where the theory predicts that they should? In this way we may find support, or not, for that approach.

The above objectives will be achieved by studying the nature of stops produced by Arabic monolinguals, English monolinguals and adult Saudi L2 learners of English. The study will be based exclusively on the acoustic analysis of the stops produced, using VOT as the acoustic cue for analysis. Our research questions are covered fully along with our predicted hypotheses, at the end the literature review in (2.12).

1.5 Key Terminology, as defined in this study

Acquisition: in this study refers to the product rather than the process of learning. We have not accessed learners at various stages in their acquisition process over the years nor asked them how they learn to pronounce stops. Rather we use the VOTs that they produce at the time that we measure them to gain a picture of how nativelike their performance is at that time. Thus, we can say what they have learnt but not how they learnt it. Given the age and general proficiency level of the learners, this may well be close to what could be considered their 'final state', in the sense of the end point of their acquisition after which little would change.

Arabic native speakers (ANS). In this study these are monolinguals who know little or no English. English native speakers (ENS). In this study these are monolinguals who know no Arabic and little of any other language.

First language (L1). For the learners and native Arabic speakers in this study this is the spoken dialect of northern Saudi Arabia (in the region of Arar), though the participants also know Modern Standard Arabic (MSA) and some Quranic/Classical Arabic (CA). The latter two varieties do not, however, form the basis of their everyday speech, which was centrally relevant to the present study: MSA is used only in the media and some formal settings, CA in religious contexts. For the English native speakers, the first language is educated English of the South of England.

Length of residence in the UK of the learners in the study (LOR).

Place of articulation (POA). In this study the key POAs in the mouth are labial, coronal and dorsal.

Second language (L2). For the learners in this study, southern UK English.

Speech Learning Model (SLM). A leading theory of how sounds are acquired by foreign learners of a language, developed by Flege (1995).

Voice onset time (VOT). This is the time (in milliseconds) between the moment of the release of the closure which occurs when a stop consonant is articulated, and the moment when the vocal folds start to vibrate. That second moment may be before the release of the closure (pre-voicing, voicing lead, negative VOT) or after it (either short lag or long lag VOT).

1.6 Structure of the study

This thesis is divided into six chapters:

Chapter 1 presents a general introduction, brief background of the study, purpose and objectives of the study; then it outlines the structure of the study.

Chapter 2 reviews the relevant literature related to this topic. It starts with providing an inventory of Arabic consonants and a comparison between stops in Arabic and English. Then, the definitions of Voice Onset Time (VOT) will be offered, and the effects on VOT of factors such as POA, following vowel setting and word context. After that, VOT studies on Arabic monolinguals (ANS), English monolinguals (ENS) and Arabic L2 learners of English as well as contrasts found in the literature between English and Arabic VOT will be reviewed. Finally, it provides a summary of the existing models of acquisition of sounds by learners of English as a second language as well as our chosen model of acquisition. The chapter concludes with offering the research questions and our hypotheses.

Chapter 3 describes the methodology used in the study. It provides the details of the participants, the instruments used for the data collection, the data collection procedures, and statistical tests used for the data analysis.

Chapter 4 presents the VOT results of the monolingual groups in turn: ENS and ANS.

Chapter 5 offers the results of the main group of the study, the Saudi Arabic L2 learners of English, It firstly presents their results descriptively in a similar way to the monolingual groups. Secondly, it compares their performance with English speakers (ENS) and Arabic speakers (ANS) in several different ways. Thirdly, it studies the relationship of length of residence (LOR) and daily use of English with the acquisition of English stops and finally, it examines the learners' performance individually.

Chapter 6 summarises the answers to the research questions, providing answers to our hypotheses, and also offers conclusions, limitations, and potential future work based on the study.

Chapter 2:
LITERATURE REVIEW

2.1 Introduction

This chapter gives an overview of the literature on topics relevant to our study. Firstly, the inventory of Arabic consonants will be presented and compared with the English inventory with respect to the stops. Secondly, definitions of voice onset time (VOT) will be offered and key factors which have effects on VOT in production will be considered. Next, studies of VOT involving Arabic as well as English monolinguals will be presented and the VOT differences emerging from those studies will be summarised. We then consider studies of VOT of Arabic-speaking English learners. Finally, we move to theories of acquisition of sounds by foreign learners and our chosen model of acquisition, the Speech Learning Model.

2.2 Phonemic Inventory of Arabic

As stated earlier this study investigates the acquisition of the VOT of English stops by Saudi L2 learners with attention to a number of factors like place of articulation, vowel environments and context etc. Therefore, it would be helpful initially to have an idea about the consonant inventory of the L1 of the learners. This section will present the inventory of Arabic consonants and the next section the inventory of the English consonants concerned in the study. Arabic is a language whose history may be traced back to long before the birth of Islam that started with Prophet Mohammad's 'Peace be upon him' prophecy. It is now spoken by almost all people of the 22 Arab countries that form the Arab world and it is also the liturgical language of over a billion Muslims around the world. Arabic has influenced a number of languages like Persian, Kurdish, Pashto, Urdu etc.

Saudi Arabic is a group of dialects spoken in the Kingdom of Saudi Arabia by more than 20 million people. The northern variety of this dialect (spoken in and around Arar in the north east of the country) is the L1 of the Arabic speakers in our study. Other modern spoken Saudi dialects which have been studied include Ghamdi in the south and Najdi in the central, Riyadh, region. To our knowledge the dialect which is our focus of attention has no other name than 'Northern' (the term also used, for example, in Alwasel (2017)). Although its region of use borders Iraq, it is considered as a dialect akin to the central Saudi Najdi dialect rather than any Iraqi dialect (Ingham 1982).

Saudi Arabic has 29 consonants and six vowels. Those consonants can be categorized as follows: one lateral /l/; one affricate /ʤ/; one trill /r/; two nasals /m, n/; two semi-vowels /w, j/; thirteen fricatives /f, θ, ð, s, z, ṣ , ʃ, x, ɣ, ħ, ʕ, h, ðˤ/ and nine stops /b, d, g, t, k, q, ʔ, tˤ, dˤ/.

Arabic has fewer vowels compared to English: three long vowels /iː/, /ɑː/, /uː/ and three short ones /i/, /ɑ/, /u/ (Al-Ani, 1970), whereas Southern British English has twelve vowels (Davenport & Hannahs, 2010). Our primary focus in the current study is on stops only; therefore, they will be the main topic of the discussion and featured in comparison with the English stops.

2.3 Inventory of Stops in Arabic and English compared

In a similar way to English, Saudi Arabic has voiceless and voiced stops. The voiceless stops found in Saudi Arabic are /t, k, q, tˤ, ʔ/: the alveolar /t/, the dorsal /k/, the uvular /q/ and the emphatic /tˤ/ together with the glottal stop (Newman, 2002). Thus, all the voiceless stops found in English are found in Arabic, as phonemes written with the same IPA symbols (though not phonetically identical), apart from the bilabial

/p/. However, some educated people produce the /p/, especially in some loanwords. Moreover, /k/ is sometimes heard as a realization in place of /g/, especially with some loanwords like *karage* for the English *garage.*

As for the voiced stops found in Saudi Arabic, they are /b d ɡ dˤ/, including the bilabial /b/, dental or in some areas alveolar /d/, the dorsal /g/, the emphatic /dˤ/; thus again all three English voiced stops /b d ɡ/ have corresponding sounds in Arabic at least as corresponding phonemes (though not in phonetic detail).

Table 2.1: Stops in English provided by Davenport & Hannahs (2010) and in Lebanese Arabic provided by Khattab (2002)

POA	Bilabial	Dental	Alveolar	Post-alveolar	Dorsal	Glottal
English	p b		t d		k ɡ	(ʔ)
Arabic	b	t d		tˤ dˤ	k ɡ	ʔ

It should be noted that Arabic /t d/ are different from the English /t d/ not only in VOT (as we shall see later) and in POA (as we can see in table 2.1 in Lebanese Arabic), but it is also reported that the Arabic ones are [+distributed], meaning produced with the blade or front of the tongue (Al-Ani, 1970), while the English ones are [–distributed], meaning produced with the tip of the tongue (Ladefoged & Johnson, 2011).

Arabic also has two post-alveolar uvularized stops not found in English that are called 'emphatic stops' /tˤ/ /dˤ/. These two stops involve secondary articulation, meaning that their production involves two manners of articulation. For example, when the English stop /t/ is produced, the tip of the tongue is brought up to the alveolar ridge, closing the airflow and producing the stop. In /tˤ/, however, the back of the tongue is then retracted after completing the articulation gesture for /t/, producing uvularization. This procedure happens similarly with the voiced emphatic /dˤ/.

Specific to Arabic when compared to English is also the uvular /q/, which in most dialects spoken in Saudi Arabia, particularly Bedouin Arabic, is substituted by the velar /g/, as in the Arabic word /galb/ ‘heart’ or /gaal/ ‘said’. Therefore, in contrast to classical Arabic and similar to English, the velar /g/ exists in Saudi Arabic (though /q/ is also used). According to Newman (2002), completely voiced stops never aspirate. However, voiceless consonants can aspirate, with the exception of the uvular /q/.

Since we are concerned with learners of English, the focus in this study will be on the stops that are found only in English, or in both English and Arabic /b d g t k p/. Those additional stops found in Arabic will not be within the scope of this study. However, even where 'the same sound' exists in both languages, there are differences. Saudi Arabic has the voiceless stops /t k/ only in aspirated form [t^h k^h] in all positions and there is no unaspirated form for these two sounds as is the case with English following /s/. Other dialects of Arabic, however, have only unaspirated voiceless stops (Khattab, 2002). According to Khattab (2002), Arabic and English also differ significantly in their phonetic realisation of the stop voicing contrast. In English, voiceless stops (if aspirated) are produced with a long-lag and voiced stops are produced with a short-lag or rarely with voicing lead (pre-voicing) (Lisker & Abramson, 1971; Weismer, 1980). That means that in English what are phonologically termed voiced stops are often not phonetically voiced (i.e. not accompanied by vocal cord vibration). The stop voicing contrast in Arabic is different in that the short lag or low long lag is used for the voiceless stops, and strong voicing lead for voiced stops (Yeni-Komshian, Caramazza & Preston, 1977; Flege & Port, 1981).

2.4 General nature of stops and Voice Onset Time (VOT) in stop production: Definitions

Laver (1994) states that stops can be uttered with a variety of phonation ranging from strong voicing to complete voicelessness, depending on their phonological identity and the neighbouring context. Also, there exist other types of phonation like creaky, breathy, whisper...etc. which are not within the scope of this study.

There also exist some natural and general characteristics for the stops. According to Khattab (2002), stops consist of three physical events, each of which is visible acoustically in a spectrogram:

- a closure phase (onset phase), when an active articulator interacts with a passive one
- a hold phase, where the closure is preserved and the air pressure increases behind it
- a release phase (offset phase), where the constriction is released

Khattab (2002) states that a fourth event may arise in prevocalic initial plosives if some turbulent noise energy appears after the burst while the glottis is partially open, as is the case in the voiceless stops. This phenomenon is *aspiration*.

Aspiration involves a delay in the start of voicing, specifically in the sound succeeding the aspirated stop where voicing does not initiate immediately, but there is a slight delay (Rogers, 2000). This delay appears as a periodic energy on the spectrogram, particularly in higher frequencies. The aspiration is represented by superscript [^{h}] in the International Phonetic Association (IPA) system as it shows some similarities with the glottal fricative [h]. According to Ladefoged & Johnson (2011), the bigger the glottal opening, the longer the amount of aspiration, that is, the amount of aspiration depends on the degree of the glottal opening during the closure.

Aspiration plays a major role in the voiceless stops at the beginning of a stressed syllable in English, and examples of initial aspirated stops in words are as follows: *put* [p^hʊt] and *can* [k^hæn] in English and *toot* [t^hu:t^h] which means '*berry'* in Arabic, etc. Nevertheless, in English at the beginning of an unstressed syllable or after a syllable-initial /s/ there is no aspiration as in *sky* [skaɪ], *booking* [ˈbʊkɪŋ] etc. The voiceless stops in Saudi Arabic are always aspirated in all positions, however, and this aspiration ranges from short-lag to long-lag, as we see in the studies reviewed later in this chapter. Therefore, aspiration plays a distinctive allophonic role in English, unlike in Arabic.

Over the last few decades, a wide-ranging amount of research has been done on the production of voicing contrasts in stops by the use of the measurement of Voice Onset Time (VOT) (e.g Lisker and Abramson 1964; Yeni-Komshian, Caramazza, & Preston, 1977; Klatt, 1975; Smith, 1979; Stevens & Klatt, 1974; Jesry, 1996; Flege & Port, 1981; Keating, Linker, & Huffman, 1983; Cho & Ladefoged, 1999; Lisker, 1975; Radwan, 1996 etc). Voice onset time is considered to be one of the most important phonetic cues by which the nature of voicing in stops may be investigated, and has been implemented in various studies of different languages. VOT is a term that was established by Lisker & Abramson (1964) in their 'classic cross-linguistic study of voicing in initial stops in eleven languages'. According to Lisker & Abramson, VOT is 'the time interval between the burst that marks release of the stop closure and the onset of quasi-periodicity that reflects laryngeal vibration' (1964, p.422).

Their study aimed to examine how well VOT helps to separate the stop categories in various languages. Measuring VOT was observed to be very effective in separating phonemic categories in the languages investigated; however, the languages varied in the number of phonological categories and in the phonetic features assigned to them.

One should, however, never forget that there exist other features than VOT which contribute to distinguishing sounds, including voiced and voiceless stops, in any language.

In the present study, we use a more recent definition introduced by Cho & Ladefoged (1999). They define VOT as 'the time between the initiation of the articulatory gesture responsible for the release of a closure and the initiation of the laryngeal gesture responsible for vocal fold vibration' (p.225). Cho & Ladefoged (1999) however used the same acoustic measurement technique for extracting the VOT values as was used by Lisker & Abramson (1964).

According to Lisker & Abramson (1964), there are three main VOT categories that describe the types of glottal and supraglottal interactions for stops in languages. Therefore, VOT can be classified into three main categories of values as follows:

- Negative VOT/ pre-voiced, aka lead voicing, that takes place when voicing starts before the stop release. A fully voiced stop has VOT values about -60 msec.
- Zero VOT value/ short-lag that takes place when voicing starts at the same time as the stop release or within its range. This category falls within the range of 0 to 30 msec and if a stop falls within this range of VOT values, it is considered voiceless unaspirated or voiced, with no early voicing in the closure period.
- Positive VOT/long-lag that occurs when voicing is delayed after the stop release (more than 30 msec after the stop release). Aspirated voiceless stops occur within this range of VOT values.

Various other ranges have been suggested in the literature to mark these three phonetic categories. Lisker & Abramson (1964) for example suggested that the short-lag extends between zero and 25 msec, but Cho & Ladefoged (1999) and Khattab (2002) advocated that the short-lag range can be between zero and 30 msec. The more recent estimates of Cho & Ladefoged and Khattab (30 msec) for the unaspirated short-lag stops will be adopted in this study.

Cho & Ladefoged (1999) in fact classified the voiceless VOT categories even further into four more categories: *unaspirated* (with a mean VOT of around 30 msec), *slightly aspirated* (with a mean VOT of around 50 msec), *aspirated* (with a mean VOT of around 90 msec), and *highly aspirated* (with a mean VOT of over 90 msec) (p.217).

Figure 2.1 shows a representation of the above three main VOT categories.

Negative (below 0ms)	**Short-lag (0-25ms)**	**Long-lag (25ms+)**
e.g. Arabic /b d g/	**e.g. English /b d g/ [p t k], Non-Saudi Arabic/t k/**	**e.g. English $[p^h\ t^h\ k^h]$**

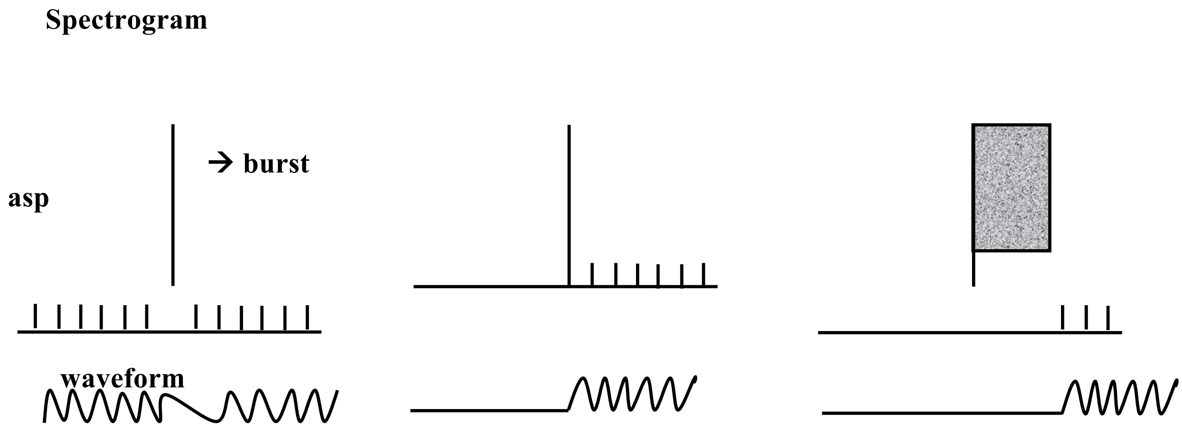


Figure 2.1: Schematic representation of spectrograms of the three main categories of VOT: negative VOT, short-lag, and long-lag, adapted from (Khattab, 2002).

2.5 Effects on Voice Onset Time (VOT)

Docherty (1992) states that there are common features that are usually found in the production of VOT in a number of languages. They are frequently attributed to some intrinsic factors such as the properties of the vocal organs, the place of articulation, the sound position in the syllable, the quality of the following vowel and so on. Some prosodic and temporal factors (e.g. speech rate and stress) have also been mentioned among those features. The timing of voicing is influenced by a number of aerodynamic and physiological aspects like speech rate, place of articulation and position in the syllable (Cho & Ladefoged, 1999) and VOT varies dependant upon such aspects. We now review those which we will pursue in our study.

2.5.1 The Effect of Place of Articulation (POA)

There is an important relationship reported by several researchers (e.g. Fischer-Jorgensen, 1954; Peterson & Lehiste, 1960; Stevens, Keyser & Kawasaki, 1986; Hardcastle, 1973; Cho & Ladefoged, 1999) between VOT and place of articulation. Voiceless stop VOT has a universal tendency to be longer in velars (or uvulars) than in alveolars and bilabials (Port & Mitleb, 1983; Lisker and Abramson 1964; Cho and Ladefoged, 1999). However, some linguists disagree that it is a universal tendency and found that /k/ has longer VOT values than /q/ (e.g. Radwan, 1996; Jesry, 1996).

Cho and Ladefoged (1999) discussed this pattern of universality for stops (though we are uncertain if they mean to include pre-voiced stops) and provided some physiological and aerodynamic explanations, by identifying the following links:

1) The further back the closure, the longer the VOT.

2) The more extended the contact area between active and passive articulators, the longer the VOT.

3) The faster the movement of the articulator, the shorter the VOT.

The current study is more concerned with the first suggested link in relation to English and Arabic, as it directly concerns POA, although as the next paragraph shows, the other two links are also involved in POA effects.

There are a number of factors which explain the length of the VOT with reference to POA that were suggested by Cho & Ladefoged (1999). The first explanation as to why VOT is longer when stops occur near the back of the mouth is the impact of aerodynamics. Cho & Ladefoged (1999) state that the cavity behind the velar stop is smaller than that behind the bilabial and alveolar stops. Therefore, the velar stops have greater VOT values than bilabials and alveolar stops, as velar stops have developed more pressure when airflow is released and because of that, voicing starts longer after the burst of the velars stops. This reasoning however is based on the assumption that vocal cord vibration starts after the burst of the stop. Since with pre-voiced stops the vibration has already commenced before the burst, it is not clear how this reasoning could apply to Arabic voiced stops.

Another factor explaining why VOT becomes longer as voicing occurs near the back of the mouth is articulatory movement velocity, as the tip of the tongue and the lips move faster than the back of the tongue. Also, the tip of the tongue moves faster than the lower lip, which might be the reason why velars yield longer VOT values than alveolar and bilabial stops (Cho & Ladefoged, 1999). This would however yield an order velar>bilabial>alveolar. The third factor is the extent of articulatory contact area, as stops with a more extended articulatory contact have a longer VOT in general (Cho & Ladefoged, 1999). As a result, velar stops again have greater VOT values than bilabial and alveolar stops. That would also imply longer VOT for +distributed stops, such as Arabic coronals.

It should be noted that the above account applies most obviously to stops with positive VOT. However, it is still unclear how these factors apply to stops with negative VOT such as the voiced stops in Saudi Arabic, and what pattern would be expected for stops with negative VOT in terms of POA. The results actually obtained for voiced stops in most of the Arabic studies are varied and do not show a clear VOT pattern as we will see later in the Arabic VOT studies.

2.5.2 The Effect of Vowel Context

The influence of vowel on the VOT of the following stop has been the subject of much debate and is still under discussion by many researchers. Some studies, like Lisker and Abramson (1967), who considered 11 languages, state that there is no major effect on VOT of the following vowel. But others, like Klatt (1975) and Weismer (1979), found that VOTs followed by tense high vowels were longer. Port (1979) examined the VOT in word-initial English plosives and had similar findings. Moreover, similar results were later found by Rochet & Fei (1991) who studied Mandarin stops and declared that vowels had a significant influence on the VOT of the preceding consonants.

Chao et al (2006) found that VOT is longer when followed by high vowels /i - u/ than low vowel /a/ in word-initial stops. Conversely, VOT was found to be longer when followed by /a/ than /i – u/ in Swedish (Fant, 1973). Schmidt (1996), Johnson & Babel (2010) and Iverson et al (2008) all agree that there is a strong effect of the vowel on L2 sounds in general.

All in all, research on the influence of vowels on the VOT is ongoing, and needs further investigation with reference to a number of languages. The current study therefore examines the influence of following vowels on VOT in English and Arabic stops.

2.5.3 The effect of word Context (words in isolation / words in sentences)

The effect of context on VOT has been studied by a number of researchers (i.e. Lisker & Abramson (1964 ,1967), Docherty (1992), Yeni-Komshian et al (1977), etc.).

Lisker & Abramson (1964, 1967) found that VOT was significantly longer in English words read in isolation than words read in sentences. However, they stated that the position of the word within the sentence, whether it is read at the beginning or in the middle of the sentence, did not make VOT differ considerably. Docherty (1992) also came to the same conclusion as he found that VOTs in isolated word contexts were longer than VOTs in words in carrier sentence contexts. However, Yeni-Komshian et al (1977) found that no major differences were found between stops produced in words in isolation and/or in sentences.

The use of what are called 'carrier sentences' is vital as it provides data on sounds which are in a more natural environment for the participants than just asking them to read words in lists, and which resembles real language use more closely. Directly gathering data from fully natural ways of speaking like conversations or reading passages however requires a huge amount of time collecting the data (Docherty, 1992), so was not attempted in the current study.

2.5.4 Other factors affecting VOT

There are other factors reported in the literature that might affect VOT. These include the position of a sound in the syllable or word and of the word in the sentence, as well as speech rate, gender, or the age of the speaker. These will not be dealt with in this current research as it would be beyond the scope of time and number of words to examine all those factors in one single study. Rather we will as far as possible control these by holding them constant. Thus, in our production study the stop is always word initial (please see chapter 3 page 88), except for the voiceless unaspirated ones in English which are in an initial cluster after /s/. Target word position when in sentences is medial, not sentence initial or final. Furthermore, only male participants are used, due to the difficulty of access to female participants by a male researcher, for cultural reasons. We will however consider some potentially acquisition related features of participants such as length of residence in the UK and the number of hours using English.

2.6 VOT studies of Arabic monolinguals

There is a very obvious scarcity of research that investigates important non-segmental structures such as VOT in speech production for any variety of Arabic, let alone research done on VOT of Saudi Arabic. However, there have been a few studies which have studied Arabic stops.

These studies have tackled the issue of VOT in a variety of dialects of Arabic (e.g. Yeni-Komshian, Caramazza, & Preston, 1977; Al Ani, 1970; Port & Mitleb, 1983; Flege & Port, 1981; Alghamdi, 1990; Radwan, 1996; Jesry, 1996; Mitleb, 2009; AlDahri, 2013). These studies carried out on Arabic differed in the methods used and the results obtained, but they all agree on the fact that voiced stops in Arabic are pre-voiced (except Mitleb, 2009 and AlDahri, 2013) and that voiceless stops fall in the short-lag region of the continuum. We will review some of those studies briefly as follows. It must be noted that not all of the studies reviewed below reported which dialect specifically they studied, but it seems to be the case that they used MSA, not colloquial or national dialects, except for Khattab (2002), Flege & Port (1981) and Alghamdi (1990) who used colloquial dialects in their studies.

Yeni-Komshian et al (1977) studied the production of stops in Modern Standard Arabic (MSA). Eight Lebanese adults were asked to read words and sentences with stops in the context of three short vowels (a, i, u). This dialect however is not widely used as a spoken medium nowadays, except in the media and in some religious ceremonies.

The researchers found that voiced stops were pre-voiced in all the productions of their subjects, while the voiceless ones fell within the short-lag region. Table 2.2 shows mean VOT values in msec obtained in the study in both contexts (words and sentences).

Table 2.2: VOT values for word initial stops in Arabic words in isolation and sentences in Yeni-Komshian et al's (1977) study

Stops	/b/	/d/	/t/	/k/
Mean VOT values in msec	-65.00	-56.66	25.00	28.33

They also found that there were no major differences between stops produced in words in isolation and words produced in sentences, therefore, they have averaged their results across contexts in Table 2.2. Moreover, they found that there is a tendency for shorter negative VOT and longer short-lag VOT in the production of stops before /i/ than the other vowels /a/ and /u/, as seen in Table 2.3.

Table 2.3: VOT values for Arabic word initial stops in three vowel contexts in Yeni-Komshian et al's (1977) study

Stops	/b/	/d/	/t/	/k/
/a/	-80	-60	20	25
/u/	-75	-70	25	30
/i/	-40	-40	30	30

Another study was also carried out, on MSA, by Jesry (1996), comparing the voicing contrast in Arabic and English. Jesry's study involved three adult Syrians reading words in MSA in a carrier sentence: " qa:la ….. ala:n" *he said …… now*. The stimuli word list used contained Arabic stops and fricatives in initial positions followed by a vowel. The vowels used in the study were the three short vowels and the three long vowels found in Arabic /i, iː, ɑ, ɑː, u, uː/, similar to the vowels used in this study. Table 2.4 summarizes the VOT values obtained in the study in all vowel environments used.

Table 2.4: VOT values for Arabic word initial stops in words in sentences in Jesry's (1996) study

Stops	/b/	/d/	/t/	/k/
Mean VOT values in ms	-68.72	-66.8	27.82	32.19

It is clear that voiced stops were produced with voicing lead and voiceless ones were borderline short-lag. Similar results were obtained by Radwan (1996) as he found that voiceless stops were in the short-lag range and voiced stops were pre-voiced.

Another MSA study by Al Ani (1970) measuring the duration of aspiration of the voiceless stops, recording himself reading words in lists, found that the VOT productions of /k/ were between 60 and 80 msec, while the VOT values of the /t/ were between 30 and 40 msec. These were then definitely in the low end of long lag rather than in the short lag region.

Table 2.5: VOT values for ***Arabic*** *word-initial stops in carrier sentences in Khattab's (2002) study*

Stops	/b/	/d/	/g/	/t/	/k/
Mean VOT values in ms	-55	-63	_	28	31

Khattab (2002) more recently investigated the VOT production of voiced and voiceless stops initially in English and Arabic by children and adults. Only the adults' Arabic VOTs will be presented here while the other two (English monolinguals and English and Arabic bilinguals) will be reported later.

23 subjects were involved in the whole study, but only four Arabic speakers who spoke colloquial Lebanese dialect of Arabic participated in the Arabic study. Results from Khattab's (2002) study again indicated that adult Arabic monolinguals produced short-lag VOT for voiceless aspirated stops and pre-voicing for voiced stops. It should be noted that the voiced velar sound /g/ does not exist in Arabic words in Lebanese Arabic, as it is usually replaced with the glottal stop, however, it is used and produced accurately in many loanwords as in *garage, gateau, English* etc (Khattab, 2002).

All the previous studies on Arabic VOT found that voicing in voiced stops starts prior to the release and was found to be in the negative VOT area, while voiceless stops were more or less in the short-lag region (positive VOT) at least in spoken MSA. As Alghamdi (1990) affirmed, VOT in Arabic dialects might vary significantly in the duration of the voiceless stops, however voiced stops remain always pre-voiced. All the above studies also confirmed the universal pattern coronal < dorsal in voiceless stop VOT, and in parallel (apart from Khattab) found labial < coronal in the negative voiced stop VOTs (in the sense that labials had more negative, less positive, VOT than coronals).

Mitleb (2009) and AlDahri (2013) however found otherwise. Although, their results are inconsistent with the previous research and they examined different dialects than the ones above, they are the most recent reports that investigated Arabic VOT and so are worth examination.

Mitleb (2009) examined stops in Jordanian Arabic (specific dialect not named), also investigating the influence of vowels and whether they play any major role in Arabic VOT or not. Mitleb analysed the alveolar and the velar stops only and excluded the labials in his analysis. His results (Table 2.6) suggested that vowels in Jordanian Arabic played a major role in VOT, as he found that VOT values (all positive) were shorter before short vowels and were longer before long vowels and the differences between them (short and long vowels) were significant. He also found that there was no significant difference between the /t/ and the /k/, contrary to the general view and the general finding of Lisker and Abramson and other studies which found that voiceless VOT values increase as the place of articulation moves back in the oral cavity. Most striking of all, he found no prevoicing with the voiced stops. However, Mitleb (2009) collected his Arabic data from four undergraduate students enrolled in a class of an English pronunciation course, which clearly might have affected his study results.

Table 2.6: VOT values for Arabic word initial stops in words in isolation in Mitleb's (2009) study

Stops	/t/	/d/	/k/	/ g /
Short Vowel	37	10	39	15
Long vowel	64	23	60	20

AlDahri (2013) investigated VOT in Modern Standard Arabic (MSA) by comparing stops like /d, t/ with their emphatic counterparts /dˤ , tˤ /. As we said earlier, studies of this dialect are somewhat artificial. AlDahri found that VOTs of /d , t/ were always longer than those of the emphatic /dˤ , tˤ/, but those of the voiceless /t/ were always longer than those of the voiced /d/. He also found that all stops (voiced and voiceless) measured were produced with positive VOT values and not a single negative value was found even with the voiced stops. He claimed that stops in MSA have no prevoicing, like English and Spanish, contradicting many well-established studies such as Yeni-Komshian et al (1977).

AlDahri (2013) and Mitleb (2009) are the only studies in the literature which found that stops in Arabic were produced with positive VOT values only and no negative VOT values were found in their analyses even for voiced stops. AlDahri (2013) however selected participants who only knew how to recite The Holy Quran properly to be recorded. In his view the best speakers of MSA are those who can recite The Holy Quran accurately. In fact, this might be true if he had selected Arabic native speakers only, but he opted to record people who were non-native speakers of Arabic, which might well have affected his results. In both AlDahri and Mitleb's studies, then, there may well have been some transfer from English evidenced in the Arabic VOTs that were found, especially for the voiced stops. Hence these studies cannot be accepted unequivocally as providing genuine data on the Arabic of monolingual speakers.

Saudi Arabic was examined by Flege & Port (1981) who compared the phonetic implementation of the stop voicing contrast in Saudi Arabic and American English. They conducted three studies. One study concerned the production of Arabic stops and the second one studied the production of English stops by Saudi learners of English and American speakers. The third study tested the learners' intelligibility in producing English stops by the American speakers. Only the Arabic study will be reported here in this section while the English one will be reviewed later in studies of VOT of L2 learners.

The learners were six adult graduate Saudi students at Indiana University, who served in both experiments (Arabic and English). Therefore, once again Arabic data were not provided by full monolinguals, similar to Mitleb's study. The students' task was to read some Arabic words inserted in an Arabic carrier sentence 'I read……and then I go home'. Stops were tested word initially and word finally in CV:C minimal pairs in the context of one long vowel only /a:/. It has to be mentioned that this study is unlike most other studies which tackled Arabic, as the productions of Arabic stops were in Saudi colloquial Arabic instead of the classical or modern standard Arabic. Our own study also followed Flege & Port's (1981) study in examining Saudi colloquial Arabic but from monolinguals of a different region (Northern region, spoken in Arar) than the one used by Flege & Port which was Najdi Saudi Arabic (from the central region of the country).

Flege & Port (1981) measured four acoustic intervals in Arabic and English: 1) initial stop duration; 2) VOT; 3) vowel duration; and 4) final stop duration. Table 2.7 shows the results that concern us.

Table 2.7: VOT values of Saudi Arabic stops in initial position, in words in sentences, from Flege & Port (1981)

Stop	/b/	/d/	/g/	/t/	/k/
Mean VOT values in ms	-85	-82	-75	37	52

Voiced stops in this study were produced 100% with glottal pulsing (prevoicing): -85 ms for /b/, -82 ms for /d/ and -75 ms for /g/. Glottal pulsing (prevoicing) occurred in a number of the voiceless /t/ and /k/ productions but in general, voiceless stops /t/ and /k/ in Saudi Arabic were found to be slightly aspirated, rather than fully short-lag as in some studies above, as the average VOT value of the /t/ was 37 msec and ranged from 20 to 65 msec. The /k/ was more aspirated (52 msec) than the /t/ and ranged from 30 to 85 msec, i.e. in the long-lag range. These VOTs are longer than the VOT values in MSA found by Yeni-Komshian et al. (1977), and indeed than those in Spanish and French (Lisker and Abramson, 1964), but shorter than the long-lag range found in English (Docherty, 1992).

Alghamdi (1990) reported similar results in his study of another Saudi dialect of Arabic (Ghamdi, from the south of the country). He measured the duration of the VOT in three positions in the word, as his subjects read a list of words inserted in a sentence. In a similar manner to that of the previous study by Flege & Port (1981), initial voiced stops were produced with prevoicing (-72 msec) for /b/, (-71 msec) for /d/ and (-69 msec) for /g/. The voiceless stops were produced outside the short-lag region with slight aspiration (32 msec) for /t/ and slightly greater (42 msec) for /k/. See table 2.8. He also found out that Ghamdi Arabic voiceless stops are always aspirated in word-initial, word-medial and word-final positions. Parallel results were also obtained for voiced stops, as pre-voicing was recorded in all positions of the word.

Table 2.8: VOT values of Saudi Arabic stops in initial position, in words in sentences in Alghamdi (1990)

Stops	/b/	/d/	/g/	/t/	/k/
Mean VOT values in ms	-72.04	-71.09	-68.7	32.32	42.12

A further study was done by Alghamdi in (2006), when he measured the VOT of 16 Saudi Arabic speakers' production of three Arabic stops (i.e. /t, k/ and the emphatic /tˤ/). This was not however a monolingual study as Alghamdi (2006) examined the role of VOT and the effect of the speakers' second language, which was English, on their L1. He found that Arabic speakers varied significantly in production of VOT in their L1 according to their fluency in English. He also found surprisingly that speakers who were more fluent in their L2 produced shorter VOT means in their Arabic voiceless stops /t,k/. However, those who were 'less fluent' in English tended to produce longer L1 VOT means. This could be explained as an effect of 'deflection' within the SLM theory that we will review later. In any event, it can be expected that the influence of phonetic variation between L1 and L2 could be extended as the speaker becomes more fluent in the L2 and language transfer could work in both directions (Alghamdi, 2006).

Therefore, to the best of the researcher's knowledge, Alghamdi (1990), Alghamdi (2006) and Flege & Port (1981) are the only studies that examined VOT in Saudi Arabic to be found in the literature at the time of our study, which shows the importance of the current study to give an up to date picture of Saudi Arabic VOT patterns. Furthermore, although, Alghamdi (1990) and Flege & Port (1981) studied colloquial Saudi Arabic, the former focused on southern Ghamdi dialect, and the latter focused on Najdi dialect, both of which are different from the dialect of Arabic speakers and L2 learners of this study (Northern Saudi dialect). Furthermore, these studies did not all gather data from true monolingual Arabic speakers.

Table 2.9 gives a summary of the VOT patterns found in the studies reviewed above.

Table 2.9: Summary of Arabic VOT values for word-initial stops found in the studies reviewed

Arabic stops	/b/	/d/	/g/	/p/	/t/	/k/
Yeni-Komshian et al (1977)	-65.00	-56.66			25.00	28.33
Jesry (1996)	-68.72	-66.80			27.82	32.19
Radwan (1996)	-71.03	-78.23			33.57	38.81
Flege & Port (1981)	-85.00	-82.00	-75.00		37.00	52.00
Alghamdi (1990)	-72.04	-71.09	-68.70		32.32	42.12
Mitleb (2009)	Short V	10	15		37	39
	Long V	23	20		64	60
AlDahri (2013)	13	15			51.65	52
Khattab (2002)	-55	-63			28	31

It is noticeable from all the previous studies that among the Arabic voiceless stops /k/ almost always was found to have longer VOT than /t/, in accordance with the universal POA claim (Cho & Ladefoged, 1999). However, for voiced stops, the picture is not quite as clear but, from the results in most of the studies above, and especially the Saudi ones, it seems that there is a tendency for voiced stops to have less negative VOT from the front to the back of the mouth /b>d>g/. Put another way, their progression follows the universal pattern in the sense that VOT becomes more positive (=less negative) from front to back. However, this will be investigated further in our study.

It is also notable that no studies have included and confronted results for Saudi Arabic stop VOT in both contexts (words/ sentences) and followed by all three vowels, but have simply chosen one context or vowel or the other. Therefore, these are gaps in the literature which our account will fill.

2.7 VOT studies of English monolinguals

English VOT patterns were investigated by several researchers (e.g. Lisker & Abramson, 1967; Klatt, 1975; Docherty, 1992; Khattab, 2002; Scobbie, 2002). All these five studies will be briefly reviewed. However, the first two studies were on American English while only the last three focussed on British English so were within a similar context to the current study.

Lisker and Abramson (1964) were the first to define VOT in their famous pioneering study of 11 different languages of which English was one. The authors measured the VOT of word initial voiced and voiceless stops in words in isolation and in carrier phrases, read by four English native speakers. The ENS produced voiceless stops with long-lag while voiced stops ranged between pre-voicing and short-lag. The researchers separated the positive VOT values from the negative ones in the presentation of their result as there was one English native speaker who was responsible for 95% of all the pre-voiced tokens while another one was responsible for the remaining 5%. In general, however, there were two separate VOT ranges for voiced and voiceless stops with no overlap found between them. Also, there was a significant difference between stops produced in words in isolation and those produced in phrases, with the former being longer. Tables 2.10 and 2.11 show the results obtained from Lisker & Abramson's study for word-initial stops in two contexts (words/sentences).

Table 2.10: VOT results for word-initial English stops in words in isolation in Lisker & Abramson (1967)

English stops	/b/	/d/	/g/	[pʰ]	[tʰ]	[kʰ]
+VOT	1.00	5	21	58	70	80
-VOT	-101	-102	-88			

Table 2.11: VOT results for word-initial English stops in carrier sentences in Lisker & Abramson (1967)

English stops	/b/	/d/	/g/	[pʰ]	[tʰ]	[kʰ]
+VOT	7	9	17	28	39	43
-VOT	-65	-56	-47			

Another study on American English was done by Klatt (1975), in which the VOT in word-initial stops followed by four vowel (i, ɛ, a, u) environments was measured. Klatt studied the effect of both POA and vowel on VOT and notably measured aspirated and unaspirated voiceless stops separately, as we will do. Her subjects were three adult male English speakers who uttered monosyllabic words inserted in a carrier sentence 'Say __ instead'. The results varied according to both POA and vowel environment. VOTs at different POA's progressed from front to back of the mouth in the usual way, and VOTs were found to be significantly longer before high vowels /i, u/ than before low ones /a, ɛ/. A number of pre-voiced tokens were recorded by some English speakers, however, Klatt ignored those tokens in her analysis as she affirmed that pre-voicing was not considered important for phonemic distinction in English (Klatt, 1975: 688).

Table 2.12 shows the results gained from Klatt's study for word-initial stops in carrier sentences. It is notable that she found that voiced stop VOTs were in a similar short lag range to those of unaspirated voiceless stops.

Table 2.12: VOT results for word-initial English stops in carrier sentences on Klatt (1975)

Stops	/b/	/d/	/g/	[pʰ]	[p]	[tʰ]	[t]	[kʰ]	[k]
Mean VOT values in ms	11	17	27	47	12	65	23	70	30

Docherty (1992) studied the VOT patterns of Southern British English and reported in detail the various features of the timing of voicing in voiced and voiceless obstruents. Docherty's subjects were five adult male students who were British speakers of Southern English doing undergraduate degrees at Edinburgh University. They were all educated and brought up in the South-East of England. He measured the VOT of stops and fricatives in different contexts (in isolation and in a carrier phrase). Voiceless stops in initial positions (not following /s/) were aspirated in British English as they were in the long-lag range: 46 msec for /p/, 66 msec for /t/ and 66 msec for /k/ in words in isolation. Voiced stops exhibited shorter VOT values (25 msec) for /b/, but with slight aspiration for /d/ (33 msec) and for /g/ (40 msec). Voiceless unaspirated stops were also in the short-lag range but interestingly with shorter VOT values than the voiced ones in words in isolation. He also tested his subjects saying those stops in carrier phrases where they were all noticeably shorter in VOT than in single isolated words (see the two tables 2.13 & 2.14 below). Docherty (1992) states that even pre-voicing was recorded for some voiced stop tokens, particularly of /b/ and /d/. Those pre-voiced tokens were excluded in calculating the means in tables 2.13 and 2.14, however.

Table 2.13: VOT results for word-initial English stops in words in isolation in Docherty's study

Stops	/b/	/d/	/g/	[pʰ]	[p]	[tʰ]	[t]	[kʰ]	[k]
Mean VOT values in ms	25.00	32.84	39.96	45.74	18.52	66.45	23.70	66.09	27.92

Table 2.14: VOT results for word-initial English stops in carrier phrases in Docherty's study

Stops	/b/	/d/	/g/	$[p^h]$	[p]	$[t^h]$	[t]	$[k^h]$	[k]
Mean VOT values in ms	15.24	20.63	26.81	41.54	15.39	64.54	22.78	62.18	26.28

The finding that some English monolinguals produced pre-voicing for voiced stops agrees with Lisker & Abramson (1967) and Khattab (2002). It shows that voiced English stops should not be regarded as only having positive short-lag VOTs: pre-voiced tokens can occasionally be produced by English native speakers.

Another study of British English is Scobbie's (2002) cited in Khattab (2002), which considered dialectal variation in the production of VOT. The subjects of this study were 12 adult British monolinguals but with unusual regional bidialectal backgrounds. The subjects were from Shetland, Scotland. Some had a regional Shetlandic accent and others had an unspecified Scottish accent. Table 2.15 shows the results of the study.

Table 2.15: VOT results for word-initial English stops in words in isolation in Scobbie (2002)

Stops	/b/	/d/	/g/	$[p^h]$	$[t^h]$	$[k^h]$
Mean VOT values in ms	-29.00	-25.00	-6.00	56.00	66.00	75.00

As can be seen from table 2.15, the VOT values for voiceless aspirated stops in the Shetlandic accent were broadly comparable to the English ones found by Docherty (1992) and other studies done on English. Their ranges varied hugely between short and long-lag, however, even though the means, as seen in table 2.15, were firmly in the long lag range. In this unusual accent of English, however the VOT pattern exhibited pre-voicing for voiced stops and long-lag for the voiceless ones, which is comparable to some of the Arabic VOT patterns found in Alghamdi (1990), and Flege & Port (1981). In Scobbie's study, the values for VOT ranged from pre-voicing to short-lag for voiced stops, and from short-lag to long-lag for voiceless aspirated stops. A substantial amount of overlap was found between the two categories of voicing, i.e. voiced and voiceless.

Table 2.16: VOT results for English word-initial stops in carrier sentences in Khattab (2002)

Stop	/b/	/d/	/g/	[p^h]	[t^h]	[k^h]
Mean VOT	5	10	28	63	70	80

Khattab (2002) examined the production of English stops in initial prevocalic position by a group of six British native speakers. The subjects read lists of English words embedded in sentences. The English subjects produced /b d g/ mainly with short-lag VOT but some pre-voiced tokens were also recorded. Voiceless stops were produced with long-lag and both voiced and voiceless stops progressed from front to back. No results about vowels or unaspirated voiceless stops were reported.

In summary, the above studies of English overwhelmingly show VOT becoming more positive (or in Scobbie's case, less negative) across POAs from front to back. Furthermore, while a few studies have compared VOTs of stops before different vowels and in different contexts for English, this does not appear to have been done for both together in British English, with attention separately to aspirated and unaspirated voiceless stops as well as voiced, all of which our study will cover.

2.8 Contrasts in the literature between English and Arabic VOT

Khattab (2002) provided some generalisations about the common VOT patterns found in Arabic and English (Table 2.17).

Table 2.17: Generalizations about Arabic and English VOT patterns in word-initial positions provided by Khattab (2002, p. 218)

Arabic VOT patterns (discounting data of AlDahri and Mitleb)	English VOT patterns (discounting Scobbie's data)
Initial VOICED stops have a predominance of voicing lead (VOT is between –60 and –90ms).	Initial VOICED stops are either unaspirated (VOT is between 0 and 25ms) or voiced.
Initial VOICELESS stops are characterized by a delay of between 25 and 60ms in voicing, relative to the release of the stop.	Initial VOICELESS (aspirated) stops are characterized by a delay of between 50 and 80ms in voicing, relative to the release of the stop.
Presence or absence of vocal fold vibration in the closure duration of stops is contrastive.	Presence or absence of vocal fold vibration in the closure duration of stops is not contrastive.

We may add that the predominant POA pattern in both English and Arabic is for VOTs to become more positive (or less negative) from front to back.

Lisker & Abramson (1964) point out that English and Arabic are categorized as languages that have the same number of voicing distinctions for stops in phonemic terms: voiceless and voiced. But there are nevertheless considerable differences in the VOT patterns of these two languages (as well as within voiceless stops in English). In English, if we leave aside Scobbie's data, and indeed a number of regional dialects of English, there are only rare vocal cord vibrations before or while producing the stops /b d g/, which are unaspirated and considered phonologically voiced. On the other hand, [p^h t^h k^h] are strongly aspirated in English. Flege & Port (1981) therefore affirm, like Khattab, that while the contrast between English voiced and voiceless stops is mainly of aspiration, it is of presence or absence of glottal pulsing during the production of the stops in Arabic.

Consequently, Ladefoged & Maddieson (1996) and Lisker & Abramson (1971) acknowledged that cues for voicing in initial stops could be obtained from a variety of timing differences between glottal and supraglottal events, i.e. VOT differences. This view is supported by a number of studies (e.g. Khattab, 2002; Jesry, 1996; Radwan, 1996; Flege & Port, 1981; Yeni-Komshian, Caramazza & Preston, 1977) which also claim that Arabic follows a binary system of presence or absence of glottal pulsing during the period of the stop closure, whereas the contrast in English is mostly one of aspiration. This is illustrated in figure 2.2. Yet once again we must not overlook the unaspirated voiceless stops in English having very similar VOT to the voiced stops and so displaying evidence that the phonological voiced-voiceless stop distinction in English must be marked by additional features (outside the scope of the present study) besides VOT.

The English VOT range for voiced stops and voiceless unaspirated stops falls almost within the range of Arabic voiceless stops, whereas the voiced Arabic stops and the voiceless aspirated English ones are each at an extreme end of the scale (see Figure 2.2).

The most common difference in word initial stop consonants found between a number of Arabic dialects and English is the short positive VOT (short-lag) for voiceless stops in Arabic and the long positive VOT (long-lag) for voiceless aspirated stops in English. In Saudi Arabic, however, as Alghamdi and Flege and Port show (see section 2.6 above), in contrast with some other dialects of Arabic, voiceless stops are more aspirated and VOTs occur rather in the low part of the long lag range. By contrast all voiceless stops [p^h t^h k^h] in initial positions in English are in the high long-lag VOT range and produced with aspiration: the voicing of the subsequent vowel starts vibrating long after the burst of the stop closure. As for the voiced stops /b g d/ in English, in initial position, and indeed the voiceless unaspirated stops [p t k], they are produced with short positive VOT. Just after the burst of the stop closure, the voicing of the following vowel starts. The pattern of VOT in both English and Arabic presented in figure 2.2 will be later examined against the results of the current study to see whether this pattern applies to results obtained in this study or not.

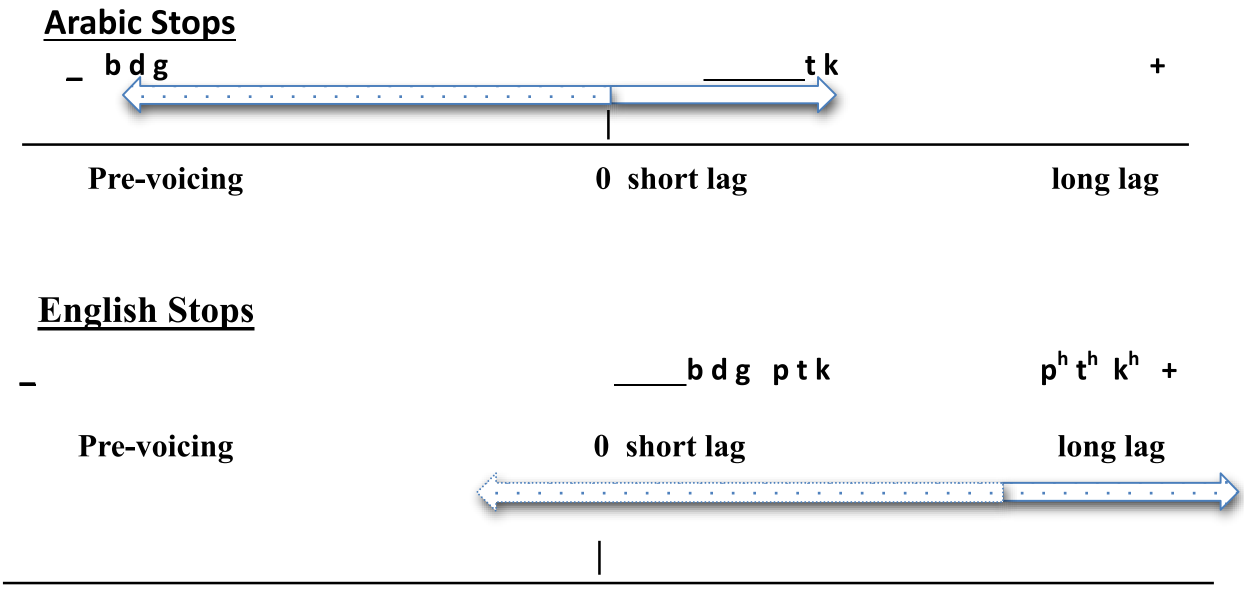


Figure 2.2: Schematic diagram of Arabic and English stops along a VOT continuum. Adapted from Deucher & Clark 1996)

2.9 VOT production studies of Arabic learners of English as a second language

There have been some studies which examined the production of VOT by adult Arabic speakers of English as a foreign language (e.g. Khattab, 2002; Port & Mitleb, 1983; Flege & Port, 1981; Flege, 1980). Some of these studies will now be briefly discussed.

One of the earlier English learner VOT studies was by Flege (1980) and Flege & Port (1981) who conducted several investigations, one of which examined a number of Saudi students' VOT productions both in English and Arabic. As discussed earlier in section 2.7, the results for Arabic were that voiced stops were produced with pre-voicing in Saudi Arabic, but voiceless ones were aspirated slightly. Two groups (six in each) of Saudi subjects along with one group (six subjects) of Americans participated in the elicitation of VOTs of English stops. The two Saudi groups varied in the length of their residence in America, as the first group had stayed for 39 months and the second group had stayed for about 8 months. All groups including the Americans were asked to produce English plosives in target words in a carrier sentence, and their VOT, vowel duration and stop closure duration were measured word-initially and word-finally.

Flege & Port (1981) measured the VOT and closure duration of the voiceless stops (aspirated only) but for some reason VOT for voiced stops in this experiment was not measured.

Table 2.18: VOT results for word-initial English stops in carrier sentences from two groups (Ar1 had stayed for 8 months in the US while Ar2 had stayed 39 months) in Flege & Port (1981)

Stop	Group	/b/	/d/	/g/	[p^h]	[t^h]	[k^h]
VOT	Ar1				14	35	41
VOT	AR2				21	30	47

Flege & Port (1981) identified that there was a significant difference between the VOTs of Saudi students and those of their American counterparts, attributable to non-English phonetic characteristics produced by the Saudi subjects (Flege & Port, 1981: 133). Saudi subjects were heavily influenced by their L1 in transferring their L1 phonetic features to their production of the English stops: compare their VOTs for /t k/ of 37 and 52 msec in Table 2.7. Although the group which had stayed for 39 months in America produced longer closure durations for voiceless stops than voiced stops at all the three places of articulation, no significant difference was found in the production of VOT between the two Saudi groups. Flege & Port (1981) concluded that the group which had the longer period of residence in America was approaching the English phonetic patterns, despite still being significantly different from them, whilst the less experienced group seemed still to be influenced by the phonetic features found in their L1 (Arabic).

Saudi learners also showed a phonetic difference between the English voiced /b/ and voiceless [p^h] even though this contrast does not exist in their L1 (Arabic). A great number of the Saudis' [p^h] productions however had pre-voicing during the period of the stop closure, and so their laryngeal control differed in producing /p/ compared to /t/ and /k/. This and a number of other first and second language studies therefore propose that it might be harder to learn to control a newly acquired pattern of glottal-supraglottal timing than one involving purely supraglottal timing (Flege & Port, 1981). A more theoretically grounded account of results like those above will be presented later, when we describe Flege's Speech Learning Model.

A later study was done using Jordanian speakers of Arabic and American native speakers of English (Port & Mitleb, 1983). There were 12 Jordanian subjects divided into two groups and one group of 5 American monolinguals. The first group of Jordanians varied in their length of residence (LOR) in America, as their LOR ranged from one year to one year and 4 months, while the other Jordanian group had never lived in America or any other English speaking country. All participants were asked to read monosyllabic English words inserted in a carrier sentence testing the production of /t-d/ and /p-b/ stops (not including voiceless unaspirated). Port & Mitleb measured the VOT of the initial stops, as well as other things which are not of concern to us.

The authors' findings were that the Jordanian speakers' English VOT productions were in the short-lag for /t/ and /p/, so a little lower than the learner values in Flege and Port (1981) above, while the Americans' VOT productions were in the long-lag region. Similar to the findings in Flege & Port's (1981) study, Jordanians were able to produce the problematic sound for many Arabs, the /p/, as well as /t/, but with shorter VOT values than their English native speaker counterparts. However, some of the Jordanian subjects with more experience of English produced longer VOTs for /p/ than for /b/, even though some of the /p/ productions were with weak voicing. As in Flege & Port's (1981) study, there was no significant difference in the production of initial English stop VOT between the two Jordanian groups.

According to Port & Mitleb (1983), new phonological distinctions can be learnt by adult foreign language learners, as shown by the Jordanians' production of a /p – b/ contrast in English, even if it differed from NS in the actual VOT values. However, control of temporal implementation rules such as precise VOT length would be more challenging as in the case with the Jordanians' shorter VOT productions for the English /t p/ than those of NS. Port & Mitleb (1983) therefore conclude that L2 learners often substitute small phonetic details of the second language with those of their first language, even when a voicing contrast is successfully made by the learners in the second language. Again, see 2.11 for further discussion of such findings from a recent theoretical standpoint.

Khattab (2002) also found that Lebanese Arabic-English bilinguals produced short-lag VOT with slight aspiration for voiceless aspirated English stops and pre-voicing for voiced stops. Thus, they produced the Arabic VOT pattern (cf. section 2.6) in their English production. Although occurrences of /p/ and /g/ are infrequent in Lebanese Arabic, the bilinguals succeeded in producing them: /p/ was produced with short-lag and some slight aspiration in some cases, and /g/ was produced with pre-voicing. We may observe that they were also produced in a parallel way to the other stops, given the universal POA pattern: /p/ had shorter positive VOT than /t k/ and /g/ had less negative VOT than /b d/. Khattab interpreted this as due to the participants applying Arabic phonetic features (see the later discussion of the Featural Model in 2.10).

Table 2.19: VOT results for English-Arabic bilingual adults for word-initial English stops in carrier sentences in Khattab (2002)

Stop	/b/	/d/	/g/	[p^h]	[t^h]	[k^h]
Mean VOT	-93	-98	-65	28	34	38

It has to be noted that Khattab did not provide exact VOT figures in her study but she provided graphs from which we estimated the above figures.

Overall, Flege (1980), Flege & Port (1981), Port & Mitleb (1983) and Khattab (2002) reported similar results for the English aspirated /p/ by speakers of Arabic, who made a phonetic difference /p – b/, but did not achieve native-like production of the VOT of newly acquired segments. One reason might be because the subjects generalised the Arabic interval difference for /t-d/ to the English /p-b/ and /k-g/ contrast (Flege, 1980; Flege & Port, 1981; Port & Mitleb, 1983). In Khattab's data the /t - d/ difference is 132 msec, while the /p- b/ one is 121 and the /k - g/ one 103 msec, which lends some support to this, but is not conclusive.

The above are all the studies we know of concerning stop VOT production of Arab learners of English (including bilinguals). Apart from their scarcity, the limited number of participants they have used and the different dialects that they have examined, there are a number of issues that have not been addressed by these Arab learner of English studies. One is that all above studies focused mainly on voicing and aspiration. They measure learners' VOT for the target of English voiced /b/ compared with voiceless aspirated [p^h] neglecting the unaspirated voiceless stops such as [p], as we have seen in their results above. Yet English allows aspiration alone to be examined through the comparison of [p^h] and [p] (*pit spit*) and voicing alone through [b] and [p] (*bit pit*), as some ENS studies did (section 2.7, e.g. Docherty 1992). Also neglected have been the effects on learner English VOT of context (isolated word versus sentence), as they either focused on one context or the other, and of following vowel quality.

Therefore, in light of these limitations of previous studies, the current study is going to examine stop VOT patterns of Saudi Arabic learners of English more thoroughly and systematically.

2.10 Models of the Acquisition process of stops by learners of English as a second language

Much acquisition research concerning sounds has been motivated by the need to understand how learner ability in a second language develops, and why second language learners, particularly adult learners, usually differ from monolingual native speakers. A wide range of theories and models have been proposed. For example, some theories concentrated mainly on the acquisition of production such as the Contrastive Analysis Hypothesis (CAH) of Lado (1957) and the Markedness Differential Hypothesis (MDH) of Eckman (1977). Some other theories focused mainly on perception, such as the Perceptual Assimilation Model (PAM) of Best (1994, 1995) and the Featural Model (FM) of Brown (1998, 2000). We will briefly review these, in respect of how they apply to our area of concern (acquisition of stops), before moving to our chosen model, which we feel synthesises and extends the best features of other approaches.

Lado (1957) claimed that the inaccuracies produced by the learners when learning the L2 are caused simply by the differences between their L1 and L2 phonemic inventories. Therefore, sounds which are dissimilar to their L1 sounds, will be hard for them to learn. Accordingly, the CAH concentrated on the contrastive analysis of the phonemic differences found between the learners' L1 and their L2. While the CAH is based on comparison of the structures of languages in a way that is neutral as to perception or production, in practice researchers focused on production data (especially written errors) to support its claims. Furthermore, it did not embrace a wider range of possible factors affecting L2 performance than simple L2 - L1 differences. Hence, though the idea that L1 affects L2 is still very much with us, the CAH is an unsuitable model today.

Eckman's (1977) MDH was a further improvement of Lado's model (CAH) and again tended to focus on production. The MDH claims that there may be markedness differences among the sounds of the L2 that are new and different from the L1 sounds. The ones that are more marked will be more difficult for the learners to acquire while the less marked ones will be relatively easier to learn. This means that there may be some directionality of difficulty when learning second language sounds. If language A has a marked feature where language B has an unmarked one, while the CAH would predict equal difficulty, the MDH would say that a speaker of language A learning language B may do better than a speaker of language B learning language A. Our study is not concerned with comparing Arab learners of English with English learners of Arabic, but if it was, the MDH might predict that Arabs learning English /p/, a very common sound in languages of the world, so unmarked, might find this not too hard to produce, despite its absence in L1. By contrast Arabic has stops that are uncommon, highly marked, which would make their learning harder for English learners of Arabic (e.g. /dˤ, tˤ/). Thus, the MDH modifies the CAH by adding to simple L1 influence the influence of considerations based on language universals (in this case markedness).

The PAM by contrast focuses on perception of differences between sounds rather than their production. In some ways, it resembles a perception oriented version of the above models in that it claims that if someone hears two foreign language sounds that are perceived as different sounds in their L1, they will distinguish them in L2 as well (no difference between languages). Where there are differences between languages, a variety of situations may arise which result in greater or less difficulty. For instance, if L2 makes a distinction that is not heard as different in the learner's L1, but one of the L2 sounds is a better fit than the other to what is heard in L1, then the learner may still have some success in differentiating them. Thus, if an Arab learner of English perceives both English /b/ and /p/ as sounding like their L1 /b/, they would surely feel that English /b/ is a better fit to their L1 category than English /p/, so might have some success in learning to differentiate /p/.

The FM also targets perception, but analyses it in terms of whether the same feature distinctions are made somewhere in the L1 as in the L2, rather than whether specific combinations of features are heard and distinguished in both L1 and L2. From this perspective, Arabs learning to hear English /p/ as different from /b/ should be helped by the fact that, although they have no /p-b/ distinction in L1, they do have the feature labial and the feature voiced-voiceless prominent in L1 in many other pairs of sounds like /t-d/ etc. English learners trying to perceive the difference between Arabic /t/ vs /tˤ/ on the other hand have no such help since plain vs emphatic (phonetically, pharyngealized) is not a feature occurring in contrastive sounds anywhere in English.

The above models have several disadvantages. One is that they focus either on production or perception but do not unite the two. Another is that they are formulated primarily for naïve monolinguals who are new to an L2, rather than L2 learners at a more advanced stage, or bilinguals. For the latter, what they have already learnt of the L2 (their developing L2 competence often termed their interlanguage) has to be considered as an influence on their production or perception, not just their L1. This makes them more models of how beginners might produce or perceive sounds rather than full models of L2 acquisition of sounds which apply over the whole period of acquisition.

The current study investigates the acquisition of L2 sounds seen in Arabs who are at an advanced stage of learning English, and sees acquisition as needing to be understood taking into account both perception and production in an interrelated way, even though the data we gathered relates just to production. Consequently, none of the previously mentioned models fit our requirement.

Syed (2013) suggests that the 'Speech learning Model' SLM is the only model that is relevant to a study of advanced learners, as well as beginners. In other words, unlike the models above, it deals with change of ability over time, which makes it 'the only extant theory that focuses explicitly on L2 speech acquisition' (Flege, 2003, p. 326). Therefore, the SLM is the main model to be adopted in this study as it includes consideration of advanced learners. Furthermore, it takes into account both perception and production. The SLM developed by Flege (1995) will be reviewed in the following section.

2.11 The Speech Learning Model (SLM)

The SLM (Flege, 1992) makes important predictions for very advanced L2 learners as well as bilinguals. It predicts that the phonetic categories held in the minds of L2 learners might be different from those of L1 monolinguals even if the learners have been learning the L2 for a long time. Such differences in categories stored in the mind will affect both their perception and production of sounds. It should be noted that the SLM refers primarily to phonetic categories, not phonemic ones. For instance, it treats the English [t] and [t^h] as different categories, and proposes that native speakers have different mental representations for them, because they are phonetically very different (e.g. in VOT). It does not suppose that native speakers have one category for them just because they are allophones of the same phoneme /t/.

The SLM also proposes that, aside from where sounds are the 'same' in both languages (meaning the same at an allophonic, phonetic level), there are several distinctive types of situations which can arise, leading to different learning routes when learning L2 segments: those for ‘new’ and various types of ‘similar’ sounds (Flege, 1987c, 1995). Flege explains that certain L2 sounds will be ‘new’ to the learners while some other sounds will be perceived as ‘similar’ to the learners’ native language (L1) and suggests, as a rough guide to the difference, the IPA symbol criterion (Flege 1992). For instance, for an Arab learner of English, [p] is new, since no phone in Arabic would be labelled with the IPA symbol [p], while [b] is similar, since both languages have a phone that would normally be written [b] in IPA. Note that for Flege English [b] would not count as the 'same' as Arabic [b] because, although they are expressed by the same IPA symbols, as we have seen there are considerable voicing differences reflected in the VOT (Arabic normally pre-voiced, English usually short lag positive VOT). For Flege what decides sameness or similarity is the allophonic properties of the sounds, not broad phonemic correspondence (cf. the CAH).

Flege also notably claims that, at an earlier stage of learning, better L2 learning of the ‘similar’ L2 sounds will be established by the learners than of the ‘new’ sounds. Nevertheless, as acquisition continues and the learner advances in learning the L2 sounds, the learner improves more on the ‘new’ than the ‘similar’ sounds and thus the learner will eventually attain a higher level of perception and production accuracy on the ‘new’ sounds than the ‘similar’ sounds.

The SLM provides the rationale behind this learning progression, arguing that learners in their initial learning stages depend on their first language phonemes to aid them in learning L2 sounds. Hence initially they appear to do better on sounds that are similar to L1 sounds than on new sounds (rather as CAH or PAM would suggest). Their performance on similar sounds, however, suffers from L1 interference. For example, in our study this would predict than Saudi learners would perceive English [b] reasonably satisfactorily, despite its VOT difference from L1, but pronounce it with pre-voicing based on L1. That pronunciation, although contributing to a 'foreign accent', would pass as successful acquisition measured just in phonemic rather than phonetic terms. Initially, however, they would have greater difficulty hearing and producing [p] as it has no L1 counterpart. The contribution of the SLM is then to follow this through time and predict that more advanced learners will in fact retain the detailed phonetic difference for similar sounds (Arabic [b] for our learners) precisely because of the similarity with L1 which makes the subtle VOT difference between an Arabic [b] and an English [b] both hard to spot and unimportant in the sense that it does not cause any communication problems. The same reasoning applies to English [t^h k^h] which are very similar to Saudi Arabic /t k/, both with aspiration in the long lag region, though in different areas of it. By contrast, new sounds (e.g. [p^h p] in our example) precisely because they are 'new' sounds unlike anything in L1, over time are more noticeable and attract more effort so that advanced learners end up more nativelike on those sounds. In Flege's account new sounds require the learner to establish a new phonetic category in their mind (called dissimilation) while the similar sounds do not: those L2 sounds are simply included with the corresponding L1 categories which the learner already possesses (called assimilation).

Essentially, Flege argues that it is the phonetic rather than phonemic similarities and differences between L1 and L2 sounds which are the major basis of learning L2 sounds, or as he labelled it creating new phonetic categories for new L2 sounds. According to the SLM, if L2 learners can notice the differences between L2 and L1 sounds, they could form new categories of L2 sounds. If this happens, that category formation of L2 sounds will enable them to defeat the interference of their L1. As a result, that will permit the learners over time to come to perceive and produce L2 sounds in a way similar to native speakers of that language.

In more detail, Flege (1995) predicts a particular performance of L2 learners. Initially they will attempt to identify each L2 sound perceptually as “a positionally defined allophone of the L1” (p. 263) and their production of the sound will follow from that perception. Then as the learners advance and gain more L2 experience, they start to distinguish the phonetic differences between L1 and L2 sounds, which is much easier to do with the new rather than the similar sounds. When learners do distinguish the phonetic differences between L1 and L2 sounds, they begin to develop a “phonetic category representation” for the new L2 sounds differing from phonetic categories already established in their L1 native sounds (Flege, 1995, p. 263).

Flege supposes that the ability to learn new sounds remains active and develops through the person’s life (Flege, 1995). This is in contrast with the critical period hypothesis, which holds that learning new sounds after/around the age of puberty would be impossible or extremely difficult (Scovel, 1988). Although, Flege considers that the learnability of L2 sounds decreases as age of learning increases. However, the idea of complete loss of ability to acquire new sounds after the critical period is not considered by Flege. Instead he proposes the idea of the filtering of L2 sounds through the L1 sound system at the early stages of learning regardless of the age at which that initial learning occurs. As indicated above, Flege (2003) adds that extensive exposure to L2 sounds could halt that filtering at later stages by enabling the learner to establish new L2 categories.

Following from these assumptions and explanations, Flege established his “Speech Learning Model” in order to provide a clearer understanding of how L2 learners perceive and produce L2 sounds in a way similar, or not, to native speakers of the target language. The SLM offered four main postulates, which are reproduced below from Flege (1995, p. 239) as follows:

1- The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning.

2- Language-specific aspects of speech sounds are specified in long-term memory representations called phonetic categories.

3- Phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as the realization of each category.

4- Bilinguals strive to maintain contrast between L1 and L2 phonetic categories which exist in a common phonological space.

The SLM also offers seven hypotheses concerning various settings of L2 acquisition, which will be referred to later as H1, H2, H3, etc. All of the seven hypotheses are repeated from Flege (1995, p. 239). All the hypotheses are listed below, although one of them, the fourth hypothesis (H4), which concerns the age of learning, is not relevant to this study. Therefore, only the other relevant ones will be reviewed afterwards.

H1: The sounds of L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level.

H2: A new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds.

H3: The greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned.

H4: The likelihood of phonetic differences between L1 and L2, and between L2 sounds that are noncontrastive in the L1, being discerned decreases as AOL increases

H5: Category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and L2 sounds (diaphones). Eventually, the diaphones will resemble one another in production.

H6: The phonetic categories established for L2 sounds by a bilingual may differ from a monolingual's if: 1) the bilingual's category is 'deflected' away from an L1 category to maintain phonetic contrast between categories in a common L1-L2 phonological space; or 2) the bilingual's representation is based on different features, or feature weights, than a monolingual's.

H7: The production of a sound eventually corresponds to the properties represented in its phonetic category representation.

Hypothesis number one (H1) predicts that sounds of an L2 will be perceived by learners at an allophonic level, rather than an abstract phonemic level, as described earlier. This means that if an L2 phoneme has allophones that differ phonetically in different specific contexts of the word, learners may perceive those different allophones differently, so with H7 go on to produce them differently. For instance, if an Arab learner of English perceived the L2 /t/ when it occurs aspirated in *team* as [t^h] as equivalent to L1 [t^h], but failed to make this connection when it appears unaspirated in *steam*, that would support Flege's contention. Just because the two sounds belong to one phoneme in L2 does not mean the learner always perceives them as the same sound and produces them the same. We will test this claim on the learners' acquisition of allophonic variation of the English voiceless stops as there are two major allophonic variants of English voiceless stops, aspirated and unaspirated voiceless stops.

H2, H3 and H6 are all associated with the establishment of a new phonetic category for L2 sounds (Flege, 1995, p. 239) (dissimilation). According to H2, learners may develop a new phonetic category (learn a new sound or distinguish successfully a similar sound) if they can distinguish the phonetic differences between an L2 sound and any equivalent L1 one. If that happens we will eventually detect it as different in their production too (H7). H3 suggests that the likelihood of establishing the new phonetic category depends on the perceptual difference between the L1 and the closest

L2 sounds: the greater the difference, the greater the learning of L2 sounds as distinct categories. However, H6 predicts that the phonetic categories of L2 sounds newly established by learners might still be different from those of the native L1 sounds produced by monolinguals of the target language due to them being 'deflected away'. This deflection of L2 learners' new phonetic categories, and hence their non-native production (H7), would arise because of the learners attempting to maintain clear phonetic contrasts between the L1 sound and its closest L2 sound. Learners may perceive an L2 sound as different from both L1 and other L2 sounds and this may lead the learner to develop a new deflected phonetic category distanced from both L1 and L2 sounds in the mind. Thus, the learner's new category, and their production based on it (H7), will not be in the location of the nearest L1 sound, but may not quite be in the position of the L2 target sound either. Although H6 of the SLM refers to bilingual behaviour in acquiring L2 sounds, it can be extended to apply to the advanced L2 learners of this study.

We will test these hypotheses (H2, H3, H6) for example on our learners' acquisition of the new labial voiceless sound /p/ in both aspirated and unaspirated forms, to see whether they show evidence of having distinguished the phonetic differences between their closest L1 sound /b/ (or possibly /t/) and their new L2 sounds [p p^h]. If, as predicted, they can distinguish the difference from L1 of such 'new' sounds, will they establish one or two new phonetic categories for them as in (H2)? Hence will they demonstrate the difference in production (H7)? Furthermore, if they establish a new phonetic representation for the L2 [p], will it be similar to the ENS [p] or 'deflected away' from it (H6) so as to create distance between the VOTs of L2 [p] and [p^h] and [b] in their own minds, especially since the TL English VOT distance of [p] from [b] is very small? The same concept applies to the other voiceless stops, which exist in their

L1 but in their aspirated form only. Can the learners perceive the difference between aspirated (long-lag) English stops and unaspirated (short-lag) ones? If they can, can they create a new phonetic category for the new short-lag category and demonstrate this in their production and will it be similar to that of English monolinguals?

Voiced stops furthermore are produced with short-lag in English (of the variety they are exposed to) while with pre-voicing in the learners' L1. Hence, we will see whether they can notice the difference between short-lag and pre-voicing also and whether they can create a new phonetic category here too. In this case of 'similarity', however, we may feel that they are less likely to notice the difference than in the case of unaspirated English [t] and Arabic [t^h], for example. According to the above hypotheses, sounds of L2 which are perceived by L2 learners as very dissimilar to their closest L1 sounds will be learned effortlessly and sounds which are considered as similar to the corresponding L1 sounds will be challenging to learn. Instead it may be that "equivalence classification" between the L1 and L2 sounds (H5) will be formed.

The above hypotheses expect that L2 learners may form a new phonetic category for each 'new' L2 sound, and for some that are 'similar' to L1, if they manage to perceive a difference. However H5 of the SLM also predicts that for 'similar' instances the outcome, if the differences are not prominent and are not perceived, might be an 'equivalence classification' between L1 and L2 sounds, which as a result blocks the formation of new phonetic categories (Flege, 1995, p. 239). In other words, if an L2 learner equates an L2 sound with an L1 sound or if he equates an L2 sound with another L2 sound, their category formation (learning) of the new L2 sound could be blocked. This happens when the learner cannot distinguish any phonetic differences between the L1 and L2 sounds at all or when the learner cannot perceive the phonetic differences quite accurately. In such a case, the learner will perceive two sounds in two different languages as one sound and will also produce the two sounds similarly (H7). The concept of equivalence classification between L1 and L2 has been studied and supported by several researchers such as MacKay et al (2001) and Flege (1987b) along with some others.

Therefore, according to the above hypotheses, a number of possible learning outcomes can be anticipated by the L2 learners according to their perception of and differentiation between L1 and L2 sounds (compare the same - similar - new distinction which we mentioned at the start). The first possible outcome is that learners create an equivalence classification of L1 and L2 sounds blocking the establishment of any new phonetic category, if the difference between the two sounds is relatively minor and they are perceived as the same or very similar. As a result learners will perceive and produce L2 sounds in a similar way to their native L1 sounds. If an L1 sound is in fact almost identical to the L2 sound (as for instance often arises with sounds like [m] in different languages), then equivalence classification is of course natural and desirable. Where there are some differences between similar L1 and L2 sounds, however, according to MacKay et al (2001), even long exposure to the L2 environment does not help the learners in learning L2 sounds if they have established a strong equivalence classification.

The second possible outcome is where learners perceive some phonetic difference between the L1 and its closest L2 sound but that perception is not well established enough for them to form a new phonetic category for that sound. This happens when the difference between the two sounds is to some extent large but it is not large enough for the learners to perceive clearly and so allow them to establish a new phonetic category for that new sound. Therefore, in such a case the learners perceive the new sound sometimes with a slight difference from their L1 sound and sometimes with no difference at all, so their production, according to Flege, will be sometimes L1 like and sometimes L2 like but it is in fact in a category that is in between those L1 and L2 sounds, neither an L1 like nor L2 like sound.

The third possibility is where the learner does not recognise any similar sound in the L1 at all so is faced with a new L2 sound. Here the establishment of a new phonetic category for the L2 sound occurs more easily, and could eventually be closer to that of the target L2 sound than in the preceding case.

The last hypothesis of the SLM, H7, states that there is eventually a correspondence between production of L2 sounds by learners and the phonetic categories which they represent in their minds, which in turn are based on their perception. In other words, it predicts that the learners in the end produce L2 sounds in the way they perceive them. In fact, this hypothesis in effect adds on a step to what each of the former ones predict. For instance, H2 stated earlier that learners may form a new category for the new L2 sound if they recognize the phonetic difference between the L1 sound and its corresponding L2 one. If we apply H7 to that, then it becomes a prediction that if learners recognize the phonetic difference between the L1 sound and its corresponding L2 one they will learn a new L2 sound as a phonetic category and that will be reflected in their production as well. In the same way, all the hypotheses 1-6 refer directly to perception and/or to category formation in the mind, but with the application of H7 become predictions also about production. The inclusion of the word 'eventually' in H7 however does warn us that there may be some time lag between perception of a difference and its reflection in production. Accordingly, during the learning process learners may experience a phase of learning where their perception is better than their production.

H7 is clearly exploited by Flege himself, and many studies in the next section, and will be used by us in this study, in that it provides a justification for examining the hypotheses by gathering relevant production data rather than by direct tests of speech perception.

2.11.1 L2 Studies conducted or interpreted through the SLM

The SLM has been supported by a number of researchers studying consonants and vowels in second language acquisition. For instance, Flege (1987b) proved that ENS L2 learners of French, who were resident in France for about 12 years, succeeded in producing the L2 French /y/, similar to French native speakers. However, those L2 learners could not pronounce the French /u/ accurately like French native speakers. Flege explained the different behaviour of the L2 learners in producing the French sounds as because the French /u/ is to some extent similar (but not identical) to the English /u/, whereas, the French /y/, which learners were able to produce, does not have any phonological equivalent in the inventory of their English L1 (at least in the varieties of English involved in this study). In the view of Flege's model (SLM), not having an equivalent for /y/ in their L1 enabled the learners to form a new category for the new sound /y/ and made them learn that sound more easily and produce it accurately (dissimilation). However, the learners perceived that the /u/ sound in French was similar to the /u/ sound that they have in their L1, and therefore were not prompted to notice the differences and put more effort into learning the similar sound: the L2 sound underwent equivalence classification or assimilation with the L1 sound. .

Consequently, Flege concludes that learners, even after living in Paris for 12 years, could not learn some of their L2 sounds accurately. Therefore, Flege affirmed that phonetic categories of L2 learners may be different from those of L2 native speakers, because of slight acoustic and featural differences between similar L2 and L1 categories of sounds. This may lead the learners to develop slightly different categories from both L1 and L2 monolinguals.

Flege and Hillenbrand (1984), in a study closer to ours, demonstrated that French L2 learners of English could not produce the voiceless aspirated /t/ with long-lag like ENS. All French voiceless stops /p t k/ are produced with short-lag VOT unlike English where voiceless stops are usually aspirated and produced with long-lag VOT (except of course the unaspirated instances following /s/ which were not included in this study). French voiceless /t/ which is unaspirated was considered as similar to the English long-lag [t^h] by the L2 learners. Flege and Hillenbrand concluded that French L2 learners who had unaspirated stops in their L1 tended to produce the long-lag voiceless stops in their L2 with short-lag VOT, using equivalence classification, and failed to create a new category. That is despite the fact that the difference in transcription would lead one to suggest that the difference might be noticeable by learners. We will look to see if this also occurs with Saudi Arab learners of English, who have an L1 where /t k/ occur not exactly with short lag, but with VOT ranging from borderline short lag to the lower area of long lag (25-60 msec, see 2.8), so still lower than the VOT of English aspirated voiceless stops.

Simon (2009) examined the production of English stop VOT by a group of Dutch learners of English. Voiced stops in Dutch are produced with pre-voicing, similar to Arabic, while voiceless stops are produced with short lag. However, as we know, the contrast is different in English, where short-lag is used for voiced stops and long lag for voiceless ones (if we again omit the unaspirated voiceless instances after /s/ etc.). Simon (2009) found that the majority of the learners' productions (93%) of English voiced stops were produced with pre-voicing, similar to the learners' L1 voiced stops.

The voiceless stops were tested in two different contexts (word-reading and continuous speech). The learners produced English voiceless stops with similar VOT ranges to ENS (long-lag) in the word-reading task, although their L1 has only short-lag voiceless stops; they produced the English voiceless stops with shorter VOT values (slightly aspirated, higher than their L1 but shorter than their L2) in spontaneous speech. Simon (2009) concludes that Dutch learners of English transferred their L1 pre-voicing to their production of English L2 voiced stops. However, they acquired the aspiration found in English voiceless stops. Simon (2009) explains that learners acquired the long-lag voiceless stops because of the salient acoustic difference between long-lag aspiration and short-lag VOT. Therefore, they were able to learn English sounds with long-lag VOT and aspiration, but not sounds with short-lag VOT, which is not prominently different from pre-voicing. It has to be noted that Simon's study was not conducted in light of the SLM model. However, its findings can be interpreted through the SLM hypotheses. With respect to the L2 voiced stops, in SLM terms the learners did not distinguish between the similar L1 voiced stops (with pre-voicing) and the L2 voiced stops (with short lag VOT), which made them create an equivalence classification and consequently that blocked the formation of a new phonetic category for the new English voiced sounds (H5). With respect to the English aspirated stops, which were close to being acquired, in SLM terms learners had perceived the difference between their L1 short lag stops and the L2 long lag ones, and were creating a new category for the L2 ones, which showed itself more clearly in the careful speech associated with word reading. This is in accordance with H3 which states that the greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned. In this case, because of the perceived phonetic difference between English (aspirated) and Dutch (unaspirated) voiceless stops, they created a new phonetic category for the new voiceless sounds (H2).

Simon's (2009) study is parallel to the current study, as Arabic voiced stops are pre-voiced, though the Saudi voiceless stops are produced with low long lag aspiration not short-lag like those of Dutch speakers. Therefore, we will be looking to see if the advanced learners in our study evidence the same performance as the Dutch ones on voiced stops, though we would not expect the same result for the voiceless aspirated ones.

2.12 Research questions and Hypotheses of the current study

Following from the account above, the research questions and hypotheses of this study could be seen as falling into two areas, motivated by the need to extend our knowledge in different ways.

First, as we pointed out in 2.6 - 2.9, there are gaps in our knowledge concerning stop VOT especially of Saudi ANS, and Saudi learners of English, in the areas of the effects of following vowels of different qualities, and the effects of context (isolated word versus sentence production). These need systematic attention. More fundamentally, even the VOT differences between aspirated and unaspirated voiceless stops have not been looked at properly for Saudi learners of English in comparison with ENS: the closest study to ours, Flege and Port (1981), discussed in 2.9 omitted that opposition so was not able to assess the separate effects on VOT of the voice opposition such as [p - b] and the aspiration opposition [p - p^h] etc.

Secondly, as we showed in 2.11, the most promising theory of the acquisition of L2 phonology these days we believe is the SLM. This study therefore wishes to further test some of the hypotheses of the SLM by gathering new stop VOT data from a dialect of Saudi Arabic not studied before, to see if the findings support SLM predictions in the way that those of Flege and Hillenbrand (1984) and Simon (2009) do in studies of learners with quite different L1s (see 2.11.1).

RQs, filling gaps in knowledge from previous studies

1. Along with the effects of POA and voice/aspiration, what effects do following vowel quality and context have on the production of stop VOT by native speakers of English and of Saudi Arabic, and by Saudi advanced adult learners of English?

2. Does POA for all those three groups, regardless of following vowel, context or voice/aspiration, always involve an increase in positive VOT / reduction of negative VOT of stops front to back, in accordance with a claimed universal trend?

3. What stops (differing in POA, voice/aspiration, following vowel and context), if any, do advanced adult Saudi learners of English produce with different VOT from Arabic native speakers? What stops do they produce with similar VOT to English native speakers?

4. Regardless of whether their actual stop VOT values are similar to those of Arabic native speakers, do adult Saudi learners of English make VOT distinctions for POA, voice/aspiration, following vowel and context that are parallel with those made by monolingual Arabic native speakers? Or, regardless of whether their stop VOT values are similar to those of English native speakers, do adult Saudi learners of English make VOT distinctions for POA, voice/aspiration, following vowel and context that are more parallel with those made by English native speakers?

Hypotheses, allowing us to test some of the postulates and hypotheses of the SLM discussed in 2.11:

RH1. Same sounds. Any English sounds that are objectively the same in VOT in English and Arabic will be perceived by advanced Saudi learners to be identical to ones in Arabic and will be placed in the same category as the L1 sounds (equivalence classification). Hence wherever in our data VOTs are in fact not significantly different between monolingual Arabic and monolingual English, we will expect that learners' VOTs will not be significantly different from those of either L1 or L2: this represents successful learning through positive transfer. (H1 H2 H5 H7)

RH2. New sounds. Since neither English aspirated [p^h] nor unaspirated [p] are found in Saudi Arabic, and there is no very similar Arabic sound which would be transcribed with the same IPA symbol, these English sounds will have been easily perceived as new, and advanced adult Saudi learners of English will have formed new categories for them and will produce them with VOT separate from [b] or [t^h], close to ENS values. (H1 H2 H3 H7)

RH3. Similar sounds, noticeable difference. English unaspirated voiceless stops [t k] do not exist in Saudi Arabic but are similar to voiceless weakly aspirated stops [t^h k^h] which do exist in Arabic. Since a difference between non-aspiration and aspiration however is relatively noticeable, advanced adult Saudi learners of English will have formed new categories for them and will produce [t k] with significantly shorter positive VOT than [t^h k^h], though perhaps not entirely similar to ENS values. (H1 H2 H3 H7)

RH4. Similar sounds, less noticeable difference: aspirated stops. Both English and Saudi Arabic have voiceless aspirated stops [t^h k^h] although Arabic ones are produced at the top of the short lag region or in the lower part of the long lag VOT, while English ones are produced in the high long lag region. Since this sort of a difference in aspiration is relatively less noticeable, advanced adult Saudi learners of English will not have formed new categories for English [t^h k^h] and will produce them with VOT much the same as ANS (equivalence classification). (H1 H2 H3 H5 H7)

RH5. Similar sounds, less noticeable difference: voiced stops. Both English and Saudi Arabic have voiced stops /b d g/ although Arabic ones are pre-voiced while English ones have VOT in the short lag region. Since this sort of a difference in voicing is relatively less noticeable, advanced adult Saudi learners of English will not have formed new categories for English /b d g/ and will produce them with VOT much the same as ANS (equivalence classification). (H1 H2 H3 H5 H7)

RH6a. Deflection on the POA scale. Saudi learners are familiar with an L1 without a voiceless labial stop. They will fill this gap in a way that locates /p/ VOT in a place where it is clearly distinct from the VOTs of their /t/ and /k/, regardless of whether that place resembles the location of ENS /p/. (Postulate 4, H6, H7).

RH6b. Deflection on the voice-aspiration scale. Saudi learners in their L1 are used to a stop VOT voice-aspiration continuum in L1 divided into only two areas/categories at each place of articulation, since they only have to distinguish [d] from [t^h] and [g] from [k^h]. On exposure to English, however, they potentially need to divide that space into up to 5 parts/categories if they differentiate between all English and Arabic sounds, e.g. for coronals: [d] pre-voiced, versus [d] short-lag, versus [t] short-lag, versus [t^h] low long-lag, versus [t^h] high long-lag. They will therefore exhibit a tendency to space out their VOT values for those sounds more evenly over the VOT continuum, maximising VOT difference between them, resulting in the VOT values for some of them being 'deflected' from NS values. (Postulate 4, H6, H7)

RH7. Effect of length of learning. Saudi learners who have had longer exposure to high quality ENS input (LOR in UK) will conform better to predictions 2 and 3 above than those with less exposure. There will be no difference on predictions 1, 4 and 5. (Postulate 1, 3).

Chapter 3: METHODOLOGY OF THE STUDY

3.1 Introduction

When gathering the data for the study, special care was taken with selecting the subjects, the materials, and the equipment, and controlling the administration of the tests and the analysis of the data, so as to ensure that the study provides accurate results without any internal or external influence which may lead to the study misrepresenting reality.

In this chapter, we describe the methodologies used for the data collection and analysis in order to answer the research questions (2.12). First, the details of the subjects who participated in the study will be presented. Second, the instruments employed to measure VOT production of the participants in the study will be described. Third, the recording procedure and the analysis of the target sounds will be described. Finally, reliability tests and the statistical means of analysing the data will be described.

3.2 Participants

Four types of participant were required for the study: English Native Speakers who spoke no Arabic (ENS), Saudi Arabic Native Speakers who spoke no English (ANS), advanced Saudi Arabic speaking learners of English (L2 learners), and the fourth group were NS judges, both Arabic and English, used to help validate the stimuli and responses of the instruments. The majority of the participants (ENS, L2 learners, and judges) were postgraduate students and employees at Essex University, UK, who were selected through the researcher's personal social network and by inviting some of them through the small ads, an electronic service provided by the university of Essex to its students and employees to share experiences. The others (ANS) were selected in Saudi

Arabia through personal contacts. Hence they are convenience samples, though we have no reason to suppose they are not broadly typical of native speakers and educated learners of the types targeted.

Following the ethical requirements of the University of Essex, written consent was obtained prior to the recording sessions from all the participants. They were informed that the study involved saying words but left unaware of the precise purpose and the nature of the current research, so that no attempts could be made to speak in a special way to 'help' the researcher. They were also promised confidentiality, in that no real names of participants would appear in the write up, and that they were entirely free to retire from the study at any time if they so wished.

Subject recruitment procedures went through many stages and took much more time than initially expected. The initial plan was to contact the potential participants via email with the researcher's contact details. Those who were interested in taking part in the study were encouraged to report back to the researcher. This worked only with English native speakers. However, the response rate from the L2 learners was less than expected, as only 3 out 31 initially agreed to participate in the study. The Arabic monolinguals were to some extent easy to find as they were close friends and members of the family back in Saudi Arabia who agreed to participate without hesitation.

A few learners refused to take part in the study because they felt they were postgraduate students and did not want to be tested and judged by another student, as this might affect their academic and social status among other students. Since the fear of their English proficiency level spreading to other students was explicitly stated by some learners, we re-emphasized to the participants that the identities of participants and their personal details would be kept confidential and that their recordings would not be used by anyone other than the researcher and only for the purpose of the research. In some cases, we had to invite some participant (from the learner group) to the researcher's house one day before the data collection session to explain more fully what the research was all about and the importance of their help in taking part. We found that the participants were more willing to take part in the study after meeting the researcher and having the purpose of the research better explained. After each recording session, we asked the participants if they could nominate other friends who might be willing to participate in the research. Although, some participants initially showed interest and agreed to take part in the study, they later missed their recording session on the day agreed between them and the researcher. As a result, the researcher needed to reschedule their recording session on another day, which caused delays to the data gathering process.

Each L2 learner was asked to fill in a form, essentially a questionnaire, which required them to provide background information such as their length of residence (LOR) in the UK, the number of hours they listen to and speak English daily and to report whether they have any speaking or hearing difficulties. No informants reported having any hearing or speaking difficulties. Therefore, participants in all of the groups have had no personal or family history of disorders in the areas of language development, speaking, hearing or reading.

The total number of the participants in this study was 101. Two groups (comprising of English and Arabic native speakers) were control groups (30 participants each) and one group of 31 learners formed the target group, together with a fourth group of 10 judges who assisted with the instruments. The control groups were employed in order to measure the native production of English and Arabic stops and were for the purpose of comparison of the VOT ranges of native Arabic and English stops with those of the learners, so as to measure how nativelike or L1-like learner VOT was.

Although the two control groups were not the main target in this study, the VOT findings for those groups are also expected to contribute to existing research on VOT and to provide a full up to date account of VOT in Arabic. Indeed, this is, to the best of the researcher's knowledge, the first study done on VOT of Northern Saudi Arabic (NSA). VOT of ENS were obtained only for the sake of comparison with the L2 learners. In the following sections, a description of each group of participants is provided.

3.2.1 Group 1 (English Native Speakers, ENS)

The first group consisted of 30 English native speakers with clear and understandable English accents. They were all selected from the town of Colchester in Essex, UK to serve as a control group (English monolinguals), since the use of participants from different areas in the UK could have led to the inclusion of various accents in the study, which was not desirable. VOTs of stops vary in various English accents as Docherty (1992) has pointed out. Therefore, we had decided to select participants from one town which is Colchester and they all spoke a southern variety of English.

Some of the participants in this group were undergraduate and postgraduate students at the University of Essex, and some were employees at the same university. Also, a few of them were neighbours and friends of the researcher. Their ages ranged from 18 to 67 years and the mean of their ages was 33.9 years (SD=11.73). Some of them were paid for their time for participating in the study, while others thankfully participated for free. Participants of this group can also be considered monolinguals, since although seven of them reported that they knew some basic French from school, they stated that they could not communicate or read in French.

3.2.2 Group 2 (Arabic native speakers, ANS)

The second group comprised 30 native speakers of Arabic. They were from the northern parts of Saudi Arabia, in the region of Arar, and therefore they were from the same area as the L2 learners of English. Thus, participants in this group and the L2 learners group spoke the same dialect of Arabic. The researcher is also a native speaker of the northern Saudi dialect. All the participants of this group were born and raised in Saudi Arabia and they were all males. Their ages ranged from 18 to 60 years and the mean of their age was 36.8 years (SD = 11.94). This group served as the second control group in the study.

Most of the participants in this group knew no other language apart from their native language, although five of them were at a very low level of English proficiency. This knowledge of English is what they had learned in the intermediate and high school, which does not equip them to engage in any sort of real conversation.

The researcher collected data from 30 informants at first. However, some participants were excluded either because later their utterances were found to be unclear and difficult to segment or showed pauses and noises when producing the target words. Therefore, data was collected again from other participants to meet the minimum number set in this study in each group. Thus, in the end 30 informants were recorded one at a time over a period of about 90 days.

3.2.3 Group 3 (English second language learners)

The third group in the study includes 31 adult L2 learners of English living in the UK at the time of the study. All of them were from the same area of Saudi Arabia from which the Arabic monolinguals were selected and all spoke the northern dialect of Saudi Arabic as their L1.

They were all male students at the University of Essex in Colchester, UK (50 miles from London). The learners were selected from this area for three reasons. Firstly, it is the place where the researcher was studying at the time of the study. Secondly, it is a regional town unlike big cities like London which would have a multi-linguistic and multi-dialectal effect on the learners; in smaller places like Colchester, the dialectal and linguistic variation in the local community is relatively controlled. Thirdly, they were selected from the same area of the United Kingdom with an aim that the VOT of the L2 learners might be judged against the VOT ranges of English stops of the local English accent to which they have been exposed for most of their stay in the UK. We have to admit, however, that the varieties of English which they would be exposed to on campus was much more varied, given that the staff and English native speaker students come from many parts of the UK, the Americas, Australasia and other former British colonies. This factor was, however, impossible to control. The students were undergraduates, postgraduates, and students learning English in the International Academy (IA) at the university of Essex. The International Academy is a language institution where international students study English in their first year before they join any programme at the university. The undergraduate and postgraduate students were from different departments (e.g. linguistics, psychology, computer science, law, biology…etc). However, the majority of this group was from the linguistics department and were students doing their masters and PhDs at that time. Since the majority of the learners were studying at least at undergraduate level, this implies that they would, in order to have been accepted for admission to the University of Essex, have attained an English proficiency level equivalent to at least IELTS 6.0, which falls between B2 (upper intermediate) and C1 (advanced) in the Common European Framework of Reference for examinations. Hence they can be assumed to be at very high intermediate or advanced level, although, no proficiency tests were applied to identify their actual level specifically for our study. Their ages ranged between 22 and 45 years and the mean of their ages was 31.8 years (SD=4.64) (see Table 3.1).

Table 3.1: Age profile of all participants.

Group	N	Minimum	Maximum	Mean	Std. Deviation
English (ENS)	30	18.00	67.00	33.90	11.73
Arabic (ANS)	30	18.00	60.00	36.80	11.94
L2 Learners	31	22	45	31.77	4.64

Their length of residence in the UK (LOR) was between 1 and 7 years and their average length of residence was 3.48 years (see Figure 3.1). LOR provides a second indication of proficiency since various studies propose specific lengths of time residing in the target language environment which can be used as indicators of being advanced learners. For example, Best & Tyler (2006) suggest a period between 6 and 12 months, while others, e.g. Flege, Schirru, & MacKay (2003), suggest a period as long as 42 years. However, most studies propose a period between 1 and 5 years (i.e. Flege (1987b); Flege (1993); Guion, Flege, Akahane-Yamada, & Pruitt (2000)) which is the period all of the learners in this study have been resident in the UK.

All the learners in this group reported that they had started learning English in a classroom setting from the age of 11. Therefore, they had started listening to English as spoken by non-native speakers at the average age of 11 years, and by English native speakers after their arrival in the UK. Their age on arriving in the UK varied between 22 and 45 years (mean=31.8 years). With respect to the 'critical period', which is widely believed to have a particular effect on the acquisition of phonology (e.g. Oyama, 1976), we would therefore say that although these learners began to learn

English just before the end of the critical period, they were not exposed to ENS phonological input until well after the critical period had ended. Hence it can be argued that they are unlikely for that reason to achieve nativelike pronunciation of English however long they reside in the UK, or however nativelike other aspects of their language such as vocabulary knowledge become. We should note, however, that, as we saw in chapter 2, Flege and the SLM do not subscribe to the view that learners who start to learn as adolescents or adults are incapable of achieving the same proficiency as learners who started earlier.

27 of the participants had either obtained masters and/or PhDs from the UK or were studying their degrees at that time, and 4 of them were studying English courses in the International Academy prior to starting their postgraduate degrees. Table 3.2 shows, length of residence of the participants and their speaking and listening to English per day according to their own reports.

A correlation analysis of the learners' length of residence (LOR) with their reported daily speaking rate (SP) and their listening (LIS) revealed that the latter were strongly correlated with each other ($r=.499$, $p=.004$): i.e. those who claimed to spend more time per day listening to English also claimed to spend more time speaking it. This would be expected where language use is typically conversational. For later analyses, we therefore used participants' mean SP/LIS scores as a combined indication of amount of current language use. LOR, however, correlated poorly with both speaking and listening, so we retained that as a separate subject variable (LOR with SP $r=.283$, $p=.122$; LOR with LIS $r=.159$, $p=.392$).

One of our research questions concerns whether those learners who have stayed longer in the UK have better performance in VOT production of the English stops compared to the learners who stayed for a shorter period. Therefore, we examined the learners' LOR (Figure 3.1) in order to consider if it made sense to divide the group into, for example, a high and low group on that basis. Clearly our sample in this study does not naturally fall into two groups, so for most purposes we used each individual participant's exact number of years of residence in calculations. However, in order to test thoroughly whether LOR had an effect on learning, we also in places compared the extremes, i.e. the 8 participants with LOR<3 with the 6 with LOR >4.

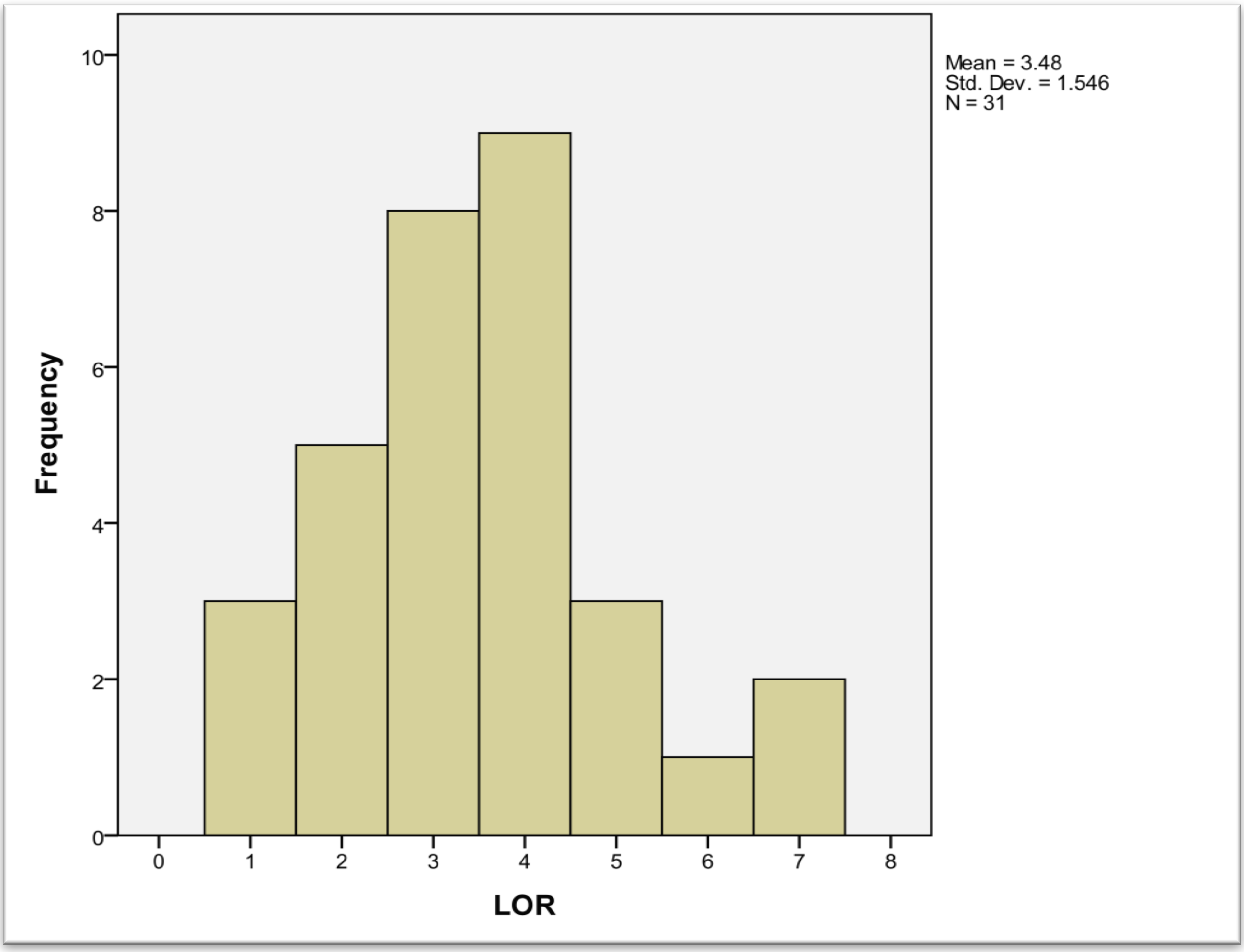


Figure 3.1: Histogram of numbers of learners with different LOR in years.

Table 3.2: Descriptive Statistics of the L2 learners.

Details of Participants	N	Minimum	Maximum	Mean	Std. Deviation
Age	31	22.00	45.00	31.77	4.64
Number of years of residence in the UK (LOR)	31	1.00	7.00	3.48	1.55
Speaking English hours/day (SP)	31	1.00	6.00	2.61	1.20
Listening to English hours/day (LIS)	31	1.00	5.00	2.84	1.00
SP/LIS average		1.00	5.5	2.73	0.96

All the participants of this group were living in Essex at the time of the study. According to their own statements, they listen to English for an average of 2.84 hours daily (minimum 1 & maximum 5 hours), speak English for an average of 2.61 hours daily (minimum 1 & maximum 6 hours) and their average time listening and speaking English daily was 2.73 hours.

3.2.4 Group 4 (Native speaker judges)

Finally it should be mentioned that a fourth group of 10 people participated in this study. These were judges who helped in the process of evaluating and validating stop productions of the participants.

Five English RP speakers were used as judges to validate the stop productions of the ENS participants (group 1). Thus, English judges listened to ENS producing the English stops to check whether they were valid and clear enough to have their VOTs measured and be judged against the learners' VOTs.

The other group of judges was five native speakers of Arabic who were used to evaluate the stop production by the ANS (group 2). They were aged 28, 33, 22, 25, 21 and were from the same region as the ANS and learners of this study.

3.3 Instruments

Two data gathering instruments were used to elicit the data from the participants as follows (See Appendix A, B and C for the full wordings of the instruments used in the study):

1. An open-ended questionnaire was given to the learner participants to elicit the required background information.

2. A production test was designed in which two lists of stimuli words were used: an English list to test the learners' and the ENS stop VOT production and an Arabic list to test ANS stop VOT production.

3.3.1 Questionnaire

Before recording the participants, the learners group only were asked to fill in a short questionnaire in English, which comprised background questions about their age, the length of their period of residence in an English speaking country, their level of study, and the number of hours they listened to and spoke English daily, etc. (see appendix A)

3.3.2 Production test

3.3.2.1 Stimuli and administration

A production test was used to measure the participants' word initial stop consonant VOT production in English and Arabic. For this purpose, two lists of monosyllabic stimulus words were used for participants to read aloud, thus eliciting the production of the target sounds: the first was for the English native speakers and the L2 learners to measure the VOT of English stops [p^h t^h k^h p t k b d g], and the second was for the

Arabic monolinguals to measure the VOT of Arabic stops [t^h k^h b d g] (see appendix B and C).

The reason for not eliciting unaspirated stops in the Arabic stimuli was the difficulty of finding unaspirated stops in our dialect of Arabic where words with plosives in initial position following /s/ like in English *speak, steal ski*…etc could not be found. Aspirated voiceless stops do exist, however, apart from the voiceless bilabial [p^h] in our dialect of Arabic (chapter 2, sections 2.2 and 2.3). Voiced plosives also exist, and indeed unlike in classical Arabic, the voiced dorsal [g] is found in our variant of Saudi Arabic. Therefore, only aspirated voiceless and voiced stops were elicited by the Arabic stimuli.

Additionally, following a number of researchers (Flege & Port, 1981; Khattab, 2002) who studied the phonetic realization of the voicing contrast in word initial stops, the same target words were also placed in a carrier sentence. The purpose of this was to measure the VOT produced by participants for stops in words read aloud in isolation and in the same words read aloud in a sentence medial context (RQ 1). The Arabic native speakers also read Arabic words with stops in initial position, where corresponding stops exist in Arabic, and the same words were also produced in a carrier sentence. The sentence used was (انا اقول ..... واروح للبيت) "I say….. and I go home". The sentence was produced in their colloquial Saudi dialect rather than the standard or classical Arabic used in most of the studies that have been done on Arabic VOT.

As we described earlier in the literature review (chapter 2, section 2.5.2), Klatt (1975) discovered that differences in VOT are further linked to the environment of the subsequent vowel. Therefore, one of the major concerns of the study is to examine the influence of the vowel on the preceding stop (RQ 1) and each of the stops was tested in three vowel contexts to determine whether vowel has any effect on VOT of preceding stops.

The lists of words used therefore contained all the stops followed by each of the three quantum vowels (Roca & Johnson, 1999), i.e. front, back and low /i:/, /ɑ:/, and /u:/, which are found in both languages, English and Arabic. These vowels were selected because of their quantum nature which makes them especially suitable to identify any effect of vowel on VOT. That is to say that they represent extreme possible vowel positions in the oral cavity. The three vowels were tested after each word initial stop in the English and Arabic words, in their long form [i:, u:, a:] in Arabic and in their tense form in English.

The target sounds in English are the voiced stops [b d g], the aspirated voiceless stops [p^h t^h k^h] and the unaspirated stops [p t k], each followed by three vowels, and the target words used in the English stimuli were: *peak, speech, teeth, steel, key, ski, beak, deal, geese, park, spark, tart, star, card, scarf, bark, dark, guard, pool, spoon, tool, stool, cool, school, boot, do* and *goose.* For Arabic the target sounds were [t^h k^h] and [b d g] and the target words were: ti:n 'fig', tɑ:b 'repent', tu:t 'berry', di:k 'roster', dɑ:r 'home', du:r 'make a circle', bi:r 'water well', bɑ:b 'door', bu:k 'wallet', ki:r 'fire', kɑ:f 'the letter K', ku:b 'cup', gi:r 'gear', gɑ:l 'said', gu:l 'say'.

All words were read by participants first in isolation, then in sentences. The order of words was randomised, with some distractor words mixed in, but participants were instructed to repeat each of the stimulus words three times in the word in isolation condition and three times in the sentence condition. The total number of target word tokens of English recorded for analysis in isolation was 81 (9 consonants *3 vowels* 3 repetitions). The same number of tokens was recorded in the sentence context. The word in isolation list (in English) in fact consisted of a total of 93 meaningful words, however, since 12 other words (e.g *fun*, *go*, *now* etc.), each occurring only once, were included as distracters (see appendix B) for the full list of stimuli). All words were checked to be known and frequently used by the participants before the recording session started.

The subjects were instructed to read the words and the sentences aloud in their normal manner, leaving a short pause between each of their productions and the next.

The Arabic monolinguals also read Arabic words containing all the voiced stops and the voiceless stops, yielding 45 target word tokens (5 consonants *3 vowels* 3 repetitions). The same procedure was followed as for English with respect to randomisation, ordering etc. The complete lists of stimuli are given in the Appendices (B and C) in the exact order as administered.

The total number of recorded sounds from the participants that needed to be analysed from words in isolation was as follows: 2430 tokens from 30 English native speakers (30*81), 2511 tokens from 31 L2 learners (31*81) and 1350 tokens from 30 Arabic native speakers (30*45), which yielded a total number of 6291 tokens. The same number of tokens was obtained as well for the production of stops in a carrier sentence totalling 12582 tokens in the whole study for both English and Arabic.

It should be noted that only stops in initial position are investigated in this study. This is for a number of reasons. Firstly, it would be beyond the study's capability to analyse VOT in more than one position (initial, medial and final), given that this study generated more than 12500 tokens in initial position alone. Secondly, the study focused on the initial position so as to compare its findings with other studies which looked at stops initially. Thirdly, stop consonants in coda position normally tend to be reduced (deleted or unreleased), which could affect the production of the learners. Moreover, onsets are known to precede codas in terms of language acquisition (Archibald, 1998).

Consequently, onsets provide a better environment to examine the production of the second language sounds.

3.3.2.2 Recording

All the recordings were made individually by the researcher in a quiet linguistics lab in the Language & Linguistics department at the University of Essex for the learners and the ENS, and in a quiet room for ANS. All the recordings were made using a high-quality microphone and a professional 2-channel mobile digital recorder (M-Audio Micro Track II). The audio recording device was set to 44.1 KHz sample rate and 16 bit depth. The sound files were transferred to a Mac OS X computer in WAV format for acoustic analysis. Each participant was recorded in a single session, which took approximately between 45 minutes and 1 hour, yielding one sound file per participant.

3.3.2.3 Verification of the participants' productions

For verification of the clarity and authenticity of the native speakers' production, ten sound files were randomly selected from the recordings in each monolingual group (Arabic and English). Those recordings were played to five other native speakers of Arabic/English who were asked to identify the words in the recordings. The purpose of this procedure was just to check whether the participants did speak clearly, and in the required variety of Arabic/English. The native speakers of Arabic identified all the words of their peer Arabic speakers correctly and stated that this was what they would normally hear in the local accent of Arar (the place in Saudi Arabia where all the learners and the monolinguals were from) and all sounds seemed to be clear and natural. The same procedure was followed for the English native speakers ENS, where five native speakers listened to ten randomly selected recordings, and all English words were identified correctly.

3.4 Overall Data Collection Procedure

The data were collected individually in two stages between 2013 and 2015.The first stage was in Colchester, Essex university, UK where a sufficient number of L2 learners and English native speakers were to be found. The second stage was in Arar, Saudi Arabia, the home town of the learners and the Arabic monolinguals, which is also the home town of the researcher. In each location, ethical consent was first obtained from participants, followed by completion of the questionnaire (3.3.1).

The production data was collected in a word-reading task as described in 3.3.2. The lists of stimuli were handed to the participants and they were asked to read those words in normal, natural speech (for the ANS, in their local Saudi accent). All the instructions were given in English to the native speakers of English, but to the L2 learners and Arabic monolinguals in their mother tongue, which is also the mother tongue of the researcher. All the participants were given some time to read the list of stimulus words and allowed to ask questions or practise saying any unfamiliar words, but they were not told the precise aim of the study. Recording of the subjects started when they indicated they were ready and took approximately between 45 minutes to 1 hour. As in the Flege and Port (1981) study, the ANS participants were asked to produce the words in their colloquial style i.e. using [g] in the way they speak normally, not in the style of modern standard or classical Arabic. Therefore, the Arabic monolinguals were asked to practise reading the words in their colloquial style to make sure that they did not read in the accent used for the classical form of Arabic.

It should be said at this point that, in addition to our study of stop VOT, which of course relates to the production/pronunciation of sounds, we initially also had aims to investigate some questions concerning perception of stops, and we did gather some very limited data on this. In the end however we have decided to omit this for a number of reasons. First, there were problems of time in dealing with this data properly and of space in the thesis in dealing with the perception literature and to report the findings. More crucially, however, we feel that the questions we could ask and answer from our perception data in the end do not contribute to the themes of the study as they have finally crystallized from the far more substantial VOT study, and are not substantial enough in themselves to merit their inclusion as an independent study. Nevertheless, we present a short summary of this in appendix I.

3.5 Data Analysis

3.5.1 VOT data analysis

In order to measure VOTs from the utterances produced by the participants, acoustic analysis was first undertaken using Praat software by Boersma & Weenink (2012) and ProsodyPro script developed by Xu (2005–2012). Participants' target words and sentences were segmented, extracted and saved as .wav files, to extract the duration of voice onset time using the aforementioned script after which they were hand checked for errors. Segmentation labels were also added to the sound files to mark word boundaries in order to differentiate easily between words and avoid confusion.

Because of the usefulness of this script in this field, an account of how it was used in detail is provided, in case a researcher in the same field may need to use it in their research. Therefore, a brief account of the script will be offered first followed by how the script is to be used.

3.5.1.1 Praat software and ProsodyPro script

Paul Boersma and David Weenink of the Institute of Phonetic Sciences, University of Amsterdam, developed the Praat program. 'Praat' is free scientific software with which sounds can be analysed, synthesized, and manipulated, and it also creates high-quality graphics for sounds to be used in research. Therefore, the VOT of the stops in this study was measured by using Praat, and some pictures of the waveforms and spectrograms of certain stops were taken from the program. Along with Praat software, ProsodyPro script developed by Xu (2005–2012) was also used to obtain the exact durations of VOTs of the target stops produced by the participants.

ProsodyPro was developed by Yi Xu who is a professor at University College London in the department of Phonetics and Linguistics. The script was developed as a useful tool initially for his own research. It was later made public in 2005 to be used by any researcher who works with speech data. From that time the script has been used in a growing number of research studies (Choudhury and Kaiser, 2012; Liu, 2010; Wu and Xu, 2010; Wang and Xu, 2011; Ambrazaitis and Frid, 2012).

The script permits researchers to thoroughly process huge amounts of speech data with great accuracy. It has greatly reduced the amount of human work by automating tasks that do not involve any human judgment, such as opening and locating sound files, taking measurements of VOT length, and saving raw results in formats ready for additional statistical and graphical analysis. However, it also allows human involvement in processes that are prone to error in automatic algorithms such as pitch detection and segmentation (Xu, 2013).

At the initial stage of this research, VOT durations were extracted by Praat by placing the computer's mouse cursor on the waveform of the sound in Praat and selecting the VOT duration manually, then typing those VOT values into Excel sheets manually, but thanks to Yi Xu who developed ProsodyPro, we later saved most of the time and work in marking the VOT durations by extracting them automatically as shown in Appendix J. Without this, analysing the productions of more than 12,000 VOT tokens, would be beyond the capability of this research.

3.5.1.2 Scoring of production data

For most purposes, we worked with the VOTs calculated as described above, and the means of the three VOTs obtained for each person for each sound in each context with each vowel. Our data, however, yielded many occasions when participants produced stops with prevoicing, i.e. negative VOT, especially, but not exclusively, from Arabic speakers on voiced stops. Therefore, in some analyses described in chapter 4, we took into account only positive or only negative VOTs. That is to say that, we counted numbers of VOTs produced by participants that were non-negative separately from those that were negative and/or calculated means just based on non-negative or just based on negative VOT productions of each person.

3.5.2 Reliability

Cronbach's alpha reliability was calculated for the production of the repetitions of the voiceless and the voiced target sounds of this study to verify that each participant's three responses in each condition were internally consistent. The test showed that the reliability of the repetitions was very acceptable, and for many of the sounds was above .9 (Table 3.3), which is considered exceptional in research.

Table 3.3: Cronbach's alpha for VOTs of production responses (three per condition).

Stop sounds		Voiceless stops						Voiced stops		
Context	Vowel	/p/	/sp/	/t/	/st/	/k/	/sk/	/b/	/d/	/g/
Word in isolation	/i:/	.905	.766	.797	.978	.836	.974	.902	.947	.927
	/ɑ:/	.905	.796	.926	.975	.885	.979	.918	.925	.917
	/u:/	.938	.908	.866	.981	.806	.980	.936	.942	.897
Word in sentence	/i:/	.927	.917	.902	.975	.793	.977	.902	.927	.928
	/ɑ:/	.935	.898	.899	.975	.836	.976	.922	.898	.896
	/u:/	.969	.892	.929	.981	.753	.976	.879	.943	.869

According to Scholfield (1995) and Larson-Hall (2010), reliability is considered excellent if the Cronbach's alpha value is 0.70 or above. As can be seen from Table 3.3, the reliability of the production of target sounds was between .75 and .98 for all the words, so in terms of value of Cronbach's alpha all were above 0.70.

The reliability of the sounds produced in the following words: *peak, park, pool, spoon, tart, steel, star, stool, ski, scarf, school, beak, bar, boot, deal, dark, do, geese, guard, goose* was above 90% (ranged between 90% and 98%) in both contexts produced (word in isolation and in a carrier sentence). Reliability was 75% and 76% in only two words of the whole data acquired: *cool* in a carrier sentence and *speak* as isolated word respectively. The highest reliability recorded was in *stool* and *school* with 98%.

The high reliability of the data has, therefore, justified us using the mean VOT of each set of responses for each condition from each participant for onward calculations.

3.5.3 Statistical tests used

A check was carried out on the distribution of the production data to ensure that it was suitable for ANOVA tests to be used to analyse it. With respect to the requirement that the distributions of responses be normally distributed, the Kolmogorov-Smirnov one sample test applied to each condition and group (138 analyses) shows that the vast majority of the data (94%) meets this requirement. However, with respect to the homogeneity of the variance of one group with that of another, which is important for ANOVA where groups are to be compared, as we will sometimes wish to do, the Levene test showed that only 17% of the data met this requirement (54 analyses). Nevertheless, ANOVA is claimed to be robust enough to give valid answers even if some of its prerequisites are violated, so we propose to use ANOVA, but with the protection that when we compare groups we will use post hoc tests for comparison of each pair of groups which are designed to deal with groups whose variances are not the same (the Games-Howell test).

3.6 Chapter Conclusion

This chapter has described in detail the groups of English, Arabic and learner participants, the construction and use of the production instruments and the analysis of the data employed during the study. In the following chapter, the results which emerge from data obtained in this study will be presented and discussed.

Chapter 4:
RESULTS FOR THE MONOLINGUALS

4.1 Introduction

In this chapter, we present the descriptive and inferential VOT results for each monolingual group (ENS and ANS) separately. For each monolingual group, we first cover production of voiceless stops, and then production of voiced stops. Both groups' results will be discussed with reference to the possible effects mentioned in the literature on VOT: effect of voice (and in English of aspiration), effect of POA, effect of following vowel, and effect of context (word in isolation versus word in sentence). The VOT results of the L2 learners and comparisons of those with results for ANS and ENS will be dealt with in the next chapter.

4.2 VOT of English stops

In this section, results of the ENS will be described in general, then we move to present each stop category, i.e. voiceless stops then voiced ones, under each of the possible effects on VOT mentioned above.

Results for ENS in this study followed the expected pattern of VOT of voiceless stops in English and several other languages. A simple picture may be obtained from Table 4.1, which provides all ENS details omitting the effect of vowel and context.

Table 4.1: VOT values of initial plosives produced by ENS (combining figures for vowels and contexts)

Stop	POA	VOT			
		Maximum	Minimum	Mean	Standard Deviation
Voiceless aspirated	Labial	111.87	36.61	67.85	15.00
	Coronal	164.76	48.82	84.20	16.96
	Dorsal	135.77	44.24	88.03	18.15
Voiceless unaspirated	Labial	40.68	1.78	14.08	5.23
	Coronal	43.73	9.34	21.27	5.99
	Dorsal	45.26	12.46	26.94	6.33
Voiced	Labial	97.13	-77.29	12.86	12.30
	Coronal	125.60	-99.16	18.95	18.50
	Dorsal	95.09	-109.84	26.30	13.93

The English speakers produced long-lag VOT for aspirated voiceless stops and short-lag for unaspirated stops. As can be seen from Table 4.1, the mean VOT values for the aspirated labial [p^h] was 68 msec (range=37 msec to 112 msec, sd=15), while the unaspirated labial [p] was produced with short-lag with a mean VOT of 14 msec (range= 1.78 msec to 41 msec, sd=5). The coronal [t^h] was produced with longer VOT values than [p^h], with mean 84 msec (range=49 msec to 165 msec, sd=17). The unaspirated [t] was produced in the short-lag region with mean VOT 21 msec (range=9 msec to 44 msec, sd=6). The longest VOT values produced by ENS were in the aspirated dorsal [k^h], as it was produced with a mean VOT of 88 msec (range=44 msec to 136 msec, sd=18), while [k] was in the short-lag region and ranged between 12 msec to 45 msec with a mean value of 27 msec. As can be seen from Table 4.1, the ENS also produced voiced stops with mean VOT in the expected short lag region. The mean VOT value of the English labial /b/ was 13 msec (range=-77 msec to 97 msec, sd=12). The voiced coronal /d/ was produced with a mean VOT of 19 msec (range=-99 msec to 126 msec, sd=19). The dorsal sound /g/ was produced by ENS's with longer mean VOT value than the other stops, as expected: 26 msec (range=-110 msec to 95 msec, sd=14).

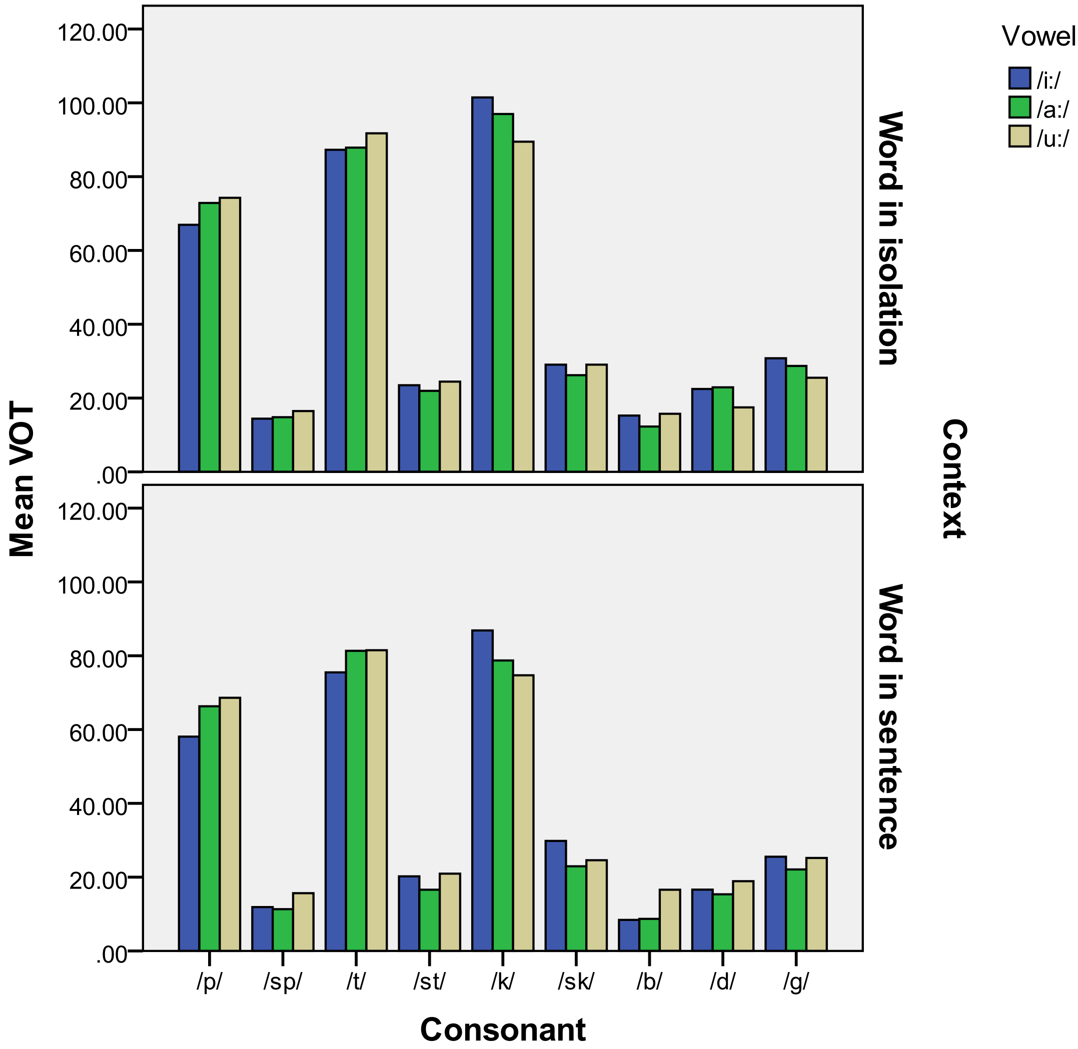


Figure 4.1: Production by ENS of all voiceless and voiced stops followed by three vowels in two contexts (words/sentences)

An initial impression of the ENS findings can also be gained from Figure 4.1, which shows visually the effects of all factors considered in the study. Just judging by eye, ENS are making a huge distinction in VOT between voiceless aspirated plosives, and the unaspirated and voiced ones, since the former have VOT means at least 40 msec longer. There are clear signs of distinctions based on place of articulation, with increasing VOT front to back. However, it is unclear what systematic impacts aspiration, vowel quality and whether the word is spoken alone or in a sentence have. Hence we proceeded to test the key comparisons.

An overall ANOVA was performed comparing effects on ENS VOT of: contexts of word use (2), places of articulation (3), voicing and aspiration contrasts (3), and qualities of following vowels (3), together with combined effects of all those (For the full table of results of this analysis see Appendix D and E).

In summary, this showed first a very clear overall main effect on VOT of place (F=153.41, p<.001), and in follow-up paired comparisons between places (performed with Bonferroni adjustment), all the pairs were significantly different (p<.001). That is to say that, considering all the data regardless of context, voice/asp differences, and following vowels, the VOT of labials differed from that of coronals (which were on average 9.7 msec longer), coronals differed from dorsals (which were on average 5.7 msec longer) and bilabials from dorsals (by 15.4 msec).

There was also a huge main effect of voicing/aspiration (F=1116.73, p<.001), and a smaller but significant one of context (F=29.00, p<.001), where VOTs in isolated words were on average 5.6 msec longer than those in words in sentences. There was no overall effect of following vowel quality (F=0.297, p=.625). There were also clear interactive effects of place by vowel (F=16.26, p<.001), context by voice/asp (F=8.58, p=.004), context by place by voice/asp (F=6.19, p<.001), and place by vowel by voice/asp (F=4.79, p=.003).

Since voice/aspiration is present in nearly all the interaction effects, and has by far the strongest overall effect, this suggests that it would be valuable to do follow-up analyses separately for different voice/aspiration options. We therefore first performed post hoc paired comparisons between the three voice/aspiration options (with Bonferroni adjustment), not differentiating on the basis of POA, context or vowel. This showed that mean VOT of voiceless aspirated plosives greatly exceeded that of voiceless unaspirates by a mean of 58.7 msec ($p<.001$) and exceeded that of voiced plosives similarly by 60.0 msec ($p<.001$). However, there was no significant difference between the VOT of unaspirated voiceless plosives and voiced ones, where the former on average were only 1.3 msec higher than the latter ($p=.499$).

This suggests that, from the statistical point of view, in further analyses of the VOT data the voiced stops and unaspirated voiceless stops could be treated as the same, in comparison with the aspirated voiceless stops. We have, however, an a priori aim to compare voiced with voiceless, since, regardless of VOT, this opposition is considered phonological in English. Also, comparison of findings with other groups will be easier if this division is observed. Hence we will nevertheless pursue the policy of examining each phonemic stop category separately, the voiceless stops (aspirated and unaspirated), and the voiced stops.

4.2.1 Voiceless stops of ENS

ENS voiceless stop results will now be described with reference to the main variables affecting VOT which we included in the study: effect of aspiration, effect of POA, effect of following vowel, and effect of context. Table 4.2 and other similar tables later all follow the usual convention of showing inferential statistics for main effects first, then two way interaction effects, then three way effects, and so on. The results show that there are significant overall differences in VOT among the voiceless stops dependant on three of the variables we included - POA, aspiration, and context (word or sentence), but not following vowel. There are also several significant interaction effects. These results will be taken up in subsections devoted to each, in descending order of size of the main effect.

Table 4.2: Overall ANOVA for ENS production data: aspirated and unaspirated voiceless stops

Effect	F	p	Partial eta squared
Aspiration	2033.32	<.001	.986
Place	252.75	<.001	.897
Context	18.22	<.001	.386
Vowel	1.42	.249	.047
Place by Vowel	25.71	<.001	.470
Place by Aspiration	23.40	<.001	.447
Aspiration by Context	8.51	.007	.227
Vowel by Aspiration	4.42	.016	.132
Place by Context	4.41	.016	.132
Vowel by Context	0.095	.909	.003
Vowel by Place by Aspiration	11.34	<.001	.281
Context by Place by Aspiration	5.22	.008	.153
Context by Vowel by Place	1.38	.245	.045
Context by Vowel by Aspiration	0.821	.445	.028
Context by Vowel by Place by Aspiration	0.777	.542	.026

4.2.1.1 Effect of aspiration on VOT of ENS voiceless stops

As we would expect from the overall result reported at the start of 4.2, the difference between aspirated stops, produced in the long lag area (mean 80 msec), and unaspirated stops, produced in the short lag area (mean 21 msec), was highly significant and had the largest effect size (.986). The simple overall difference was 59 msec. However, the effect of aspiration on VOT differed somewhat in combination with all three of the other variables, POA, vowel and context (see the significant interactions in Table 4.2), so will be considered further with them below. This is in line with other studies (i.e Lisker & Abramson, 1967; Klatt, 1975; Docherty, 1992; Khattab, 2002 .etc) which find that aspirated and unaspirated stops in English are significantly different from each other and produced with two different lags (short-lag for unaspirated and long-lag for the aspirated ones).

4.2.1.2 Effect of POA on VOT of ENS voiceless stops

With an effect size of .897, almost as great as that of aspiration, POA also has a large overall effect on VOT in voiceless stops as well as effects in interaction with vowel and aspiration considered in a later section.

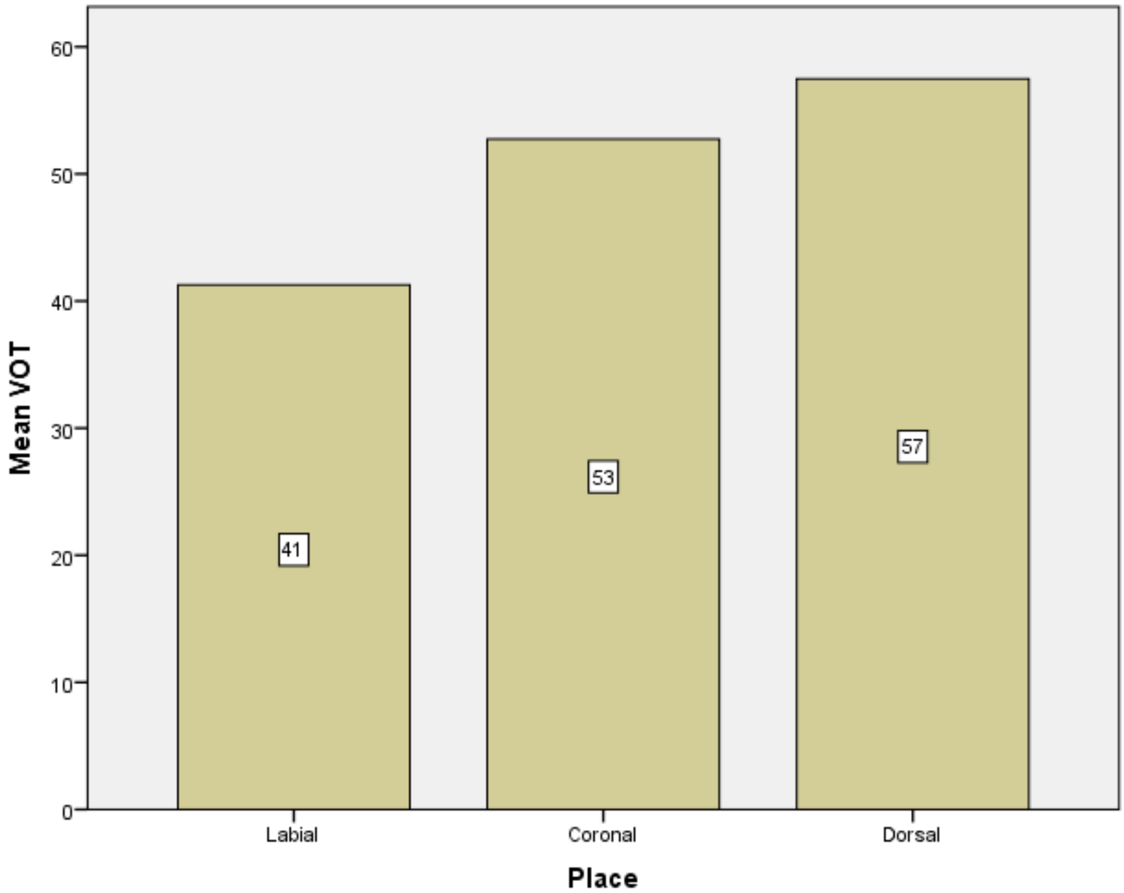


Figure 4.2: Production by ENS of word initial voiceless plosives, by place

The overall effect of place, seen in Figure 4.2, regardless of aspiration, vowel and context, with VOT increasing from front to back of the mouth, is again well-documented and in accordance with previous research as shown in section 2.9 in the literature review. This confirms the effect of place of articulation on VOT found by Cho & Ladefoged (1999) and several other researchers.

Follow-up tests showed that VOT differed significantly between all the pairs of places of articulation (p<.001). Figure 4.2 shows the overall effect of places of articulation in English voiceless stops.

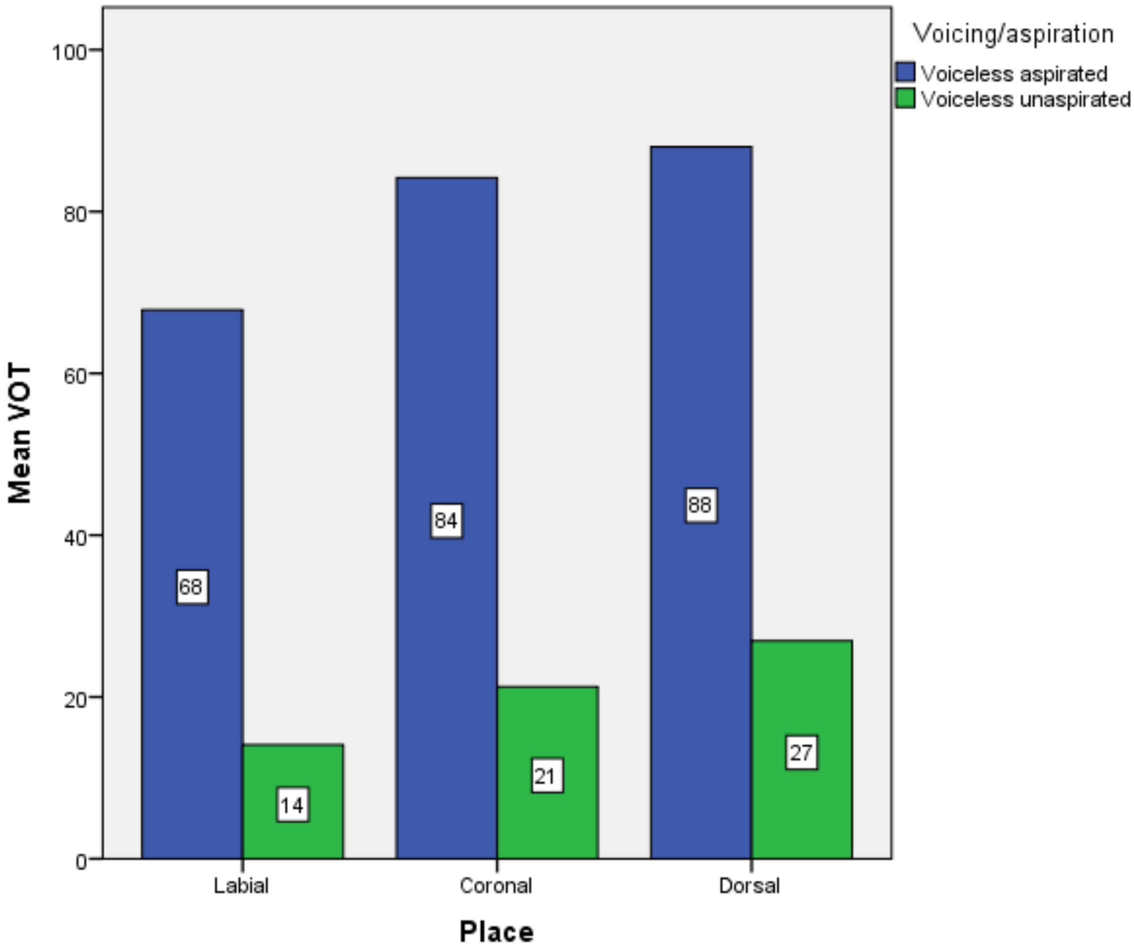


Figure 4.3: Production by ENSs of voiceless plosives, by place and aspiration

The place effect however significantly interacts with aspiration in affecting VOT (Table 4.2: p<.001). As Figure 4.3 shows, this is due to the slightly different rising pattern of VOT from front to back in the aspirated and unaspirated stops. In the unaspirated stops VOT increases steadily by around the same amount (6 or 7 msec) between each POA from front to back. In the aspirated stops, however, the increase between [p^h] and [t^h] is four times that between [t^h] and [k^h] (16 msec compared with 4 msec). This difference between [p^h] and [t^h] was also observed in Lisker and Abramson (1964), Klatt (1975) and Docherty (1992). The difference between [p^h] and [t^h] was 20 msec in Docherty's findings, 18 msec in Klatt's findings and 12 msec in Lisker and Abramson's findings while differences between [t^h] and [k^h] were smaller, within 10 msec in all of these studies, excluding Klatt who strangely found [k^h] was produced with lower VOT values (59 msec) than [t^h] (65 msec).

4.2.1.3 Effect of context on VOT of ENS voiceless stops

As Table 4.3 shows, results for the English native speakers in this study followed the expected pattern of the English VOT of voiceless stops spoken in words in isolation found in other studies (e.g Lisker & Abramson (1967), Klatt (1975), Docherty (1992), Scobbie (2002)). The English native speakers produced predominantly long-lag VOT for aspirated voiceless stops and short-lag for unaspirated stops, as we have seen earlier. More specifically, our finding is also in line with Docherty's and Klatt's findings where they tested ENS in two different contexts and found that VOTs in words in isolation (Table 4.3) were higher than those in words produced in a carrier phrase (Table 4.4). The overall effect for context differentiated by aspiration may be seen in Figure 4.4, and differentiated by POA in Figure 4.5.

In words in isolation, the mean VOT value for the aspirated labial [p^h] was 71 msec (range=40 msec to 112 msec, sd=16.45). The unaspirated [p] was however produced with short-lag VOT with a mean VOT of 15 msec (range= 4 msec to 41 msec, sd=5.59). By contrast, in words in sentences, the voiceless [p^h] was produced with a lower mean VOT value of 64 msec (range=37 msec to 94 msec, sd=12.52), 7 msec less in sentences than in isolated words. The unaspirated [p] was produced with 13 msec VOT, 2 msecs less in sentences than in isolated words (range=1.78 msec to 27 msec, sd=4.61).

In words in isolation again the coronal stop was produced with longer VOTs than the labial sound in both unaspirated and aspirated forms. Thus, the aspirated [t^h] was produced with mean VOT of 89 msec (range= 52 msec to 165 msec, sd=17.69) and the unaspirated [t] was produced in the short-lag with mean VOT of 23 msec (range=10 msec to 44 msec). In sentences, however, the ENS produced the aspirated coronal [t^h] with mean VOT 79 msec (range=49 msec to 123 msec, sd=15). The unaspirated [t] was produced with mean VOT of 19 msec (range=9 msec to 35 msec, sd=5.35).

Finally in words in isolation the dorsal sound [k^h] was produced with even higher VOT ranges than the coronal stop with mean VOT of 96 msec (range=63 msec to 136 msec, sd=16.42), and with mean VOT 28 msec in its unaspirated form (range=13 msec to 45 msec, sd=6.13). In sentences, however, the English [k^h] was produced with a lower mean VOT of 80 msec ranging between 44 and 115 msec (sd=16.29) and the [k] was pronounced with mean VOT of 26 msec ranging between 12 msec and 39 msec (sd=6.35).

Clearly, all VOT values for the ENS voiceless stops increased across places of articulation (labial < coronal < dorsal) in both contexts as well as both aspiration forms (unaspirated and aspirated). However, VOTs in isolated words were always longer than those in sentences.

Confirming statistically the overall effect of context on ENS voiceless stop VOT, the main effect of context is highly significant though smaller in effect size (.386) in comparison with place and aspiration as main effects (Table 4.2). However, while

VOTs were generally longer in words in isolation than in sentences, the difference was significantly different for aspirated compared with unaspirated stops (as reflected in the significant interaction effect of context and aspiration together p=.007). As we can see in Figure 4.4, the difference between the two contexts in mean VOT is 10 msec for aspirated stops but only 3 msec for unaspirated stops. This is however commensurate with the generally shorter VOT of the unaspirated stops: the smaller the VOT, the smaller the context dependent difference in VOT becomes.

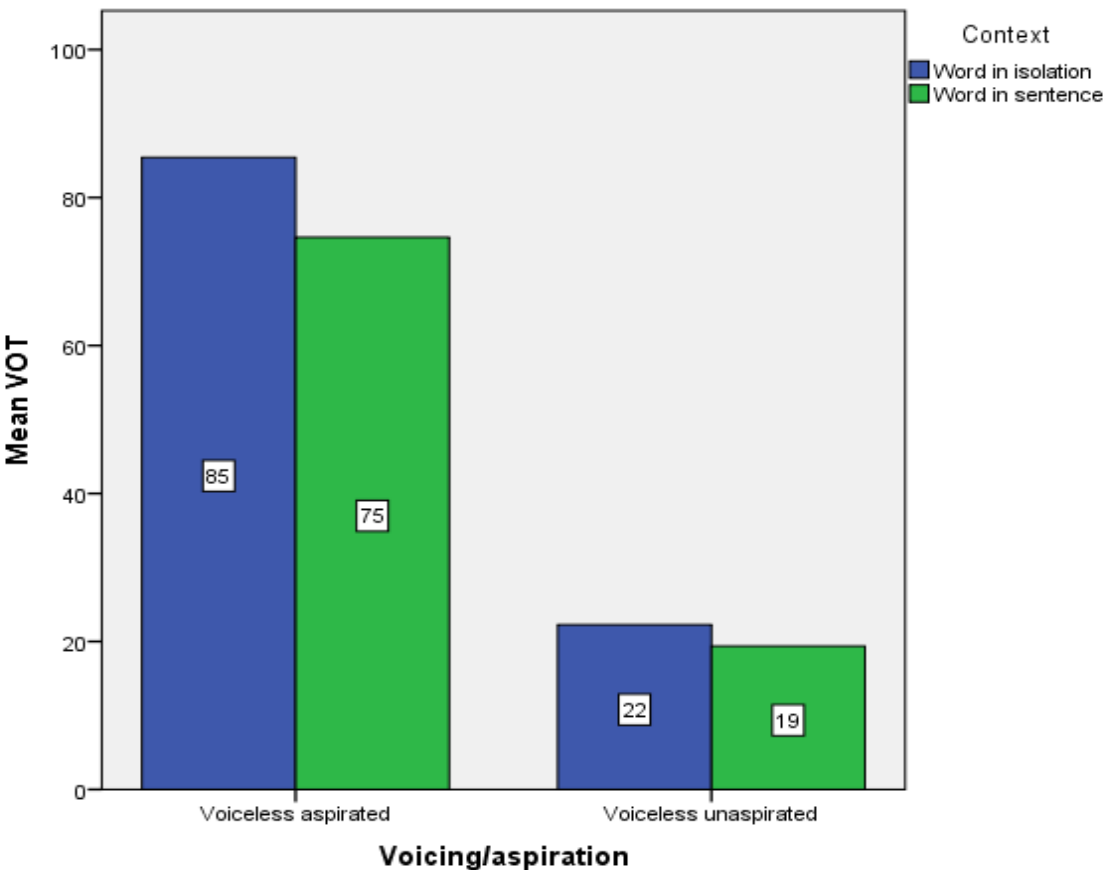


Figure 4.4: Production by ENSs of word initial voiceless plosives, by context and aspiration

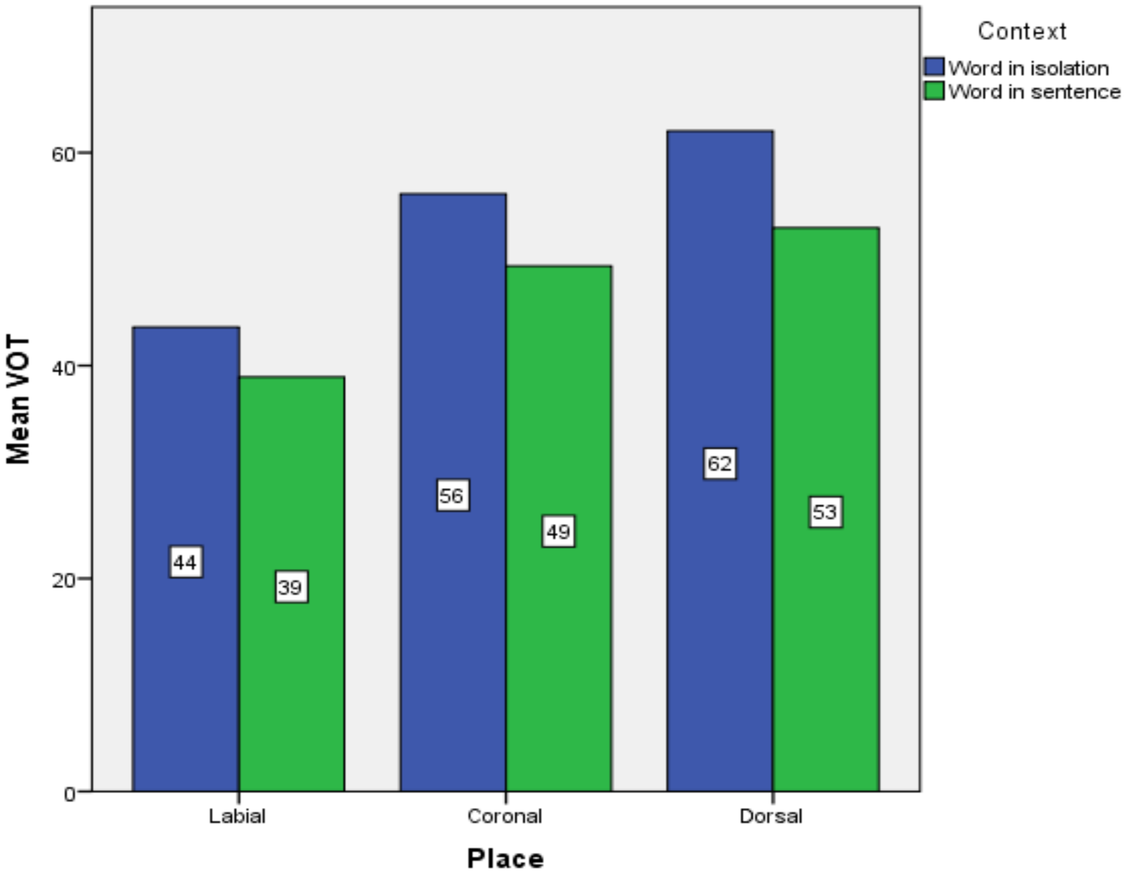


Figure 4.5: Production by ENSs voiceless plosives, by context and place

The context effect also varies in relation to place (p=.016) due to the same principle. As Figure 4.5 shows, the differences between contexts increase across POA, front to back, in parallel with the increase in VOT in general associated with each POA.

Table 4.3: VOT values of English monolinguals by place and aspiration (not differentiating following vowels): in words in isolation

Stop	POA	VOT			
		Minimum	Maximum	Mean	Standard Deviation
Voiceless aspirated	Labial	39.67	111.87	71.36	16.45
	Coronal	51.87	164.76	88.96	17.69
	Dorsal	62.55	135.77	95.97	16.42
Voiceless unaspirated	Labial	4.32	40.68	15.22	5.59
	Coronal	9.66	43.73	23.29	5.94
	Dorsal	12.71	45.26	28.10	6.13

Overall, VOTs of initial voiceless stops in words read by ENS in carrier sentences were considerably lower than those spoken in isolation. Similar results for English were found by Docherty (1992) and Klatt (1975), who both confirmed that VOT increases in words uttered in isolation and decreases with continued speech.

Table 4.4: VOT values of English monolinguals by place and aspiration (not differentiating following vowels): in carrier sentences

Stop	POA	VOT			
		Minimum	Maximum	Mean	Standard Deviation
Voiceless aspirated	Labial	36.61	93.57	64.33	12.52
	Coronal	48.82	122.81	79.44	14.83
	Dorsal	44.24	114.92	80.09	16.29
Voiceless unaspirated	Labial	1.78	27.46	12.95	4.61
	Coronal	9.34	35.09	19.25	5.35
	Dorsal	12.46	38.65	25.78	6.35

4.2.1.4 Effect of following vowel on VOT of ENS voiceless stops

Table 4.5 shows the ENS voiceless stop results in the fullest detail, by POA, aspiration, vowel and context. However, as Table 4.2 shows, there was no overall significant difference in VOT in production of voiceless stops by ENS dependent upon the following vowel, when aspirated and unaspirated are not separated. Furthermore, although there were interaction effects of vowel with other variables, as we will see below, there were no vowel effects in interaction with context. Thus, the context differences we have noted above did not vary systematically depending on the vowel. For these reasons figures for the two contexts may be combined in the account which follows.

For example, the vowels /iː, ɑː, uː/ have a similar effect after [p^h], as the mean VOT values of [p^h] fell between 58 msec and 69 msec before the three vowels in words in sentence contexts and were between 67 msec and 74 msec in isolated words. Similarly, the mean VOT values of [p] were between 12 msec and 15 msec in all three vowel contexts in words in sentences and were between 14 msec and 16 msec in isolated words. Thus, vowel related differences between voiceless stop VOT means for ENS were all small: between 6 and 12 msec for the aspirated stops and between 2 and 7 msec for the unaspirated ones.

The significance test results (Table 4.2) therefore show that there is no simple overall effect of vowel on VOT of English voiceless stops, however there are significant effects in combination with aspiration (p=.016) and place (p<.001) and aspiration and place together (p<.001). As Figure 4.6 indicates, this arises because the effect of vowels on stop VOT across POAs is quite different for the aspirated and the unaspirated stops. For aspirated stops the VOT effect of vowels rises across the three vowels /iː ɑː uː/ for preceding labials, and coronals, but falls for dorsals. By comparison in the unaspirated stops, the VOT effect of following vowel is that /ɑː/ always induces the lowest VOT of the three vowels; /iː/ is associated with lower VOT than /uː/ in labials, and to a lesser degree in coronals, but rises with POA front to back so that it is associated with higher VOT than /uː/ for dorsals.

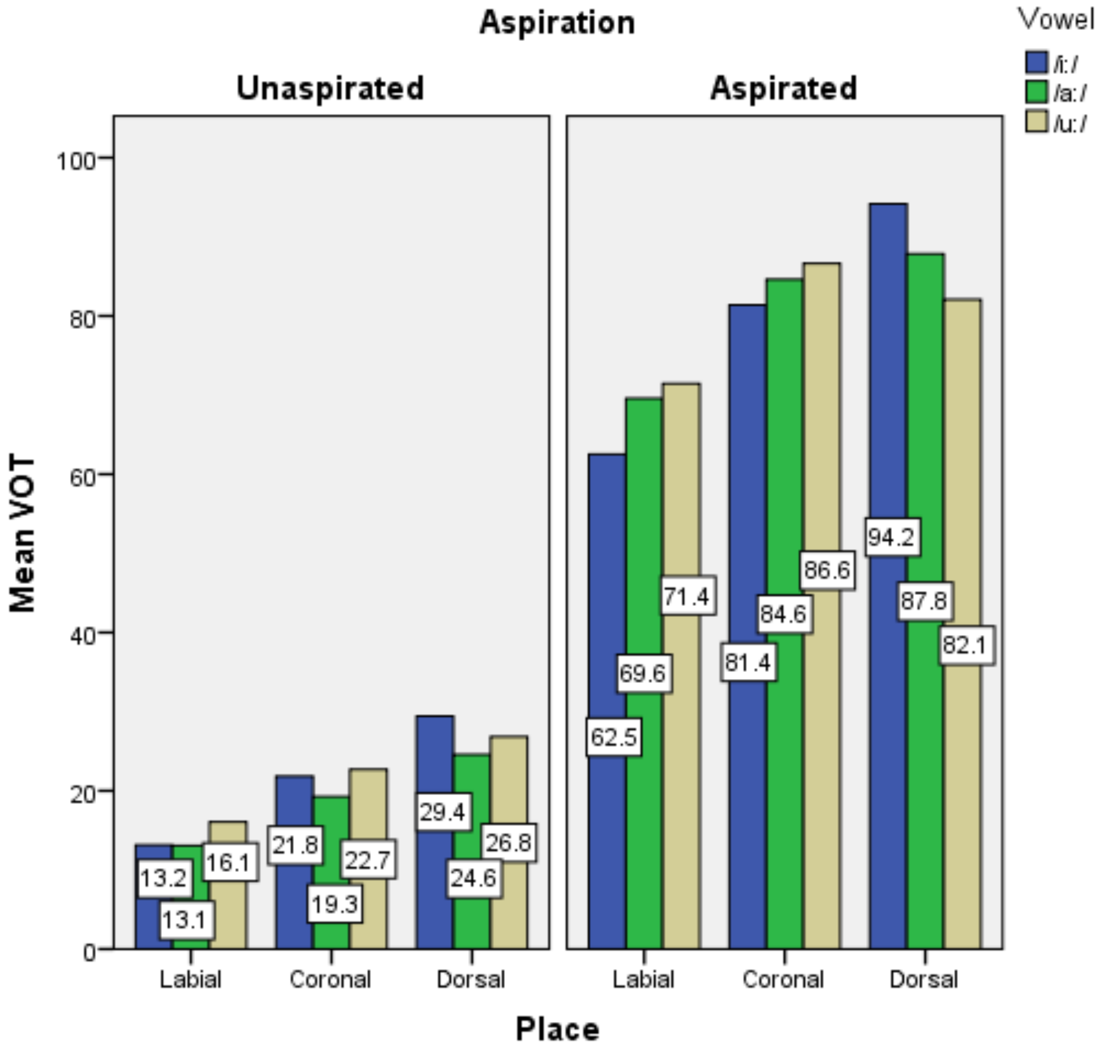


Figure 4.6: Production by ENSs of word initial voiceless plosives, by place, aspiration and vowel (contexts combined)

There was no interaction effect of vowel, aspiration, and place with context (p=.542), however. This signals that the vowel related VOT patterns seen in Figure 4.6 are not significantly different in words in isolation from those in words in sentences. The sequence of vowels with increasing VOT in English voiceless plosives spoken in sentences is the same sequence as in isolated words: (u: > ɑ: > i:) in the aspirated labial and coronal and the other way round for the aspirated dorsal (i: > ɑ: > u:). For the unaspirated stops in words in sentences, the order of increasing the duration of VOT was (u: > i: > ɑ:) for labial and coronal and (i: > u: > ɑ:) for the dorsal, which does not exactly match the sequences for words in isolation, but this variation is not significant.

Table 4.5: VOT values of ENS voiceless stops by POA, aspiration and following vowel, in two contexts (words/ sentences)

Context			Words in isolation		Words in sentences	
POA	Vowel	Aspiration	Mean	SD	Mean	SD
Labial	/i:/	Unaspirated	14.41	3.83	11.89	4.02
		Aspirated	66.95	15.98	58.08	10.05
	/ɑ:/	Unaspirated	14.81	4.40	11.32	4.19
		Aspirated	72.86	17.93	66.30	12.76
	/u:/	Unaspirated	16.47	7.74	15.66	4.51
		Aspirated	74.26	14.91	68.61	12.41
Coronal	/i:/	Unaspirated	23.47	6.27	20.22	5.39
		Aspirated	87.27	20.07	75.49	13.54
	/ɑ:/	Unaspirated	21.95	5.17	16.60	4.00
		Aspirated	87.86	19.05	81.33	15.88
	/u:/	Unaspirated	24.45	6.25	20.95	5.61
		Aspirated	91.74	13.50	81.50	14.67
Dorsal	/i:/	Unaspirated	29.05	5.99	29.82	4.48
		Aspirated	101.47	16.01	86.86	15.56
	/ɑ:/	Unaspirated	26.19	6.70	22.94	5.06
		Aspirated	96.97	16.47	78.72	17.59
	/u:/	Unaspirated	29.05	5.37	24.59	7.18
		Aspirated	89.47	14.97	74.70	13.57

From the results in Table 4.5, it is obvious that the low vowel /ɑ:/ has caused the least lengthening effect on VOT in all the unaspirated stops [p,t,k] both in sentences and in isolated words in English. The high vowel /i:/ is associated with relatively high VOT of the dorsal stop, in both forms (aspirated and unaspirated), and in both contexts, as well. This is also true of dorsals with /i:/, as we will see, in the ANS data. The back vowel /u:/ by contrast is associated with the highest VOTs of the ENS labial and the coronal voiceless stops both aspirated and unaspirated and in both contexts (words and sentences).

It is widely found from previous research (Klatt, 1975; Weismer, 1979; Rochet & Fei, 1991; Chao et al., 2006) that word-initial stops have longer VOTs when followed by high vowels than by low vowels. This has only been confirmed in our findings with the unaspirated stops, while in the aspirated stops, one of the two high vowels is always associated with the shortest VOT, shorter than that of the low vowel /ɑ:/.

The high back vowel /u:/ in our study is in line with the literature, however, as it yielded higher VOT than /ɑ:/ in five stops [p], [t], [k], [p^h], [t^h] though not the dorsal [k^h], see Figure 4.6. Although, the widely accepted phenomenon that high vowels increase VOT more than low vowels has not been confirmed 100% in this study, it can be safely claimed that high vowels in general generate longer VOT values than low ones in English. However, what is more important to conclude is that, considering all the descriptive and inferential statistics presented here, the vowel has no effect on the VOT of the preceding stop in English when voiceless aspirated and unaspirated stops are lumped together, but it does have a significant effect, in combination with POA, when the aspirated and unaspirated are considered separately (as reflected in Figure 4.6).

These results agree with Lisker and Abramson's (1967) study, as they state that there is no major effect on VOT of the following vowel, but disagree with a number of studies (e.g. Klatt, 1975; Weismer, 1979; Port, 1979; Rochet & Fei, 1991; Chao et al, 2006; Schmidt,1996; Johnson & Babel, 2010; and Iverson et al, 2008), as they all found a strong effect of vowel on the adjacent consonant.

4.2.2 Voiced stops of ENS

The same presentation strategy of describing ENS voiceless stops is now implemented with the voiced ones, that is all ENS results will be described with regards to the main effects on VOT of voiced stops i.e. effect of POA, effect of context and effect of vowel.

English VOT patterns for voiced stops found in this study agree with other studies of English (e.g. Lisker & Abramson, 1967; Klatt, 1975; Docherty, 1992) which all found that English voiced stops were produced in the short-lag area (Table 4.1, Figure 4.1). As we can see from Table 4.1, some tokens of the English native speakers voiced stops were produced with pre-voicing in all places of articulation (labial, coronal and dorsal), similar to ANS as will be seen later. This finding is also in line with other studies like Lisker & Abramson (1967), Docherty (1992) and Khattab (2002) who all found that some tokens recorded by their ENS participants were produced with pre-voicing.

Table 4.6: Overall ANOVA for ENS production data: voiced stops

Effect	F	p	Partial eta squared
Place	46.82	<.001	.661
Context	7.24	.013	.232
Vowel	0.276	.615	.011
Place by Vowel	1.99	.153	.077
Vowel by Context	1.10	.311	.044
Place by Context	0.122	.824	.005
Context by Vowel by Place	0.287	.759	.012

We next performed an ANOVA within the voiced stop data, in order to see which of our variables were having a significant impact on VOT variation. As Table 4.6 shows, the only significant differences in VOT of voiced stops produced by ENS were overall effects of place and context. There were no significant effects of vowel either alone or in combination with any other factor. We now consider POA, context and vowel in turn.

4.2.2.1 Effect of POA on VOT of ENS voiced stops

By far the greatest effect was for POA (.661), which had the same effect as for voiceless stops in that VOT was successively longer from front to back (Figure 4.1), with very similar values to those of the unaspirated voiceless stops. Differences between all three pairs of places of articulation were highly significant (p<.001) and the differences are: bilabial < coronal with mean VOT difference 6 msec, coronal < dorsal mean difference 7 msec, bilabial < dorsal mean difference 13 msec (Figure 4.7).

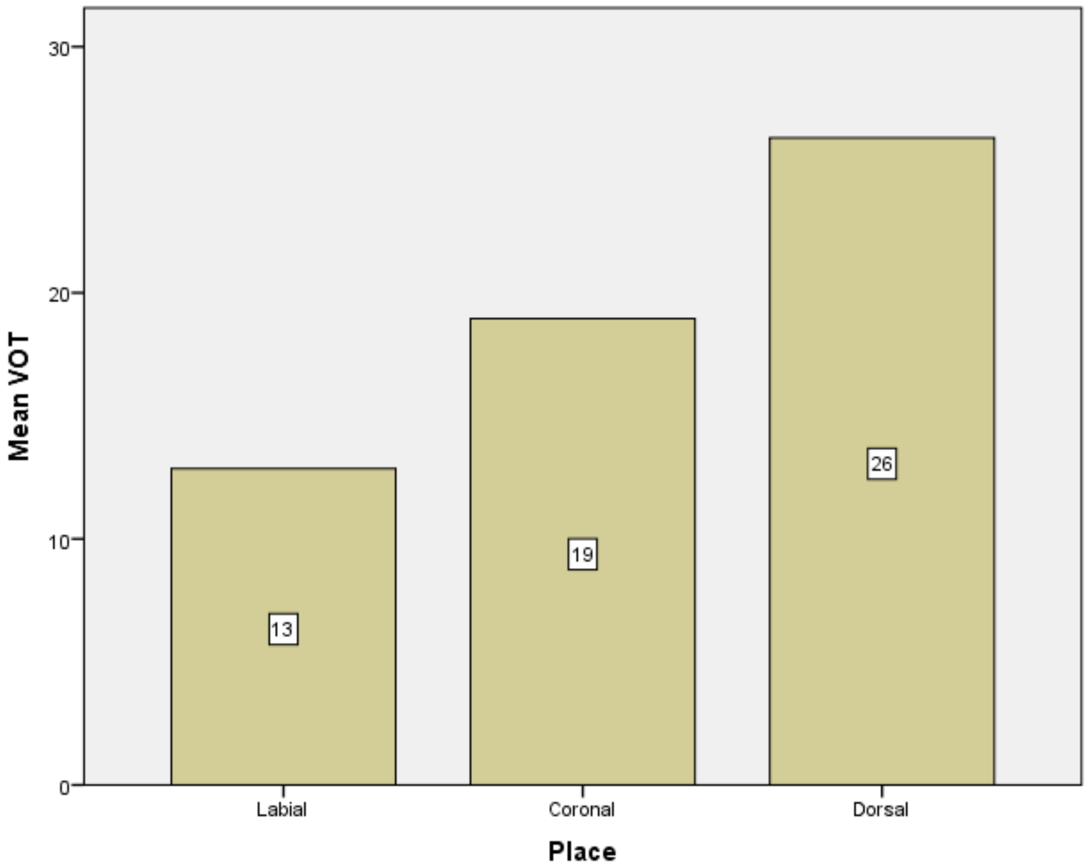


Figure 4.7: Production by ENSs of word initial voiced plosives, by place

4.2.2.2 Effect of context on VOT of ENS voiced stops

Context had a smaller effect (p=.013, effect size .232), similar to that for the voiceless unaspirated stops, in that VOT of voiced stops in words in isolation was on average 4 msec longer than that of words in sentences.

Looking at ENSs' results in Table 4.7 below, the mean VOT value in words in isolation of the English labial /b/ was 14 msec (range=-21 msec to 97 msec, sd=11.9), the coronal /d/ was with a mean VOT of 21 msec (range=-99 msec to 126 msec, sd=23.66), and the dorsal /g/ was produced with an average VOT value longer than the other two stops, 28 msec (range=-110 msec to 95 msec, sd=18.46).

Table 4.7: VOT values of initial voiced stops of ENS (not differentiating following vowels): in words in isolation

Stop	POA	VOT			
		Minimum	Maximum	Mean	SD
Voiced	Labial	-21.36	97.13	14.42	11.09
	Coronal	-99.16	125.60	20.96	23.66
	Dorsal	-109.84	95.09	28.33	18.46

When ENS pronounced their native voiced stops in words in a carrier sentence (Table 4.8), they produced shorter VOT values than in isolated words (Table 4.7) as we have seen earlier also with voiceless stops. Again, some ENS produced some of their tokens with negative VOT values (pre-voicing). The labial stop was produced with a mean VOT of 11 msec (range=-77 msec to 28 msec, sd=13.34). The /d/ was pronounced with an average VOT of 17 msec (ranging between -59 msec and 31 msec, sd=11.11). The dorsal voiced stop was produced with mean VOT value of 24 msec (ranging between 11 msec and 42 msec, sd=6.41). All ENS mean VOT results were in the short lag as predicted, however only the labial and the coronal were produced with negative VOTs in some of the ENS tokens as we can observe from the minimum values in Table 4.8.

Table 4.8: VOT values of initial voiced stops of ENS (not differentiating following vowels): in words in sentences

Stop	POA	VOT			
		Minimum	Maximum	Mean	Standard Deviation
Voiced	Labial	-77.29	27.97	11.20	13.34
	Coronal	-59.50	31.02	16.96	11.11
	Dorsal	10.68	41.91	24.28	6.41

4.2.2.3 Effect of vowel on VOT of ENS voiced stops

The detailed figures for voiced stops, including vowels, may be seen in Table 4.9. No VOT effects of following vowel were significant for the voiced stops, unlike for the unaspirated voiceless stops which are in other ways very similar in VOT. Nevertheless, some descriptive observations may be made.

Table 4.9: VOT values of ENS voiced stops by POA and following vowel, in two contexts (words/ sentences)

		Words in isolation		Words in sentences	
		Mean	SD	Mean	SD
Labial	/i:/	15.25	15.69	8.42	16.91
	/ɑ:/	12.28	5.78	8.70	13.30
	/u:/	15.74	9.51	16.59	5.92
Coronal	/i:/	22.45	26.41	16.62	15.63
	/ɑ:/	22.90	11.29	15.36	10.20
	/u:/	17.47	29.49	18.91	5.31
Dorsal	/i:/	30.78	9.63	25.54	5.78
	/ɑ:/	28.70	15.11	22.09	5.53
	/u:/	25.50	26.65	25.20	7.38

We can see, for example, that descriptively the difference between contexts is maintained in the VOT of voiced stops before /i:/ and /ɑ:/ much more clearly than before /u:/, where it all but disappears or slightly reverses.

It is perhaps most instructive to compare the effects of vowels on preceding voiced stops, as seen in Figure 4.8, with the corresponding unaspirated voiceless stop results displayed in the left hand side of Figure 4.6.

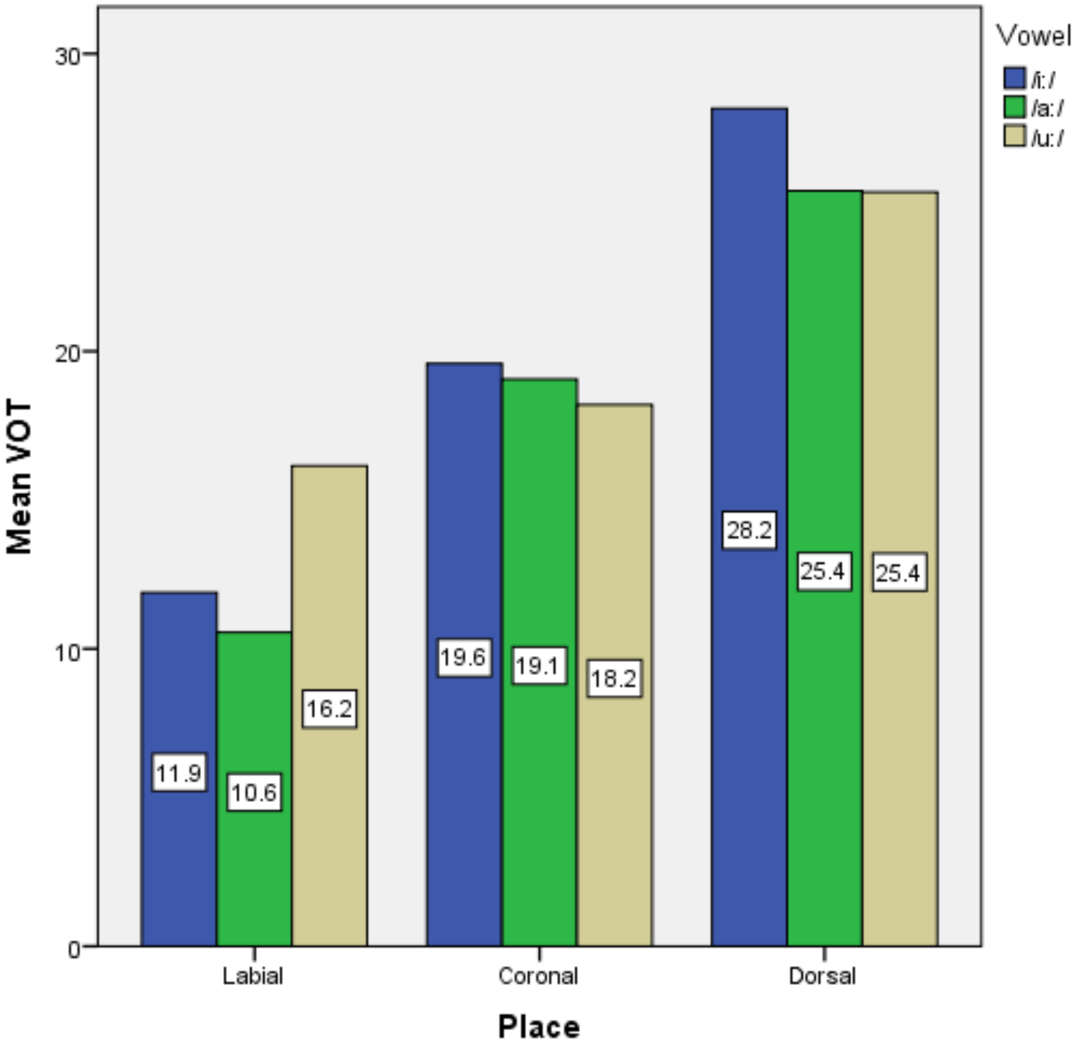


Figure 4.8: Production by ENSs of word initial voiced plosives, by place and vowel (contexts combined)

We can see that the patterns in the left hand side of Figure 4.6 are not totally dissimilar from those in Figure 4.8. In both, the high vowel /i:/ induces a shorter VOT than /u:/ in labials, but the VOT increases front to back over the POAs more steeply before /i:/ than before /u:/ so ends up higher than before /i:/ in the dorsals. With the voiced stops, however, /ɑ:/ is not always associated with lowest mean VOT as it is (just) for the unaspirated stops.

4.2.3 Summary of ENS results

For all stops ENS show a significant effect of place on mean VOT regardless of aspiration or voice: labial < coronal < dorsal. Therefore, VOT increases as POA of the stop moves further back in the mouth, which confirms the universal place of articulation effect on VOT found in English and across other languages.

There is also an effect everywhere of context, with words in isolation having longer mean VOTs than those in sentences. The smaller context differences for unaspirated and voiced stops than for aspirated ones presumably reflect the fact that the former stops have shorter VOTs than aspirated ones in the first place.

Our ENS also make an overwhelming distinction between voiceless aspirated plosives on the one hand (with VOT in the long-lag area) and unaspirated and voiced plosives on the other (in the short-lag VOT area). The latter do not differ significantly in VOT.

Following vowel does not have a simple overall effect on VOT. Rather, it relates to VOT differently for aspirated voiceless stops than for unaspirated/voiced ones, and in each case, it follows a different pattern across places of articulation.

The English patterns of VOT obtained in this study in general accord with previous results obtained by Lisker & Abramson (1967), Klatt (1975), Docherty (1992), Scobbie (2002), which all agree that VOTs of English voiceless stops [p^h,t^h,k^h] in initial position are in the long-lag region and produced with aspiration, as voicing of the subsequent vowel starts vibrating long after the burst of the stop closure, while voiceless unaspirated stops [p,t,k] are produced in the short-lag region.

4.3 VOT of Saudi Arabic stops

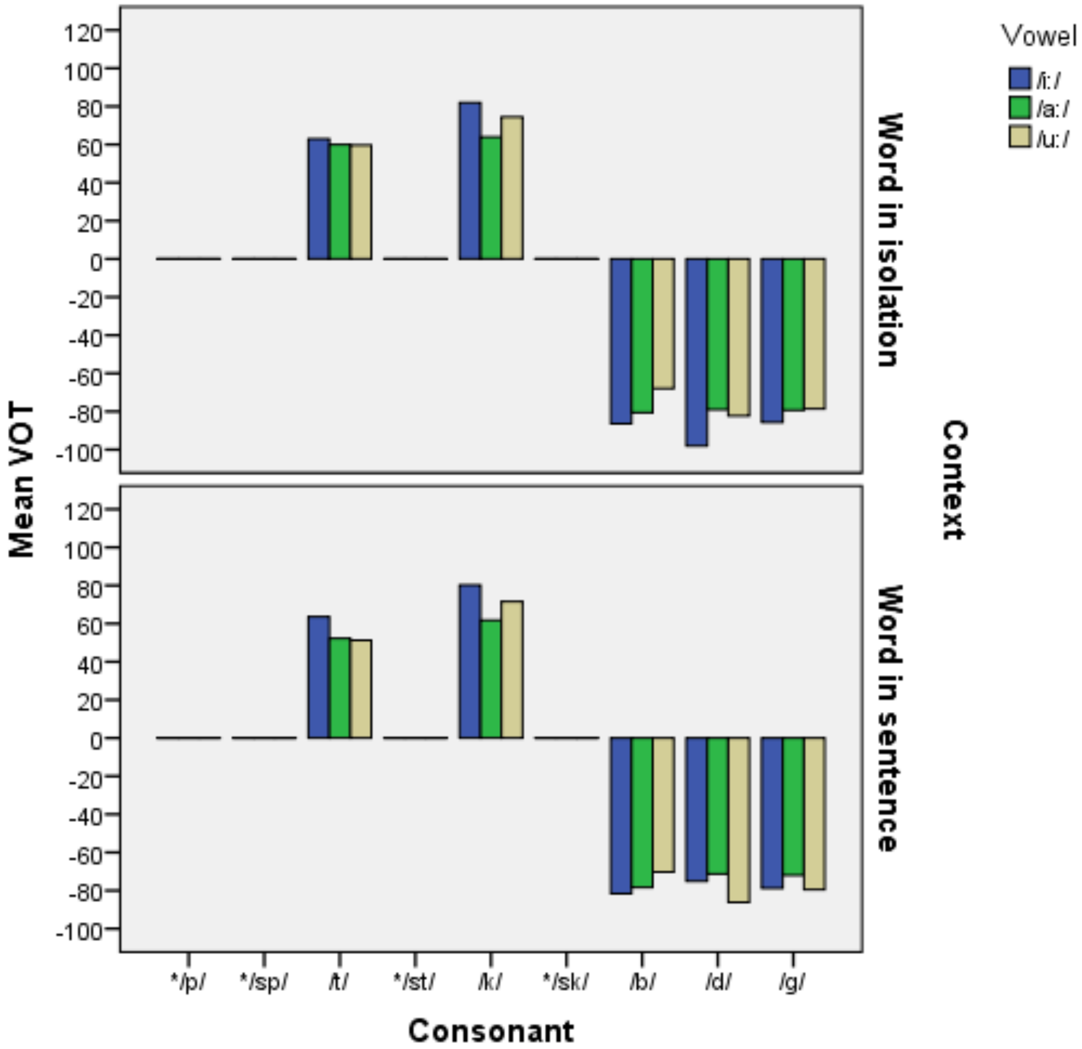
Next, results of the ANS will be described. First, a general review of their results is provided, then we move to describe each stop category separately: voiceless stops then voiced ones, with reference to each effect separately (of POA, vowel and context).

Table (4.10) below presents the mean VOT values (in msec) of Arabic voiceless stops /t k/ and voiced stops /b d g/ for ANS with the range (minimum and maximum), and the standard deviation, excluding vowel and context. As seen earlier with ENS, all stops elicited in this study were tested before three vowels [i, ɑ, u], however, for simplicity we first display ANS VOT results in general, in relation to voice and POA (Table 4.10), considering the other effects, i.e. vowel and context, later.

Table 4.10: VOT values of ANS initial stops by voice and POA (not differentiating contexts or following vowels)

Stop	POA	VOT			
		Maximum	Minimum	Mean	Standard Deviation
Voiceless	Labial	.	.	.	.
	Coronal	116.20	25.63	58.28	15.91
	Dorsal	131.33	34.88	72.31	14.87
Voiced	Labial	-17.99	-135.26	-77.54	24.77
	Coronal	-11.85	-156.94	-81.97	26.45
	Dorsal	-2.29	-146.96	-78.96	24.86

An initial overall idea can also be gained from Figure 4.9. Clearly, there is a huge difference between Arabic voiced and voiceless stops, as the former show negative VOT values (pre-voicing) while the latter were produced towards the other end of the continuum with positive long-lag VOT. However, differences between places of articulation and between vowels and between contexts are not so clear, therefore we started with an overall ANOVA analysis. However, due to the nature of the Arabic consonant system which lacks /p/, this, of necessity, included only two places of articulation: contexts of word use (2), places of stop articulation (2), voicing and aspiration options (2), and qualities of following vowels (3), together with combined effects of all those (For the full table of results of this analysis see Appendix F).



(Note: Asterisked sounds do not exist in this variety of Arabic.)

Figure 4.9: Production by ANS of all voiced and voiceless stops

As we would expect from Figure 4.9, the ANOVA shows a colossal overall main effect of voicing/aspiration ($F=2529.17$, $p<.001$) in stops with a mean difference of 145 msec in VOT between the Arabic voiceless and voiced stops. The overall effect of POA, in voiceless and voiced together, was similarly highly significant ($F=23.06$, p<.001). This result however is just for coronals and dorsals, considering together both contexts, both stop categories (voiceless and voiced) and all vowels, and shows that dorsals /k, g/, with mean VOT -6.7, are less negative (i.e., in a sense, more positive) in VOT than coronals /t, d/, with mean VOT -23.7, by around 17 msec. This difference is consistent with the quasi-universal pattern of positive VOTs increasing across POAs from front to back of the mouth.

There were also a number of significant interactions between the two stop categories and other variables. The strongest of these both involve voice: place by voice (F=25.45, p<.001) and vowel by voice (F=24.77, p<.001). There was no overall effect of vowel or of context but these did appear in some further significant interactions: context by vowel (F=7.58, p=.001), place by vowel (F=4.96. p=.010), context by place by vowel (F=4.57, p=.014) as we will see later.

Given the above, and due to the extreme differences between voiced and voiceless stop in Arabic, we will now pursue the results in more detail for each stop category (voiceless and voiced) separately, which also then allows us to include all three places of articulation of the voiced stops in the analysis. Therefore, the analyses performed for the following sections are:

Voiceless stops: ANOVA of 2 places of articulation x 3 vowel qualities x 2 contexts (isolation vs in a sentence)

Voiced stops: ANOVA of 3 places of articulation x 3 vowel qualities x 2 contexts (isolation vs in a sentence)

It should be noted that the comparison of Arabic voiced stops with voiceless has been widely reported in previous studies and is very well known in the literature of other dialects of Arabic. We repeat this analysis here as a way of establishing the position of our selected dialect (that of northern Saudi Arabic) since this is, to the best of our knowledge, the first study to be done on stop consonants of this dialect within Saudi Arabic.

4.3.1 Voiceless stops of ANS

Table 4.11 shows that, within the ANS voiceless stops, there is a highly significant overall effect on VOT of POA, regardless of vowel or context (effect size .782). There is also a considerable simple effect of following vowel, regardless of POA or context (effect size .608), and a combined effect of POA and vowel together. There are no significant effects of context alone or in combination with anything else. We will therefore now examine each effect in descending order of size.

Table 4.11: Overall ANOVA for ANS production data: voiceless stops

Effect	F	p	Partial eta squared
Place	104.25	<.001	.782
Vowel	44.95	<.001	.608
Context	1.84	.185	.060
Place by Vowel	12.21	<.001	.296
Vowel by Context	.195	.151	.063
Place by Context	.855	.363	.029
Context by Vowel by Place	.134	.269	.044

4.3.1.1 Effect of POA on VOT of ANS voiceless stops

Considering the mean VOT values for Arabic voiceless stops produced by ANS (Table 4.10), we can see that VOTs increased from coronal to dorsal, and the inferential analysis (Table 4.11) shows a significant difference in POA between the two places of articulation ($F=104.25$, $p<.001$), with dorsal higher than coronal by 14 msec in VOT.

The Arabic speakers produced the coronal voiceless stop (disregarding vowel and context) with mean VOT of 58 msec (range=26 msec to 116 msec, sd=15.91). The voiceless dorsal stop was produced significantly higher with mean VOT of 72 msec (range=35 msec to 131 msec, sd=14.87). Therefore, VOTs of Arabic voiceless stops showed the expected universal POA pattern, which we also found with ENS earlier, where VOT ranges increased from coronal to dorsal.

Furthermore, we can see from the ANS VOT data (Table 4.10) that Saudi Arabic speakers aspirate all voiceless stops as their VOT ranged between 35 msec and 131 msec for the dorsal and between 26 msec and 116 msec for the coronal. Much of the VOT of their voiceless stops therefore ranged in the aspiration (long lag, over +30 msec) region, apart from very few productions which were below 30 msec and even those were close to it (26 msec).

Our results for /t/ and /k/ in Saudi Arabic correspond with Flege and Port's (1981) results, as they too found voiceless stops in Arabic were aspirated, though with shorter VOT values than our study VOT values: 37 msec for /t/ and 52 msec for /k/. Alghamdi's (1990) study results were also similar to Flege and Port (1981), and to Jesry (1996), though voiceless stops were produced with even less aspiration by his subjects, 32 msec for /t/ and 42 msec for /k/. Flege and Port's (1981) and Alghamdi's (1990) are the only studies investigating Saudi Arabic to be found in the literature, to the best of the researcher's knowledge. Therefore, VOT values of Saudi Arabic voiceless stops in this study were found to be longer than VOTs found in all the previously mentioned studies.

The reason for the difference in VOT lengths could be that, although Flege and Port (1981) and Alghamdi (1990), like the present researcher, studied the colloquial Saudi variety, not MSA or Classical Arabic, there are detailed differences in the dialects used within KSA. Alghamdi studied the Ghamdi dialect, spoken in the southern part of Saudi Arabia, and Flege and Port studied the Najdi dialect, spoken in the central region (Riyadh and its surrounding areas). By contrast, the present researcher used the dialect spoken in the northern part of Saudi Arabia. VOT, however, increased from coronal to dorsal in all of the previous studies as well as the current study.

Our study then confirms that no unaspirated voiceless stops are found in this dialect of Arabic. Unlike other dialects of Arabic (i.e. Syrian, Lebanese, Iraqi, etc.), our northern Saudi Arabic speakers produce their voiceless stops with aspiration.

4.3.1.2 Effect of vowel on VOT of ANS voiceless stops

The ANS voiceless stops differ significantly overall depending on following vowel, regardless of POA or context ($p<.001$). The pattern is: /iː/ > /uː/ > /ɑː/, with mean VOTs respectively 72.2, 64.2, and 59.5 msec. That is, high vowels generate longer VOT in the preceding voiceless stop than the low vowel.

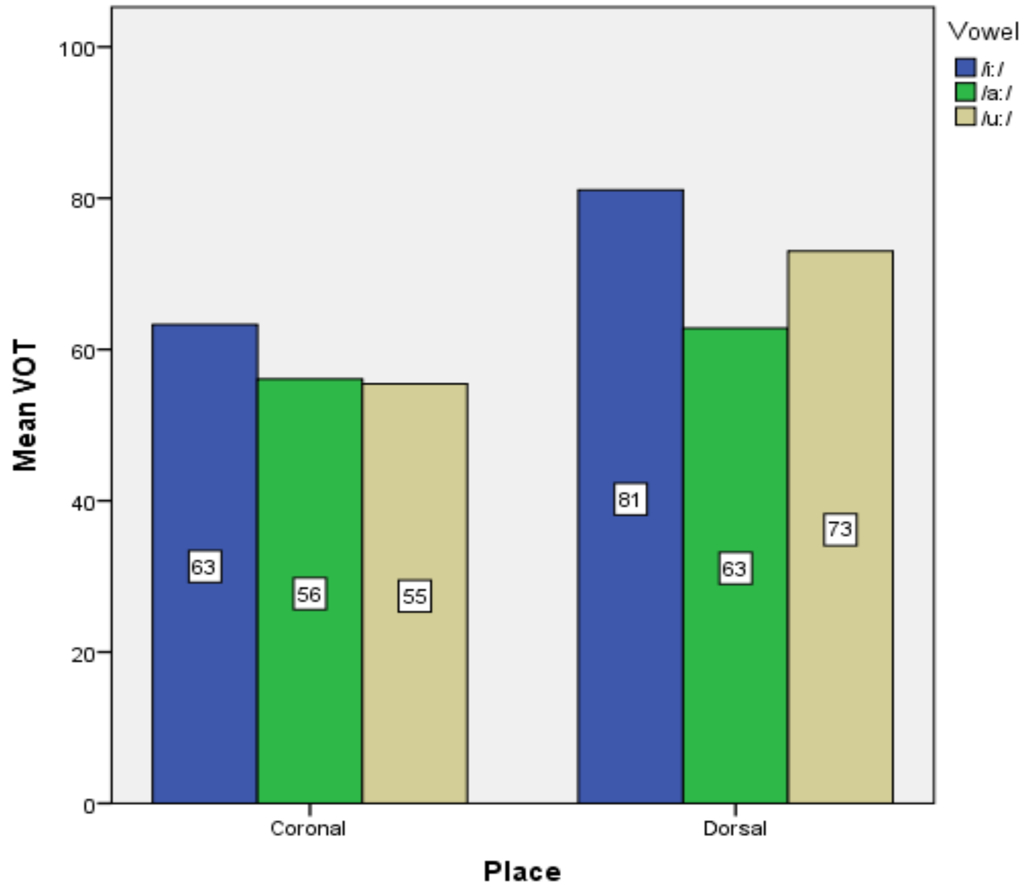


Figure 4.10: Production by ANSs of word initial voiceless stops, by place and vowel

The significant vowel by place interaction effect (Table 4.11) however shows that the difference between vowels is not the same at each place of articulation. We can see this in Figure 4.10 and Table 4.13, where the VOTs before all three vowels are higher for dorsals than coronals, but the vowel related pattern of differences is not the same for each stop. /iː/ yields higher VOT of both the preceding stops than do the other vowels, but while /uː/ gives the second highest VOT with dorsals (as in the overall pattern for both POAs together), it does not clearly do so with coronals, where /uː/ and /ɑː/ yield almost identical VOT. This is consistent with Giannini and Pettorino (1982) who found that the plain stop sounds in Arabic have a longer duration when they occur before /i/ than before other vowels environments.

A follow-up significance test of post hoc paired comparisons of vowels within each place, with Bonferroni adjustments, was conducted (Table 4.12). From this we learn that among the coronals the only vowel related significant difference in VOT was between /ti:/ and /tu:/ while all the vowel related VOT differences were highly significant for dorsals.

Table 4.12: VOT differences of voiceless stops produced by ANS, related to following vowels

Place:	Post hoc paired comparisons (Bonferroni adjusted)					
	/i:/ - / ɑ:/		/i:/ - /u:/		/ɑ:/ - /u:/	
	Mean diff.	p	Mean diff.	p	Mean diff.	p
Coronal	7.19	.127	7.82	.011	0.63	1.000
Dorsal	18.27	<.001	8.09	<.001	-10.18	<.001

Table 4.13: VOT values of voiceless stops of ANS by POA and vowel, with two contexts combined

Stop	POA	Vowel	VOT			
			Maximum	Minimum	Mean	Standard Deviation
Voiceless	Coronal	/i:/	99.67	25.63	63.29	14.82
		/ɑ:/	116.20	30.09	56.09	18.33
		/u:/	88.94	27.97	55.47	13.18
	Dorsal	/i:/	131.33	45.77	81.09	15.30
		/ɑ:/	89.00	43.41	62.82	10.82
		/u:/	104.75	34.88	73.01	12.22

4.3.1.3 Effect of context on VOT of ANS voiceless stops

As we saw in Table 4.12, context had no significant effect on VOT of ANS voiceless stops, either as a main effect or interacting with another factor. Table 4.14 below shows that descriptively VOTs are slightly longer in stops in isolated words than in sentences. This therefore follows the same pattern as ENS, with similar SDs, but the differences between means between contexts are smaller in Arabic (respectively 5 and 2 msec for ANS coronal and dorsal versus 9 and 16 msec for ENS).

Table 4.14: VOT values of initial voiceless stops of ANS, by POA and Context

Context	POA	VOT			
		Maximum	Minimum	Mean	Standard Deviation
Word in isolation	Coronal	116.20	27.97	60.85	17.11
	Dorsal	131.33	37.12	73.47	15.52
Word in sentence	Coronal	91.53	25.63	55.71	14.23
	Dorsal	116.87	34.88	71.15	14.19

4.3.2 Voiced stops of ANS

As Table 4.15 shows, remarkably, ANS voiced stops exhibit no overall significant difference in mean VOT based on POA or context. The Following vowel however yields significant differences in VOT both as an overall effect regardless of place and context, and in interaction with both of those variables.

Table 4.15: Overall ANOVA for ANS production data: voiced stops

Effect	F	p	Partial eta squared
Vowel	7.62	.001	.208
Place	1.84	.168	.060
Context	1.72	.200	.056
Vowel by Context	5.76	.010	.166
Place by Vowel	4.92	.001	.145
Place by Context	1.12	.332	.037
Place by Vowel by Context	1.56	.188	.051

4.3.2.1 Effect of POA on VOT of ANS voiced stops

As can be seen from Table 4.10, all VOT results (minimum, maximum and means) for Arabic voiced stops are negative values, which means that Arabic speakers pre-voice all the voiced stops in their native language. Although POA has no significant overall effect on VOT, the differences descriptively are as follows. Arabic voiced labial /b/ was produced with mean VOT of -78 msec (range=-135 msec to -18 msec, sd=24.77). The coronal stop was produced with a mean VOT value of -82 msec (range= -157 msec to -12 msec, sd= 26.45). The Arabic dorsal /g/ was produced with mean VOT of -79 msec (range= -147 msec to -2.29 msec, sd=25). There are some particular vowels that increase the pre-voicing in voiced stops of Arabic, but we will discuss this under the effect of vowel on VOT in 4.3.2.3.

These VOT values of voiced stops are, in general, similar to the results found by Yeni-Komshian et al (1977), Jesry (1996), Radwan (1996), Flege and Port (1981) and Alghamdi (1990). All the above-mentioned studies found that Arabic voiced stops are produced with pre-voicing by Arabic speakers. Flege and Port (1981), who studied Najdi Saudi Arabic, found that their subjects had a mean VOT of -85 msec for /b/, -82 msec for /d/ and -75 msec for /g/. Alghamdi's (1990) study results for Ghamdi Saudi Arabic were also to some extent similar to those of the current study (-72 msec for /b/, -71 msec for /d/ and -68.7 for /g/). Those studies, however, evidenced a pattern of decreasing pre-voicing in VOT across POAs from front to back b > d > g. If we regard *less* pre-voicing as corresponding to *more* positive VOT, then this matches the universal pattern that VOT becomes more positive across POAs from front to back, as seen in the ENS voiced and voiceless and the ANS voiceless data. The current study of a northern dialect of Saudi colloquial Arabic however does not totally support that trend, since pre-voicing decreases d > g > b (-78 msec for /b/, -82 msec for /d/ and -79 for /g/). In other words, /b/ is exceptional: in order to fit the expected pattern it would have to have the greatest pre-voicing (e.g. -83), not the least. However, these differences are very small and not in any case statistically significant. Furthermore, the pattern of increasing or decreasing VOT across POA in voiced stops in Arabic is not so clear in the literature compared with the pattern for voiceless stops where there is to some extent a consensus that VOTs increase from front to back $p < t < k$. Also, the wider range of studies conducted on Arabic (e.g. Yeni-Komshian, Caramazza, & Preston, 1977; Al Ani, 1970; Port & Mitleb, 1983; Flege & Port, 1981; Alghamdi, 1990; Radwan, 1996; Jesry, 1996; Mitleb, 2009; AlDahri, 2013) yielded mixed results on POA in voiced stops, which suggests the importance of verifying this in more future research on Arabic.

4.3.2.2 Effect of context on VOT of ANS voiced stops

ANS VOT does not differ significantly overall between contexts for the voiced stops, just as it did not for the voiceless stops. Descriptively, however, Table 4.16 shows that pre-voicing is always slightly less in sentences than in words in isolation. This interestingly suggests that the context-related pattern is not that VOT is more positive (including less negative) in words in isolation than in sentences, but that it is more extreme (more distant from zero VOT) in words in isolation than in sentences. This would be consistent with an interpretation that words in sentences are spoken faster than words in isolation, so time consuming aspects of individual sounds, such as VOT, get shorter in sentences, regardless of whether they extend from the core event (plosion) in a pre- (negative) or post- (positive) direction.

Table 4.16: VOT values of initial voiced stops of ANS, by POA and Context

Context	POA	VOT			
		Maximum	Minimum	Mean	Standard Deviation
Word in isolation	Labial	-20.33	-135.26	-78.28	28.49
	Coronal	-11.85	-156.94	-86.48	31.83
	Dorsal	-2.29	-146.96	-81.19	31.50
Word in sentence	Labial	-17.99	-114.78	-76.79	20.51
	Coronal	-16.95	-128.41	-77.46	18.78
	Dorsal	-38.11	-112.09	-76.73	15.53

4.3.2.3 Effect of vowel on VOT of ANS voiced stops

As Table 4.15 shows, vowels produced an overall significant effect on the VOT of preceding ANS voiced stops, regardless of place and context, though with a relatively small effect size (.208). This followed the pattern of decreasing pre-voicing in the order iː > uː > ɑː with mean VOTs respectively -84.3, -77.5, -76.7 msec. This matches the overall order of the vowels in terms of decreasing positive VOT in the voiceless stops. However, the overall difference here between /uː/ and /ɑː/ is clearly minute. Post hoc paired comparisons of the significant main effect for vowels shows that voiced stops before /iː/ (mean VOT -84 msec) have significantly greater negative VOT than those before /uː, ɑː/ which are similar (mean -77 msec): the preceding stop VOT differs significantly between /iː/ and /ɑː/ (p=.002), and between /iː/ and /uː/ (p=.027). The VOT difference between /ɑː/ and /uː/ is however not significant (p=1.00). Thus we can only assert definitely that /iː/ is associated with longer negative VOT than the other vowels.

Furthermore, as in the voiceless ANS data, the significant interaction effects of vowel with POA reveal that the overall pattern of vowel related differences is not evidenced equally at all the places of articulation: it appears more clearly for stops articulated further back in the mouth.

As Figure 4.11 and Table 4.17 show, the iː > uː > ɑː sequence of decreasing (successively less negative) VOT appears in the voiced coronals and dorsals but not the labials, where VOT becomes less negative in the order iː > ɑː > uː.

It therefore appears that ANS tend to produce greatest prevoicing in VOTs before high vowels, particularly before /iː/, as both voiceless stops (/t/ and /k/) and all voiced stops (/b/, /d/, /g/) were produced with greater VOTs before /iː/ than the other two vowels (/ɑː/ and /uː/). The sequence of size of the VOT, whether positive or negative, is to some extent clearer (iː > uː > ɑː) for /k/, /d/ and /g/ but it is not clear with the coronal /t/ where VOTs were produced with very similar values before /ɑː/ and /uː/ (56 msec and 55.47, see Figure 4.10). The voiced labial shares the same pattern by increasing VOT before /iː/, but not /uː/, therefore the pattern is iː > ɑː > uː.

Overall then, the only clear finding concerning ANS vowels is that the high front vowel /iː/ is associated with the greatest negative VOT for all three voiced stops, and the greatest positive VOT for the two voiceless stops.

Table 4.17: VOT values of voiced stops of ANS by POA and vowel, with two contexts combined

Stop	POA	Vowel	VOT			
			Maximum	Minimum	Mean	Standard Deviation
Voiced	Labial	/iː/	-41.67	-125.09	-84.01	21.56
		/ɑː/	-17.99	-135.26	-79.47	26.91
		/uː/	-24.33	-135.26	-69.12	23.53
	Coronal	/iː/	-11.85	-156.94	-86.58	28.10
		/ɑː/	-16.95	-132.72	-75.09	23.32
		/uː/	-26.00	-138.82	-84.24	26.71
	Dorsal	/iː/	-2.29	-144.93	-82.21	25.53
		/ɑː/	-5.80	-146.96	-75.60	24.25
		/uː/	-8.33	-142.89	-79.07	24.77

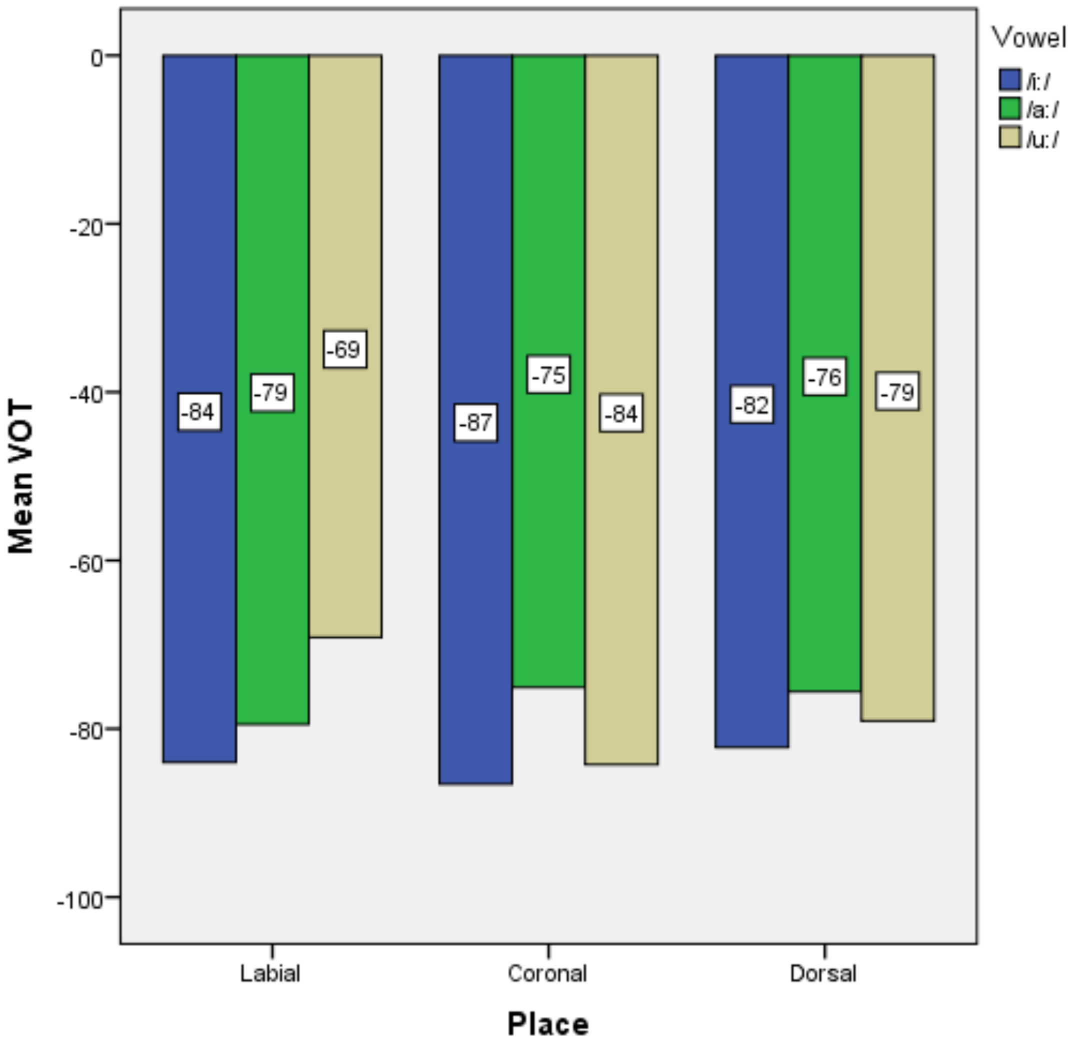


Figure 4.11: Production by ANS of word initial voiced stops, by place and vowel

Finally, if we reorganise Figure 4.11 as Figure 4.12, we can see the interactive effect on VOT of vowel and place together in a different way, from the point of view of how the sequence of VOT across POAs differs depending on the following vowel. We can discern that the overall (non-significant) sequence of decreasing pre-voicing d > g > b is only in fact evidenced before the back vowel /u:/. The other two vowels exhibit different and less well distinguished orders across POAs.

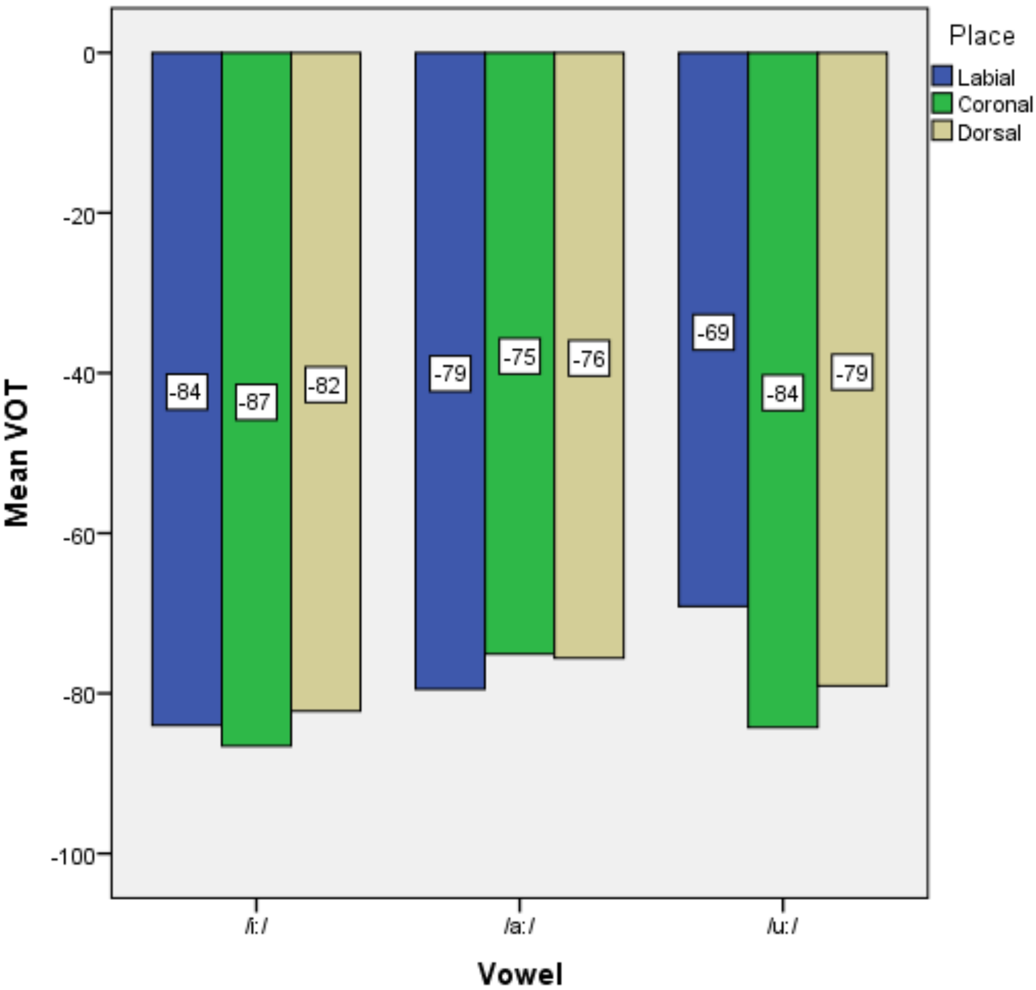


Figure 4.12: Production by ANS of word initial voiced stops, by vowel and place

We have not come across a study that measures vowels in a similar way to the current study, since we have measured effects of vowel height and position (front versus back), not length. The closest study with respect to vowels, Mitleb (2001), found that vowel length does trigger a significant effect on VOT values, but he only analysed the effect of the vowels /a/ and /a:/ before /t/, /d/, /k/ and /g/. Hence our study provides new findings in this area.

4.3.3 Summary of ANS results

Our ANSs make an overwhelming distinction between voiced and voiceless stops: the former are produced with negative VOT (means approximately -60 to -90) and the latter with positive VOT (means in the region +50 to +80) indicative of aspiration.

ANS do not significantly differentiate VOT in production of stops depending on context, unlike ENS.

For voiceless stops they make VOT distinctions primarily for place of articulation, increasing from coronal to dorsal.

For voiced stops, they do not clearly distinguish in VOT between the three places of articulation.

Vowel related differences in VOT of both voiced and voiceless stops are widely made, with generally greatest negative VOT (for voiced stops) or greatest positive VOT (for voiceless stops) associated with following /iː/.

4.4 Chapter conclusion

This chapter has provided the ENS and ANS VOT findings with which our learner VOT values can be compared, in the next chapter, so as to ascertain whether they resemble more closely the values in their L1 or L2. In particular, differences related to stop POA, voicing/aspiration, following vowel and isolated word/sentence context, together with interactive effects of combinations of those factors, have been detailed.

The ENS and ANS VOT results which were obtained all closely replicate the findings of previous studies, where comparisons are able to be made. This applies mainly to findings for the effects of POA, voicing/aspiration and context in English, and for voicing and POA in Arabic. This therefore supports the validity of the methods of data gathering and analysis which we used to measure VOT.

We have, however, also added to previous knowledge, particularly in the area of detailed vowel effects in both languages in relation to POA, and context effects in Arabic and in providing detailed figures for many of the interactive effects.

Chapter 5: Results for the Saudi Learners of English

5.1 Introduction

In this chapter, all results of the L2 learners will be examined. We will answer the research questions in turn with appropriate inferential statistical analysis, interpretation and discussion. The study's findings will be explained in the light of Flege's (1995) Speech Learning Model 'SLM' wherever appropriate.

First we will review the learners' results on their own, mostly in a parallel way to those of the ENS and ANS in chapter 4, but additionally paying attention separately to learners who produced voiced stops with positive VOT. Next, the three groups' production results (learners, ANS, ENS) will be compared in two key ways in order to assess how nativelike the learners are in their VOT production. We then consider how far any relationship could be found between learner length of UK residence and daily use of English and the nativelikeness of their VOTs. Finally, we look separately at a few notable individual learners.

5.2 Learner production of stops

In this section, we will present the learners' results overall, then we will examine each stop category (voiceless and voiced) separately with reference to the four factors potentially having effects on VOT (POA, following vowel, context, and aspiration (for voiceless only)).

An initial impression relevant to the learners' results can be gained from Figure 5.1. Just judging by eye, learners seem to be making a huge distinction in VOT between voiced and voiceless stops, because the former have negative VOT. There are some signs of distinctions based on place of articulation, however, it is unclear what impacts aspiration, vowel quality and context (whether the word is spoken alone or in a sentence) have.

Table 5.1: Designed VOT values of initial voiceless plosives produced by learners (disregarding vowels and contexts).

Stop	POA	VOT			
		Maximum	Minimum	Mean	Standard Deviation
Voiceless aspirated	Labial	140.82	-129.97	42.17	47.05
	Coronal	143.40	22.33	63.71	21.81
	Dorsal	131.67	33.67	76.71	18.14
Voiceless unaspirated	Labial	106.28	-48.00	32.23	24.13
	Coronal	138.32	24.33	58.09	17.89
	Dorsal	111.00	26.32	65.37	18.56
Voiced	Labial	151.20	-191.74	-42.07	58.12
	Coronal	153.51	-183.33	-77.18	53.42
	Dorsal	110.52	-188.73	-59.93	48.51

Table 5.1 presents a simpler picture, leaving out vowel and context differences, of the mean VOT values (in msec) of the production of L2 learners' voiceless stops /p t k/ and voiced stops /b d g/, with the range (minimum and maximum), and the standard deviation.

We can see at once that voiced stops are all produced on average with pre-voicing, similar to ANS rather than ENS (see further 5.3), though some positive VOT responses were recorded by the learners. Furthermore, Saudi Arabic L2 learners of English do not have voiceless (aspirated and unaspirated) labial, unaspirated coronal, or unaspirated voiceless dorsal stops in their native language, as we have seen earlier in their ANS counterparts' results (4.3). Aspiration therefore might be a challenging distinction for them, as mentioned in the literature (2.3). The mean VOTs of the L2 learners of all voiceless stops were in general lower than the ENS mean VOTs for the aspirated stops, higher than ENS for the unaspirated stops. As we will explore in more detail in 5.3, the learners did not differentiate the aspirated and unaspirated voiceless stops with VOT differences of at least 50 msec as the ENS do, but produced VOTs somewhere in between the ENS values, primarily in the low long-lag region.

In order to test which differences were significant, an overall ANOVA was first performed comparing effects on learner VOT of: contexts of stops (2), places of stop articulation (3), voicing and aspiration (short-lag and long-lag) (3), and qualities of following vowels (3), together with combined effects of all those (for the full table of results of this analysis see Appendix G). In summary, this showed very clear overall main effects of place ($F=13.72$, $p<.001$), voicing-aspiration ($F=320.38$, $p<.001$), and vowel quality (13.92, $p<.001$), but not of context ($F=.104$, $p=.750$). There was also a very clear interactive effect of place, vowel and voicing-aspiration together ($F=6.92$, $p<.001$) together with interactions of place and vowel ($F=6.88$, $p=.001$), place and voicing-aspiration ($F=25.17$, $p<.001$) and vowel and voicing- aspiration ($F=10.43$, $p<.001$). Therefore, all the factors POA, vowel, voicing-aspiration, excluding only the context, had a highly significant effect on the VOT of the stops produced by the learners.

Further analysis showed that while there was a significant overall difference of VOT between aspirated and unaspirated voiceless stops (mean difference 9.34 msec, $p<.001$), the dominant differences are between voiced stops and unaspirated voiceless ones (mean difference 112.38, $p<.001$) and between voiced and aspirated voiceless (mean difference 121.73, $p<.001$). Comparison of this with ENS results will be covered in section 5.3.2 of this chapter.

For the present account, however, given the overwhelming difference between the voiced and voiceless results, we divide our detailed coverage by analysing and discussing the results for the voiced and voiceless stops separately.

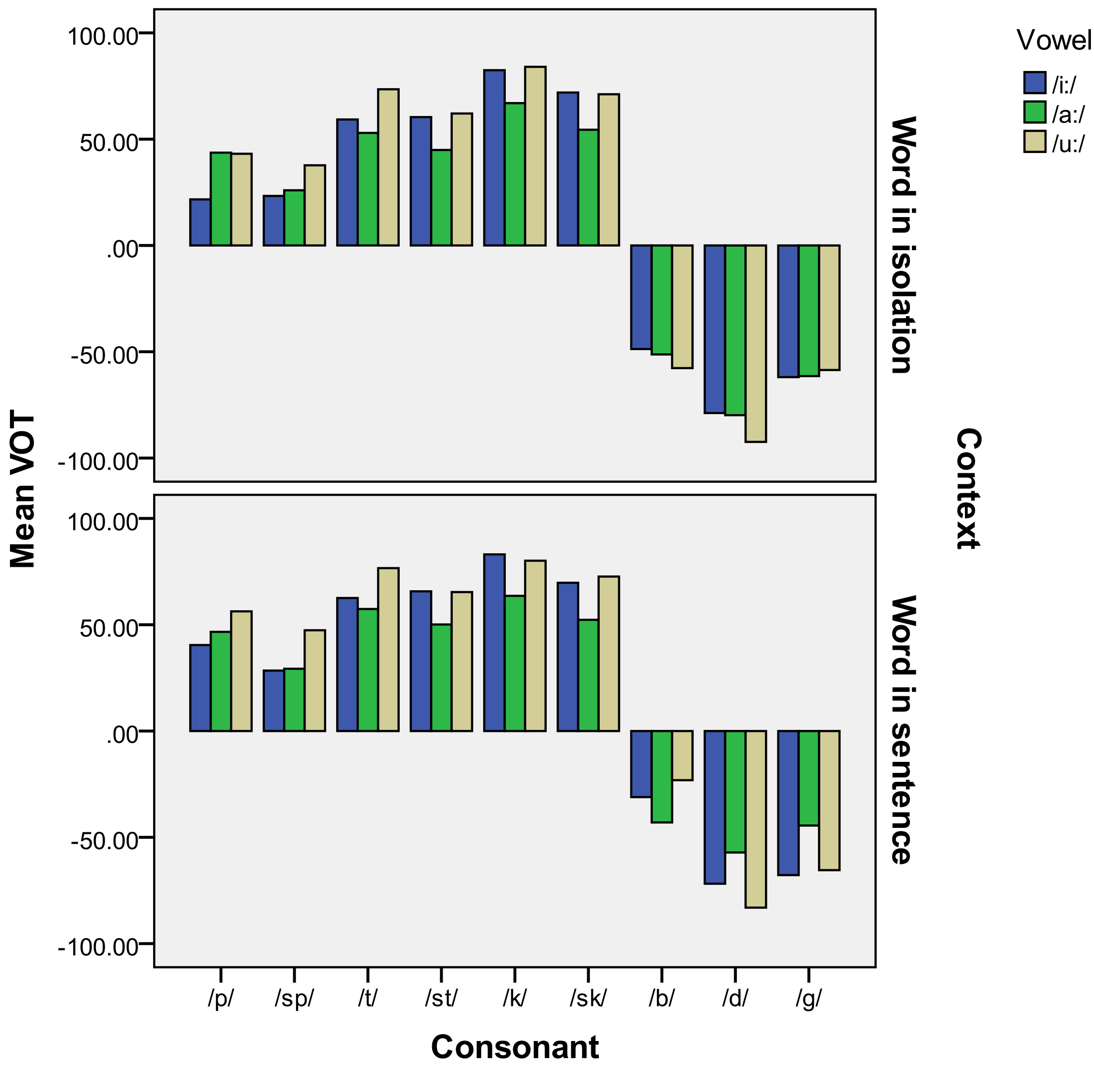


Figure 5.1: Production by learners of word initial stops.

5.2.1 Learner production of voiceless stops

We performed an ANOVA analysis for the voiceless stops separately (Table 5.2). This shows strong main effects on VOT not only of place of articulation, but also of following vowel and aspiration (all p<.001), together with interactive effects of POA and vowel, and of vowel by aspiration, and of all the three factors combined (F=18.31, p<.001). Context creates no overall effect, just a small interactive effect (p=.031) with place of articulation. We shall now proceed to examine the effect of each key factor in turn.

Table 5.2:Overall ANOVA for learner production data: voiceless stops.

Effect	F	P	Partial eta squared
Place	66.24	<.001	.688
Vowel	62.89	<.001	.677
Aspiration	29.13	<.001	.493
Context	1.57	.220	.050
Place by Vowel	18.98	<.001	.387
Vowel by Aspiration	9.63	<.001	.243
Place by Context	4.20	.031	.123
Place by Aspiration	1.04	.321	.033
Vowel by Context	.428	.654	.014
Aspiration by Context	.217	.646	.007
Vowel by Place by Aspiration	4.69	.006	.135
Context by Vowel by Aspiration	1.74	.183	.055
Context by Place by Aspiration	1.15	.311	.037
Context by Vowel by Place	.480	.695	.016
Context by Vowel by Place by Aspiration	.903	.464	.029

5.2.1.1 The effect of POA on VOT of learner voiceless stops

As we see in Figure 5.2, the voiceless stops, disregarding aspiration, vowel and context, are produced with increasing VOT front to back. All places are significantly different from each other ($p<.001$), and descriptively the increase is greater between labial and coronal than between coronal and dorsal. This pattern therefore fits the universal expectation for the trend of differences in VOT across POAs which we found also evidenced for ENS and ANS voiceless stops (chapter 4). As we shall discuss in 5.3, the precise VOT values differ from those of ENS more than from those of ANS. It is noticeable, however, that they produce /p/ with VOT in an unremarkable location relative to the VOT of the other stops, despite it being an entirely new sound for them in English. Thus, perhaps they show the impact of a language universal POA-VOT relationship.

The VOT values obtained in our study are, however, higher, so more native-like, than those obtained for example by Flege and Port (1981), even for their learner group that had lived the longest in the US: means for voiceless aspirated stops front to back 42, 64, 77 msec in our study (Figure 5.3), but only 21, 30, 47 msec in theirs.

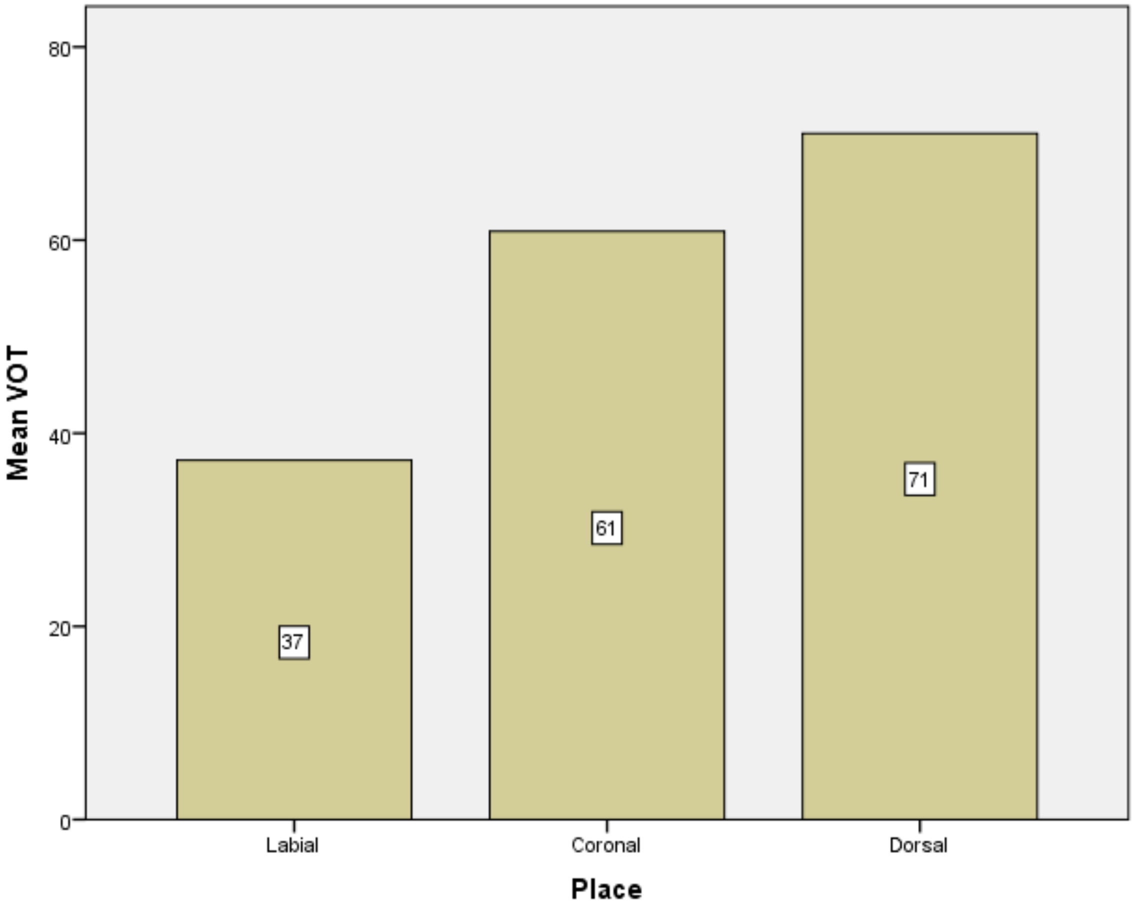


Figure 5.2: Production by learners of word initial voiceless stops (regardless of aspiration, context and vowel).

5.2.1.2 The effect of aspiration on VOT of learner voiceless stops

As we saw in Table 5.2, there is a strong overall effect of aspiration on VOT, with unaspirated voiceless stops overall produced 9 msec lower in mean VOT than aspirated (mean VOTs respectively 52 and 61 msec). Thus, although the learners on average produced long-lag VOTs for the unaspirated stops as well as the aspirated ones, those values produced for the unaspirated were significantly shorter than the ones produced for the aspirated stops.

In order to understand the effects more clearly we may inspect Figure 5.3 which shows a slightly simplified picture, since it omits vowel (as well as context) effects. However, it clearly shows how our learners produced VOTs which were systematically greater for aspirated than for unaspirated voiceless stops at each POA. As we have seen (Table 5.2), there is no significant place by aspiration interaction effect, however, meaning that the size of the difference between aspirated and unaspirated stops was not significantly different between different POAs.

In detail, as Table 5.1 shows, the mean VOT value of [p^h] was aspirated with mean VOT of 42 msec (range= -130 msec to 141 msec, sd=47). Thus, the learners' VOTs were shorter than those of ENS and some learners even produced negative VOTs (prevoicing), similar to their Arabic L1 voiced labial. The unaspirated [p] was also produced with slight aspiration but with lower VOT values than [p^h] with mean 32 msec (range=-48 msec to 106 msec, sd=24). The learners therefore show no sign of systematically marking any aspiration contrast in terms of long lag versus short lag production, although they do make a statistically significant VOT contrast within the aspirated/long lag range. Both means were in the long lag range (+30 msec VOT) but learners produced the labial stop in individual utterances across all three different VOT ranges: negative, short-lag and long-lag.

With respect to the voiceless coronal /t/, the learners produced [t^h] with an average VOT of 64 msec (range=22 msec to 143 msec, sd=22) and produced [t] with VOT of 58 msec (range=24 msec to 138 msec, sd=18). Both stops [t^h] and [t] were again produced with aspiration but a significant difference.

The dorsal sound was quite similar to the previous ones, as learners produced [k^h] with mean VOT of 77 msec (range= 34 msec to 132 msec, sd=18) and produced [k] with mean VOT of 65 msec (range= 26 msec to 111 msec, sd=19). Thus, again both stops have VOT in the long lag range but [k^h] was produced with 12 msec higher VOT than [k] (a significant difference). As seen in 5.2.1.1, VOT values are also greater in coronals than labials and greater in dorsals than coronals, which shows the pattern that VOT increases as the place of articulation moves further back, as seen with ENS.

In sum, we may regard the significant difference found between the learners' VOTs of the aspirated and unaspirated English stops as an indication that the learners have learnt a distinction not present in their L1. Whether, in terms of the SLM, a new category has been created, and if so, which one is the new one (the aspirated or unaspirated or both) we will leave to be discussed later, as also the issue of why neither of them are in the ENS VOT locations.

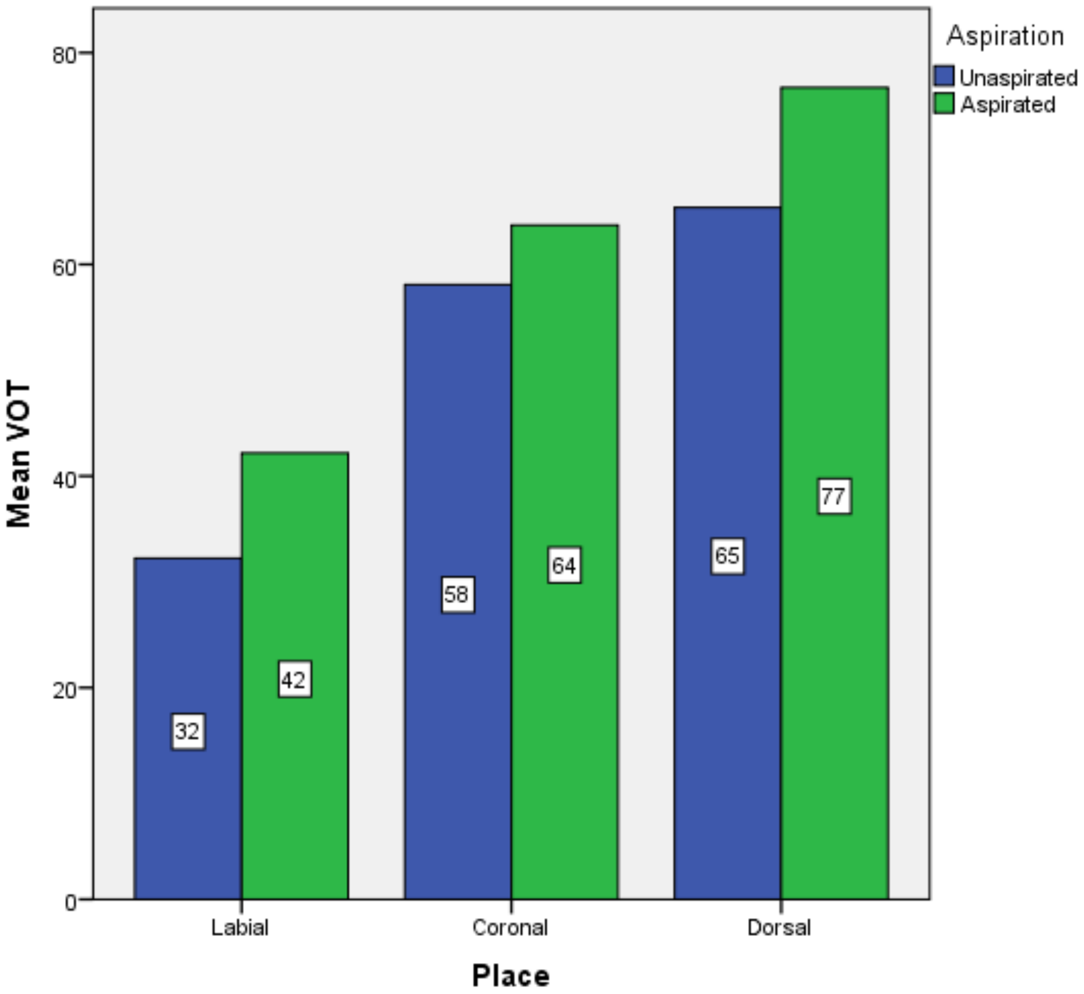


Figure 5.3: Production by learners of word initial voiceless stops: place in relation to aspiration.

As we have seen, the interaction effect of place with aspiration (disregarding context and vowel) was not significant. This means that, although mean VOT differences between aspirated and unaspirated were not descriptively identical at each POA, they did not differ significantly beyond the variation that might arise due to sampling. By contrast, the interaction effect of place with vowel on mean VOT (disregarding context and aspiration) was highly significant, as we can be seen in Figure 5.4. In particular, the effect of /i:/ is not the same at each POA, while that of the other vowels is more uniform across POAs. This will be illuminated more clearly in the next section.

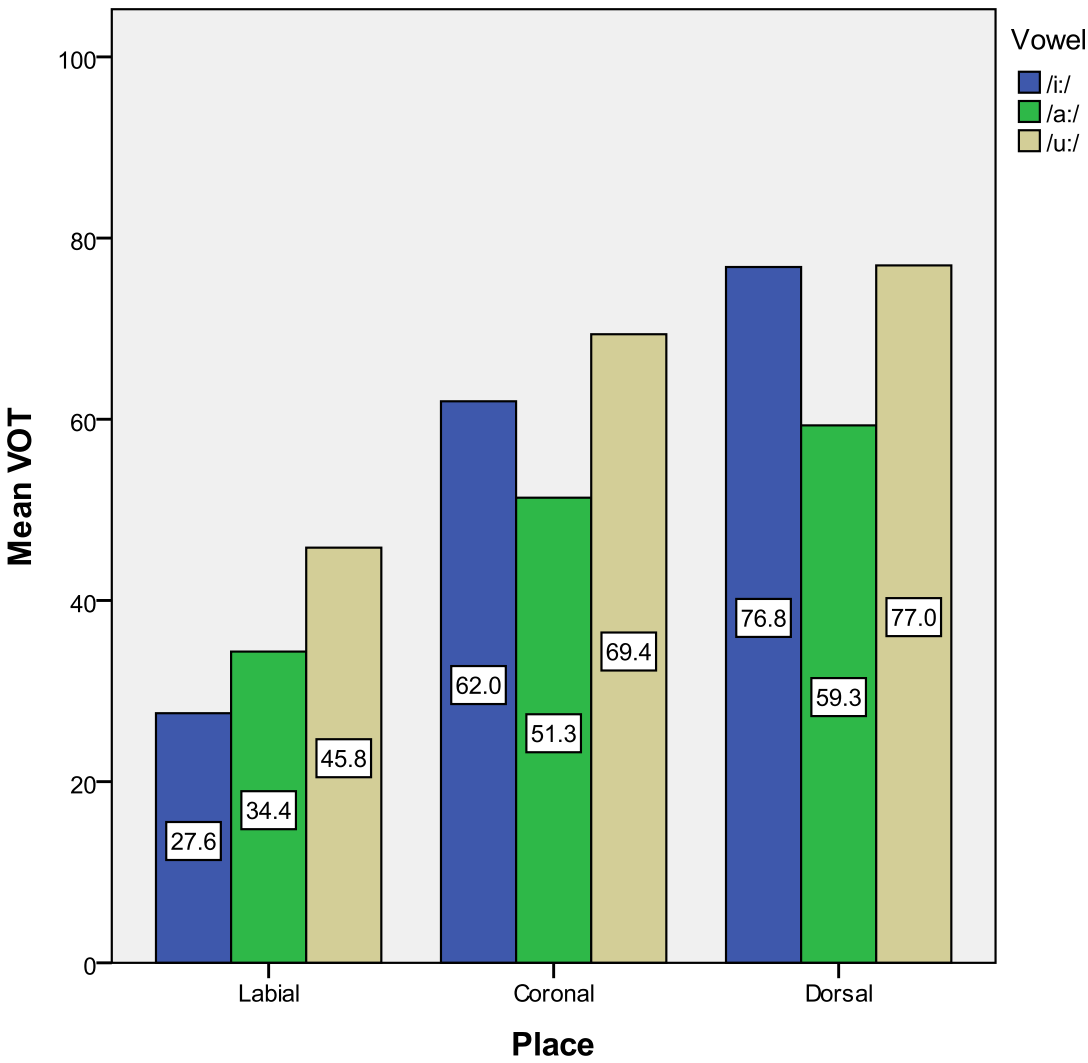


Figure 5.4: Production by learners of word initial voiceless stops: place in relation to following vowel.

5.2.1.3 The effect of vowel on VOT of learner voiceless stops

As we saw in Table 5.2, the effect of vowel is highly significant overall. Indeed, the effect size (.677) is greater than that of aspiration (.439) and second only to that of POA (.688).

The overall pattern for the vowels, regardless of place of articulation (or aspiration or context), is that mean VOT is longest for a voiceless stop before the high rounded back vowel, second longest for the high unrounded front vowel, and shortest for the low back vowel (see Figure 5.5). These overall differences between vowel effects on

VOT, disregarding place (and aspiration and context), are all significant: /iː - ɑː/ p=.002; /iː - uː/ p<.001; /ɑː - uː/ p<.001. As we also found earlier with ENS results less strongly (4.2.1.4), the low vowel /ɑː/ yields the lowest VOT of preceding voiceless stops in English. The high vowels [iː uː] produced higher VOTs, especially the back vowel /uː/.

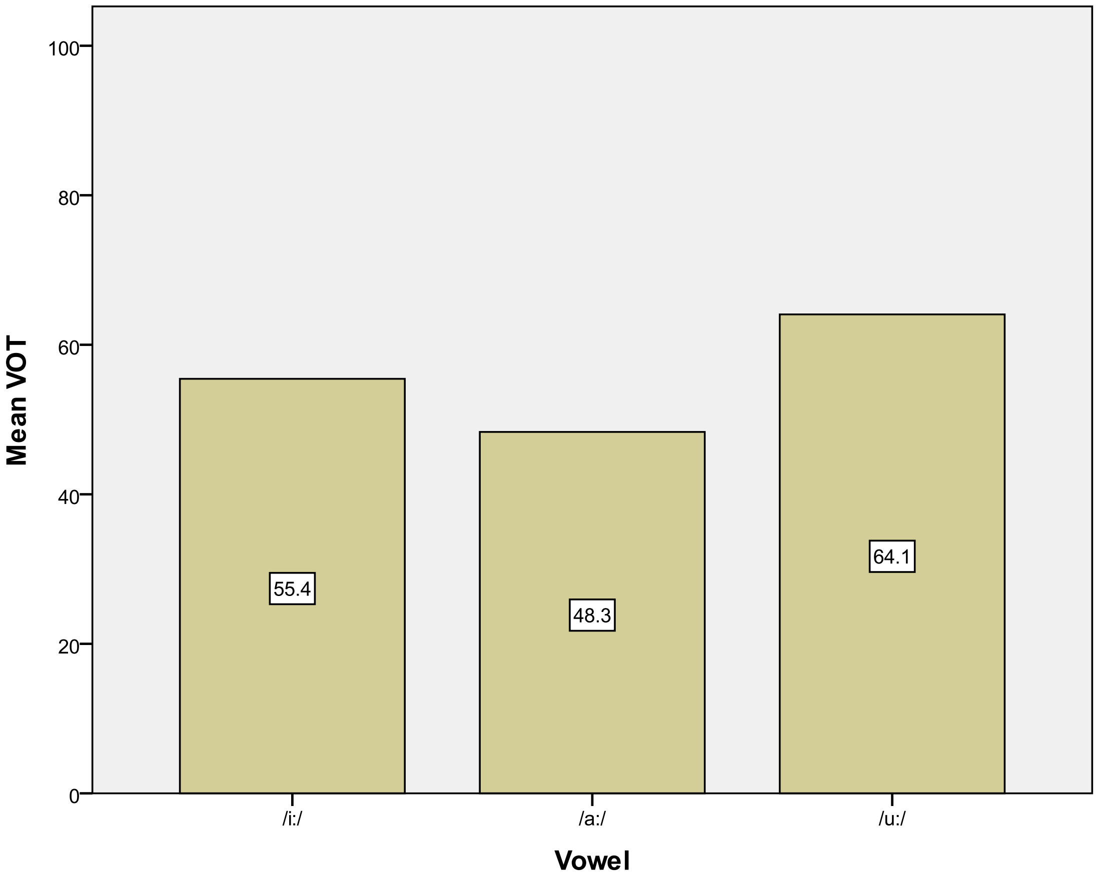


Figure 5.5: Production by learners of word initial voiceless stops: effect on stop VOT of following vowel (regardless of POA, context or aspiration)

Since the three-way interaction effect on VOT of place combined with vowel and aspiration is also highly significant (p=.006), we next consider Figure 5.6.

It is at once apparent that, while VOT increases front to back across places of articulation regardless of which of the three vowels follows, the precise pattern of increase differs between vowels. For the labial [p], VOT is successively longer depending on whether /iː/, /ɑː/ or /uː/ follows. VOT before the low vowel /ɑː/ increases less strongly from [p] to [t] and [k] than does VOT before /iː/ or /uː/, yielding the overall pattern of Figure 5.6 where /ɑː/ is lowest.

Table 5.3: VOT differences between voiceless stops produced by learners in relation to quality of the following vowel.

Vowel	Main effect between place		Post hoc paired comparisons (Bonferroni)					
			/p - t/		/p - k/		/t - k/	
	F	p	Mean diff.	p	Mean diff.	p	Mean diff.	p
/iː/	107.90	<.001	-34.43	<.001	-49.25	<.001	-14.82	<.001
/ɑː/	35.85	<.001	-16.99	<.001	-24.97	<.001	-7.98	<.001
/uː/	29.98	<.001	-23.56	<.001	-31.15	<.001	-7.60	.003

However, the post hoc comparisons (Table 5.3) show that the learners are distinguishing all the places of articulation of the voiceless stops, whatever the following vowel is.

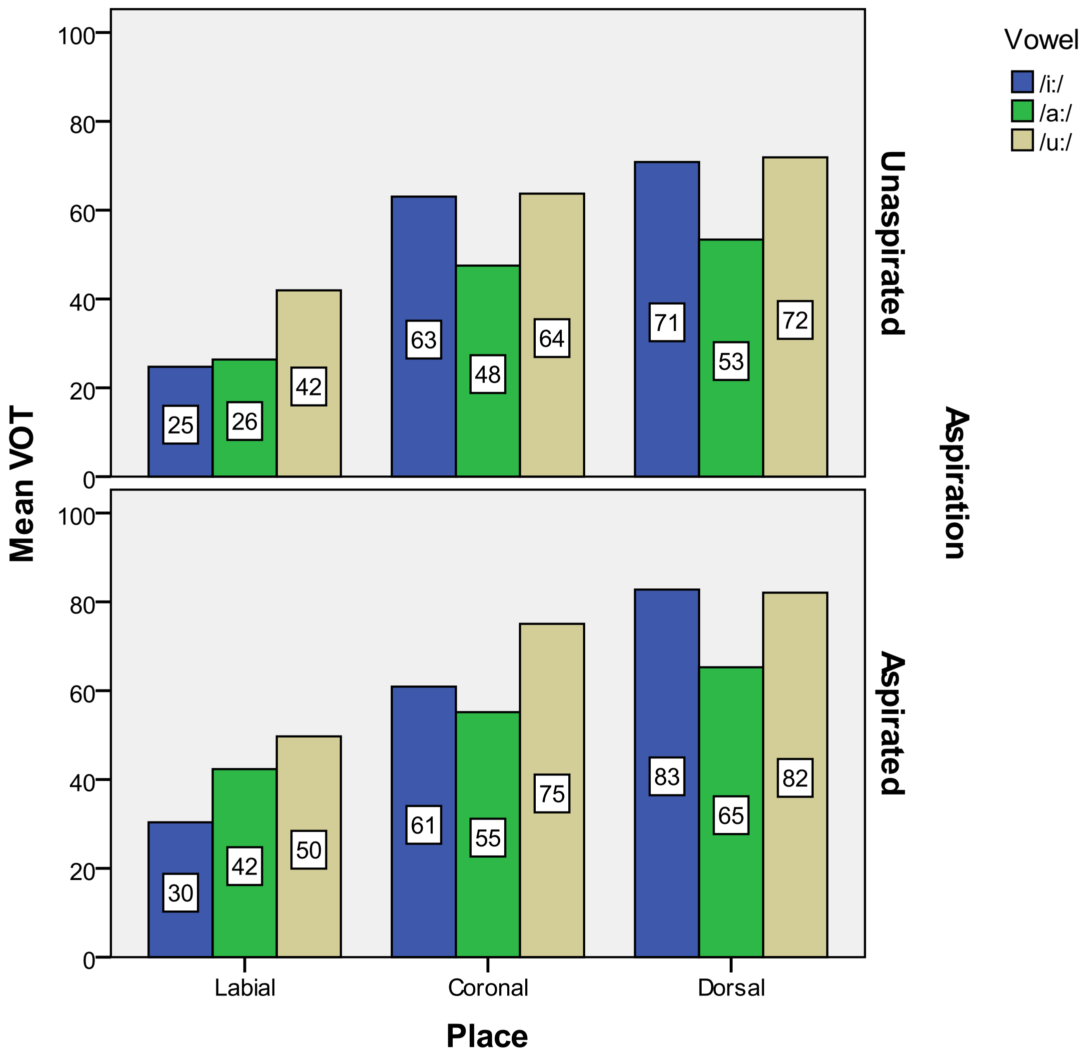


Figure 5.6: Production by learners of word initial voiceless stops: place, aspiration and following vowel .

We have already seen that VOT of aspirated stops was generally higher than that of unaspirated, and that is evidenced here in all instances except coronal /tʰi: - ti:/. However, we see here that, beyond this, there is a slight difference in the way VOT progresses front to back with each vowel depending on aspiration. Although the general pattern of Figure 5.5 is repeated for both levels of aspiration in Figure 5.6, and indeed the pattern before /u:/ is very similar, regardless of aspiration, there are subtle differences where /i:/ and /ɑ:/ are involved. This is what created the significant vowel by place by aspiration effect (Table 5.2).

5.2.1.4 The effect of context on VOT of learner voiceless stops

The learners do not differentiate globally between saying English stops in isolation or in sentences, as there was no main effect of context on voiceless stop VOT produced in two different contexts. Although, the average VOT values in the two tables below indicate that learners produced higher VOT ranges in sentences than in isolated words (overall means 58 vs 55 msec), there was no significant difference between them. Descriptively however this result is contrary to both ENS and ANS for whom isolated words generated longer VOTs.

Table 5.4: VOT values of L2 learners of English (not differentiating following vowels): in words in isolation.

Stop	POA	VOT			
		Maximum	Minimum	Mean	Standard Deviation
Voiceless aspirated	Labial	114.33	-129.97	36.14	52.19
	Coronal	121.00	22.33	61.89	21.33
	Dorsal	131.67	33.67	77.81	19.66
Voiceless unaspirated	Labial	88.33	-48.00	29.19	25.03
	Coronal	103.62	24.33	55.79	16.88
	Dorsal	111.00	26.67	65.83	17.92
Voiced	Labial	151.20	-167.67	-52.53	60.84
	Coronal	153.51	-183.33	-83.69	59.27
	Dorsal	110.52	-188.73	-60.65	53.25

There was however a small interactive effect ($p=.031$) of context with place of articulation. As we can see from Figure 5.7, the progression of VOT /p<t<k/ (regardless of aspiration and following vowel) occurs in both word and sentence contexts but is steeper in the former than the latter, i.e. it starts lower and ends higher. Therefore, this has the effect that the overall finding of longer VOT in sentences than in words in isolation is supported only by labials and coronals, but not in dorsals, as seen below.

Table 5.5: VOT values of learners of English (not differentiating following vowels): in sentences.

Stop	POA	VOT			
		Maximum	Minimum	Mean	Standard Deviation
Voiceless aspirated	Labial	140.82	-108.11	47.80	41.17
	Coronal	143.40	32.67	65.54	22.25
	Dorsal	111.36	42.33	75.60	16.51
Voiceless unaspirated	Labial	106.28	-14.32	35.04	23.04
	Coronal	138.32	28.99	60.40	18.65
	Dorsal	104.24	26.32	64.91	19.27
Voiced	Labial	82.17	-191.74	-32.40	54.02
	Coronal	26.99	-147.47	-70.68	46.26
	Dorsal	22.67	-165.27	-59.22	43.60

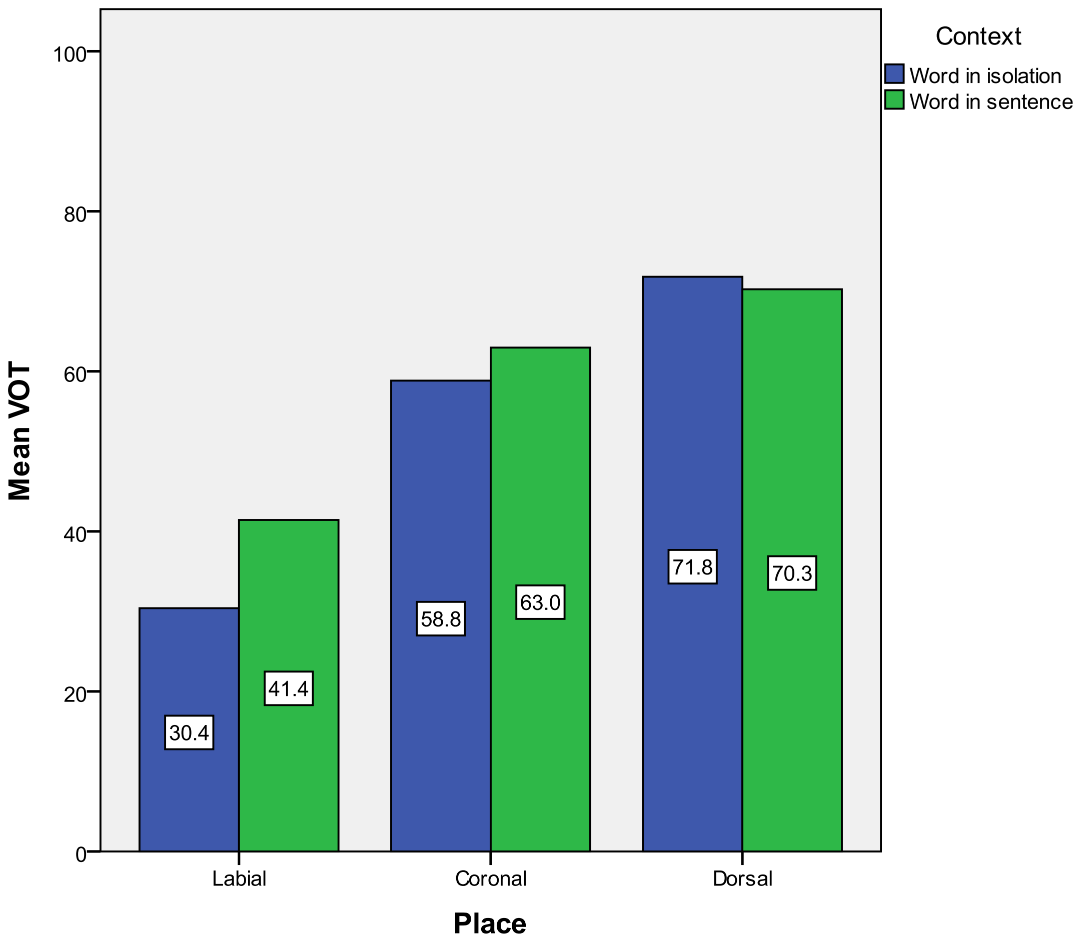


Figure 5.7: Production by learners of word initial voiceless stops: place of articulation in relation to context of production.

5.2.2 Learner production of voiced stops

Results for the voiced stops of the learners in this study are similar to those found in Flege and Port (1981) and Alghamdi (1990), as they reveal that the majority of the VOT values produced by Saudi Arabic learners of English for the voiced stops are pre-voiced, similar to those of ANS. We therefore conclude that the L2 learners in our study were greatly affected by the VOT norms of their L1 in producing the English voiced stops. As we will discuss later, this could suggest that, in terms of the SLM, an 'equivalence classification' (H5 of the SLM) has been created for the Arabic and English voiced stops by the learners.

For instance, the learners produced the English labial /b/ with a mean VOT of -42 msec (range= -192 msec to 151 msec, sd= 58). The English coronal by contrast recorded the most negative VOT values among the voiced stops (which will be discussed later), as it was produced with a mean VOT of -77 msec (range= -183 msec to 154 msec, sd= 53). The voiced dorsal was produced by the learners with a mean VOT value of -60 msec (range= -189 msec to 111 msec, sd= 49) (see Table 5.1 above).

A repeated measures analysis of variance on the VOT of the voiced stops of L2 learners (Table 5.6) shows that, overall, the place of articulation difference was significant but the following vowel effect and word context were non-significant. Therefore, learners were not making a clear overall VOT distinction between contexts or vowels, but they do distinguish in some way places of articulation ($F=26.17$, $p<.001$), with a moderate effect (partial eta squared =.466). However, they do so in a way that differs somewhat depending on the following vowel ($F=5.264$, $p=.001$) and that further varies dependent on the word context ($F=3.080$, $p=.019$).

Table 5.6: Overall ANOVA for learner production data: voiced stops.

Effect	F	p	Partial eta squared
Place	26.17	<.001	.466
Vowel	1.41	.251	.045
Context	.672	.419	.002
Place by Vowel	5.26	.001	.149
Place by Context	1.87	.163	.059
Vowel by Context	1.10	.340	.035
Place by Vowel by Context	3.08	.019	.093

5.2.2.1 The effect of POA on VOT of learner voiced stops

Far and away the strongest effect was for place of articulation alone (Figure 5.8). Follow up tests confirm that all three places of articulation differ significantly from each other, disregarding vowel and context variation (Table 5.7). In other words, the learners are showing an overall negative VOT for all three voiced stops in order from /b/ the least through /g/ to /d/ the most.

This means that the general universal pattern of increasing VOT from front to back of the mouth b < d < g found in the voiceless stops, and for ENS and some of the ANS studies (i.e. Radwan, 1996; Khattab, 2002) in the voiced ones too, was not found with learner voiced stops. Whether we interpret the expected trend for pre-voiced stops as being of increasing pre-voicing across POA front to back, or of increasing positive (i.e. less negative) VOT front to back, in neither case does the result fit the expectation. This appears to be a new finding since we know of no other studies of VOT of voiced stops in Saudi learners of English.

The pattern found was of increasing pre-voicing in VOT labial< dorsal< coronal b < g < d. This order is similar to what was found with voiced stops of their Arabic counterparts (ANS), albeit for the ANS the pattern was very weak and non-significant (4.3.2.1).

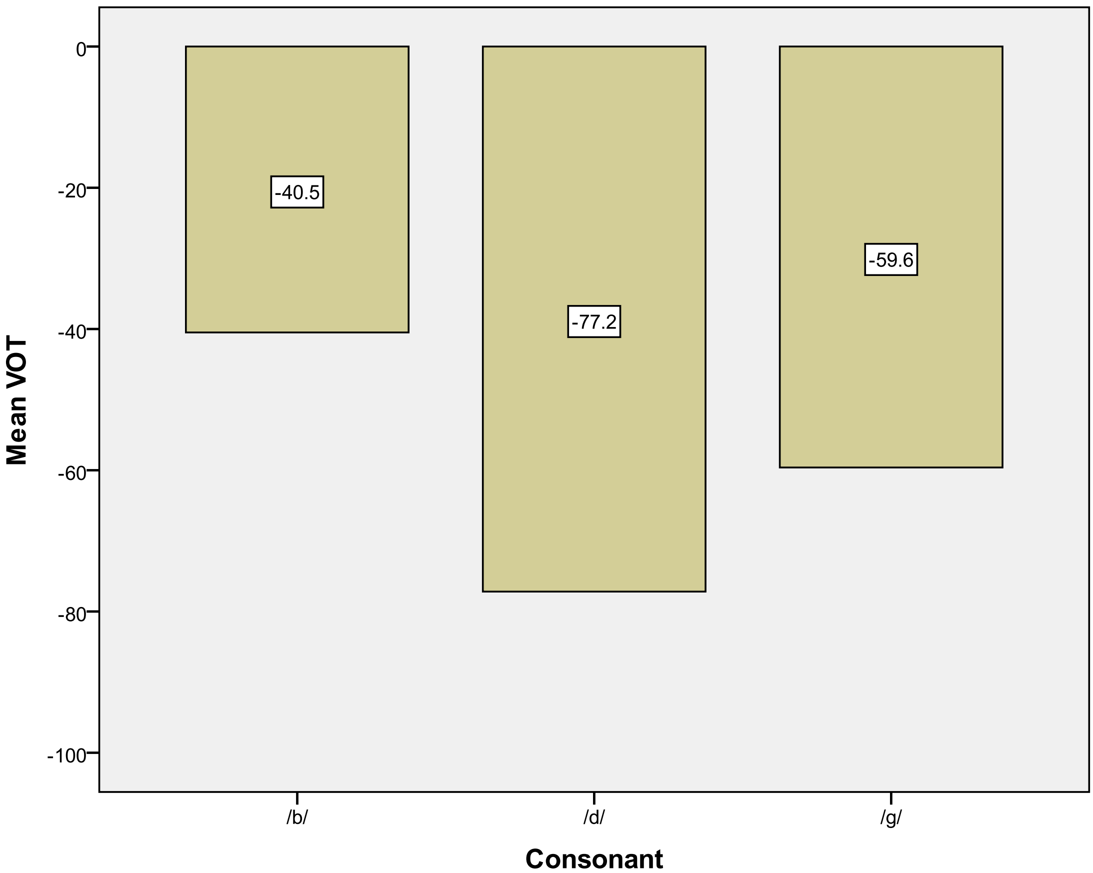


Figure 5.8: Production by learners of word initial voiced stops, by place.

Table 5.7: differences between places of articulation voiced stops produced by learners.

Main effect between places		Post hoc paired comparisons (Bonferroni)					
		/b - d/		/b - g /		/d - g/	
F	p	Mean diff.	p	Mean diff.	p	Mean diff.	p
26.17	<.001	36.70	<.001	19.12	.006	-17.57	<.001

5.2.2.2 The effect of vowel on VOT of learner voiced stops

The results of the analysis of variance confirm that difference of vowel alone, regardless of context and place of articulation, does not lead to a significant overall difference in learner voiced stop VOT (means /iː/ -60, /ɑː/ -56 and /uː/ -63 msec). These findings are different from a number of studies like Klatt (1975), Weismer (1979), Port (1979), Rochet & Fei (1991), Schmidt (1996) Johnson & Babel (2010), Chao et al (2006), and Iverson et al (2008) who all found that there was a strong effect of the vowel on the acquisition of L2 sounds in general. But this study is similar to Lisker and Abramson (1967) who found that there was no major effect on learner VOT of the following vowel.

As Table 5.6 shows, however, there was a significant effect of vowel on VOT of voiced stop that differed between POA of the stop. From Figure 5.9, we can see that the pattern of VOTs of places of the articulation (pre-voicing b<g<d) differs somewhat depending on the following vowel, and whether the word is produced in isolation or in a sentence, yielding the significant interaction effects noted above. Difference of vowel seems to have an effect on how clearly the differences between the three places of articulation appear.

We can see descriptively from Figure 5.9 that in isolated words the overall pattern of differences between places seen in Figure 5.8 is maintained for each vowel separately. Just there is a tendency for the gap in VOT between /b/ and /g/ to become smaller as we progress from /i:/ to /ɑ:/ to /u:/, and indeed with /ɑ:/ and /u:/ the /b - g/ difference is not significant (Table 5.8).

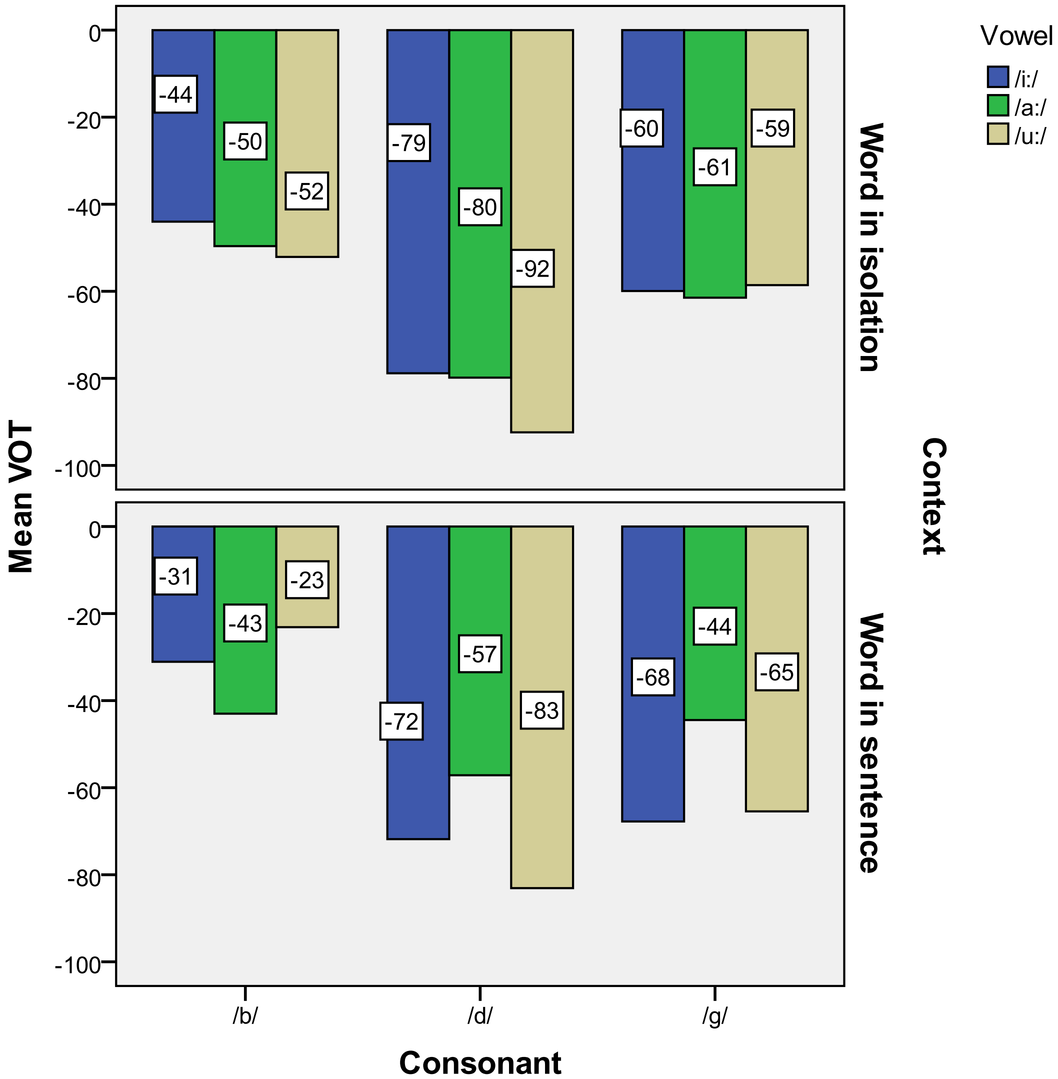


Figure 5.9:Production by learners of word initial voiced stops by place, vowel and context.

Where words are produced in a sentence context there are stronger vowel effects. Although descriptively the /d<g<b/ hierarchy of VOT is preserved, in some instances there are close to zero differences. Indeed /d/ and /g/ are not significantly different in VOT with any of the three vowels following, and with /ɑː/ following, none of the three places of articulation differ in VOT.

Table 5.8: VOT differences between places of articulation of voiced stops produced by learners.

Context	Vowel	Main effect between places		Post hoc paired comparisons (Bonferroni)					
				/b - d/		/b - g/		/d - g/	
		F	p	Mean diff.	p	Mean diff.	p	Mean diff.	p
Word	/iː/	13.67	<.001	34.83	<.001	15.94	.043	18.89	.016
	/ɑː/	8.57	.001	30.22	<.001	11.85	.530	-18.37	.027
	/uː/	8.66	<.001	40.29	.004	6.48	1.00	-33.82	.005
Sentence	/iː/	10.29	<.001	40.75	.001	36.69	.017	-4.06	1.00
	/ɑː/	1.82	.172	14.14	.355	1.46	1.00	-12.68	.347
	/uː/	23.29	<.001	59.95	<.001	42.32	.001	-17.63	.076

Overall, then, we can see that in fact our learners are not making significant VOT distinctions between the three places of articulation in all instances, as they differentiated them in only 11 out of the total 18 voiced CV combinations (3 vowels x 3 POA x 2 contexts = 18) in Table 5.8. Particularly pairs /b - g/ and /d - g/ are least often distinct (only in half the CV combinations covered) while /b - d/ are most often significantly differentiated (all but once highly significantly). The distinctions are less clearly made in sentence contexts (only in 4 out of 9) than in isolated word contexts (7 out of 9), perhaps because speakers know that the hearer can depend on other, grammatical and semantic, contextual clues to decode words correctly in sentences. In terms of following vowels, VOT distinctions of place were clearest before /iː/ (5 out of 6 were significant), then /uː/ (4 out of 6) and least /ɑː/ (2 out of 6). Indeed, in sentence contexts there is no significant distinction in VOT between /bɑː dɑː gɑː/ at all.

The learners differ from ANS in some major respects on the voiced stops: for ANS the overall VOT differences between POAs are not significant while overall vowel differences are, while the reverse is true of the learners (see further 5.3.2). Nevertheless there are some similarities in that there are significant interaction effects of POA with vowel in both. That is to say, both show evidence of POA differences in VOT being significant with some following vowels but not others (e.g. /b/ and /g/ differ when followed with /iː/ but not with /ɑː/).

5.2.2.3 The effect of context on VOT of learner voiced stops

There was no significant effect of context other than the detailed interaction effect just described above. However, as Tables 5.4 and 5.5 show, descriptively the VOTs exhibit the phenomenon of being more negative in words in isolation than in sentences. This matches the same unexpected finding for learner voiceless stops where VOT was lower (less positive) in isolated words than in sentences. This therefore makes the learner production of voiced stops in sentences apparently slightly more ENS-like than that in words in isolation (see further 5.2.2.4).

5.2.2.4 Positive VOT of learner voiced stops

In order to assist the comparisons that we will discuss in more detail in 5.3 between learner English VOTs and ENS VOTs, we further conducted an analysis focused solely on learner responses which evidenced positive VOT values for voiced stops. In doing this we followed the standard method (Lisker and Abramson, 1964): the non-negative and negative VOT responses of the stops were calculated separately for each participant, and summarised in a variety of ways, separately for POA and context.

We have already seen from the maximum values in Tables 5.4 and 5.5 that, in both isolated word and sentence contexts, some learners on some occasions produced the voiced sounds with positive VOT values, which suggests that some learners perhaps managed to produce the English voiced stops in the short-lag region, like the ENS.

Table 5.9: VOT of English voiced stops produced by the learners with positive VOT only, by POA and context regardless of vowel (omitting all negative VOTs).

Context	POA	VOT			
		Minimum	Maximum	Mean	Std. Deviation
Word in isolation	Labial	.00	133.29	17.64	28.34
	Coronal	.00	111.88	32.81	38.48
	Dorsal	.00	88.91	32.68	25.34
Word in sentence	Labial	.00	103.50	30.46	21.15
	Coronal	18.81	30.45	24.20	3.59
	Dorsal	7.48	39.00	27.64	8.44

Table 5.10: VOT of instances of English voiced stops produced by the learners with pre-voicing (negative VOT) only, by POA and context regardless of vowel.

Context	POA	VOT			
		Minimum	Maximum	Mean	Std. Deviation
Word in isolation	Labial	-166.75	-11.00	-82.74	38.18
	Coronal	-153.78	-34.24	-99.68	30.33
	Dorsal	-131.12	-36.63	-84.02	24.79
Word in sentence	Labial	-145.01	-20.75	-69.66	30.41
	Coronal	-129.54	-21.70	-85.11	28.55
	Dorsal	-161.13	-30.50	-82.36	29.87

Table 5.9 shows the average zero or positive VOTs of the learner voiced stops separately from the mean negative VOTs in Table 5.10.

From the maximum values in Table 5.9 we see that the learners produced the English voiced stops, most prominently /b/, not just with unaspirated short lag but sometimes with aspiration (long lag VOT). The ENS however also did this on occasion (see Table 4.7), though their means were short lag. In the non-negative VOTs there is no clear context effect, just as there was not for learners overall. How similar the values are to ENS values will be examined further in 5.3.1: at this point the learners seem to be descriptively just above, or in the higher end of, short lag.

As Tables 5.9 and 5.10 show, the learners produced some tokens of the English voiced stops /b d g/ with pre-voicing and others with positive VOT. It is not possible to say from that table, however, how many individual participants produced voiced stops with pre-voicing or indeed how many were able to produce all the three repetitions with positive VOT, as they all produced three tokens and some of them produced some tokens with positive value and the other tokens pre-voiced. However, there were in fact some learners who produced all three repetitions of the English voiced stops in a specific vowel context without pre-voicing (see Table 5.11).

Table 5.11: Number of learners out of 31 producing English voiced stops with positive VOT in all three repetitions for a condition, by place, vowel and context.

Context	POA	/i/	/a/	/u/	Total
Word in isolation	Labial	8	5	8	21
	Coronal	2	2	1	5
	Dorsal	4	2	2	8
Word in sentence	Labial	10	10	10	30
	Coronal	2	2	2	6
	Dorsal	0	4	0	4

In table 5.11 it is apparent that more learners systematically produced positive VOTs for the labials than the other two stops. They also overall produced more VOTs in the positive range in sentences rather than isolated words, though this was not systematically supported by the dorsals. Once again, of course, we must be aware that some of the utterances taken into account here (especially for words in isolation in general and for /b/ in sentences) were in the long rather than short lag region so not fully nativelike.

In addition to participants who produced all repetitions of the target English sounds with pre-voicing, or all with positive VOT, some participants produced a specific stop with positive VOT in some of the repetitions but not others. Table 5.12 shows how many tokens the learners produced with positive VOT and their percentage out of the total trials.

Table 5.12: Percent of individual trials/repetitions when English voiced stops were produced by L2 learners with positive VOT, by POA and context.

Context	POA	Total tokens	Percent (out of 279)
Word in isolation	Labial	93	33.33 %
	Coronal	35	12.54 %
	Dorsal	57	20.43 %
Word in sentence	Labial	117	41.94 %
	Coronal	48	17.20 %
	Dorsal	61	21.86 %

For learners, there were 279 repetitions (31 participants*3 vowels*3 repetitions) for each of the stops, in each context (word and sentence), so a total of 558 trials in the two contexts, and 270 repetitions by ENS and ANS (30 participants*3 vowels*3 repetitions) for each stop in each context.

The learners again showed more positive VOT in producing labials, followed by dorsals. As we found earlier, the learners' performance also exhibited more positive VOT when pronouncing the voiced stops in a sentence context than when saying them in individual words. Thus the greatest amount of positive VOT for voiced stops in terms of individual trials was recorded for labials in sentence contexts, although, as usual, we cannot simply say that /b/ in words in sentences has been the best learnt, given that some of the trials with positive VOT are in the long lag rather than short lag region.

We have seen above that learner VOT for voiced stops is more often positive in sentence than in word contexts. This could be interpreted as follows. Arguably our learners, living as they do immersed in a target context, in real life are usually hearing and using English in normal connected speech, not isolated words (which would be more common in an artificial classroom environment). Therefore, they get more nativelike input/exposure for sounds in words in sentence contexts than in isolation. As postulate 1 of the SLM recognises, good quality input is crucial to achieving nativelike performance, and this continues to have an effect on learners even in adulthood (see 2.11). Therefore, it is perhaps for this reason that learner utterances in sentence contexts exhibit more instances of positive VOT, so to an extent are more ENS-like.

Also, surprisingly, the learners showed more positive VOT in producing labials than the other voiced stops, as evidenced both in Table 5.11 and 5.12. Although not all the instances are short lag, so nativelike, this might indicate that the chances of learning are greater in labials, although the reason for the labial taking the lead here remains unclear, as does the overall finding that /b/ has the lowest rather than highest negative VOT (Figure 5.8). Possibly it could be related in some way to the fact that, in L2 English, the voiceless counterpart /p/ of their /b/ is a new sound for them, and although on average they produce it with a very different, positive long lag, VOT value, there are lingering instances where their production of tokens of /b/ are not always clearly distinguished from those of /p/.

We conclude with a graph showing the mean VOT values of learners if we in fact limit the data solely to their individual responses for each condition which were non-negative and short lag (Figure 5.10). It is notable that, when only this subset of responses is considered, VOT is now longer rather than shorter in sentence contexts, so less like ENS in that respect. While the VOTs in word context present a steady increase in positive VOT by POA front to back as the corresponding ENS figures do, those in sentence context do not. Overall, these values are shorter than ENS ones for words in isolation, and longer than ENS ones for words in sentences except for /g/.

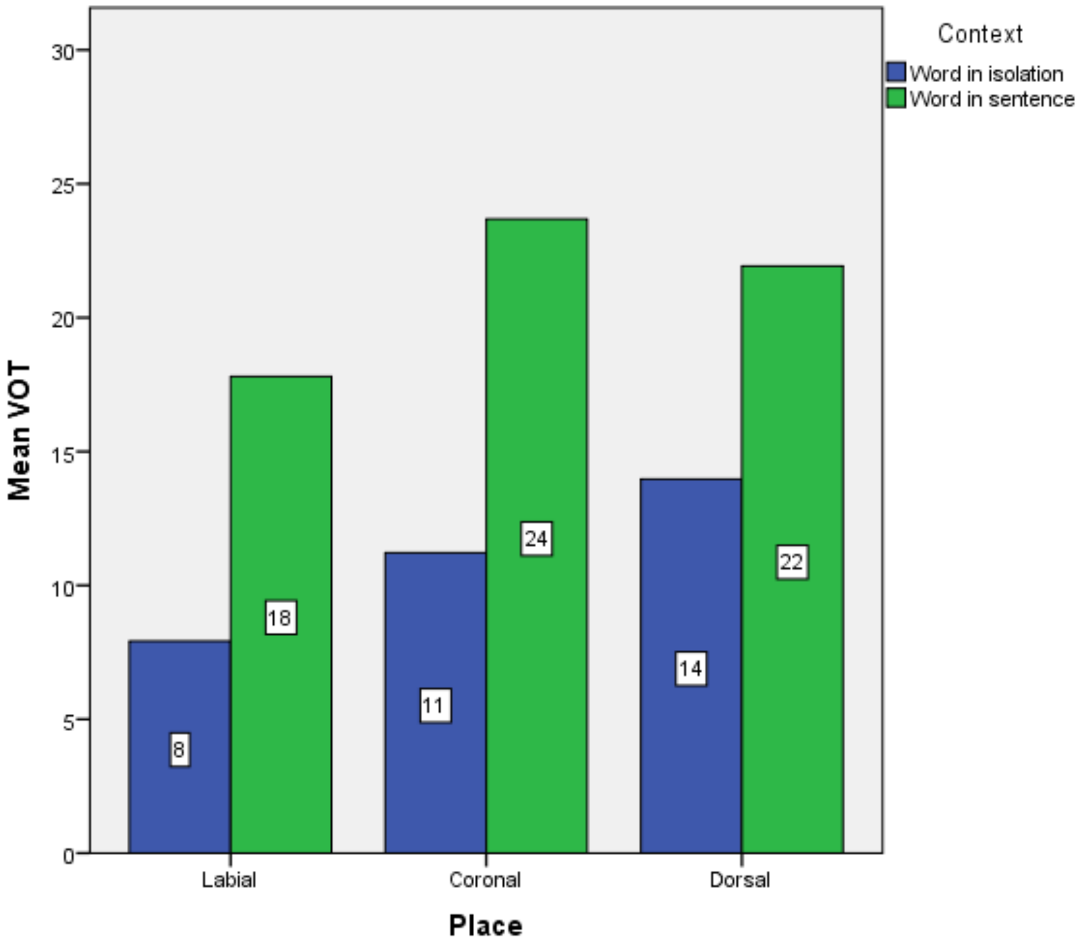


Figure 5.10: English voiced stop VOT tokens produced by the learners in the short lag region only, by POA and context regardless of vowel (omitting VOT tokens less than zero, and 30 msec or higher).

5.2.3 Summary

Our learners make an overwhelming distinction between voiced and voiceless stops: the former are produced with negative VOT (means approximately -20 to -80) and the latter with positive VOT (means in the region +20 to +80).

For voiceless stops they make VOT distinctions primarily for place of articulation, aspiration, and following vowel.

VOT is longer for aspirated than unaspirated voiceless stops, though for both mostly in the long-lag area, and also increases with place of articulation from front to back, following the universal expected pattern.

Vowel related differences in VOT of voiceless stops are widely made, but not in the same way at each place of voiceless stop articulation.

For voiced stops, learners distinguish primarily between the three places of articulation, most consistently between labial and coronal, which have the greatest difference from each other in VOT, with dorsal in between.

Place of articulation of voiced stops is most clearly distinct in VOT in isolated word contexts, and before high vowels /iː/ and /uː/ rather than low /ɑː/.

If we consider only learner responses for voiced stops that exhibit positive VOT, we find that they are more often positive in sentence contexts than isolated words but still fail to increase progressively across POAs front to back. If we consider only learner responses for voiced stops that exhibit positive short lag VOT, there is a better progression front to back, like ENS, but the precise ENS values are not matched.

5.3 Comparison between learners, ENS and ANS in their production of stop VOT

In this comparison we aim to bring together and supplement what we have already said about the ways in which learner production of VOT of stops is or is not in line with ENS production, or like that of ANS, for the corresponding stops.

This comparison shows whether the learners' have in any sense acquired L2 stop VOT and if there is any evidence of impact of exposure to L2 input on their VOT or if they are still affected by their L1 stops. It also identifies which aspects of stops are more 'difficult' for learners, as well as showing the similarities and differences between learners and monolingual Arabic and English groups.

In this section, we also revisit several of the predictions of our hypotheses derived from the SLM, to see if our learner data accords with those, and so supports the validity of the SLM as a model of acquisition of L2 sounds.

It makes sense first to again inspect the overall graphs for ENS and ANS production (Figures 4.1 and 4.9) in comparison with Figure 5.1. It is immediately obvious that the learner VOTs are descriptively more similar to the ANS ones in Arabic words than those of ENS in English words. For voiceless aspirated stops the learner VOTs are positive but at a similar low level to those of ANS rather than the high VOT level of ENS aspirated stops. For voiceless unaspirated stops, with no L1 counterpart, the learners are higher than ENS. For voiced stops, learner VOTs are negative, like ANS, not positive like those of ENS.

We now pursue the similarities and differences between learners and the two monolingual native speaker groups (ENS and ANS), following up two different interpretations of what 'similarity' can mean: similarity in terms of closeness of learner VOT to that of an NS group for the 'same' sound, and similarity in terms of how far learners make the same distinctions between sounds as NS do, using VOT differences, even if the actual VOT mean values are rather different in different groups (perhaps due to SLM 'deflection').

5.3.1 VOT differences between the three groups in each context by POA, voice/aspiration and following vowel

In order to ascertain precisely where learners' VOTs were or were not similar to ENS or ANS ones, we performed comparisons between groups for every specific condition (Tables 5.13. 5.14). Where only the learners and the ENS could be compared, we used the independent groups t test. Where all three groups could be compared, we used one-way ANOVA with Games-Howell post hoc paired comparisons. Although the English unaspirated voiceless stops in a sense have no counterpart in Arabic, we include comparisons of them with the Arabic aspirated stops because the SLM is interested in any L1 sound that might be identified by learners with an L2 sound. In these tables the highlighted results are those which most clearly evidence acquisition having occurred. These are instances where learners' VOT was on average not significantly different from that of ENS, and/or was significantly different from that of ANS.

Table 5.13: Group differences in VOT production: words in isolation (italicised comparisons are between English unaspirated and Arabic aspirated).

V	Place	Voicing/ Aspiration	CV:	Learner - ENS		Learner-ANS		ENS - ANS	
				Mean diff.	p	Mean diff.	p	Mean diff.	p
/i:/	labial	v-less asp.	/pi:/	-46.26	<.001				
		v-less unasp	/spi:/	8.88	.051				
	coronal	v-less asp.	/ti:/	-28.01	<.001	-3.71	.560	24.31	<.001
		v-less unasp	/sti:/	36.91	<.001	*-2.59*	*.764*	*-39.50*	*<.001*
	dorsal	v-less asp.	/ki:/	-19.02	<.001	.417	.995	19.44	<.001
		v-less unasp	/ski:/	42.90	<.001	*-10.09*	*.036*	*-55.52*	*<.001*
	labial	voiced	/bi:/	-63.97	<.001	37.67	.007	101.6	<.001
	coronal		/di:/	-101.28	<.001	19.27	.259	120.6	<.001
	dorsal		/gi:/	-92.72	<.001	23.71	.109	116.4	<.001
/ɑ:/	labial	v-less asp.	/pɑ:/	-29.21	.002				
		v-less unasp	/spɑ:/	11.14	.005				
	coronal	v-less asp.	/tɑ:/	-34.93	<.001	-7.00	.398	27.92	<.001
		v-less unasp	/stɑ:/	22.96	<.001	*-15.03*	*.004*	*-37.99*	*<.001*
	dorsal	v-less asp.	/kɑ:/	-30.02	<.001	3.02	.698	33.04	<.001
		v-less unasp	/skɑ:/	28.24	<.001	*-9.50*	*.011*	*-37.74*	*<.001*
	labial	voiced	/bɑ:/	-63.55	<.001	29.27	.056	92.82	<.001
	coronal		/dɑ:/	-102.74	<.001	-.835	.997	101.9	<.001
	dorsal		/gɑ:/	-90.16	<.001	17.81	.233	108.0	<.001
/u:/	labial	v-less asp.	/pu:/	-31.13	.005				
		v-less unasp	/spu:/	21.23	.001				
	coronal	v-less asp.	/tu:/	-18.26	.002	13.83	.023	32.09	<.001
		v-less unasp	/stu:/	37.62	<.001	*2.43*	*.841*	*-35.20*	*<.001*
	dorsal	v-less asp.	/ku:/	-5.44	.471	9.59	.086	15.03	<.001
		v-less unasp	/sku:/	42.07	<.001	*-3.32*	*.723*	*-45.39*	*<.001*
	labial	voiced	/bu:/	-73.42	<.001	10.23	.739	83.65	<.001
	coronal		/du:/	-109.86	<.001	-10.04	.740	99.83	<.001
	dorsal		/gu:/	-84.08	<.001	20.05	.209	104.1	<.001

Table 5.14: Group differences in VOT production: words in sentences (italicised comparisons are between English unaspirated and Arabic aspirated).

V	Place	Voicing/ Aspiration	CV:	Learner - ENS		Learner - ANS		ENS - ANS	
				Mean diff.	p	Mean diff.	p	Mean diff.	p
/i:/	labial	v-less asp.	/pi:/	-17.63	.006				
		v-less unasp	/spi:/	16.55	<.001				
	coronal	v-less asp.	/ti:/	-12.92	.015	-1.03	.974	11.89	.007
		v-less unasp	/sti:/	45.50	<.001	*2.11*	*.881*	*-43.39*	*<.001*
	dorsal	v-less asp.	/ki:/	-3.77	.609	2.93	.715	6.70	.195
		v-less unasp	/ski:/	39.92	<.001	*-10.42*	*.017*	*-58.10*	*<.001*
	labial	voiced	/bi:/	-39.49	.002	50.56	<.001	90.06	<.001
	coronal		/di:/	-88.44	<.001	3.25	.927	91.69	<.001
	dorsal		/gi:/	-93.31	<.001	11.00	.339	104.3	<.001
/ɑ:/	labial	v-less asp.	/pɑ:/	-19.64	.009				
		v-less unasp	/spɑ:/	17.97	<.001				
	coronal	v-less asp.	/tɑ:/	-23.93	<.001	5.15	.410	29.09	<.001
		v-less unasp	/stɑ:/	33.52	<.001	*-2.13*	*.817*	*-35.65*	*<.001*
	dorsal	v-less asp.	/kɑ:/	-15.12	.001	1.88	.824	17.00	<.001
		v-less unasp	/skɑ:/	29.37	<.001	*-9.40*	*.020*	*-38.78*	*<.001*
	labial	voiced	/bɑ:/	-51.70	<.001	35.40	.005	87.1	<.001
	coronal		/dɑ:/	-72.50	<.001	14.05	.260	86.55	<.001
	dorsal		/gɑ:/	-66.55	<.001	27.46	.012	94.01	<.001
/u:/	labial	v-less asp.	/pu:/	-12.32	.202				
		v-less unasp	/spu:/	31.74	<.001				
	coronal	v-less asp.	/tu:/	-4.87	.603	25.34	<.001	30.21	<.001
		v-less unasp	/stu:/	44.42	<.001	*14.08*	*.002*	*-30.34*	*<.001*
	dorsal	v-less asp.	/ku:/	5.42	.271	8.54	.026	3.13	.601
		v-less unasp	/sku:/	48.07	<.001	*1.09*	*.965*	*-46.99*	*<.001*
	labial	voiced	/bu:/	-39.72	<.001	47.21	<.001	86.93	<.001
	coronal		/du:/	-102.0	<.001	3.04	.937	105.0	<.001
	dorsal		/gu:/	-90.64	<.001	14.07	.174	104.7	<.001

Out of the total 54 comparisons between L2 learners and ENS, the vast majority (89%, 48 comparisons) showed a significant difference in VOT between the two groups, so only in 11% of specific conditions compared could learners be said to be native-like in VOT. By contrast, out of the total 42 comparisons between the learners and ANS, only 14 comparisons showed a significant difference (33%). The majority of these interesting instances were for words uttered in sentence contexts.

This confirms overwhelmingly that indeed the learners resemble ANS far more than ENS and are still transferring VOT habits from their L1. This in turn does not provide much support for the SLM proposal which claims (Postulate 1) that established phonic categories for L1 sounds develop throughout the person's life span reflecting the properties of all L1 or L2 sounds, and that learning of L2 sounds is possible in adult life. In order more fully to interpret these findings as signs of learning, we have to focus on the highlighted results in Tables 5.13 and 5.14. First, there are two instances, /k^hi:/ and /k^hu:/ in sentences, where ENS and ANS do not differ, so would presumably be classified in the SLM as 'same' sounds (see 2.11). The prediction (RH1) is that learners will perceive such sounds as the same and simply identify the L2 sound with their L1 sound without forming an equivalence classification rather than a new category for it. They will then go on to produce the target sound correctly. This would be what is more generally termed positive transfer. Indeed, in both cases learners do not differ significantly from ENS, which supports this. For /ki:/ learners do not differ significantly from ANS either, with a mean VOT between those two, about 3 msec greater than ANS and the same less than ENS. For /ku:/ however, learners differ significantly from ANS because their mean VOT does not lie between the ENS mean (74.7 msec) and the ANS mean (71.6 msec) but is higher than the ENS mean (at 80.1 msec). This therefore throws doubt on whether the SLM prediction is fully supported in this instance, since if the learners truly had only one category for [k^h] before /u:/ in words in sentences, then there would surely be no significant VOT differences between any of the three groups. We must, however, bear in mind that the ANS values we are using are not necessarily those of the learners themselves. The SLM recognises that, when learners learn new L2 sounds, their L1 sound values may shift. We have, however, no way of detecting if that occurred in our learners.

A second SLM prediction (RH2) would be that the VOT of /p/ in any combination of conditions would be well acquired, because 'new' sounds such as this are, with time, easier to perceive and establish a suitable new category for than sounds that are perceived as 'similar' to L1 sounds. In all, there are 12 combinations of /p/, aspirated and unaspirated, with different vowels in the two contexts. Of these, however, learners differ significantly from ENS in all but two instances, which does not strongly support the SLM. The two instances which do conform to expectation are [pi:] in isolated words and [p^hu:] in sentences. In all other instances /p/ follows the pattern seen in the learner stops across all POAs of being in the aspirated cases significantly lower in VOT than ENS and in the unaspirated instances being significantly higher. It would seem therefore that the learner VOT production for voiceless labial stops largely runs parallel with that for their other voiceless stops rather than being specially more target-like due to being totally 'new'. We will see in 5.3.2 however that the data does support the idea of 'deflection' applying here (RH6a). The unaspirated voiceless English stops [t] and [k] are again possibly 'new' sounds, though not as clearly so as /p/. On the IPA symbol test, they are new if one transcribes the Arabic voiceless stops as [t^h] and [k^h], but 'similar' if one transcribes the Arabic stops as [t] and [k]. We have already seen that the Arabic voiceless stop VOT in fact spans the higher end of short lag and the lower end of long lag so that is not a simple choice to make. In the case of sounds that are 'new' and 'similar but with an easy to perceive difference', the SLM however makes more or less the same prediction (RH3), that with time a new category will become established, separate from the L1 and close to the L2. In fact, the 12 instances of voiceless unaspirated coronal and dorsal stops all exhibit significant differences from those of ENS by having greater positive VOT. They differ from the corresponding ANS voiceless stops in being generally lower in VOT, significantly different in half the instances. This therefore suggests that indeed the learners may have begun to establish a new category for voiceless unaspirated stops, though it is not identical to the target in mean VOT: rather it is in most cases between the ANS and ENS means, and nearer to the ANS one. A curious exception is [tu:] in sentences, where the learner mean VOT does not lie between the ENS and ANS means but significantly above the ANS one. Overall, however, the results partially support an SLM interpretation that this is an instance where learners perceive a difference between the L2 and L1 sounds and have more or less established a new category for the L2 sound, though it remains to be explained why it is closer to the L1 than the L2 position on the VOT continuum. In 5.3.2 we will consider whether deflection has been operative here.

The aspirated stops [t^h] and [k^h] represent instances, in SLM terms, where the sounds would be imagined to be perceived by learners as similar to L1 sounds, since they all have some aspiration in the long lag area, albeit the amount of aspiration is less in Arabic (lower mean VOT than English). Where such a lack of prominent perceptual difference exists, the SLM predicts (RH4) that learners will very likely end up with the L2 sounds being identified with the L1 sounds in an 'equivalence classification', and not being seen as a new category. Of the 12 relevant instances available in our study in fact 7 do show this pattern: significant difference from ENS mean VOT and no significant difference from ANS mean VOT. However, five instances go against the SLM prediction of an equivalence classification. Notably they all involve high vowels, not /ɑ:/.

First, learner [t^hu:] in isolated words has VOT significantly different from both ENS and ANS, which would suggest new category formation. Learner [k^hu:] in isolated words and [k^hi:] in sentences are not significantly different from either ANS or ENS. The second of those, however, is simply reflecting the fact mentioned above that ANS do not differ from ENS here, so an equivalence classification by the learner of the L2 sound with the L1 sound would likely be with a VOT in any case not far from the monolingual target VOT. Finally, [t^hu:] and [k^hu:] in sentences exhibit the reverse of the expected pattern: significant difference from ANS but not from ENS, which is more what we would expect in the SLM for a 'new' sound, which these are not. Overall then, although there is a majority of results that support the predicted equivalence classification of aspirated English /t k/ with Arabic /t k/, there is also a substantial minority which do not.

Finally, the voiced stops would again by the SLM be regarded as candidates for equivalence classification (RH5), since the same IPA symbols are used both in English and Arabic, and the detailed phonetic differences (negative VOT in Arabic and short lag positive VOT in English) are not perceptually prominent. Indeed, the data largely supports this, in that 13 out of the 18 relevant instances present the equivalence classification pattern of significant difference from ENS but not from ANS. The five exceptions all exhibit significant differences from ANS as well as ENS and have mean VOTs in between the two, suggesting that they have begun to move away from equivalence classification. One of the exceptions is /gɑ:/ in sentences, but notably all the others involve /b/ (/bi:/ in words and sentences, and /bɑ:/ and /bu:/ in sentences). This finding fits in with the results of the positive VOT analysis in 5.2.2.4 where there was some evidence of voiced labials exhibiting positive VOT more than coronals or dorsals.

The reason why /b/ should be especially prone to depart from the expected equivalence classification is not transparent. According to the SLM, it would be due to learners noticing the difference between Arabic and English /b/ in input more than the interlingual differences of /d/ and /g/. Such noticing is expected to be more prominent with 'new' sounds, but /b/ is not of course a new sound for Arabic learners of English. We may however speculate that the reason could be as follows. The voiceless counterpart of /b/, /p/ is a new sound, and according to the SLM learners will have noticed /p/ a lot. Part of noticing a new sound is trying to establish a new phonetic category for it, however, and that means paying attention to how that sound differs from near neighbouring sounds. Thus, the novelty of /p/ will have therefore incidentally drawn learner attention to /b/ as well, so as to try to locate the difference between them, which is a well-known problem for Arab learners of English (e.g. *gap* confused with *gab*). This extra incidental attention generated for /b/ and not for /g d/ may have been just enough for learners to identify the Arabic difference from English in pre-voicing more for /b/ and start to develop a new category for /b/ closer to the ENS VOT value. This appears to be a scenario not previously discussed in the SLM literature.

5.3.1.1 Comparisons based on separate means for the positive VOTs of voiced stops and the negative VOTs of voiced stops

Following on the findings for learner production of voiced stops with positive VOT separately (5.2.2.4), we next compared ENS and learners just with respect to positive VOT production. We wished to ascertain whether ENS and learners differ significantly on mean VOT of voiced stops based solely on responses that exhibited positive VOT, for each context and place of articulation separately (not differentiating vowels). We therefore used mean VOTs for each learner using only the repetitions of each learner that were non-negative (Table 5.9). Any participants who made all nine responses supporting one mean (e.g. for /b/ in sentence context) with negative VOT were omitted in the relevant analyses. The independent t test was then used to compare the two groups.

Table 5.15: Results of the t-test comparison between ENS and L2 learners in terms of mean positive VOT for voiced stops, in two contexts (disregarding negative VOT).

Context	Sound	Group	N	Mean	t	p
Words	/b/	ENS	30	14.85	-0.453	.655
		L2 learners	22	17.64		
	/d/	ENS	30	24.39	-0.740	.474
		L2 learners	12	32.81		
	/g/	ENS	30	30.39	-0.356	.726
		L2 learners	17	32.68		
Sentences	/b/	ENS	30	12.62	-3.965	.001
		L2 learners	23	30.46		
	/d/	ENS	30	18.61	-3.570	.001
		L2 learners	12	24.20		
	/g/	ENS	30	24.28	-1.658	.107
		L2 learners	22	27.64		

When comparing the learners with ENS, we can see from Table 5.15 that 22 learners produced the English voiced labial in isolated words with positive VOT at least some of the time, with mean value of 18 msec. This is a perfect ENS-like VOT mean, and was not significantly different (p=.655) from the ENS mean (15 msec). However, 23 learners produced the same sound in sentence contexts with a higher, long-lag, mean VOT (31 msec). In this case the difference of 18 msec above the ENS short-lag mean was highly significant (p=.001). Only 12 participants produced some of their voiced coronals in single words with positive VOT, and their mean VOT was slightly aspirated (33 msec), which was above the short lag ENS mean (24 msec). Despite the size of the difference (9 msec), it was not significant due to the small number of learners qualifying to be compared and their variability. The corresponding VOT difference for words in sentences was smaller (learners 24 msec versus ENS 19 msec), and indeed learners here were within the target short lag range. With 12 qualifying learners, the difference from ENS emerged as highly significant (p=.001). The voiced dorsal was produced in isolated words by 17 L2 learners with a mean VOT of 33 msec. This was not significantly greater than the ENS mean which was 30 msec (p=.726), though it is notable that here the ENS mean was itself slightly above the short lag limit. In sentences, 22 learners qualified, with mean VOT of 28 msec. This was also with the ENS range and not quite significantly different from the ENS mean (p=.107).

The most striking finding here is that, whether the differences are significant or not, the learner mean VOT is always higher rather than lower than that of ENS. We have seen elsewhere (5.2) that, on overall VOT for voiced plosives, the learners are far lower than ENS in VOT due to extensive use of prevoicing. It is therefore somewhat surprising to find that, if we exclude instances of prevoicing and only consider where the learners produce positive VOT, we find in fact that they use longer positive VOT than ENS. This is however consistent with our findings for unaspirated voiceless plosives (which in ENS are not significantly different in VOT from voiced plosives), where the L2 learners also exceed the VOT length of ENS. The L2 learners therefore seem to have a general difficulty with positive VOT in being unable to operate with it in the short lag range. Either they produce higher positive VOT, in high short lag and low end of long lag, or negative (pre-voicing).

That point must, however, be qualified by the observation that word versus sentence context also plays a role. In words in isolation the learner means are not significantly higher than ENS VOT means, even though generally in the long lag region. In sentences, the learner VOT means tend to be significantly different from ENS ones, even when in the short-lag. This is independent of how many people recorded positive VOT responses, which tended to be more in sentence contexts (5.2.2.4: Table 5.12).

Overall, we can see that a good number of learners produced the voiced stops with positive VOT at least on some repetitions, and were able to suppress the pre-voicing found in their L1. Of course, they have not reached the overall target nativelike mean and range found in the L2, and in fact they exceeded it on occasions. Still it is a good indication that they have the ability to suppress pre-voicing and they could produce instances close to the phonetic category found in the L2. Therefore, there is evidence of the potential to learn new categories closer to those of the L2 than those which they currently possess for voiced stops, though of course no proof that they will in fact ever establish those fully, however long they stay in the UK. The pre-voicing in the production of English stops by the learners was also compared with the pre-voicing of the ANS in Arabic stops. We therefore created mean VOTs for each person using only the tokens of each person that had negative VOT. Any participants who made all nine responses supporting one mean (e.g. for /b/ in sentence context) with positive VOT were omitted in the relevant analyses. The independent t test was then used to compare the two groups.

Table 5.16: Results of the t-test comparison between ANS and L2 learners in terms of mean negative VOT (pre-voicing) for voiced plosives, by place and context (disregarding zero and positive VOT).

Context	Sound	Group	N	Mean	t	p
Word in isolation	/b/	L2 learners	28	-82.74	-.401	.691
		ANS	30	-79.40		
	/d/	L2 learners	30	-99.68	-1.581	.119
		ANS	30	-88.36		
	/g/	L2 learners	30	-84.02	-.048	.962
		ANS	30	-83.73		
Word in sentence	/b/	L2 learners	25	-69.66	1.294	.205
		ANS	30	-78.12		
	/d/	L2 learners	31	-85.11	-1.183	.244
		ANS	30	-78.51		
	/g/	L2 learners	31	-82.36	-1.024	.313
		ANS	30	-76.73		

The results (Table 5.16) show that when we only consider negative VOTs, there was no significant difference between the two groups ($p>.1$) in the voiced labial, coronal and dorsal stops in either isolated word or sentence context.

It is notable that while none of the differences was significant between the learners and the ANS, all but one (/b/ in sentence) descriptively show that, when engaging in pre-voicing, the learners on average actually use slightly longer negative VOTs than ANS. This means that the learners, excluding instances where they produced L2 voiced stops without pre-voicing, produced them with longer than L1 pre-voicing. Therefore, although the negative VOTs for the English voiced stops produced by the learners were not significantly different from the negative VOTs of Arabic native speakers for Arabic stops, learner responses that have negative VOT show signs of not just having transferred their L1 VOT norms for the production of their English L2 voiced stops, but exaggerating them.

It is then only when learner VOTs for voiced plosives are considered as a whole, not separating positive and negative, that learners appear to be moving their VOTs somewhat in the direction of ENS (5.3.1).

To conclude the comparison between learners and ENS on positive voiced stop VOT data, we considered what the picture would be if, instead of working with mean VOT values, we worked simply with the three conventionally recognised VOT regions - long lag, short lag, pre-voiced - regardless of what the scores were within that. In other words, the criterion for being nativelike would then not be that learner VOT values were not significantly different from ENS VOT values, but rather that learner VOT fell in the same region, out of those three, that the ENS mean fell in, regardless of how far apart they might be within that region.

Tables 5.17 and 5.18 display this for tokens at the level of numbers of individual trials falling in each of the three regions (recalling that each score for a person for one condition, such as [bu:] in words in isolation, is a product of three identical trials). This contrasts with, for example, Figure 5.10 which showed mean VOTs of tokens within a region rather than numbers of VOT tokens falling within a region.

Table 5.17 shows the tokens that were produced with positive and negative VOT by the learners and confirms that more responses were produced of English /b/ than /d g/ with short lag VOT. Table 5.18 shows that at all POAs there were more short lag tokens in words in sentences than in isolation.

Table 5.17: The number of voiced stop production tokens produced with and without pre-voicing by the learners in both contexts.

POA	Negative VOT	Positive VOT		Total positive VOTs	Percentage of correct/ native like production (i.e. short lag region)
	Prevoicing	Short lag	Long lag		
Labial	348	154	56	210	27.6 %
Coronal	476	69	13	82	12.4 %
Dorsal	440	65	53	118	11.6 %

Table 5.18: The number of voiced stop production tokens produced without pre-voicing by the learners, by context.

Context	POA	Short lag (like ENS)	Long lag
Word in isolation	Labial	65	28
	Coronal	23	11
	Dorsal	22	35
Word in sentence	Labial	89	28
	Coronal	46	2
	Dorsal	43	18

5.3.1.2 Conclusion on the closeness of learner English voiced stop VOT to ENS VOT

The findings of this study for learner voiced stops are similar to those of many other studies with respect to the occurrence of extensive prevoicing, but new with respect to almost everything else, due to other studies not having covered learner voiced stop VOT in the same detail. In Shimizu (2011), Thai learners of ESL also produced voiced stops of English with pre-voicing because in their L1 they have such fully voiced stops. Similarly, Simon (2009) found Dutch learners of English producing English stops with pre-voicing.

One of the important reasons that the L2 learners of English transfer L1 VOT of pre-voiced stops to their L2 stops is that the native speakers of English do not have this difference between short-lag VOT and negative VOT of voiced stops (Bell-Berti and Raphael, 1995), and, crucially for the SLM, learners are unable to perceive this difference between languages. Some native speakers do however sometimes pre-voice the voiced stops of English (Lisker & Abramson, 1967; Docherty, 1992; Scobbie, 2002; Simon, 2009), as we also found. Normally, however, pre-voiced stops of learners are perceived as correct by listeners, or sufficiently understandable, so the communication process is not disrupted. Hence, consistent with the SLM, learners have no strong reason to notice and learn the nativelike VOT for the voiced stops of English (RH5).

The exceptional VOT of /b/ compared with /d g/ however stands out as a new and interesting finding, albeit no single explanation presents itself. As we suggested above, discussing RH5 in 5.3.1, there must be some reason why the Arabic learners pronounce English /b/ with different ranges of VOT from /d g/. According to our interpretation, the reason for producing English /b/ more often with both a short and long-lag VOT must be in some way related to the fact that Arabic does not have a /p/ consonant. This could affect learner production in a number of ways, however, and we cannot be certain which of them in fact are operative.

For instance, as the SLM predicts, the presence of new sound /p/ in English is very noticeable for Saudi learners, and we suggested that some of the attention it attracts might rub off on its voiced but not new counterpart /b/, resulting in learners noticing the latter's ENS VOT value more and moving their own English value for /b/ closer to that. Alternatively, the well known perceptual confusion of the two English sounds by Arabic speakers might lead to their English /b/ sometimes being spoken as if it was /p/, which would also move the mean VOT of /b/ into the positive VOT range, closer to that of ENS.

Brown (2000) by contrast argues that L2 learners can acquire a new sound if the nature of the contrast exists in their L1. In the present case, the L1 has either voiced stops which are actually pre-voiced or it has voiceless stops at coronal and velar places which are normally weakly aspirated, but it does not have a voiceless stop at the labial position. Hence, although the native language does not have the voiceless stop [p], the learners are nevertheless aware of the contrast between pre-voiced and voiceless stops in other positions, and of three POAs. As the learners are aware of such contrasts, they extend them to the labial position and shift their category of English [b] from pre-voiced to voiceless labial. At coronal and dorsal positions, they already have a voiceless stop. So, they produce [d] and [g] with an idea that they should not assimilate them with the voiceless aspirated coronal and dorsal of their L1. They maintain a categorical difference between their English voiced stop and the voiceless aspirated stop of the L1. But since they do not have any category for voiceless labial in their L1, they extend the category of English labial [b] in the wider range between 0 and 151 msec which encompasses both voiced and voiceless labials of English. This analysis is plausible if we consider that the restructuring of categories and adjustment of L2 is being processed under the strong influence of L1 consonants. The whole issue would, however, be much clearer if we knew what the learners' own L1 VOT values were for the voiced stops in our study.

We conclude that overall, the learning of the English voiced stops by the Saudi learners was poor but since they managed in some instances to shift their L1 laryngeal setting from pre-voicing to positive VOT, and even sometimes short lag VOT, learning of English L2 stops is possible and is occurring, especially for /b/.

5.3.2 Comparison of learners with ENS and ANS in terms of what key distinctions are made, regardless of VOT differences between groups

While comparison of learners' VOT performance with ENS or ANS performance through significance tests, as made in the previous section, is a common way of assessing how nativelike learners are, and where their difficulties lie, it is not the only way. Such an approach takes the view that if learners have acquired nativelike ability to produce plosives, they will produce them with very similar absolute VOT values to ENS. Suppose however that learners consistently produced VOTs around 30 msec lower than ENS. Very likely they would be significantly different from ENS in VOT on every stop possibility, so on that criterion not nativelike. But surely one could argue that they are nevertheless in a sense nativelike in keeping exactly the same distinctions that ENS make, but just employing a systematically shorter level of VOT to make the distinctions.

This focus on whether and how L1 and L2 sounds are distinguished as separate categories from each other in a particular phonetic space (in our case the VOT continuum), regardless of nativelikeness of their precise VOT values, is what the SLM deals with under the term 'deflection' (our RH6). Recall SLM H6 (2.11): 'The phonetic categories established for L2 sounds by a bilingual may differ from a monolingual's if: 1) the bilingual's category is 'deflected' away from an L1 category to maintain phonetic contrast between categories in a common L1-L2 phonological space; or 2) the bilingual's representation is based on different features, or feature weights, than a monolingual's.' In our study, we cannot examine the second type of possible deflection since we are only concerned with measuring one feature of stops, VOT, not others such as distribution, closure duration etc. We can however examine the first by seeing whether stop conditions that are significantly differentiated by VOT in ENS or ANS production are also separated in learner production, and by examining how learners space out L1 and L2 categories in one phonetic space (the VOT continuum).

We may begin by comparing what significant differences we found earlier in 5.2 within the learners with what we found for ENS and ANS on a parallel analysis (ch4). We may summarise the strongest effects as in Table 5.19. With effect sizes (partial eta squared given in brackets) larger figures indicate a stronger effect, on a scale 0-1. They are a valuable guide to where the most important contrasts are, especially when, as in our data, many effects are highly significant ($p<.001$). High significance alone does not tell us the size of an effect, simply that it is an effect that is not likely to be a chance effect due to sampling.

Table 5.19: Main distinctions marked by VOT in learner, ENS and ANS data, with effect size (partial eta squared).

	Learners	ENS	ANS	Which pair is similar?
Main effects	p (effect size)	p (effect size)	p (effect size)	
Context	ns	<.001 (.569)	ns	Learner-ANS
Place of articulation	<.001 (.407)	<.001 (.875)	<.001 (.443)	Learner-ANS
Voicing/aspiration	<.001 (.941)	<.001 (.981)	<.001 (.989)	Learner-ANS (see text)
Following vowel	<.001 (.410)	ns	ns	ENS-ANS
Two way interaction effects				
Place by vowel	<.001 (.256)	<.001 (.425)	.010 (.146)	Learner-ANS
Place by voicing/asp.	<.001 (.557)	<.001 (.298)	<.001 (.467)	Learner-ANS
Vowel by voicing/asp.	<.001 (.343)	ns	<.001 (.461)	Learner-ANS
Context by place	ns	ns	ns	None
Context by voicing/asp.	ns	.004 (.281)	ns	Learner-ANS
Context by vowel	ns	ns	.001 (.207)	Learner-ENS
Three way interaction effects				
Place by vowel by voicing/asp.	<.001 (.257)	.003 (.179).	ns	Learner-ENS
Context by vowel by place	.048 (.112)	ns	.014 (.136)	Learner-ANS
Context by place by voicing/asp.	ns	<.001 (.220)	ns	Learner-ANS
Vowel by context by voicing/asp.	ns	ns	ns	None

From Table 5.19 it is immediately obvious that far and away the strongest effect on VOT in all three languages (counting learners' English as a separate language) is that of voicing/aspiration, which exceeds even that of place of articulation of a stop, which comes a strong second in ENS but third for effect size in learners' English and fourth in Arabic. In other words, one thing that all participants show, is that they use VOT differences primarily to mark voicing/aspiration differences rather than anything else. This could be seen indeed as a language universal characteristic of VOT. However, this does not mean that they all do it in the same way.

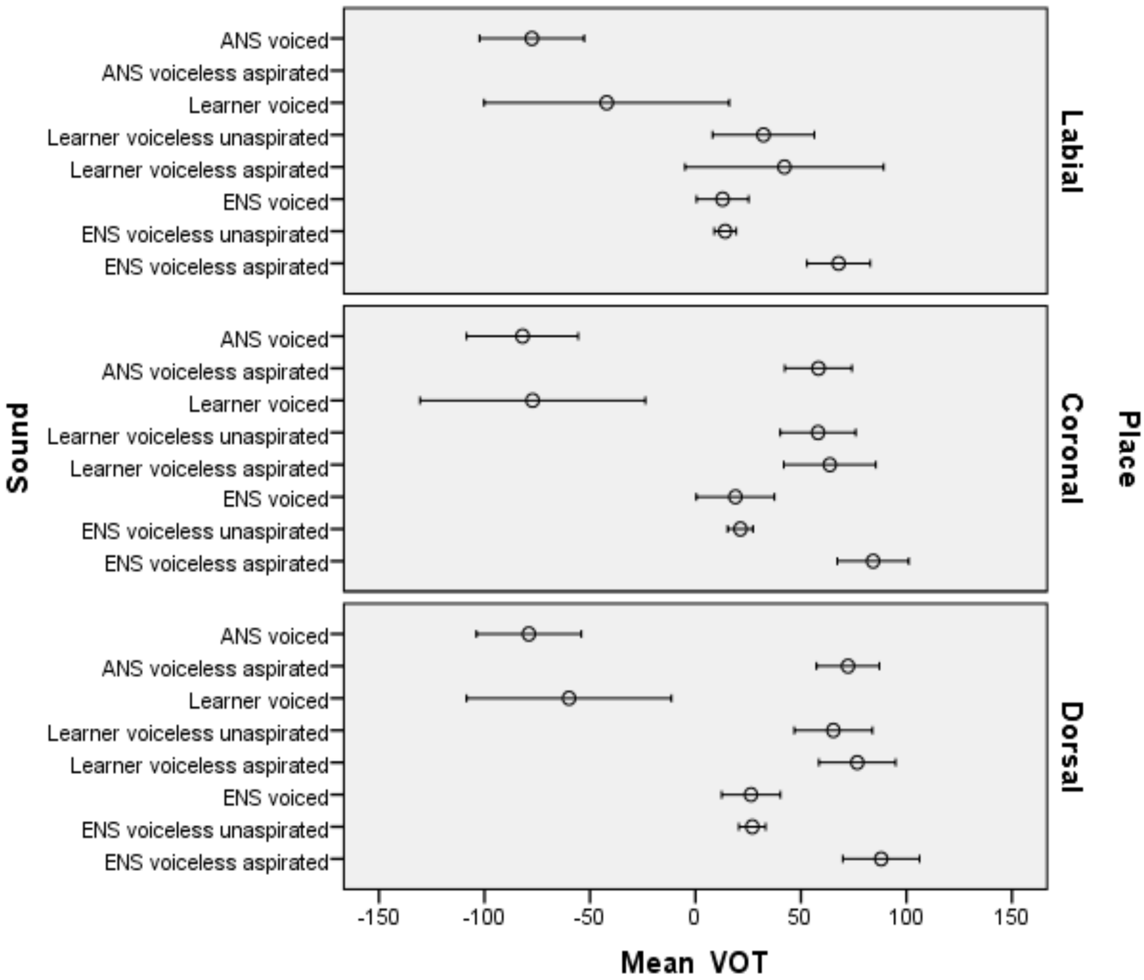


Figure 5.11: Error bar graph of voice and aspiration differences for each POA in the three groups (horizontal bars indicate the SD for each mean in the horizontal VOT continuum).

As Figure 5.11 shows, in Arabic of course the difference marked is purely one of voicing, since no voiceless unaspirated option exists (mean difference between voiceless aspirated and voiced is a huge 145.76 msec). In learner English, also a post hoc analysis reveals (as described in 5.2) that although VOT difference between aspirated and unaspirated voiceless plosives is significant (mean difference 9.34 msec, $p<.001$), the overwhelming distinction marked is between voiceless unaspirated and voiced (mean difference 112.38, $p<.001$) and between voiceless aspirated and voiced (mean difference 121.73, $p<.001$). In Figure 5.11, for coronal and dorsal, it is apparent that learners exhibit a major distinction between voiced and voiceless, as does L1, and that their English aspirated and unaspirated stop categories are very close to each other and to L1 voiceless stops in VOT location. This does not therefore reflect full deflection in the sense of trying to 'maintain phonetic contrast between categories in a common L1-L2 phonological space' which would require L1 voiced, L2 voiced, L1 voiceless, L2 voiceless unaspirated and L2 voiceless unaspirated to be spaced out over the available horizontal VOT space more evenly (RH6b). Instead learner categories are bunched into L1 and L2 voiced on the one hand, on the left, and L1 and L2 voiceless on the other, at least for coronal and dorsal. This suggests that equivalence classification wins out over establishment of separate categories with deflection.

A similar analysis of ENS shows, however, that the key difference marked is not between unaspirated voiceless and voiced stops (mean difference only 1.30 msec, $p=.909$), but between aspirated and unaspirated voiceless stops (mean difference 58.72, $p<.001$), and between aspirated voiceless and voiced (mean difference 60.02, $p<.001$). This emerges very clearly in the bottom three rows of each panel in Figure 5.11. In short, ANS, and learners, make VOT distinctions primarily based on voicing, using both the negative and positive sides of the VOT scale, while ENS make a distinction based on aspiration, within the positive half of the scale. This therefore distances learners from English quite apart from specific VOT values of their sounds.

An interesting feature of the ENS result is of course that it appears to fail to differentiate between voiced and unaspirated voiceless sounds altogether. There is no deflection here in the sense of ENS spacing out their distinct phonetic categories on the available VOT continuum. This draws attention to a limitation of our study that we have already mentioned; clearly ENS must be creating some space to differentiate between voiced stops and unaspirated voiceless stops on other phonetic aspects of stops than VOT, but these are beyond the scope of the present study. However, if ENS are doing this, learners may be doing this as well and perhaps distinguishing between their three voiceless stops (L1, L2 aspirated, L2 unaspirated) on other phonetic dimensions. Once again this is beyond our scope to ascertain.

Next, ANS and learners share the characteristic of not making VOT distinctions of place of articulation very strongly, regardless of other dimensions, while ENS do make a clear overall main effect distinction (ENS effect size for POA is twice that of the others). This appears clearly in Figure 5.12 where we can see that, especially for voiced and unaspirated voiceless stops, the SDs of ENS production make the differences between POA far more distinct than those of ANS and learners. For ANS and learners the interactive effect of place with voicing/aspiration is stronger than the main effect of place, meaning that a different pattern of differences between places of articulation is found for different voicing/aspiration options.

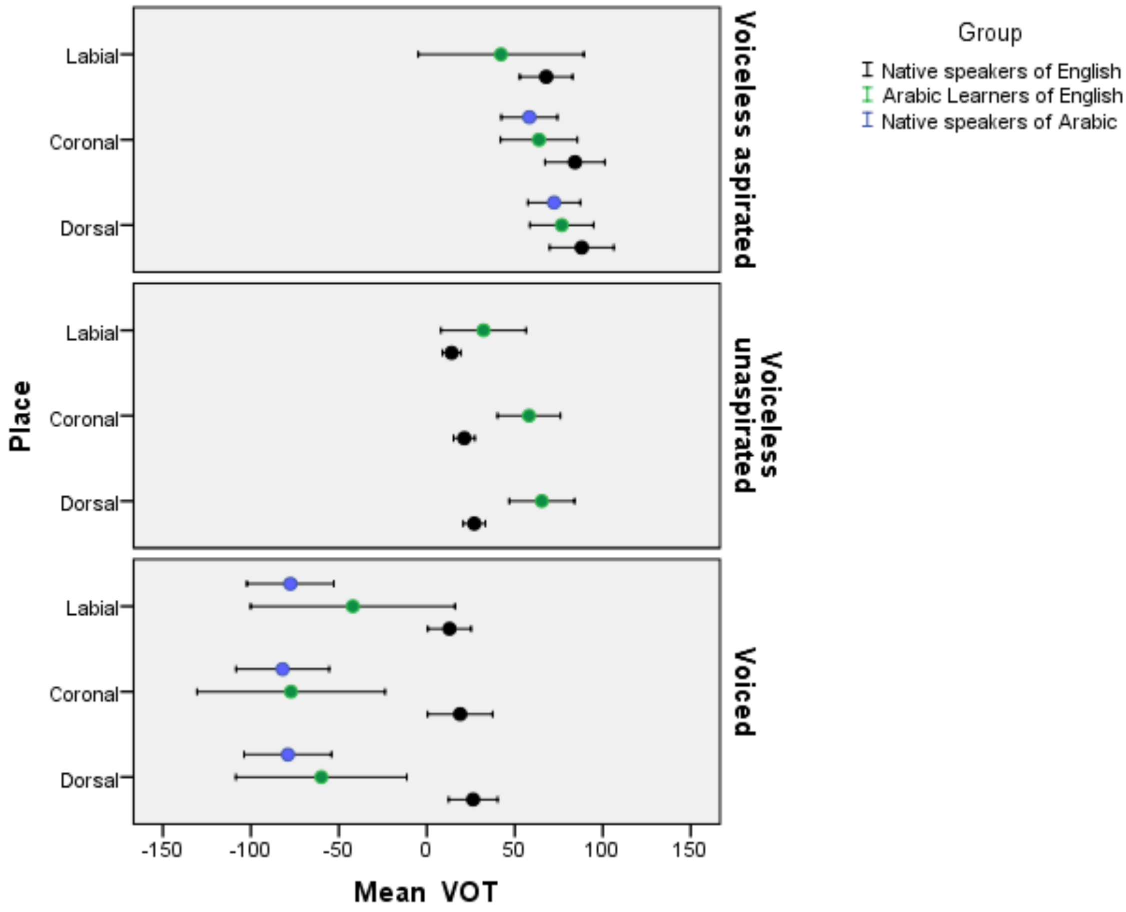
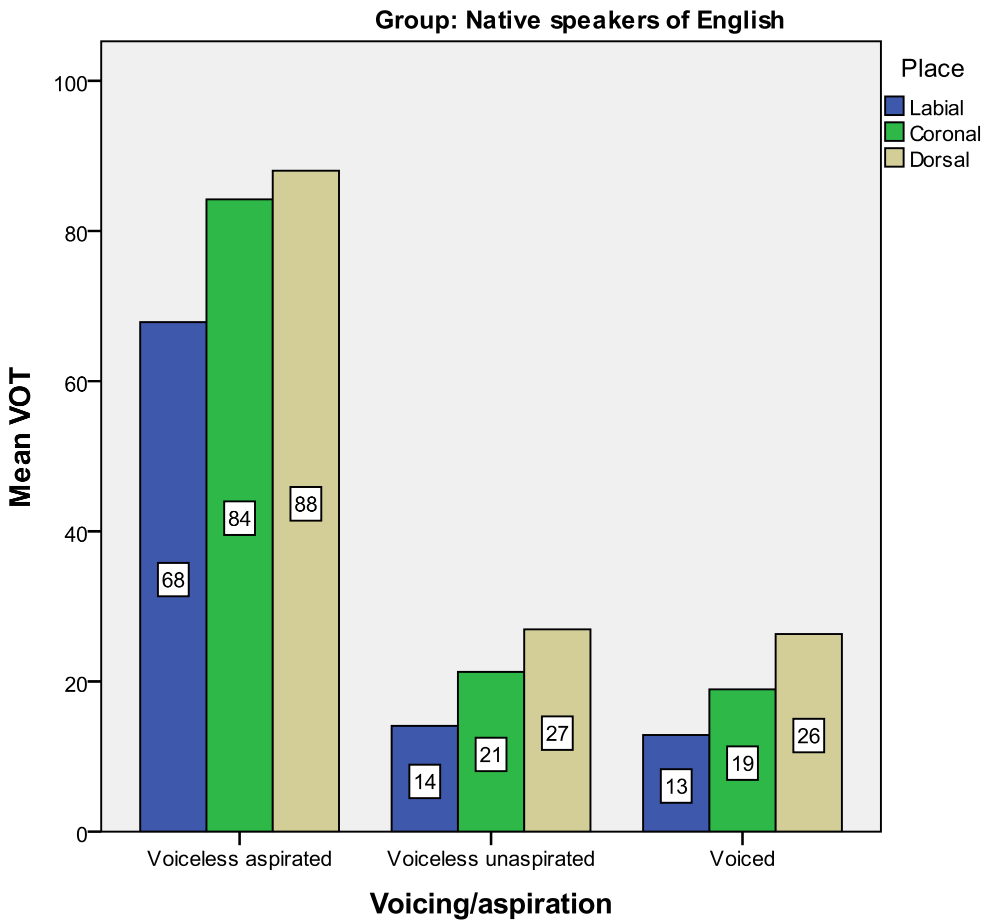
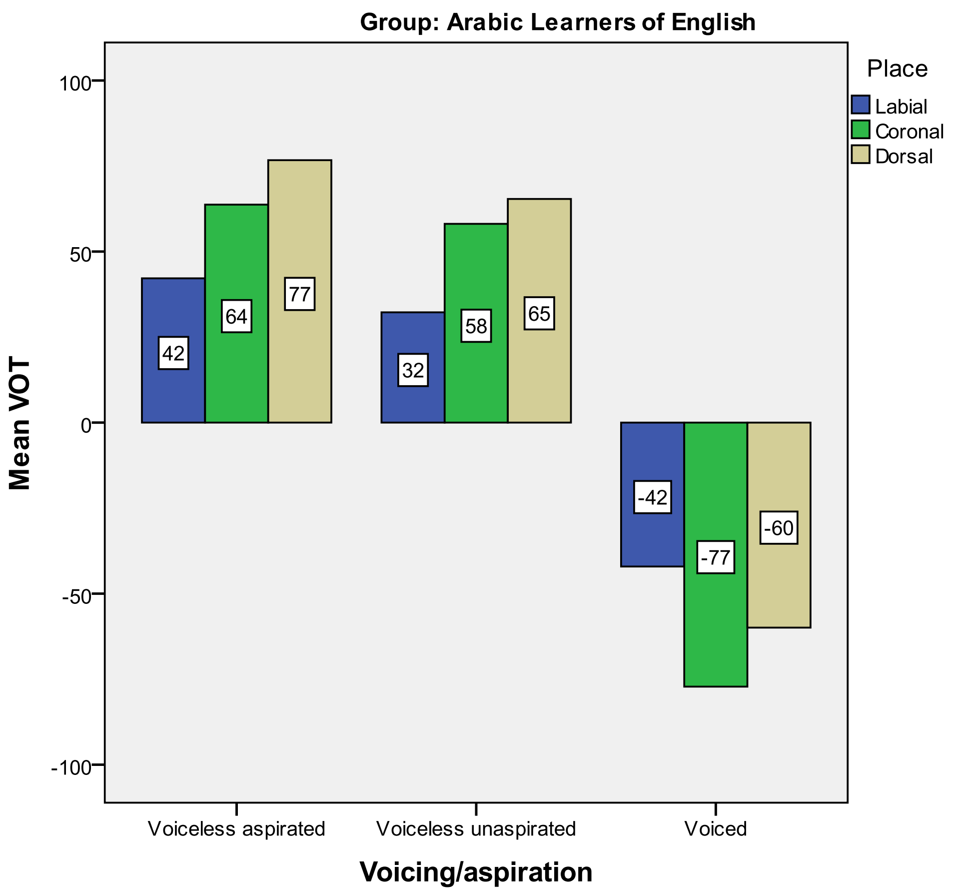


Figure 5.12: Error bar graph of POA differences for each voice and aspiration option in the three groups (horizontal bars indicate the SD for each mean).

As we can see in Figure 5.12, in ENS there is the same clear pattern of VOT getting longer from front to back labial<coronal<dorsal regardless of aspiration or voicing. This is not apparent in Arabic where this pattern holds for voiceless /t < k/ but not for the voiced stops. Again, learners' English was more similar to Arabic of ANSs, in that the /labial<coronal<dorsal/ pattern holds for voiceless plosives only but not for voiced ones, where there is a clear coronal<dorsal<labial pattern, more prominent than that in Arabic in fact. In short, learners' /b/ was not pronounced with the very negative VOT, such as -80ms, needed to fit a general labial<coronal<dorsal pattern.

This is why the learner and ANS results show a stronger effect of place difference in VOT that is different for voiced than voiceless plosives (the place by voicing/aspiration interaction effect) than for overall place difference regardless of voicing (place main effect). For ENS however, the interaction effect of place with voicing/aspiration, although significant, is far smaller (effect size .371) than the main effect of place (effect size .875). In fact, the interaction arises solely because the progression of VOT length over voiced and unaspirated voiceless stops is in more evenly spaced steps front to back, while for the voiceless aspirated plosives there is a much larger increase between labials and coronals than between coronals and dorsals. Figure 5.13 shows these effects in a different style. A different point that does emerge clearly from the POA results, however, concerns deflection (RH6a) in the location of aspirated and unaspirated /p/ by learners in their VOT space. This is perhaps the one area where our data supports the operation of 'deflection', in the sense that learner instances of /p/, though mostly located in positions which differ significantly from those of ENS (Table 5.13, 5.14), are (as we saw in 5.2) significantly differentiated from /t/ and /b/. We can further see in Figure 5.12 how the VOT of /p/ is systematically distanced below that of /t/ and /k/, in accordance with the quasi-universal progression of VOT increase over POAs front to back, and of course distanced higher than /b/, albeit the SDs for /p/ tend to be higher than those for /t k/. This means that learner production of VOT for /p/ is more varied around the mean value than that for /t k/. This is entirely to be expected with what is a new sound.


Group: Native speakers of English
Place
Labial
Coronal
Dorsal
Mean VOT
100
80
60
40
20
0
68
84
88
14
21
27
13
19
26
Voiceless aspirated
Voiceless unaspirated
Voiced
Voicing/aspiration


Group: Arabic Learners of English
Place
Labial
Coronal
Dorsal
Mean VOT
100
50
0
-50
-100
42
64
77
32
58
65
-42
-77
-60
Voiceless aspirated
Voiceless unaspirated
Voiced
Voicing/aspiration

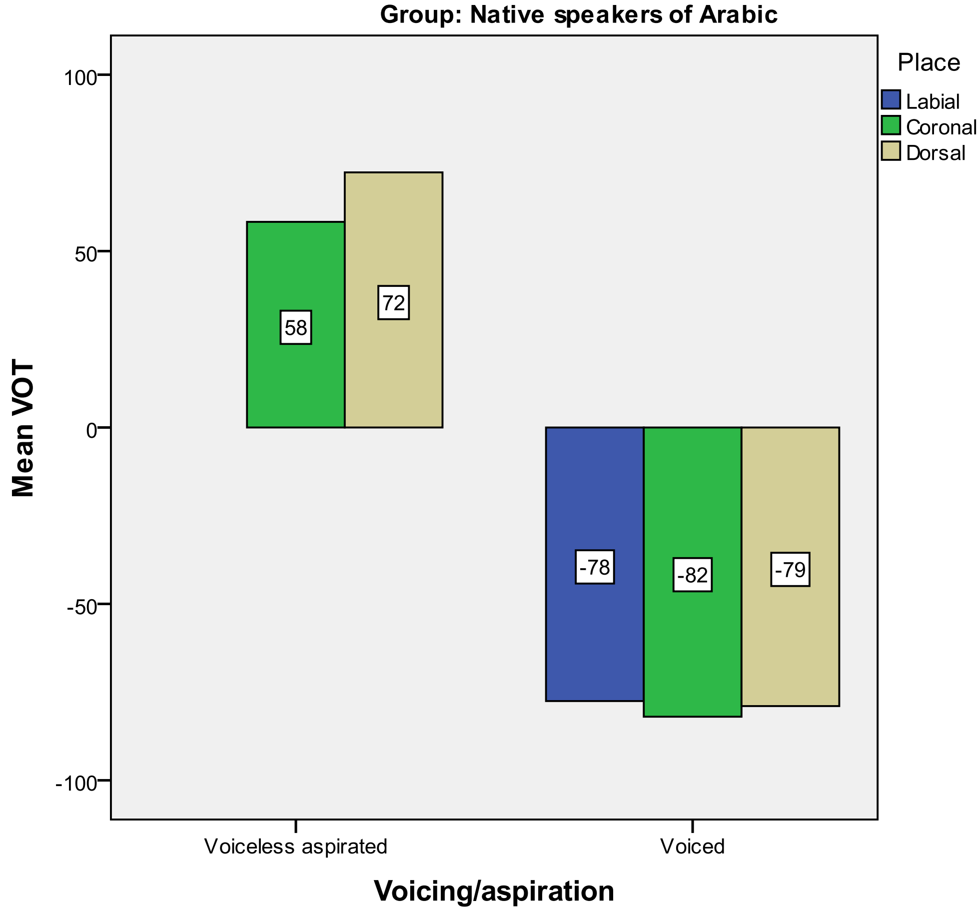
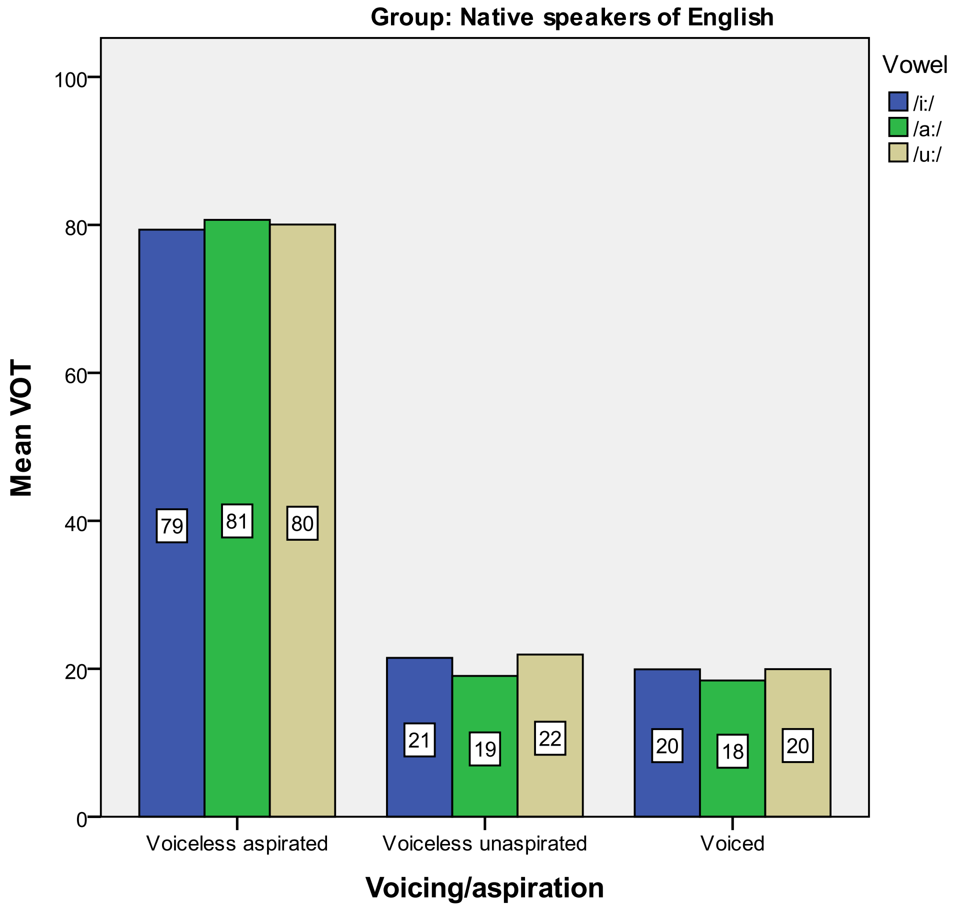


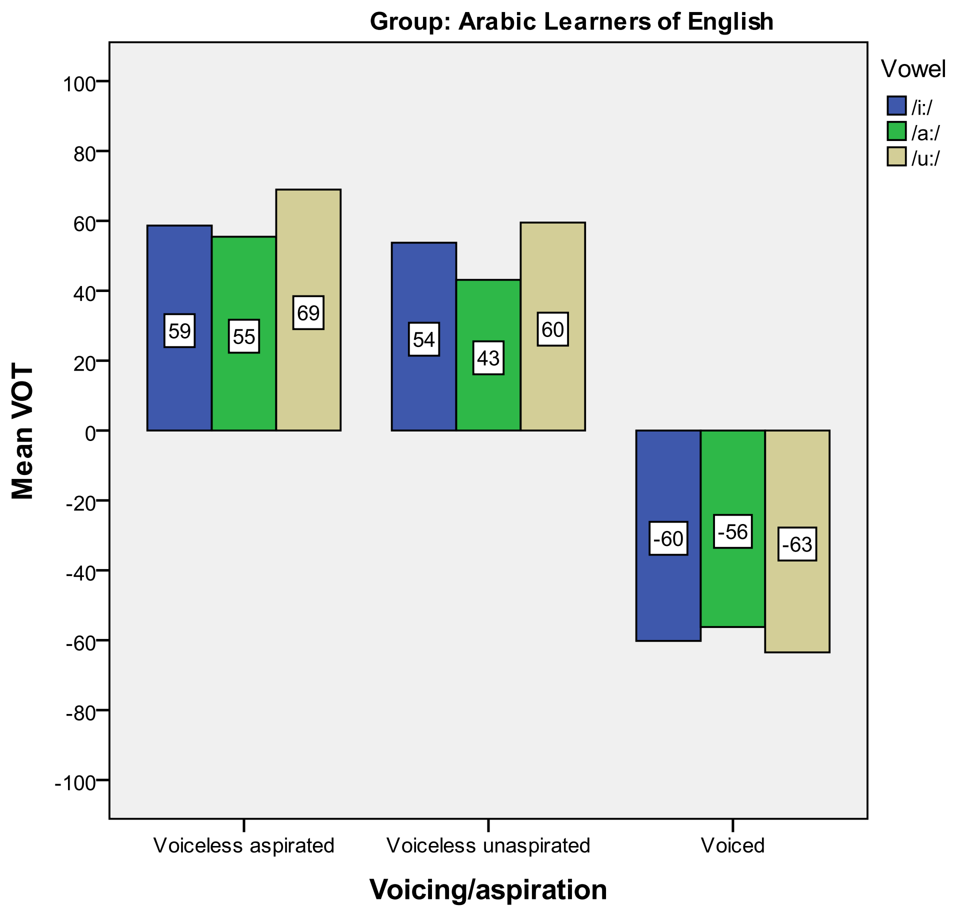
Figure 5.13: Place by voicing/aspiration VOT effects in the three groups (disregarding vowel and context).

Another way in which learners and ANS make the same distinction compared to ENS concerns overall context related variation in VOT. Neither learners nor ANS made any significant overall distinction of VOT between contexts (words/sentences), while ENS made quite a substantial one (effect size .569). For ENS VOT was systematically higher for words in isolation by 5.85 msec on average. One might interpret this on the assumption that ENS possess a stylistic capability to make an overall adjustment of VOT to suit word or sentence reading, in effect reading faster, so with shorter VOT, in sentences. Learners either have no such overall capability or follow the apparent L1 habit of making no difference, and read out words similarly regardless of context.

There are further effects in Table 5.19 where the learners' result prima facie appears more close to the ANS than the ENS one, in the sense that ANS and learners have a non-significant effect while ENS have a significant one, regardless of the actual VOT levels involved, or the learner and ANS effects are of a similar size very different from the size of the ENS effect.

We will look now at the effect with the next strongest effect size, the interaction effect on VOT of vowel by voicing/aspiration. Table 5.19 shows that there is no significant effect of this sort for ENS while learners and ANS both have one (effect sizes respectively .343 and .461). In order to understand this, let us inspect Figure 5.14. Although the effect was subtle, what the figures are telling us here is that for ENS, while there is some small fluctuation in VOT depending on vowel within each voicing/aspiration condition, there is no significant effect whereby the difference between vowels follows a different pattern, say, for voiced stops versus voiceless aspirated ones. The large difference between VOTs generally between voiceless aspirated stops and the rest was discounted in this calculation. By contrast, there do exist such interactive effects, in a similar form, for learners and ANS. Here visually we can see that for both groups there is a pattern of the type: with voiceless stops /ɑː/ induces shorter stop VOT than high vowels /iː, uː/ while with voiced stops /ɑː/ evokes higher, i.e. less negative, stop VOT compared with /iː, uː/. Thus, the differences between vowel effects on VOT are systematically different dependent on voicing, albeit the effect in Arabic depends more on the difference between /ɑː/ and /iː/ while that in learner English depends more on the difference between /ɑː/ and /uː/.


Group: Native speakers of English
Mean VOT
100
80
60
40
20
0
79
81
80
21
19
22
20
18
20
Voiceless aspirated
Voiceless unaspirated
Voiced
Voicing/aspiration
Vowel
/i:/
/a:/
/u:/


Group: Arabic Learners of English
Mean VOT
100
80
60
40
20
0
-20
-40
-60
-80
-100
59
55
69
54
43
60
-60
-56
-63
Voiceless aspirated
Voiceless unaspirated
Voiced
Voicing/aspiration
Vowel
/i:/
/a:/
/u:/

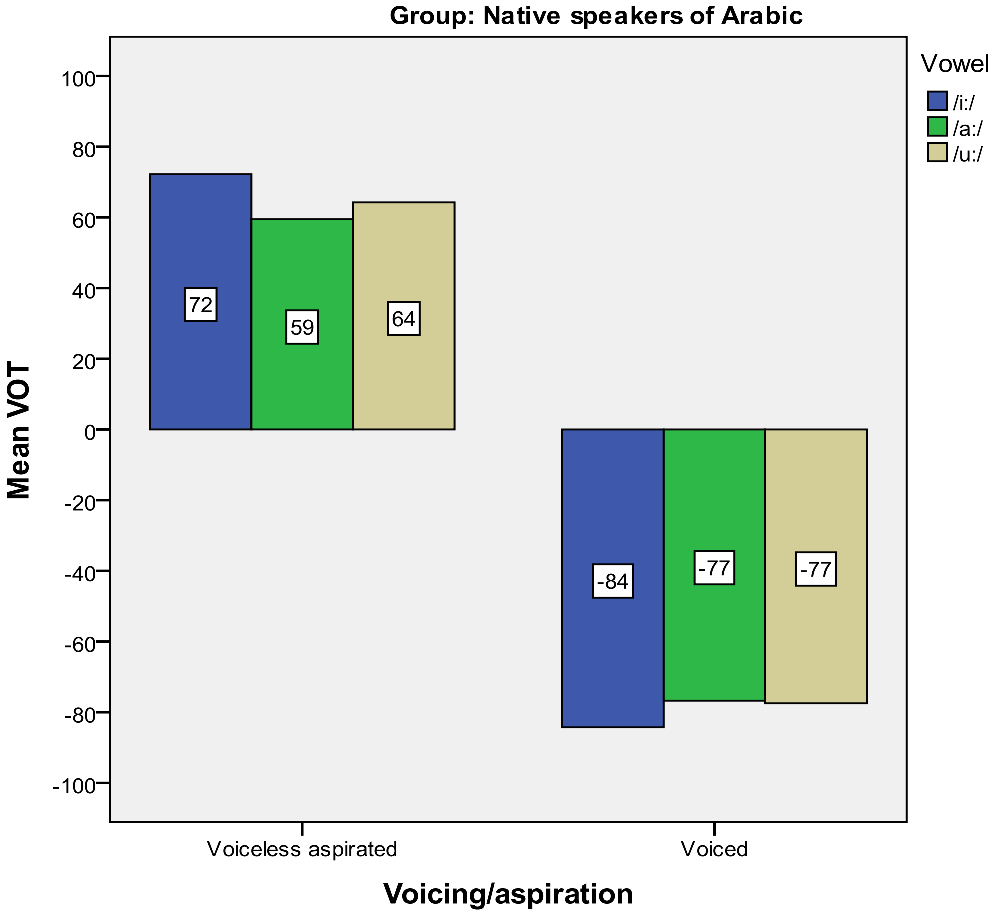
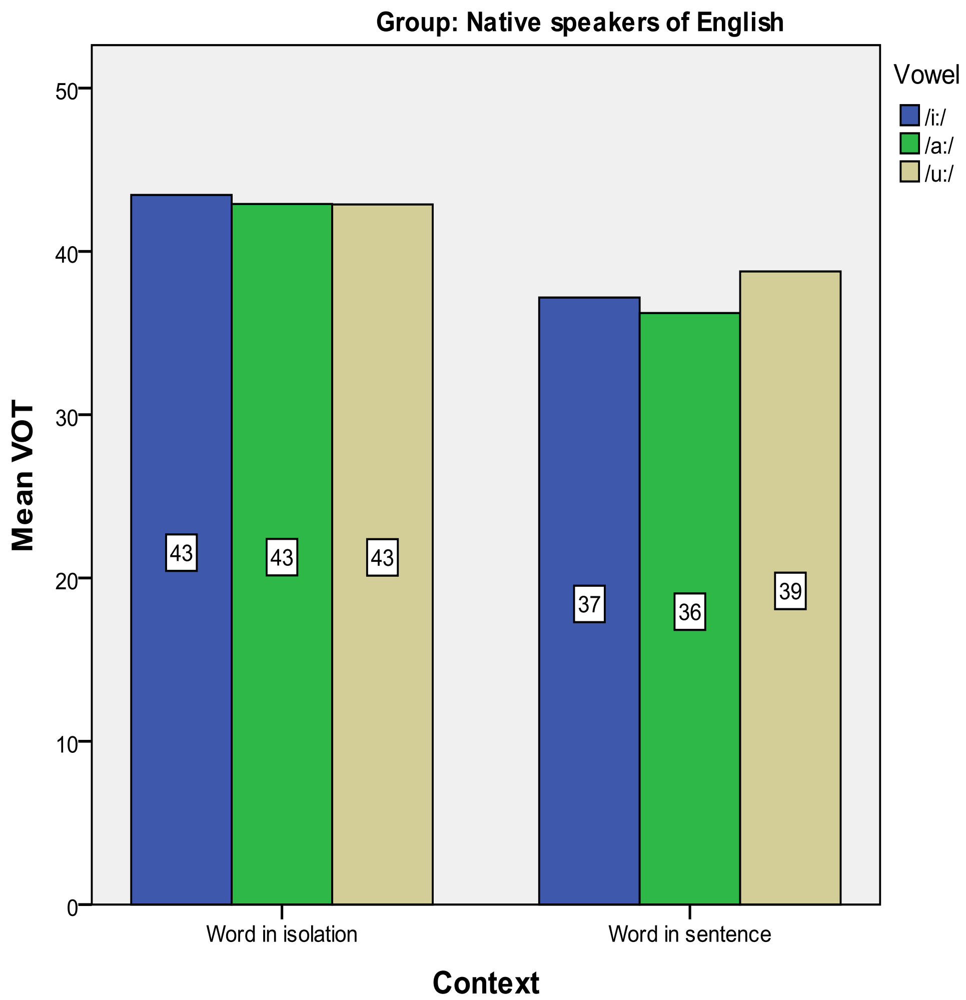
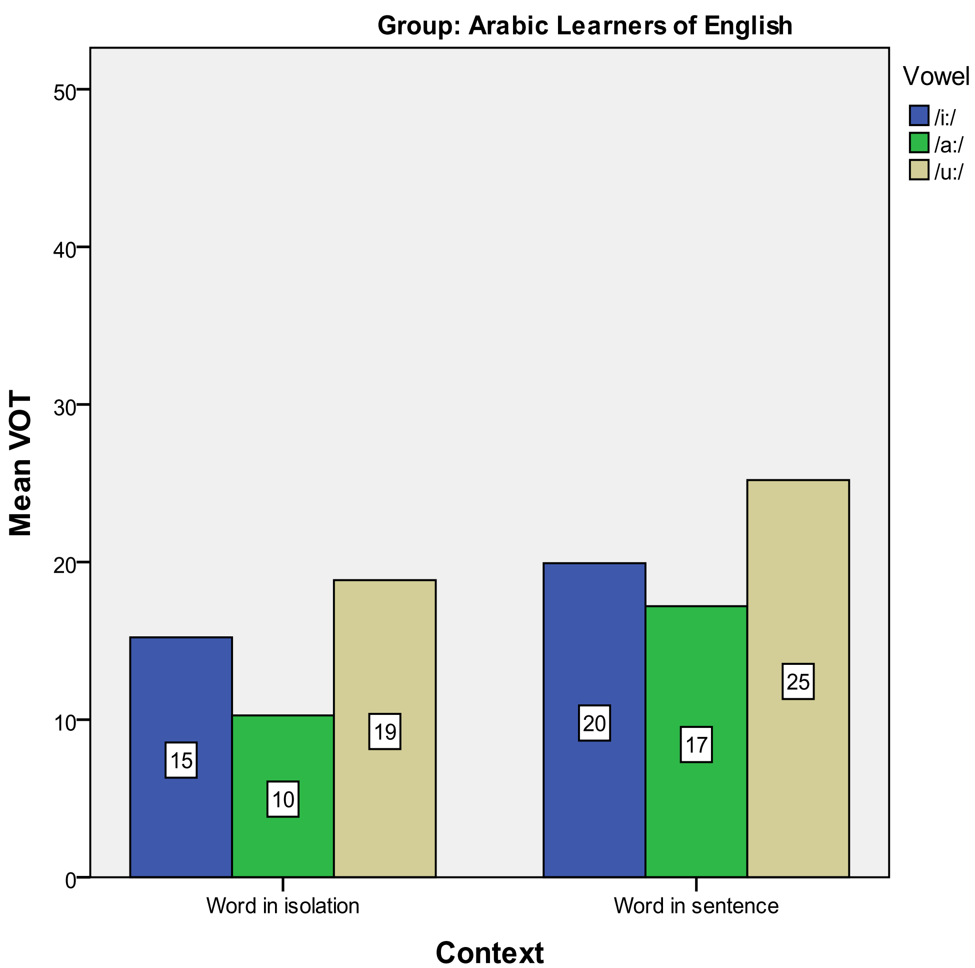


Figure 5.14: Vowel by voicing/aspiration VOT effects in the three groups (disregarding place and context).

Overall, out of the 14 effects in table 5.19, only 5 did not show signs of distinctive similarity between learner and ANS in contrast with ENS. Of those 5 only two appear to show learners and ENS more similar, with ANS the odd one out, and those two all involve relatively small effect sizes. One is the context by vowel interaction effect, which we can see in Figure 5.15.


Group: Native speakers of English
Vowel
/i:/
/a:/
/u:/
Mean VOT
0
10
20
30
40
50
43
43
43
37
36
39
Word in isolation
Word in sentence
Context


Group: Arabic Learners of English
Vowel
/i:/
/a:/
/u:/
Mean VOT
0
10
20
30
40
50
15
10
19
20
17
25
Word in isolation
Word in sentence
Context

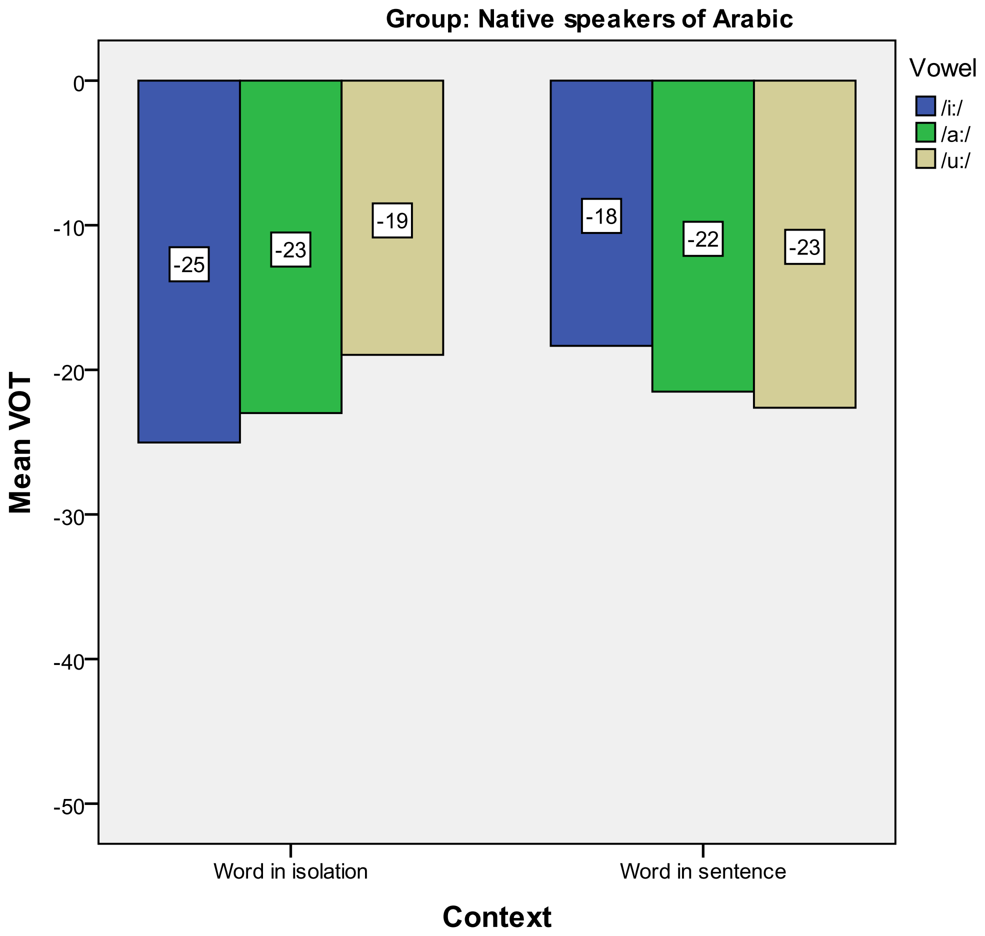


Figure 5.15: Context by vowel VOT effects in the three groups (disregarding place and voicing/aspiration).

As Table 5.19 shows, neither learners nor ENS show a significant interaction effect here while ANS do. This is highly visible in the graphs above. For ENS, the three vowels yield almost the same VOTs for all plosives considered together within words in isolation, and again within words in sentences the VOTs are almost the same, though all somewhat lower than those in isolated words. This therefore reflects the overall result we have seen for ENS that there is a VOT difference depending on context (VOTs shorter in sentence context) but generally none depending on vowel. Furthermore, there is no pattern of differences between vowels being different in different contexts (i.e. no context by vowel interaction effect). For learners, the pattern initially looks different in that there are clear differences between vowels (for which the main effect is significant): the low vowel /ɑː/ leads to shorter stop VOT than the high vowels /iː, uː/, with /uː/ inducing the longest positive VOT. However, once again that pattern is repeated in the same way, albeit at a generally slightly different level, in both word and sentence contexts. Once again, there is no difference between the vowel-related VOTs that is different depending on context, so as for ENS, no significant context by vowel interactive effect. When we turn to ANS however, quite aside from all the VOTs being negative (due to an overall negative mean when all places and voicings are combined), it is clear that the differences between vowels are not parallel in pattern in each context, but in fact the reverse. In words in isolation we have / iː < ɑː < uː / while in words in sentences we have /iː > ɑː > uː/. (In words in isolation there is a significant difference between /iː/ and /uː/ only, p=.036; in sentences, there is a significant difference only between /iː/ and /ɑː/, p=.025).

Therefore, out of all this, we can understand that the context related differences between vowels are the same for learners and ENS, in the sense that there are no significant distinctions of that sort made by either group, in contrast with ANS who do make a significant context related distinction between vowels.

5.3.3 Summary of the findings concerning nativelikeness of learner VOT

The learners' VOTs overwhelmingly resemble those of ANS more than ENS, whether one considers similarity of actual VOT or similarity in terms of the same distinctions been significant (regardless of VOT level)

If we take nonsignificant difference of VOT from ENS as a sign of nativelikeness, then there are few stop CV: combinations where the learners are nativelike, all for voiceless and mostly aspirated plosives. There appears to be a small core of incipient nativelike production based around /ku:/ and other voiceless aspirates especially + /u:/ assisted by their L1 VOT values being in a few instances not different from ENS ones.

There are only a few instances where learners significantly differ from ANS VOTs, involving primarily unaspirated voiceless or voiced stops, though if we analyse only positive VOT values produced for voiced stops there are signs that production close to that of ENS is possible, especially for /b/.

Learners' mean VOTs for each condition descriptively do mostly fall between the generally shorter or negative ANS ones and the longer ENS ones, suggesting slight movement in the desired direction and that creation of new phonetic categories may be occurring in some cases.

RHs 1-5 which we formulated based on SLM expectations however not all confirmed in an unqualified way.

If we take making the same significant distinctions as ENS do between stop conditions such as POA, voice/aspiration, following vowel etc., regardless of VOT size, as being a sign of nativelikeness, then again there is little evidence of nativelikeness in the learners' data. For all, voice/aspiration is the main dimension differentiated by VOTs, but while for ENS the main distinction is between voiceless aspirated on the one hand and unaspirated voiceless and voiced on the other, for ANS and learners it is between voiceless and voiced.

There was support for RH6a (SLM H6 and Postulate 4) concerning deflection occurring in the SLM sense of categories being 'deflected' away from an L1 category "to maintain phonetic contrast between categories in a common L1-L2 phonological space", at least as far as the location of voiceless bilabials relative to other POAs were concerned, but not so clearly where the aspirated - unaspirated distinction for stops is concerned (RH6b). We noted that it is always possible in the latter case that learners possessed a "representation ... based on different features, or feature weights, than a monolingual's" (Flege, 1995, p. 239), which it was outside the scope of the present study to be able to detect.

ENS make strong overall distinctions in VOT based on context (isolation longer), and place (rising from front to back) but not vowel. Learners and ANS by contrast make no overall distinction based on context, and for place show the progressive ENS pattern (increasing VOT front to back) only in the voiceless stops: for the voiced stops the pattern is coronal<dorsal<labial (in terms of VOT becoming less negative) or labial<dorsal<coronal (in terms of VOT becoming more negative).

64% of the main and interaction effects examined show similarity between learners and ANS rather than ENS in the distinctions significantly made, and their effect sizes.

The main similarity of pattern lies only in the fact that all three groups produce longer positive VOT for aspirated than unaspirated or voiced plosives, and all coincide in some parts of the data at least in presenting a VOT increasing across places of articulation, front to back.

5.4 The relationship of length of residence (LOR) and daily use of English in the acquisition of VOT in production

RH7 concerns whether the length of residence (LOR) in the UK or mean daily use of English (speaking and listening) relate to the learners' VOT in production in a way such that those with greater LOR etc. are further advanced in some areas of acquisition at least. This is implied by Postulates 1 and 3 of the SLM (see 2.11 in Chapter two). Other SLM hypotheses suggest that progress might be seen more where L1 - L2 differences between sounds are more noticeable, which in our case means acquisition of /p/ and possibly of the unaspirated voiceless stops (cf. SLM H2 and H3 and our RH2 and RH3). As we saw in 5.2 and 5.3, in order to be more nativelike on voiceless aspirated and voiced stops, learners need to produce longer positive VOTs than they typically do, so we would seek a positive correlation between VOT and LOR. In order to be more nativelike on voiceless unaspirated stops, however, they need to produce shorter positive VOTs, so for them we would expect a negative correlation between LOR etc. and VOT.

Correlations were therefore examined between learners' LOR and reported mean daily use of English on the one hand, and VOT in each place by voice-aspiration by vowel condition (27 for word in isolation and 27 for word in sentence) on the other. We additionally treated LOR as two extreme groups and used the t test to see if the group with longer LOR differed anywhere from that with shorter LOR. In general, the analysis of LOR as groups produced fewer significant differences than that treating LOR as scores and calculating correlations.

5.4.1 Voiceless stops

Effects of LOR were more evident for the voiceless stops than the voiced. There were significant correlations with VOT in almost a third of the possible 36 voiceless conditions. In word alone contexts: [kiː] r=-.380, p=.035; [t^hɑː] r= -.364, p=.044; [tɑː] r=-.561, p=.001. The same three were significant on the group analysis, with the addition of [tiː], t=2.59, p=.024. All these relations were negative, which means that greater LOR yields shorter VOT. Shorter VOT than the learner norm is indeed more nativelike for [kiː] and [tɑː] and [tiː] but not for [t^hɑː].

In sentence contexts there were the following significant correlations with LOR: [t^{h}iː] r=.466, p=.008; [tiː] r=.373, p=.039; [k^{h}iː] r=.400, p=.026; [t^{h}ɑː] r=.397, p=.027; [k^{h}ɑː] r=.511, p=.009; [t^{h}uː] r=.555, p=.001; [k^{h}uː] r=.439, p=.014. Out of those only [k^{h}ɑː] and [t^{h}uː] were significant on the analysis by LOR groups. Here all the correlations were positive, meaning that greater LOR is associated with longer VOT. Since all but one concerned aspirated plosives, for which longer VOT is indeed more nativelike, this means that in the great majority of instances where a significant correlation was found, the correlation showed that LOR produced some beneficial effect.

The overall effect however is not mainly where the SLM would predict, since, as we understand the SLM, it would argue that the difference between L1 and L2 in voiceless aspirated stops is hard to perceive and would result in an equivalence classification that is hard to shift (RH4). It is rather with the unaspirated stops that L1-L2 difference should be more apparent (RH3) and hence more likely to show acquisition along with greater LOR. It is noticeable that the majority of the correlations were for production in sentence contexts rather than isolated word contexts, perhaps because that is where the more spontaneous competence of the learners showed itself, and it is spontaneous competence that greater LOR would do most to develop.

It is also noticeable that none of the correlations involved the labials [p^{h}, p]. These are of course sounds missing in the L1 of the learners' (Arabic), so this perhaps suggests that these are sounds especially resistant to any change due to LOR, though the SLM would in fact predict the contrary, due to new sounds being especially noticeable. However, there was some correlation between the VOT of labials [p^{h}, p] and speaking and listening as we shall see below.

Mean daily language use also showed more relationship with VOT of voiceless than of voiced stops. There were significant correlations with VOT in less than a quarter of the possible 36 conditions however. Once again the majority were in sentence context production. In word alone contexts we have: [p^hiː] r=.458, p=.010; [p^huː] r= .485, p=.006. In sentence contexts: [piː] r=.464, p=.009; [t^hiː] r=.388, p=.031; [tiː] r=.393, p=.029; [t^hɑː] r=.380, p=.035; [kɑː] r=.359, p=.047; [t^huː] r=.462, p=.009. All these correlations were positive, which means that, for the aspirated stops, learners who used English more per day were more nativelike than those who used it less. However, the greater English users were less nativelike on the unaspirated plosives following /s/, where a decrease in mean VOT would be closer to ENS. This suggests perhaps that the learners treated aspirated and unaspirated voiceless stops in English rather similarly (and of course they do not mark a phonemic distinction). We have seen earlier in (5.2) that they did make a significant distinction between aspirated and unaspirated stops. It was however small in VOT size compared with the difference made by ENS.

This all suggests that both use of English daily and LOR have some impact on acquisition, in a minority of all the possible combinations of conditions. This impact however is mainly on improvement of pronunciation of the aspirated voiceless stops, relative to ENS VOT: 11 out of 12 significant correlations were positive. This is not where the SLM would predict improvement to be most likely, however, since both L1 and L2 have long lag VOT, so the difference should be hard to notice (RH4). On the unaspirated stops, which the SLM would predict to be more likely to improve, due to their greater noticeability relative to L1 (RH3), we found fewer correlations and they were not predominantly negative, as is required to show acquisition: only 3 out of 7 significant correlations were negative. Bilabial voiceless stops also did not show a tendency to change with LOR or use more than stops at the other POAs, as the SLM would predict, due to novelty (RH2): only 3 out of 19 significant correlations involved voiceless bilabials.

5.4.2 Voiced stops

Here LOR does not correlate significantly with VOT in any of the 18 conditions except one: /buː/ in sentence contexts where r=-.436, p=.014 (and the group analysis yielded a similar result). I.e. learners with greater LOR produced more negative VOTs in this one instance. This is surprising since less negative VOTs would be more nativelike. In short, this suggests that length of residence has almost no impact on the VOT production of voiced stops by L2 learners, implying that exposure to L2 in this area is powerless to change their phonetic habits. This conforms to the SLM expectation (cf. RH5), since the difference between Arabic voicing and English voicing, though very different in VOT, is considered hard to perceive. This also implies that learners had created an enduring equivalence classification which prevented them from creating a new category for these sounds in English even with longer exposure to L2 input.

Our current daily language use measure (combining listening and speaking) also failed to show any significant relationship with VOT of voiced stops in a correlation analysis except one: /diː/ in word contexts: r=.384, p=.033. Here greater daily use relates to less negative VOT, as one might expect, but again the overall paucity of correlations supports the SLM interpretation (previous paragraph).

We also looked separately at correlations involving only the voiced stop data with positive VOTs, which we presented results for earlier. Just one significant result stood out. Mean daily use of English correlated significantly and positively with positive VOTs for /b/ in words (r=.542, p=.009). This means that those who reported using more English orally on a daily basis were more likely to produce /b/ in words with VOTs higher rather than lower within the positive VOT range. For the most part, however, it seems that while the learner category for each voiced stop covered quite a range (with a large SD including some nativelike positive VOT instances), its core remained close to that of the corresponding L1 category for the voiced stops in equivalence classification, and there was little sign of any shift in that with increasing LOR or mean daily use.

5.4.3 Summary of the LOR and language use findings

As Table 5.20 shows, there were 12 significant correlations of VOT with LOR and 9 with daily language use. Of the former three quarters were in the direction showing acquisition, of the latter two thirds. This suggests that LOR is slightly more influential in affecting acquisition than mean daily use of English.

Table 5.20: Significant correlations of learner length of residence and reported mean daily use of English with VOT in production (raw numbers and percent of maximum possible number of correlations).

	LOR			**Daily language use**		
	Voiceless aspirate	Voiceless unaspirate	Voiced	Voiceless aspirate	Voiceless unaspirate	Voiced
In the direction of nativelike VOT	6 (33%)	3 (17%)	0 (0%)	5 (28%)	0 (0%)	1 (6%)
Away from nativelike VOT	1 (6%)	1(6%)	1 (6%)	0 (0%)	3 (17%)	0 (0%)
Total possible	18	18	18	18	18	18

Aspirated stops evidenced the greater number of significant correlations of LOR and daily use with VOT, which were predominantly positive, indicating movement towards ENS norms was occurring, somewhat unexpectedly in the light of SLM expectations that an equivalence classification would have been established here. There was even movement, as LOR increased, towards higher values of the VOT of [k^hu:] in sentence contexts, although this combination, as we saw in Table 5.14, was not produced overall with VOT significantly different from that of ENS in any case.

Unaspirated voiceless stops by contrast less often showed significant correlations, and they were more evenly divided between movement towards and away from ENS norms, contrary to SLM expectation which would predict movement towards ENS norms. Voiced stops, as predicted by the SLM, attracted few correlations of any sort, evidencing no effect of LOR or daily use on nativelikeness: they may be seen as fixed in an equivalence classification, albeit the variation in VOT produced, as we saw in 5.2.1.1, embraces nativelike positive VOT instances which show that the potential for acquisition is present. Finally, correlations mostly involved coronal and dorsal stops rather than labial, counter to SLM expectation.

Finally, the impact of LOR and daily use was greater on VOT of stops in words produced in sentence contexts rather than words in isolation. Perhaps this was due to the learners' exposure being in fact mainly to connected speech (sentence contexts), not to single isolated words.

5.5 Individual VOT production performances

All the other production results we have generated examined the learners and ENS as groups, and have found many significant differences between learners and ENS. It therefore came to our mind to explore if, despite the non-nativelikeness of the learners as a group, there were individuals who had managed to achieve near nativelikeness of

VOT. If that occurred, then we might learn something from examining the individual nature and background experiences of such a learner.

In order to achieve this, we generated VOT scores for each learner reflecting how far they differed from ENS VOT on each of the 54 stop conditions measured (3 places by 3 aspiration/voicing options by 3 following vowels by 2 contexts). That is, we adopted the notion of similarity/difference of 5.3.1. Each score was made by subtracting the ENS mean VOT from the learner's individual VOT and dividing by the standard deviation (SD) of the ENS VOT. If just the difference between learner VOT and ENS mean VOT were used as an indication, this would not take account of the fact that the ENS themselves do not perfectly agree on one VOT range for a given stop in a particular condition. Clearly if ENS were not varying much as a group, then we might consider a learner with quite a small difference from the ENS mean in VOT as not being nativelike. But if the ENS are varying considerably, we would require a bigger learner difference to be convinced that the learner was not ENS-like. It is for that reason that the SD, reflecting ENS variation, was included in the calculation as well. In this way we take into account not only the simple difference between the learner and the ENS average in VOT, but also how varied the ENS responses were. The more the ENS themselves varied in their VOT around the mean for ENS as a group, the more the difference between learner and ENS mean gets reduced.

Note that this sort of score only reflects how far away from a mean a particular person's score is, not whether it is higher or lower than the target mean. This kind of score, termed a z score, is often interpreted using the threshold value of 2. If the learner obtains a score 2 or more, calculated in this way, it means the learner is not a likely member of the ENS group, i.e. is not nativelike. If the learner obtains a difference z score below 2, this suggests he could be a member of the ENS group, i.e. nativelike.

Although this is a standard statistical procedure, we have not seen it used before in VOT studies. The same is true of most of the types of analysis and graph used below.

5.5.1 Findings for individuals overall

Looking at Figure 5.16 for all stops taken together in all conditions, we can see that the difference scores of the ENS were all well below 2, as one would expect. Although they vary, none do so by more than 2 SDs. In Figure 5.17, however, we see that none of the learner scores were within the range of 2 SDs. The closest to being nativelike overall was learner 29 (z score 3.54) who has in fact only been in the UK for three years and reports relatively low daily exposure to English. The least was learner 9 whose background was not dissimilar.

As a further check we performed a cluster analysis using K-Means Cluster in SPSS. This is a procedure where one supplies the program with the score information about each participant, but does not tell it that one has two groups of people. One then allows the program to look at the profiles of scores of each person (in our case, the 54 'difference from ENS mean' scores of each person) and, purely on the basis of similarity of profile, find two groups. This as expected found two groups coinciding with the ENS and the learners, but with one exception. Learner 29 had a profile which the procedure found similar to the ENS, and he was placed in that group rather than with the learners.

Learner 29 in fact is an unusual participant in a number of respects. He was a second year PhD student at the department of Language and Linguistics at the University of Essex. He obtained his BA degree in English language and Literature and his masters degree in Linguistics both in Saudi Arabia. When he came to study in the UK, he also did another MA on linguistics before proceeding to the PhD. He describes himself as a very sociable person and has lots of friends and attend lots of gatherings and conferences. In addition to all that, he was a university teacher back in Saudi and had probably 8 years of teaching experience.

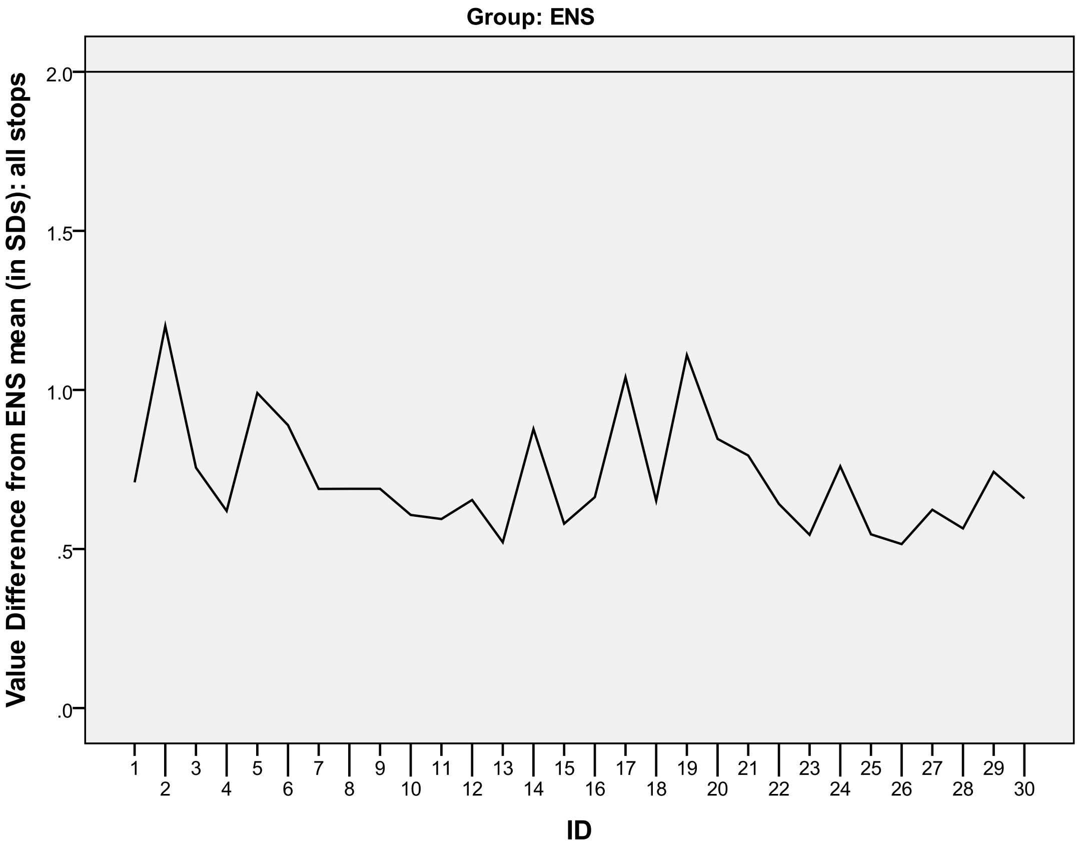


Figure 5.16: Z score differences between individual ENS VOTs and mean ENS VOT: all stops.

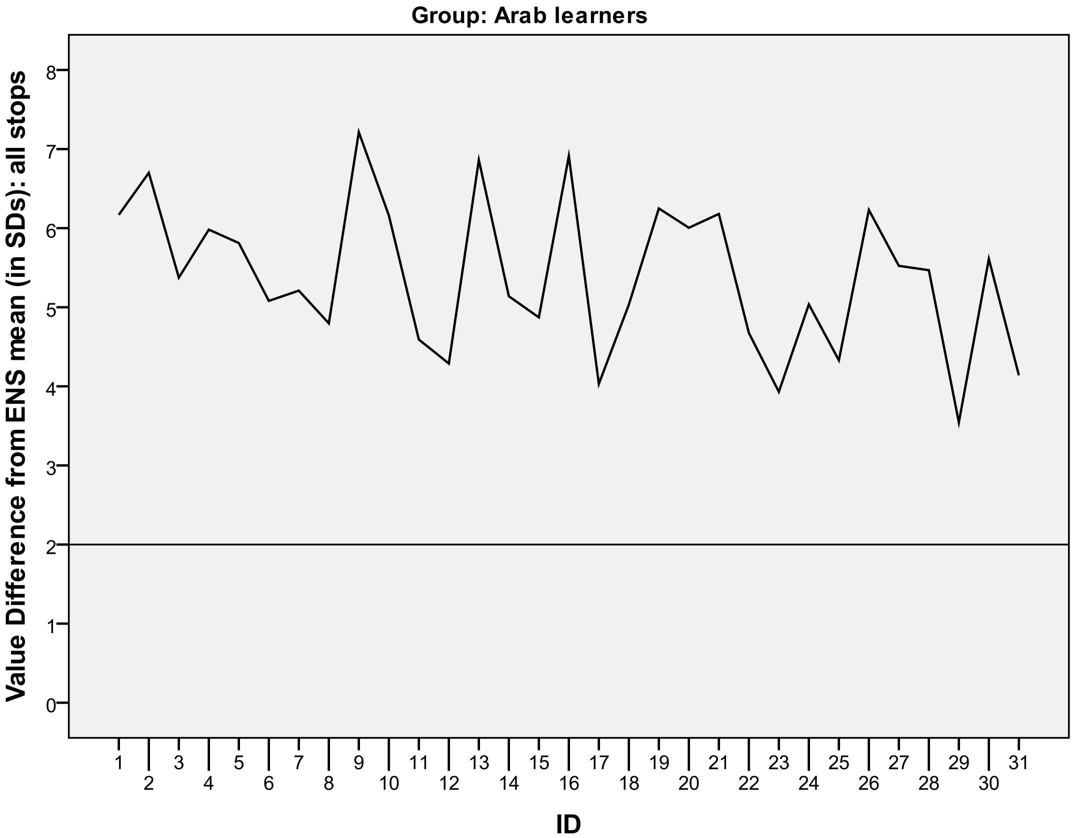


Figure 5.17: Z score differences between individual learner VOTs and mean ENS VOT: all stops.

5.5.2 Findings for individuals on the three English voice-aspiration categories separately

If we look separately at VOTs for aspirated, unaspirated and voiced stops, we see little systematic difference in ENS variation related to which category of stop is concerned (Figure 5.18). The learners however exhibit clearer distinctions (Figure 5.19). Visually there is a clear progression: learners are more distant from ENS on voiced stops, slightly closer on unaspirated voiceless stops, but clearly closer on aspirated voiceless stops, and indeed some individuals fell within the 2 SD area, like ENS.

This therefore supports in a different way what we had learnt from 5.3.1. It supports what we found earlier in the group analyses, that responses centred around /ku:/, in areas where Arabic is more similar to English in VOT, were most nativelike. Nevertheless, we can see that while most of the ENS variation for aspirated stops was below 1 SD away from the mean (Figure 5.18), most of the learner variation was greater than 1 SD away from the ENS mean (Figure 5.19). Still this supports the conclusion that the aspirated stops are the locus of most acquisition.

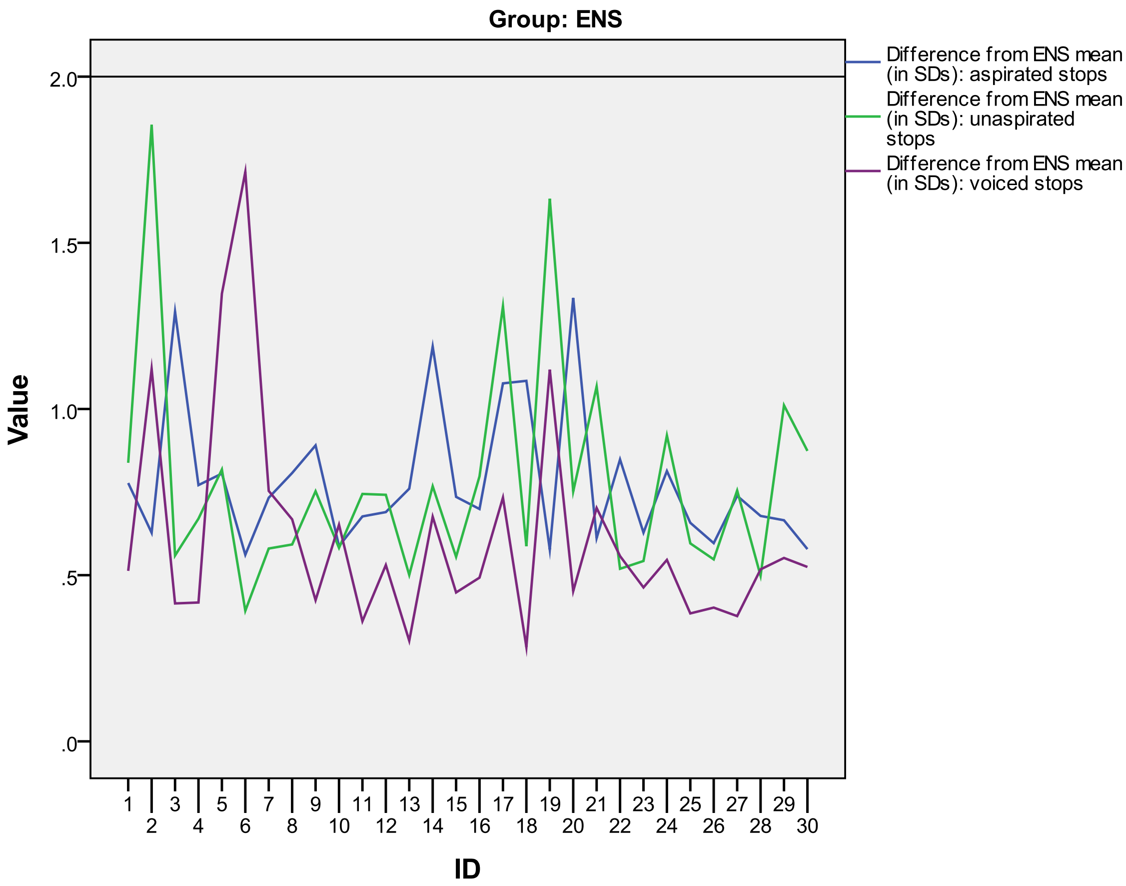


Figure 5.18: Z score differences between individual ENS VOTs and mean ENS VOT: unaspirated, aspirated and voiced stops.

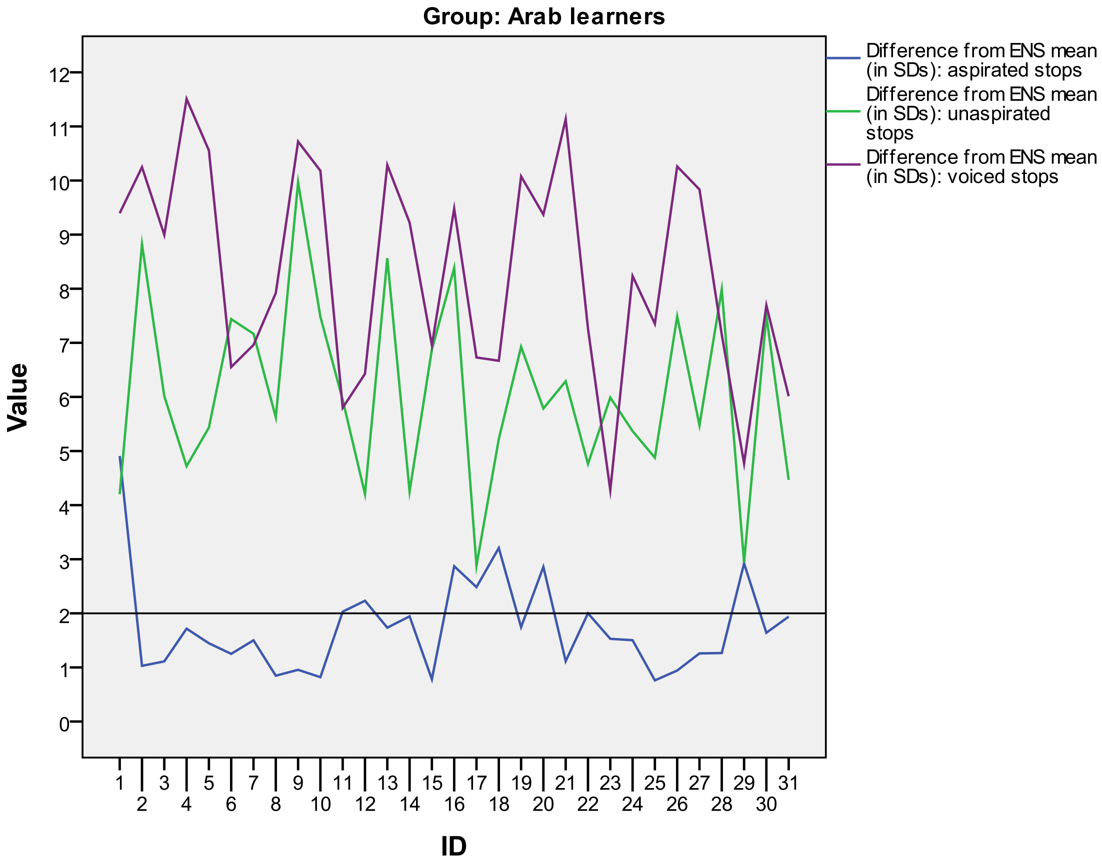


Figure 5.19: Z score differences between individual learner VOTs and mean ENS VOT: unaspirated, aspirated and voiced stops.

Among the learners, individuals number 8, 9, 10, 15 and 25 were the most nativelike on the aspirated stops (though not on the unaspirated where 29 and 17 stood out, nor the voiced where the best were 29 and 23). Of those learners only number 15 has extensive UK residence (6 years). We also examined the correlations in z scores between the three aspiration/voicing conditions. For the ENS we found no significant correlations, supporting the interpretation that native speaker variation in precise VOT produced simply varies randomly, as one might expect.

With the learners, however, this was not the case and two of the three correlations were significant. There was a positive relation between nativelikeness of voiced and unaspirated stops, suggesting that a learner who was closer to the ENS mean on one of those was also closer on the other (r=.390, p=.030). This is also reflected by the parallelism in rises and falls of the top two lines in Figure 5.19. This perhaps then reflects that, as might be expected with learners, those who are better at acquiring one type of sound will be also better at acquiring others. Or it could more specifically reflect the fact that ENS VOTs for voiced and unaspirated voiceless stops are virtually identical. Hence a learner who is able to spot that identity (presumably subconsciously) will improve for both if he improves for either one.

There was however no significant correlation between nativelikeness scores (z scores) of learners for aspirated and voiced stops (r=-.171, p=.357). These categories however are at extremes of the VOT continuum and have nothing in common, so maybe that is to be expected.

Interestingly a significant negative correlation emerged between nativelikeness of aspirated and unaspirated stops (r=-.454, p=.010). The significant negative correlation indicates that learners who were closer to ENS means on aspirated stops were further away on unaspirated ones and vice versa. While initially unexpected, this however can be explained by the fact that, as we saw in 5.3.1, learner VOTs for both these categories tended to fall between ENS values for aspirated (higher) and unaspirated (lower). Hence, of course, if their VOT is closer to one it will be further from the other. Those who lower their VOTs to approach closer to the unaspirated target by the same token move away from the high aspirated target. We can see this clearly for cases 1 and 29 in Figure 5.19 for example.

5.5.3 Findings for individuals on the three POA categories separately

Looking at Figure 5.20, it is apparent that learner VOTs for bilabial stops (regardless of voice/aspiration) usually exhibit lower differences from ENS than the other POAs.

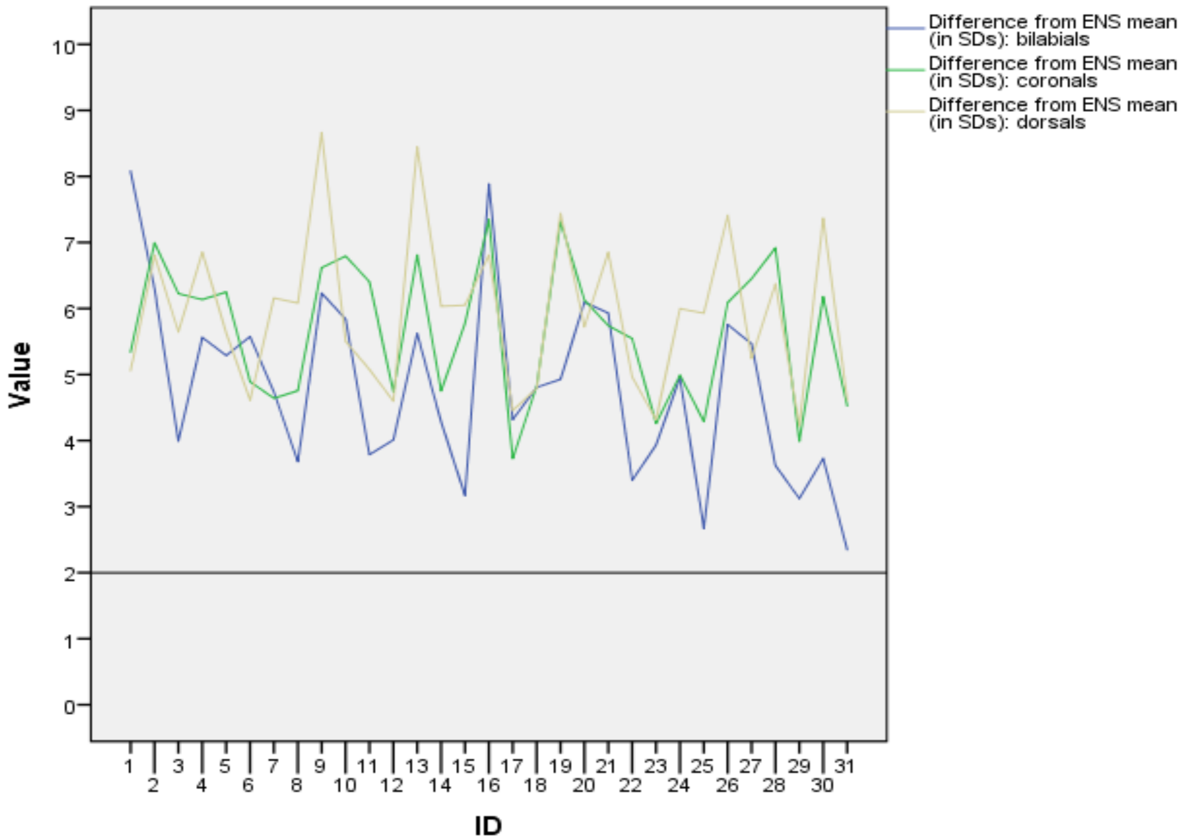


Figure 5.20: Z score differences between individual learner VOTs and mean ENS VOT: POAs.

Learner 31 gets close to the threshold value of 2. This again is consistent with signs elsewhere in 5.3 that bilabials were slightly better acquired than stops at other POAs, which was indeed predicted by the SLM for the voiceless bilabials at least, but also appeared for the voiced ones. All three POAs were significantly correlated with each other. The strongest relationship was between coronals and dorsals. Learners who were more ENS-like on one of those tended to be nativelike on the other as well (r=.644, p<.001). The correlations with bilabials were slightly lower, signalling again that the bilabial POA was the 'odd one out' of the three POAs in our study, as we have already seen (coronal-bilabial: r=.490, p=.005; dorsal-bilabial: r=.357, p=.049).

5.5.4 Conclusion on individual performances

Individual learners do stand out as being more nativelike in VOT than others for individual stop conditions. In many conditions their VOTs are all, however, distant from the ENS mean by far more than ENS individual VOTs are.

One exception is aspirated stops, where some learners differ from the ENS mean by a similar amount to the ENS themselves.

One learner stood out as having a general VOT profile more similar to that of the ENS than that of the learners.

Chapter 6:
DISCUSSION AND CONCLUSION

6.1 Introduction

In this chapter, firstly, we will answer each of the research questions. Secondly, we summarise all the findings concerning our SLM based hypotheses. Thirdly, the chapter concludes by presenting the limitations of the study and some suggested areas for future research.

6.2 Summary answers to Research Questions

6.2.1 **Research Question 1**:

Along with the effects of POA and voice/aspiration, what effects do following vowel quality and context have on the production of stop VOT by native speakers of English and of Saudi Arabic, and by Saudi advanced adult learners of English?

Since this research question involves the three groups of the study, its answer is divided into three sub-sections as follows.

6.2.1.1 Research Question 1 answer for ENS

Since a great deal of work has been done on English, we were merely expecting to replicate the findings of other studies here and indeed that is substantially what happened. As we have seen in 4.2, apart from some vowel findings, results for the English native speakers in this study followed the expected pattern of the English VOT found in the literature (e.g Lisker & Abramson (1967), Klatt (1975), Docherty (1992), Khattab (2002), and with respect to voiceless stops, Scobbie (2002)).

The ENS as expected produced long-lag VOT for aspirated voiceless stops and short-lag for unaspirated stops. The unaspirated stops (with an average VOT of 21 msec) were significantly shorter than the aspirated ones whose average was 80 msec and the difference between the two stop categories was of 59 msec.

The voiced stops were also produced with the expected short-lag VOT pattern found by other studies with some pre-voiced tokens found from a number of ENSs. Pre-voicing occurred occasionally with the coronals and the dorsals, and more in words in isolation than in sentences (see 4.2.2.2). However, there was no significant difference between the VOT of unaspirated voiceless stops and voiced ones, where the former on average were only 1.3 msec higher than the latter (p=.499).

This last phenomenon in English has long been discussed (e.g. Lisker, 2002) since it presents a dilemma. If voiceless unaspirated stops closely resemble voiced stops in VOT, but are very different in VOT from voiceless aspirated ones, as is the case, then why are voiceless unaspirated instances treated as belonging to the voiceless phonemes and not the voiced ones? E.g. why is the second sound in *spit* regarded as an allophone of the first sound of *pit* and not of the first sound of *bit* which has almost the same VOT? Of course, the conventional English spelling favours the usual classification which goes against VOT, but to an extent that is an arbitrary choice. Young English L1 children are often found to use spellings such as *sd-, sg-, sb-* (Treiman, 1985), e.g. *sgie* for *sky*, evidencing that, without knowledge of the accepted spelling, NS do not unambiguously perceive the stop as voiceless in this context. A possible answer is that there exist other phonetic variables which distinguish [b] from [p] beyond VOT, such as amount of air pressure released at the burst, or length of closure, and these mark what phonologists call the voice distinction in English more clearly in this case, and maybe also are more prominent in the end for the ordinary person hearing the sounds. These other variables were of course outside the scope of our study.

The context played a significant role in producing VOT in English, as VOTs of stops in isolated words were on average 5.6 msec longer than stops produced in words in sentences, across all POAs etc. More specifically, our finding is also in line with Docherty's (1992) and Klatt's (1975) findings where they examined ENS producing stops in two different contexts (words in isolation and words in sentences) and found that VOTs in words in isolation were higher than those in words produced in carrier phrases.

There was however no overall effect of following vowel quality on VOT of all English voiced and voiceless stops. Nevertheless, there were significant interaction effects. For example, VOT before /i:/ seems to increase across POAs from front to back more steeply than does POA before the other two vowels. Furthermore, a following vowel relates to VOT differently for aspirated voiceless stops than for unaspirated/voiced ones, and in each case, it follows a different pattern across places of articulation. Although, there was some variation in the sequence of vowel patterns with increasing VOTs, and as we stated this variation was not significant as a simple overall effect for the stops, there was a tendency for high vowels /uː, iː/ to evoke higher VOT than the low vowel /ɑː/, for the unaspirated or voiced stops at least (see 4.2.1.4).

These findings are broadly in line with Lisker and Abramson's (1967) study, as they found that there is no major influence on VOT of the subsequent vowel. Other studies however have found that there was an effect of vowels on preceding stops in English.

It is generally established from previous research (Klatt, 1975; Weismer, 1979; Rochet & Fei, 1991; Chao et al., 2006) that word-initial stops have longer VOTs when followed by high vowels than by low vowels. This has only been confirmed in our findings with the unaspirated stops and to an extent the voiced stops, while in the aspirated stops, one of the two high vowels is always associated with the shortest VOT, shorter than that of the low vowel /ɑː/ (see again Figure 4.1). The reversal which occurs for the aspirated dorsals is particularly striking. While for labial and coronal aspirated stops /i:/ evokes the lowest VOT, compared with the other vowels, for the dorsals it evokes the highest. We have no explanation for why no other study seems to have identified this before.

6.2.1.2 Research Question 1 answer for ANS

Arabic native speakers produced Arabic voiced stops with negative VOT. The VOT values of voiced stops found in this study are in line with other studies which investigated Arabic and found that Arabic speakers pre-voice their voiced stops (i.e. Yeni-Komshian et al, 1977; Jesry, 1996; Radwan, 1996; Khattab, 2002; Flege and Port, 1981; and Alghamdi, 1990). Although Yeni-Komshian et al (1977), Khattab (2002), Jesry (1996) and Radwan (1996) examined non-Saudi dialects of Arabic (Lebanese and Syrian and MSA) they still found the same results for voiced stops. More specifically with Saudi Arabic, Flege and Port (1981) and Alghamdi (1990) also found that Saudi Arabic speakers pre-voice voiced stops.

ANS voiceless stops were mostly with positive long-lag VOT. Saudi Arabic speakers aspirated nearly all voiceless stops as their VOT ranged between the upper end of the short-lag region (28 msec) and high aspiration (131 msec). However, most of their VOT in voiceless stops fluctuated in the lower end of the long-lag aspiration area (between 40 msec and 70 msec). Since the only short-lag VOT productions among the ANS stops were for the voiceless stops, not the voiced ones, the VOT difference between their voiced and voiceless stops was colossal (consistent with Yeni-Komshian et al, 1977; Jesry, 1996; Radwan, 1996; Flege and Port, 1981; and Alghamdi, 1990; Khattab, 2002).

Our results for voiceless stops in Saudi Arabic correspond with other studies conducted on Saudi Arabic such as Flege and Port (1981) and Alghamdi (1990), as both studies found that voiceless stops in Arabic were aspirated. However, both studies found shorter VOT values than our study VOT values and Alghamdi (1990) found even shorter VOT values than Flege and Port (1981). Consequently, VOT values of Saudi Arabic voiceless stops in this study were found to be longer than VOTs found in all the previously mentioned studies.

We explained (4.3.1.1) that the reason for such variation in VOT of stops could be due to the different dialects which other researchers have looked at in their studies, as Flege and Port looked at the central Najdi dialect and Alghamdi looked at the southern Ghamdi dialect. Therefore, we seem to have detected some variations in VOT even between groups of people who speak different sub-dialects within the same country.

The overall effect of POA when both voiced and voiceless stops were considered was also significant, but this was due only to the two voiceless stops, since there was no significant difference between the three voiced stops. The previous Saudi studies provide no comparable significance figures to compare with.

There was no significant overall effect of context although, descriptively VOTs were marginally longer in stops in words than in sentences. This therefore follows the same form as the English speakers, with similar SDs, but the differences in means between contexts are smaller in Arabic (respectively 5 and 2 msec for ANS coronal and dorsal versus 9 and 16 msec for ENS). Underlying this, descriptively, however, there is a striking pattern that the voiceless stops have longer positive VOT in words in isolation than in sentences, while for voiced stops the pre-voicing found in the Arabic voiced stops is always again somewhat longer in words in isolation than in sentences. This interestingly suggests that the context-related pattern is not that VOT is more positive (including less negative) in words in isolation than in sentences, but that it is more extreme (more distant from the burst/zero VOT) in words in isolation than in sentences (see further discussion of this distinction below). This, as we mentioned earlier, would be consistent with the interpretation that words in sentences are spoken faster than words in isolation, so duration of VOTs gets shortened in sentences, irrespective of whether they are in the positive or negative area of the VOT continuum.

There was no overall effect of vowel, but there were some significant interaction effects particularly involving vowel quality when looking at each stop category (voiced and voiceless) separately. Indeed, there was a substantial effect for the following vowel on Arabic VOT, as in both voiced and voiceless stops ANS produced significantly different VOTs depending on the succeeding vowel, regardless of other factors.

For example, within the ANS voiceless stops and along with the highly significant effect on VOT of POA, there is a considerable simple effect of following vowel on Arabic VOT, as their voiceless stops differ significantly overall depending on following vowel. High vowels generated longer VOT in the preceding stop than the low vowel in general, though it was really only the voiceless dorsals that showed this clearly. In any case this does not match what we found above for the voiceless aspirated stops in English.

Similarly, following vowel had reasonably small but significant effect in voiced stops, regardless of other factors. Moreover, the pattern found of decreasing overall pre-voicing (though not replicated in all POAs etc.) was iː > uː > ɑː with no significant difference between /uː, ɑː/. This also matches the overall order of the vowels in terms of decreasing positive VOT in the voiceless stops. Hence, we can assert that high vowels and particularly /iː/ is associated with longer VOT than the other vowels.

It has come to our attention, however, that discussion with regard to whether patterns match or not between voiced and voiceless sounds in Arabic runs into conceptual difficulties due to the voiced stop VOT values all being negative. In the last two paragraphs, we made the assumption that one can say that patterns match if the sound which has the larger positive VOT for voiceless stops has the largest negative one for voiced, and the sound with the lowest positive value for voiceless has the smallest negative value for voiced. On this argument, the vowel pattern is similar for voiced and voiceless stops in our variety of Arabic. That interpretation in effect uses the standard of how far the VOT is from the burst of the stop (where VOT=0). We claim that a consistent pattern is found if voiceless sounds which have VOTs further from the burst on the positive side correspond to the voiced sounds whose VOT is further from the burst on the negative side.

However, it is alternatively possible to take the view that what should be regarded as matching a larger positive VOT among negative VOTs is the most positive of the negative VOTs, in other words the least negative (not as above the most negative), and so forth. On this simple numerical view, which has no regard for the burst, the negative VOTs associated with voiced stops before different vowels in the ANS data are a mirror image of those associated with voiceless stops: /i:/ would have to generate the most positive VOT among voiceless stops but the most positive (i.e. least negative) among the voiced stops. We have not found this issue discussed in the literature.

It should be noted that it is because the match between vowel effects in voiced and voiceless Arabic stop VOT is of the former type, not the second, that when both voiced and voiceless stops are combined, the apparent influence of vowel on VOT disappears and overall effect of vowel becomes non-significant.

Yeni-Komshian et al. (1977) examined Arabic VOT in the context of three vowels, as tested in the current study, and found (in a different dialect from ours) that there was a tendency for the high vowel /i:/ to generate shorter negative VOT and longer positive VOT in the production of some of the Arabic stops than the other vowels /a/ and /u/ (see Table 2.3). This of course does not match our finding with respect to the voiced stops. To our knowledge, no other studies reported results on the influence of vowel on Arabic VOT, which also shows the importance of our study findings on post-vocalic effects in Arabic VOT.

6.2.1.3 Research Question 1 answer for the L2 learners

Learners of English produced both aspirated and unaspirated stops in long-lag region with a small but significant difference between them (mean 8.6 msec). Voiceless aspirated stops were produced in the lower long-lag region. Unaspirated stops were also produced with aspiration but in lower ranges than the aspirated stops. Voiced stops were produced predominantly with negative VOT though some short-lag tokens were recorded.

Results for the voiced and voiceless stops of the learners in this study are similar to Flege & Port (1981) and Alghamdi (1990), so confirm their finding that the majority of the VOT values produced by Saudi Arabic learners of English in voiced stops were pre-voiced and voiceless stops were aspirated, though with higher aspiration found than in both of those studies.

An effect of context on VOT was not found. The learners produced stops in both contexts similarly. This finding is in line with Yeni-Komshian et al (1977) who investigated Lebanese MSA where they found no major differences between producing stops in words in isolation or in sentences. Lastly, the following vowel had a major effect on the VOT of the preceding voiceless and voiced stops of the learners. But when voiceless and voiced stops were analyzed separately, no overall effect of vowel on voiced stops was found. There were differential vowel effects at different places of articulation, however, for both voiced and voiceless stops. This finding however, disagrees with Lisker and Abramson (1967) where they found no major effect on VOT of the following vowel but in agreement with many other studies such as Klatt (1975), Weismer (1979), Port (1979), Rochet & Fei (1991), Chao et al (2006), Schmidt (1996), Johnson & Babel (2010), Iverson et al (2008) etc. As all these studies established that an influence of vowel was found on the preceding stop. The former five studies found that high vowels generate longer VOTs than the low ones while the latter three ones found the opposite finding (i.e. low vowels generate longer VOTs than high vowels).

The overall learner pattern for the vowels (regardless of POA, aspiration and context) is that mean VOT was longer for voiceless stops before the high vowels /u:, i:/ and shortest for the low vowel (see Figure 5.5). These overall differences between vowel effects on VOT, disregarding POA (and aspiration and context), were all significant: /iː - ɑː/ $p=.002$; /iː - uː/ $p<.001$; /ɑː - uː/ $p<.001$, see (5.2.1.3), although in fact in detail this was only supported by the coronals and dorsals not the labials.

However, the vowel had no significant effect with the voiced stops produced by learners in English, as learners were not making a clear overall VOT distinction between vowels. Difference of vowel alone, regardless of other factors like context and POA, does not lead to a significant overall difference in learner voiced stop VOT (means /iː/ -60, /ɑː/ -56 and /uː/ -63 msec), see (5.2.2.2). Descriptively, however, the low vowel had the shortest negative VOT, thus matching, in the 'distance from burst' sense (see discussion above) the lower positive VOT of the low vowel for the voiceless stops.

6.2.1.4 Vowel effects in the three groups overall

In all three groups studied, the vowel effects were always the most complex and engaged us in uncovering a depth of detail that we have not found matched in reports in the literature for any of the three types of participant.

Our findings for the learners' voiceless stops are broadly in line with other mostly L1 studies which found that there was an effect of the vowel on the VOT of stops, Klatt (1975), Weismer (1979), Port (1979), Rochet & Fei (1991), Schmidt (1996) Johnson & Babel (2010), Chao et al (2006), and Iverson et al (2008), but not with Lisker and Abramson (1967) who found that there was no major influence of vowels on speakers VOT, although Lisker and Abramson's study was on learners of English from other languages not on Arab learners of English, as we described in the literature review (see 2.7). Those who found a difference tended to find high vowels generated significantly longer VOTs than low vowels, but the accounts generally do not extend to negative VOT instances as ours do.

Port (1979) who examined the VOT in English stops, Rochet & Fei (1991) who studied Mandarin stops, and Chao et al (2006) who studied English and Mandarin VOT found that vowels had a significant influence on the VOT of the preceding consonants and VOT was longer when followed by high vowels /i - u/ than low vowel /a/ in word-initial stops. However, VOT was found to be longer when followed by the low vowel /a/ than the high ones /i – u/ in Swedish (Fant, 1973). Our findings in the end do not fully support either of these positions in any of the groups in an unqualified way.

All in all, our research shows that following vowel effects are not simple in any of the three groups studied. In particular, which vowel evokes the higher VOT of the preceding stop often cannot be stated validly as a general pattern but varies depending on voice and POA of the stop. Research on the influence of vowels on the VOT is ongoing, and needs further investigation with reference to a number of languages particularly with Arabic and Arab learners of English.

6.2.2 Research Question 2:

Does POA for all those three groups, regardless of following vowel, context or voice/aspiration, always involve an increase in positive VOT / reduction of negative VOT of stops front to back, in accordance with a claimed universal trend?

Again, the answer to this RQ involves all the three groups of the study, therefore, its answer is divided into three sub-sections as follows.

6.2.2.1 Research Question 2 answer for ENS

The English voiceless stops followed the universal VOT pattern found by Chao and Ladefoged (1999) and others based on place of articulation, with increasing VOT front to back. All voiceless stops, whether aspirated or not, were produced with significantly different VOT values from each other p < t < k. The same was true for the voiced stops b < d < g.

Additionally, the POA effect was also well-preserved in the two contexts, as all VOTs for voiceless (unaspirated and aspirated) and voiced stops increased across places of articulation (labial < coronal < dorsal) in both contexts (word in isolation and in sentence).

Several studies conducted on English VOT all confirmed the POA effect of increase from front to back, however, the only exception was Klatt (1975) who strangely found that the dorsal sounds were shorter in VOT than the coronal ones.

6.2.2.2 Research Question 2 answer for ANS

The ANS data straightforwardly confirmed the universal POA trend of increasing positive VOT with respect to the voiceless coronal and dorsal stops, but of course could not tell us anything about labials. Dorsals were on average 14 msec longer than coronals in VOT. This /t<k/ pattern was consistent with all the Arabic studies reviewed in the literature (i.e. Alghamdi, 1990; Khattab, 2002 etc.).

The voiced stops did not present a straightforward picture, however. POA had no significant overall effect on VOT, with similar overall means at each POA: /b/ -78, /d/ -82, /g/ -79 msec. Even descriptively this does not present any trend front to back.

We note however that the literature did not in fact provide us with a clear picture of what the expected trend would be for pre-voiced stops, in the sense of what pattern would be regarded as matching the trend found for positive VOT of voiceless stops in languages. Following on from the discussion above in relation to vowels, we could argue two ways. We could adopt the 'distance from burst' stance, arguing that because on the positive side /p/ (when present in languages) has the lowest VOT, closest to zero, and /k/ the highest, then on the negative side /b/ should have the smallest negative value, again closest to zero, and so on. That would yield a predicted sequence /b>d>g/ (e.g. -60, -70, -80). On the other hand, disregarding the burst, we could take the view that because /p/ has the lowest VOT on the positive side (in languages that possess it), /b/ should have the lowest VOT on the negative side in the sense of the greatest negative value, and so on, yielding a predicted sequence /b<d<g/ (e.g. -80, -70, -60). In fact, of course the observed descriptive sequence in our data for ANS is neither of those.

It should be noted that some of the studies in our review in fact found a decreasing pattern of pre-voicing from front to back in voiced stops, following the second suggestion above. Interestingly Scobbie (2002), in the unusual dialect of English which he studied, where voiced stops were pre-voiced, also found this pattern (see Table 2.15) with mean VOT values -29, -25, -6 msec. In the Arabic studies, again this pattern of decreasing pre-voicing across POAs from front to back was found in Yeni-Komshian et al, (1977), Jesry (1996), Flege and Port (1981), and Alghamdi (1990). Radwan (1996) and Khattab (2002) however found an irregular pattern like ours.

Overall, then, the pattern of increasing or decreasing VOT across POA in voiced stops in Arabic is not universally established in the literature compared with voiceless stops where there is some degree of an agreement that VOTs increase from front to back.

The voiced POA VOT data from the current study of a northern dialect of Saudi colloquial Arabic does not completely fit any of the previous trends found with other studies. If we argue that the pattern we should expect to find is the one matching the posititive sequence in terms of distance from burst, then the coronal /d/ could be seen to be the odd one out: it should be less negative around -78.5. There could be other reasons for this. One is that the POA of the Arabic /d/ in our dialect is in fact dental not alveolar, however, even if that is the case that would still leave /d/ in the middle between the labial and the dorsal in POA. Hence that should not affect any VOT trend. Another reason for why Arabic /d/ was the highest in pre-voicing VOT among the other voiced stops could be the distribution feature as indicated in the literature, since the Arabic /d/ is said to be [+distributed] which may make it generate more pre-voicing VOT than the English [-distributed] (cf. Cho and Ladefoged (1999), second point cited in 2.5.2).

The interpretation we prefer, however, is that the expected sequence for pre-voiced POA is the simple one, starting from /b/ with the most prevoicing, with successively less negative/more positive VOTs at /d/ and /g/. Under this scenario, in the ANS data /b/ is the odd one out as its VOT is too positive/not sufficiently negative (i.e. it should be around -85 msec, not -78). However, there is a potential explanation for this in the absence of /p/ for the ANS. This could be argued to have removed the need for /b/ to mark its distinction from a voiceless counterpart in the usual way, parallel to /d/ and /g/. Instead it is free to move to a less extreme (less pre-voiced) position in the VOT continuum for labials, without danger of a distinction from a voiceless counterpart getting lost. The difficulty with this explanation, however, is that on the whole the other Arabic studies cited in chapter 2 did not show this.

This all points to the need for further investigation of this issue in future research on Arabic VOT, and indeed in languages with prevoiced stops in general.

6.2.2.3 Research Question 2 answer for L2 learners

The overall main effect of place of articulation on VOT was very clear and highly significant in the production of English stops by the L2 learners. A separate analysis of the voiceless stops from voiced stops, disregarding other variables (aspiration, vowel and context), showed that positive VOT increased front to back for the voiceless stops (both aspirated and unaspirated), following the expected trend. All places were also significantly different from each other and the pattern therefore accords well with the universal expectation for voiceless stops which we also showed for ENS and ANS in voiceless stops. Only, the precise VOT values of the learners diverge more from ENSs than from ANSs.

For the voiced stops, there was also a significant effect for POA, and all three places of articulation differed significantly from each other, disregarding vowel and context variation. However, the pattern did not fit either of the hypothetical expected POA trends for pre-voiced stops which we discussed above. Rather it exhibited the pattern with most negative VOT for /d/, followed by /g/ then /b/. This therefore resembles the ANS finding except that for the learners it was highly significant rather than just a barely detectible descriptive observation. The same arguments and potential explanations therefore apply here as were advanced in 6.2.2.2.

6.2.3 Research Question 3.

What stops (differing in POA, voice/aspiration, following vowel and context), if any, do advanced adult Saudi learners of English produce with different VOT from Arabic native speakers? What stops do they produce with similar VOT to English native speakers?

We proceed now to answering RQ3, which is about similarities or differences between learners and other monolingual groups in precise VOT means. This is the most fundamental question in any study which, as ours does, wishes to ascertain if a group of learners have in fact reached the desired end point of becoming nativelike in whatever is being acquired. If that was the case, we would expect many non-significant differences between learner VOT values and those of NS of the L2, and many differences between learner VOT and those of L1 NS. We calculated the differences between learners and ENS on 54 comparable specific combinations of our variables (e.g. VOT for production of a particular stop voiced, with a specific POA, before a specific vowel, in words in isolation), and between learners and ANS on 42 comparable combinations of conditions.

The vast majority of the comparisons (89%) showed a significant difference in VOT between the learners and ENS, and in only 11% of the specific conditions could the learners be regarded as being in this sense native-like in VOT. The combinations where the learners were nativelike, were all for voiceless and mostly aspirated stops. There appears to be a small core of incipient nativelike production based around /ku:/ and other voiceless aspirates especially + /u:/. A couple of these instances, however, are where the VOT of ENS and ANS do not differ (/ku:, ki:/ in sentences). Hence, these can be seen as instances where, if the learners do not change from their L1 values, they nevertheless produce VOTs like those of ENS. In short, positive transfer is available here, which is not the case in the vast majority of comparisons, where ENS and ANS VOT values differ.

By contrast, in 67% of the comparisons learners were not significantly different from ANS. Only 33% of the comparisons showed a significant difference between the learners and ANS. The majority of these differences between learners and ANS were for words spoken in sentence contexts, involving primarily voiced plosives. Learner mean VOTs for each condition descriptively do however mostly fall between the generally shorter or negative ANS ones and the longer ENS ones, suggesting slight movement in the desired direction. Furthermore, comparison of the 33% learner-ANS differences with the 11% learner-ENS non-differences suggests that learners have departed from their ANS values to a greater extent than they have arrived at ENS values.

In overview, what we predominantly found is that the voiceless aspirated stops of learners had VOTs which were positive but at a similar low long-lag VOT range to ANS rather than the high long-lag VOT range of ENS. For voiceless unaspirated stops, with no L1 counterpart, the learners were higher than ENS and for voiced ones, the learner VOTs were negative (pre-voiced), like ANS, not positive (short-lag) like those of ENS, even though we did find some individual learners producing tokens with positive VOT.

On this criterion, then, little acquisition has occurred, lending support to the idea that VOT is something very difficult for a learner to change (Flege, 1980). However, although we did not test it, we must assume that, functioning as they do successfully in an L2 environment both for study and daily living purposes, our learner participants nevertheless produce stops in ways that are recognisable to ENS for normal communication purposes. Possibly, then, for such communication, fully nativelike VOT is simply unnecessary, since there are other features of stops which learners may produce in a more nativelike way (intensity, velocity, etc.), which are beyond the scope of our study, but may be sufficient to support effective communication.

We also looked briefly at the learners' individual results, which we found were consistent with the group results with the exception of only one learner who was to some extent native like in more of his English L2 VOT productions. The learners always pre-voice their English voiced stops and produce their voiceless aspirated stops with shorter VOTs than ENS, like their L1 stops. However, some learners managed to produce some more authentic L2 English stop VOT tokens, particularly for /b, p/.

6.2.4 **Research Question 4**.

Regardless of whether learner actual stop VOT values are similar to those of Arabic native speakers, do adult Saudi learners of English make VOT distinctions for POA, voice/aspiration, following vowel and context that are parallel with those made by monolingual Arabic native speakers? Or, regardless of whether their stop VOT values are similar to those of English native speakers, do adult Saudi learners of English make VOT distinctions for POA, voice/aspiration, following vowel and context that are more parallel with those made by English native speakers?

We now consider the target of learner acquisition in a different sense, that of whether, regardless of absolute VOT values produced, learners make significant VOT distinctions between all the same sounds that ENS make distinctions between. As we described in many places, ENS fundamentally distinguish significantly in VOT between six sounds, disregarding vowel and context factors. These are the voiceless aspirated stops at three POAs, and the voiced/unaspirated voiceless stops at the same three POAs. They also make an overall VOT significant difference between contexts, while the vowel related differences are more complicated.

Considering now what our learners have learnt to do, up to the time when we measured them, again we find that overall what they distinguish is overall less like ENS, and more like ANS, although the difference is perhaps less extreme than when they are judged on the criterion of producing the same absolute VOT (6.2.3).

On the positive side the learners have of course learnt to produce a voiceless /p/, which they distinguish significantly in VOT both from /b/ and /t/ etc., despite the lack of this sound in L1. They also distinguish /p/ from the other voiceless stops by its lower rather than higher positive VOT, again consistent with ENS in relative terms, although that might be argued to be a language universal effect rather than evidence of learning the L2. More broadly, they make some sort of significant VOT distinction marking POA difference for all the stops, including the voiceless unaspirated ones which are new to Arabic learners, even though for voiced stops it is /d/ that has the lowest/most negative VOT rather than /b/ as in English. They also resemble ENS in some minor ways, such as not exhibiting a context by vowel interaction effect on VOT, which ANS do.

On the negative side, however, they differ from ENS in major ways. First, whereas ENS distinguish in VOT between voiceless aspirated stops on the one hand and voiced/voiceless unaspirated ones on the other, learners make a significant difference between all three voice/aspiration options. Thus, they distinguish 9 rather than 6 basic stops, in VOT terms. We may assume of course that at the phonemic/phonological level they are making the same distinctions as ENS (and ANS, apart from lack of /p/), just between the three voiced and three voiceless stop phonemes, but of course in the present study we have been concerned rather with what distinctions they mark with significant differences in VOT. That is not the same issue as what phonemic distinctions they mark, since those phonemic distinctions (e.g. voiced - voiceless and POA) may well be marked additionally, and perhaps more clearly, by other phonetic features of stops which are outside the scope of our study.

Second, as we said, the order of the voiced stops marked by learner significant differences in VOT is not the same as that in English but if anything more like ANS. Third, learners make no context related VOT difference which ENS do and again ANS do not. Furthermore, descriptively learners produced slightly longer VOTs in sentences than isolated words, the reverse of ENS.

Thus, learners may be seen to be working with a system of VOT-marked, phonetically distinct, categories rather different from that of ENS.

6.3 Summary reports on the hypotheses

The study was conducted in part with the aim of testing the predictions of the SLM, from which we formulated a set of hypotheses for our own study (2.12) (RH1 to RH7). Therefore, in the following sections, we will recapitulate the findings of the study in light of our predictions based on SLM.

6.3.1 **Research Hypothesis 1.**

Same sounds. Any English sounds that are objectively the same in VOT in English and Arabic will be perceived by advanced Saudi learners to be identical to ones in Arabic and will be placed in the same category as the L1 sounds (equivalence classification). Hence wherever in our data VOTs are in fact not significantly different between monolingual Arabic and monolingual English, we will expect that learners' VOTs will not be significantly different from those of either L1 or L2: this represents successful learning through positive transfer. (SLM H1 H2 H5 H7)

The first hypothesis (RH1) is related to the acquisition of ‘same sounds’ which are virtually identical in both L1 and L2 (2.11). This hypothesis, based on the SLM, predicts that any English sounds that are perceived by the learners to be identical to ones in their Arabic, will be placed in the same category as the L1 sounds (equivalence classification). Therefore, when the two monolingual groups are in reality not significantly different from each other in particular sounds, the prediction (RH1) is that learners will not also be significantly different from either of those L1 or L2 groups and will perceive such sounds as the same and simply identify the L2 sound with their L1 sound without forming a new category for it, and go on to produce the target sound correctly.

We had not really anticipated in advance, based on the literature, that our data would yield any instances at all where ANS and ENS VOTs were the same. However, as described earlier there were in fact two specific instances where this occurred, so this hypothesis could be tested (/ki: ku:/ in sentence contexts). In both cases learners were not significantly different in VOT from ENS, thus supporting RH1.

Learners did however differ significantly on /ku:/ from ANS, because their mean VOT was higher than that of both groups (80 msec), not between the ENS mean (75 msec) and ANS mean (72 msec), though not significantly different from ENS. This might be seen as slightly detracting from full support for RH1 since if the learners truly had only one category for [k^h] before /u:/ in words in sentences, then there would indubitably be no significant VOT differences between any of the three groups. A possible explanation, however, could be that the L1 value of the learners had shifted slightly in the process of learning English and was no longer in the VOT position which we measured from monolingual ANS. It is recognised in the SLM as well as more widely (Cook, 1992) that when someone learns an L2, their L1 can change slightly. Possibly in this instance the learners' L1 VOT value had shifted slightly and would not be significantly different from their L2 value for /ku:/. Since however we only measured Arabic VOTs from a monolingual group of ANS, not from the VOT production in Arabic of our learner participants, this must remain speculation.

6.3.2 **Research Hypothesis 2**.

New sounds. Since neither English aspirated [p^h] nor unaspirated [p] are found in Saudi Arabic, and there is no very similar Arabic sound which would be transcribed with the same IPA symbol, these English sounds will have been easily perceived as new, and advanced adult Saudi learners of English will have formed new categories for them and will produce them with VOT separate from [b] or [t^h], close to ENS values. (SLM H1 H2 H3 H7)

The second hypothesis (RH2) is about the acquisition of 'new sounds' (2.11), and claims that L2 learners can acquire a new phonetic category for an L2 sound if they perceive some clear difference between the new L2 sound and any potential corresponding L1 sound. As RH2 states, English /p/ fits this profile.

Therefore, according to the prediction of this hypothesis (RH2), the VOT of the English /p/ in any combination of conditions would be easily acquired, because 'new' sounds such as this are, with time, easier to perceive and establish a suitable new category for than sounds that are perceived as 'similar' to L1 sounds.

The results of the learners for the acquisition of the new voiceless labial sounds do indeed evidence success in the sense that they have established a new category, distinct from /b t/ etc. The actual VOT values they have ended up with, however, are significantly different from those of ENS in many instances. They were not different in only two instances, which does not completely support the SLM predictions: [pi:] in isolated words and [p^hu:] in sentences. In all other instances, /p/ followed the pattern seen in the learners' stops across all POAs of being significantly lower in VOT than ENS in the aspirated stops and being significantly higher in the unaspirated ones. The learner VOT productions for the new voiceless labial stops seemed therefore to be mainly running parallel with other voiceless stops rather than being especially more target-like due to being totally 'new'. This could perhaps in part be explained as due to 'deflection', another SLM concept which we will consider below (RH6a).

6.3.3 **Research Hypothesis 3**.

Similar sounds, noticeable difference. English unaspirated voiceless stops [t k] do not exist in Saudi Arabic but are similar to voiceless weakly aspirated stops [t^h k^h] which do exist in Arabic. Since a difference between non-aspiration and aspiration however is relatively noticeable, advanced adult Saudi learners of English will have formed new categories for them and will produce [t k] with significantly shorter positive VOT than [t^h k^h], though perhaps not entirely similar to ENS values. (SLM H1 H2 H3 H7)

Research Hypothesis 3 of our study (RH3), is about 'similar sounds' (2.11) that exist between L1 and L2 sounds but that have some noticeable difference between them. One case of such sounds in our study, as the RH states, is the English unaspirated voiceless stops [t k], which do not exist in Saudi Arabic but are similar to the voiceless weakly aspirated stops [t^h k^h] that do exist in Arabic.

These unaspirated voiceless English stops [t] and [k] could also be possibly classified as 'new' sounds, though they are not as clearly new as the labial sound /p/. These sounds are new on the IPA symbol test if we transcribe the Arabic voiceless stops as [t^h] and [k^h], but could be 'similar' if we transcribe the Arabic stops as [t] and [k]. However, we have already seen that the Saudi Arabic voiceless stop VOT in fact is found to be in the higher end of short lag but mainly in the lower end of long-lag. Hence we prefer the [t^h k^h] transcription for them.

The learner results showed that they did succeed in establishing distinct VOT categories for [t k] significantly lower than those for their English [t^h k^h], although the values were all significantly higher than ENS ones for [t k]. The learners' English voiceless unaspirated stops however are not all significantly different in VOT from the corresponding L1 sounds. In fact, of the 12 relevant instances only half are significantly different from L1. Thus, the data here places the English unaspirated voiceless stops on the borderline between equivalence classification with L1 and a new phonetic category, while the SLM would seem to predict the latter. Generally, the results again partially support RH3, though it remains to be explained why the values are closer to the L1 than the L2 position on the VOT continuum. Again, we will talk about the idea of deflection under (RH6a).

In conclusion on RH3 we should say that an alternative line of argument, which we considered but did not adopt, was that since the VOTs of unaspirated voiceless stops do not differ from those of voiced stops in English, learners would treat them in the same way as the SLM predicts for voiced stops (RH5). Clearly this did not happen.

6.3.4 **Research Hypothesis 4**.

Similar sounds, less noticeable difference: aspirated stops. Both English and Saudi Arabic have voiceless aspirated stops [t^h k^h] although Arabic ones are produced at the top of the short lag region or in the lower part of the long lag VOT, while English ones are produced in the high long lag region. Since this sort of a difference in aspiration is relatively less noticeable, advanced adult Saudi learners of English will not have formed new categories for English [t^h k^h] and will produce them with VOT much the same as ANS (equivalence classification). (SLM H1 H2 H3 H5 H7)

The fourth hypothesis of our study RH4, is again about 'similar sounds' that exist between L1 and L2 but where those sounds have some less noticeable difference in their phonetic realisations. In this case the hypothesis predicts that learners will not have formed new categories for them and will produce them with VOT much the same as L1 sounds, so the mechanism of ‘equivalence classification’ will be operative and will block the formation of new categories for those L2 sounds. In this discussion, we have to exclude the two cases of 'same' classification covered under RH1.

The results for the learners on English aspirated stops [t^h] and [k^h] showed that out of the 10 relevant production conditions for these sounds, 7 showed significant difference in VOT from L2 stops produced by ENS and no significant difference from L1 stops produced by ANS, which is in line with our prediction in RH4. As learners could not perceive the difference between their L1 aspirated sounds and the L2 English ones, therefore, an equivalence classification was made which prevented them from learning.

However, there were three other cases involving [tʰu:] and [kʰu:] in words and sentences which in various ways contradicted predictions of equivalence classification by being significantly different from ANS and/or not significantly different from ENS. This would suggest the establishment of new category. Such instances suggest that the learners managed sometimes to perceive the small phonetic difference found between L1 and L2 in voiceless aspirated stops even though it was thought to be unnoticeable within the long-lag range. Therefore, learning had occurred in these few conditions. Really one would have to say that with the voiceless aspirates there exist, for Arab learners, a cline of possibilities in SLM terms, depending on the detailed conditions. This throws doubt upon the SLM enterprise of trying to digitize in categories what is really an analog scale of learning possibilities (Table 6.1).

Table 6.1: Acquisition of voiceless aspirates

Actual VOT	ANS = ENS = L	ANS ≠ ENS = L	L = ANS ≠ ENS
Hypothesis	RH1	RH4	
SLM acquisition category	Same in L1 and L2	Similar in L1 and L2 with a slightly noticeable difference	Similar in L1 and L2 with nonoticeable difference
SLM predicted outcome	Equivalence classification	New category forming	Equivalence classification
Specific stop conditions found in each category	ki: ku: (WS)	tu: (WS) ku: tu: (WI)	ti: kɑ: tɑ: (WS) ki: ti: kɑ: tɑ: (WI)
Acquisition success	Nativelike	Close to nativelike	Non-nativelike

Note: WS = word in sentence context; WI = word in isolation context

6.3.5 **Research Hypothesis 5**.

Similar sounds, less noticeable difference: voiced stops. Both English and Saudi Arabic have voiced stops /b d g/ although Arabic ones are pre-voiced while English ones have VOT in the short lag region. Since this sort of a difference in voicing is relatively less noticeable, advanced adult Saudi learners of English will not have formed new categories for English /b d g/ and will produce them with VOT much the same as ANS (equivalence classification). (SLM H1 H2 H3 H5 H7)

Our fifth hypothesis (RH5) is also about similar sounds with a less noticeable difference but for a different stop category i.e. voiced stops. As indicated before, both English and Saudi Arabic have voiced stops /b d g/, although Arabic ones are pre-voiced while English ones have VOT in the short-lag region. According to the SLM, English voiced stops are classified as similar sounds to the Arabic ones because they share the same IPA symbols in both languages, despite the voicing differences (negative VOT in Arabic and short-lag VOT in English).

This difference is predicted to be perceptually not prominent to L2 learners. Therefore, the hypothesis predicts equivalence classification will occur between L2 voiced stops and the closest L1 voiced ones and therefore, Saudi learners of English will not have formed new (short-lag) categories for English /b d g/ and will produce them with VOT much the same as their L1 pre-voiced stops.

The study results predominantly supported that in the majority of the learners' productions: 13 out 18 instances showed the predicted equivalence classification pattern of significant difference from L2 stops but not from L1 stops). This agrees with Simon's (2009) study, where she also found that Dutch learners of English produced pre-voiced stops similar to L1 in their English L2.

However, the other five cases, where the learners were significantly different from ANS, were also significantly different from ENS and their VOT means were in between the two groups, implying that they have started to move away from equivalence classification towards creating a new category for these English sounds.

Notably almost all these cases involved the labial sound /b/ (mainly in sentence context).

This result fits in with the general VOT analysis, where we found learner voiced labials presented less negative VOT than the voiced coronal and dorsal stops. The reason why the labial in particular departed from the expected equivalence classification is not transparent. From the SLM perspective, it would be argued that learners detected the difference between Arabic voiced /b/ and English voiced /b/ in input more than the interlingual differences of /d/ and /g/, although that greater detection is anticipated only with 'new' sounds. Obviously /b/ is not a new sound for Arabic learners of English, but we might speculate that it gained extra attention perhaps due to it being the voiced counterpart of the obviously new sound /p/. This sort of secondary newness effect is, however, as far as we know not recognised in the SLM.

6.3.6 **Research Hypothesis 6a**.

Deflection on the POA scale. Saudi learners are familiar with an L1 without a voiceless labial stop. They will fill this gap in a way that locates /p/ VOT in a place where it is clearly distinct from the VOTs of their /t/ and /k/, regardless of whether that place resembles the location of ENS /p/. (SLM Postulate 4, H6, H7).

The sixth hypothesis makes predictions about 'deflection', which is the mechanism the SLM invokes to explain why learner L2 sounds may occupy positions in the phonetic continuum which are not those either of L1 NS nor L2 NS, which we have seen above occurs in our findings in a number of instances. The type of deflection relevant to us is that which is said by the SLM to occur due to the learner's perceived need to maintain separation between the phonetic categories that they are working with in whatever languages they speak (in cases where they have not treated them as identical through equivalence classification).

We first consider this for VOT position on the POA dimension of L2 stops (RH6a). The learners' result for the new labial sound /p/ provides support for this prediction. When learning English, learners located the VOT of their new /p/ categories (aspirated and unaspirated) in places where they are clearly distinct from the VOTs of their aspirated and unaspirated English /t/ and /k/, and their L1 /t/ and /k/, regardless of whether those places resemble the locations of ENS aspirated and unaspirated /p/ (which they did not).

6.3.7 Research Hypothesis 6b.

Deflection on the voice-aspiration scale. Saudi learners in their L1 are used to a stop VOT voice-aspiration continuum in L1 divided into only two areas/categories at each place of articulation, since they only have to distinguish [d] from [t^h] and [g] from [k^h]. On exposure to English, however, they potentially need to divide that space into up to 5 parts/categories if they differentiate between all English and Arabic sounds, e.g. for coronals: [d] pre-voiced, versus [d] short-lag, versus [t] short-lag, versus [t^h] low long-lag, versus [t^h] high long-lag. They will therefore exhibit a tendency to space out their VOT values for those sounds more evenly over the VOT continuum, maximising VOT difference between them, resulting in the VOT values for some of them being 'deflected' from NS values. (SLM Postulate 4, H6, H7)

This hypothesis is also about deflection, but for the stops on the voice-aspiration dimension, as the RH describes.

Our data does not in this case provide such a straightforward answer as to whether the SLM hypothesis is supported. We have already shown that the learners have on a majority basis formed an equivalence classification for their L1 and L2 voiceless aspirated stops, which therefore occupy one place at one end of the VOT voice/aspiration continuum. The same is true on a majority basis for L1 and L2 voiced stops, which therefore for the most part occupy another place at the other end of the continuum. That leaves out only L2 voiceless unaspirated stops, which despite being in the target not distinct from voiced stops in VOT are in the learner's phonetic space located in between L1/L2 aspirated VOT location and the L1/L2 voiced location. They are not however significantly different from all of those, nor placed in the obvious intermediate location which would be the short lag, as the distancing scenario of SLM deflection might suggest. Rather they are indeterminately distinct from L1 aspirated stops (50%) and in the low long-lag area (but significantly different from their L2 aspirated stops and voiced stops and from L1 voiced stops).

If we disregard what differences were significant and simply inspect Figure 6.1, it is very clear that the learners simply divide the available voice/aspiration VOT continuum on the vertical axis into two parts, along the lines of L1: voiceless versus voiced. This does not fit the deflection scenario which, given the prior equivalence classifications, would yield three parts: voiceless aspirated L1/L2, voiceless unaspirated L2, voiced L1/L2. Therefore, there was good support for RH6 concerning deflection occurring in the SLM sense of categories being ‘deflected’ away from L1 and L2 categories "to maintain phonetic contrast between categories in a common L1-L2 phonological space", at least as far as the location of voiceless bilabials relative to other POAs were concerned (RH6a), but not so clearly where the aspirated - unaspirated distinction for voiceless stops is concerned (RH6b). We recognise however that it is always possible in the latter case that learners possessed a "representation based on different features, or feature weights, than a monolingual’s" (Flege, 1995;

p.239), which it was outside the scope of the current study to be able to find, since we were concerned with only one feature.

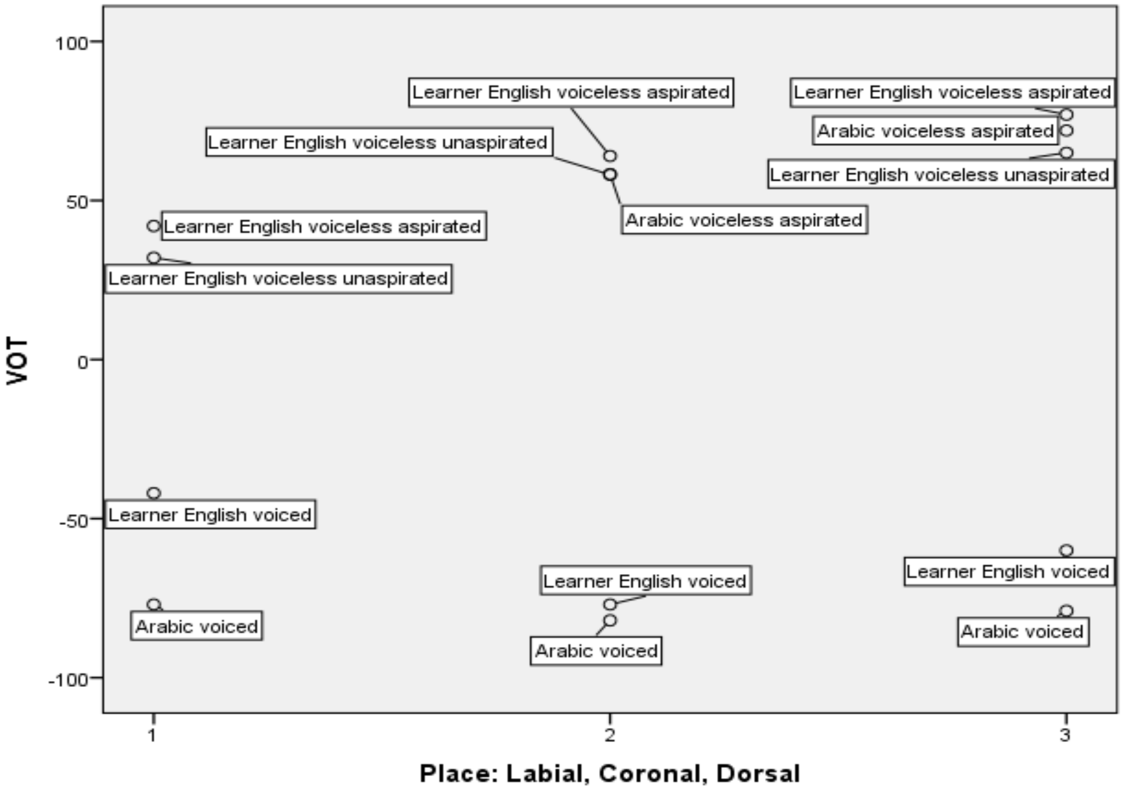


Figure 6.1: Distribution of stops by voice/aspiration in learner VOT space (vertical axis)

6.3.8 **Research Hypothesis 7**.

Effect of length of learning. Saudi learners who have had longer exposure to high quality ENS input (LOR in UK) will conform better to predictions 2 and 3 above than those with less exposure. There will be no difference on predictions 1, 4 and 5. (SLM Postulate 1, 3).

RH7 predicts that the length of residence (LOR) of the learners in the UK or experience of English daily use (their listening and speaking) relate meaningfully to the learners' production. In other words those who had greater LOR and/or greater use of English daily, were predicted to be more native like in some areas of acquisition than those with less exposure and less LOR. The SLM particularly predicts that the more experienced learners will do better on new (more noticeable) sounds i.e /p/ and the unaspirated voiceless stops but there will be no difference between learners on the voiceless aspirated and voiced stops.

As explained earlier in 5.2 and 5.3, in order for the learners to be more nativelike on voiceless unaspirated stops, they need to produce shorter positive VOTs than they typically do, so we would expect a negative correlation between LOR/use of English and VOT. But for voiceless aspirated and voiced stops, they need to produce longer positive or less negative VOTs than they normally do in L1, so we would seek a positive correlation between VOT and LOR/use of English.

In fact, LOR had a greater positive effect on nativelikeness than did daily use, though neither had a really strong impact (no greater than in 25% of potential specific stop conditions which could have been affected). In more detail, neither variable had much impact on the VOT of voiced stops, there was some positive impact on the unaspirated voiceless stops, but the strongest effect was on the aspirated voiceless stops [t^h k^h], especially in sentence contexts. This last fact makes sense since learners' exposure was predominantly to words in connected speech, not in isolation.

The more detailed predictions of RH7 were not fully supported. On the positive side, the prediction that L1-L2 differences that were hard to notice would not be affected by quantity and quality of input was strongly supported with respect to the voiced stops. This confirms the idea that once an equivalence classification has been formed, it is hard to shift. However, the most noticeable L1-L2 difference, involving the new sound /p/, which would be predicted to be strongly affected by exposure, was also barely affected. Furthermore, the most beneficially affected stop category was the voiceless aspirated (other than [p^h]) which would be regarded as less noticeably different from L1 than the voiceless unaspirated. Yet the latter was assisted by exposure less than the former.

*Table 6.2: Summary of results for the hypotheses **RH***

Our Research hypotheses (**RHs**)	**Verdict**
1	Supported
2	Supported
3	Weakly supported 50%
4	Predominantly supported 70%
5	Predominantly supported 72%
6a	Supported
6b	Not unambiguously supported
7	Weakly supported

For further comment on this table see 6.5.2.

6.4 Limitations of the study

Inevitably some limitations arose or came to be noticed during our study.

Although we wanted to find pure monolinguals to provide the ENS and ANS groups, in fact some Arabic monolinguals had a basic knowledge of English and some English monolinguals had some basic French. It is however quite difficult these days to find participants who genuinely have no knowledge whatsoever of any other language, and we do not believe this impacted negatively on the study.

With respect to the instrument, arguably it could have been better to use a range of different words for each condition rather than one word repeated three times, so as to be more representative of the language, and perhaps counter differential word familiarity effects. However, as it was, the study already involved measuring and analysing over 12000 VOT tokens and adding to that was hardly feasible.

Possibly the main limitation, within the scope of what we attempted, was that, with hindsight, it would have been valuable to have additionally obtained the learner participants' responses on the Arabic words. This would, however, have added massively to the number of VOT tokens to handle and so was not really feasible, as just indicated above. The benefit, however, would be that we would have been able to judge the SLM in a slightly more refined way. In our assessment of the SLM-based hypotheses we necessarily had to use our data to check, from the significances of the differences between means, whether learners were or were not producing English VOTs similar to Arabic, suggesting an equivalence classification had been created, or in some other way that suggested a new phonetic category had been formed (i.e. similar to English or to neither language). Our standard to compare with for Arabic was the ANS monolingual group. However, the SLM, like acquisition research more generally, is concerned in the end with the Arabic of the learner himself, in comparison with the English of the learner himself, in deciding what phonetic categories the learner has established, and as the SLM recognises, and studies based on it have sometimes followed up (e.g. Flege, 1987), the learner's Arabic may not be exactly the same as that of the monolingual Arabic speaker. If we had data on the learners' own Arabic VOTs we do not believe the findings of an equivalence classification for voiced stops would change. However, it is possible that we might obtain a slightly different perspective on the voiceless stops (RH3, 4, 6b), and see more clearly whether in fact the learner L1 mean VOTs have moved from the ANS means either in the direction of English unaspirated or aspirated voiceless VOTs.

6.5 Implications and suggestions for future work

Our study has made a number of contributions, each of which in turn suggests further work.

6.5.1 Description of VOT of ANS, ENS and Arab learners of English

With respect to description of the VOT of our Arabic monolinguals, we found some differences in VOT even from other Saudi dialects reported in the literature, e.g. in the precise degree of aspiration of the voiceless stops. This points to the need for further work on different dialects within Saudi Arabia. It also suggests that in other Arabic speaking countries also the assumption should not be made that the national vernacular varieties are uniform (e.g. Egyptian Arabic, Syrian Arabic etc.).

With respect to POA related VOT variation, while our findings replicated the common finding of increasing VOT from front to back for the voiceless stops in all three groups, in the learners and Arabic monolinguals this did not appear for the voiced stops. The literature also seems divided on the issue of POA related order effects on voiced stops in Arabic, and indeed on voiced stops with pre-voicing in general. Indeed, we found that the pattern of POA related VOT is mostly talked about only in relation to increasing positive VOT front to back and it is not clear how precisely this possible universal order is supposed to apply to pre-voiced stops.

With respect to all the three groups, we further showed that vowel related differences in stop VOT are among the least well studied, quite complex, and often involve interactions of vowel with place and/or voice. It is not possible simply to generalise that high vowels always engender longer stop VOT than low vowels, for example. While we have contributed useful information on this in a uniform way for each of the three groups studied, we feel that there is scope for a lot more investigation of this area of VOT.

The effect on VOT of word/sentence context has also not been systematically studied as well as it should. While the difference we found for English monolinguals between longer VOT for words in isolation and shorter for words in sentences replicated other studies and might have been expected to be universal, perhaps dependent upon the different speed of utterances associated with those modes, in fact it was not found for the Arabic monolinguals or the learners. This needs more investigation along with measurement of velocity of utterance, which was not covered in the present study, to unravel the factors involved and reasons for language differences more thoroughly. It would also be valuable to examine genuine spontaneous speech and how speakers produce their VOTs in natural conversation.

6.5.2 Acquisition of L2 English VOT by Arabic speaking learners and the SLM

Turning now to acquisition, our study prompted a number of thoughts about future directions.

First, we demonstrated that even quite advanced learners, with exposure to English in a target context, have not been able to acquire nativelike VOT in almost any respect, if non-significant difference from English native speaker VOT is the criterion for acquisition. This chimes with ongoing research in syntactic SLA where there is also debate as to whether some aspects of L2 grammar are simply impossible for learners ever to become nativelike in. In other words, their 'end state' of acquisition for such features can never be that of the L1 monolingual. The SLM however does not seem to claim that there are any such limitations, for learners of any age, so this does not support the SLM view that, with enough quality input, learners of any age can progress to nativelikeness.

If only VOT region (lag), rather than precise VOT value, is considered, the picture is not much better. While both learners and English monolinguals do use (different parts of) the long-lag range for voiceless aspirated stops, the learners fail to use the short-lag range for voiceless unaspirated and voiced stops as English monolinguals do, instead preferring the long-lag and pre-voiced ranges respectively (Figure 6.1).

Our study, however, was not of Arabic L2 learners at the most advanced possible proficiency level in terms of vocabulary and grammar (i.e. equivalent to IELTS 8 or 9), so studies of such very advanced learners are still worth considering to test if nativelike VOT of a second language truly is beyond the ability of second language learners to achieve. Furthermore, our participants were not learners who started to learn L2 English at a very young age, and they would not have been hearing authentic L2 pronunciation in input from native speakers before the end of the critical period/puberty, a point in time which some experts believe to be a crucial turning point for ability to fully acquire sounds (though not Flege and the SLM). Indeed, English VOT has been found to be possible to acquire in a fully nativelike way by Lebanese children in the UK (Khattab, 2002).

Our adult learners did perform slightly better, however, if we pay attention not to the precise VOTs they produce but to what sounds they distinguish with significantly different VOTs. However, even in this sense they are far from being overall nativelike. For POA of voiceless stops they made the same distinctions as monolingual target speakers, in that relative sense. However, in the word/sentence context distinction they made fewer distinctions than English monolinguals. In others, such as POA for voiced stops, and vowel related distinctions, differences were often made on a different basis by learners than ENS, as we have seen earlier. Interestingly, in one area they differed by actually making more VOT distinctions than monolingual speakers of English do. That is to say that learners were making significant VOT distinctions between both voiced and unaspirated voiceless stops, and between unaspirated voiceless and aspirated voiceless stops, while English monolinguals make no significant VOT difference between the former.

We are not aware of how far that last phenomenon has been documented before although it does not seem to be ruled out by the SLM. The SLM refers mainly to whether or not learners notice or make distinctions between L1 and L2 sounds, rather than between different L2 sounds (whether phonemically distinct or not).

Perhaps the other main comment that emerges concerning the SLM (beyond its lack of success on RH7) is that, as the data from the voiceless stops show (especially the unaspirated stops), the facts do not always clearly support either an equivalence classification or new category formation (in the more recent jargon: assimilation versus dissimilation, Flege 2007), but something in between. We feel the SLM should evolve to deal with the acquisition process more as a continuum rather than as a set of discrete possible outcomes. Intermediate states are to be expected since acquisition is a process, so a model should accommodate that.

We have no way of telling whether, with time, our participants' voiceless unaspirated stops would develop from their current ambivalent VOT position either to definitely merge with the other voiceless stops, or to become definitely a separate category, or indeed whether they would stay where they are in between those possibilities. Longitudinal studies need to be conducted in order to track such changes, and so provide data from which the SLM can be developed in this direction. Due to the problem of length of time needed for such studies, and the inevitable attrition of participants when a researcher attempts to revisit them and remeasure their VOT or whatever year after year, such studies are rare. Yet in the end the acquisition of sounds is a process that goes on over time, and needs to be studied as such so that it can be fully understood, not just, as it is largely currently studied (including our present work), based on examining the product of acquisition at some snapshot moment with a group of participants. Flege (1987) for instance reported learners with intermediate VOT values between L1 and L2 locations, but, being a snapshot study, could not deal with the issue of whether this was a change in progress or a steady state for his learners.

6.5.3 Broader suggestions

At the more general level, the experience of conducting our study has drawn to our attention some wider issues in the L2 acquisition research concerning sounds in general, and VOT in particular.

One is the question of what is the appropriate unit to study in studies of the acquisition of sounds. Our study has been, in a sense, a study of learner success in acquiring English stop VOT rather than in acquiring English stops. In this way it followed a long tradition, including a number of studies by Flege himself (summarised in Flege, 2007). There has been a history of research on stops, and their acquisition, which has focused on their VOT, partly perhaps precisely because so many studies have focused on this before, and perhaps also because VOT is fairly straightforward (if time consuming) to measure accurately with existing equipment. This has led to acquisition of stops having become almost synonymous with acquisition of VOT.

However, as was drawn to our attention at a number of points in our study, clearly there is more to a stop, and to the learning of it, than its VOT. In particular, phenomena associated with the unaspirated voiceless stops in English make this clear. If there was nothing more to a stop than its VOT, how could English native speakers tolerate an allophone of a voiceless stop phoneme like [k] having the same VOT as a voiced phone [g] representing a different phoneme, at the same POA etc.? And why would phoneticians usually give a different transcription to each of those and not write [sgi:], for example for *ski*? And indeed, in our study, it was clear that although the stop VOT that learners would have heard in *ski* was the same as that in *geese*, they were inclined to produce the unaspirated stop with the VOT of their L1 aspirated stop and the voiced stop with the VOT of their own L1 voiced one.

This prompts us to suggest that a more fully rounded approach needs to be taken to the study of the acquisition of stops by L2 learners. Oddities such as those above can only be investigated properly if a wider range of features which differentiate stops are measured, time consuming though that no doubt is. Some other features such as length of closure, amplitude and spectrum of burst, tongue extension (the 'distributed' feature) etc. are familiar in pure phonetic research (Ali et al., 2001), and it seems that, in order to understand learner acquisition of stops, in the end all relevant features will need to be studied together. It might be that our learners, while not very nativelike on VOT, are much more nativelike on other features of stops and, if all relevant stop features are taken into account, on balance might be considered much more nativelike than they appear just using VOT as the measure.

A related issue is that of what nativelikeness really should be regarded as consisting of in this domain. We have largely assumed that nativelike VOT would be VOT produced with values not significantly different statistically from the values of English native speakers, and such an approach is common in such studies (Flege, 2007, and most of those reviewed in chapter 2). The same approach could be applied to any other phonetic feature measured. However, for many practical purposes, such as real life language learning and teaching, it could be argued that many such fine distinctions, even if statistically significant, are not perceivable by native speakers, so of little practical importance. Although it is quite difficult to design studies where one can target a particular sound, such as a stop, rather than just spoken words or longer stretches of spoken material, we would urge more studies of native speaker judgment of foreignness of learner sounds, so as to establish, for instance, just how different a learner stop has to be in VOT from the NS norm before a native speaker actually notices a difference (i.e. detects a foreign accent). Again, on such a criterion, our participants might appear much more nativelike than they do on the strict objective statistical significance criteria commonly used.

Finally, our study also followed a common tradition in VOT research in being concerned with production, since VOT is in essence something that is produced when people speak. Yet clearly learner production must depend to a great extent on what they have previously heard and how they perceived it. Indeed, the SLM places great emphasis on successful perception of sounds as being a required antecedent to category formation and ensuing nativelike production. For Flege, it is which L1 sound a learner perceives an L2 sound as being similar to which drives the whole acquisition process. Yet this crucial part of the process is not often researched: a rare instance is Flege and Eefting (1987) who had participants listen to a whole range of [da] to [ta] sounds covering the whole VOT spectrum and checked what sounds they claimed to hear. This could then be compared with what they produced. While we considered this for our study, in the end it was just too much to attempt in addition to everything else.

We would speculate, however, that pure sound perception would not prove to be the only factor underlying production, as Flege claims. Returning to the odd case of English unaspirated voiceless stops, it seems very likely that learner production would be guided to a great extent by the fact that they are exposed to the written forms of words like *ski* and *sport*, and indeed will never see a written sequence *sg* or *sb* or *sd* word initial. In a listening test, if our Arab learners had been exposed to English voiceless unaspirated tokens without [s] as cue before them, who knows whether they would have identified them as similar to their L1 voiced or voiceless aspirated sounds, though in production our study shows that they clearly chose the latter.

6.6 Envoi

Clearly many avenues offer themselves for further study of learner stops and their acquisition. It has been said that "Research is an organized method for keeping you reasonably dissatisfied with what you have" (Charles F. Kettering), and indeed we cannot end this thesis with the feeling that our work is perfect. Nevertheless, we hope we have managed to illuminate a few corners of the topic of stop VOT and its ramifications in ENS, ANS and above all Arabic learners of English.

APPENDICES

Appendix A

University of Essex

Participant Information Sheet and Consent Form

Project Title: An investigation of the acquisition of English by Saudi L2 learners of English

Project Description

The project studies aspects of the acquisition of different words including nouns verbs and adjectives in English by Saudi Arabic speakers who are learners of English. It also looks at how English and Arabic interact in acquisition.

What does participating involve?

Your first task is to produce some words of English reading them from lists and your second task is to listen to a recording and write down the words that you have listened to. All responses will be audio-recorded by the researcher. The recordings will be used for analysis and discussion without disclosing the participant's identity. For the sake of the research some participants will be asked certain questions about their English background.

Taking Part

Please write the appropriate answer: Yes or No

I have read and understood the project information given above: __________

I have been given the opportunity to ask questions about the project: _____

I agree to take part in the project. Taking part in the project will include being interviewed and audio-recorded: ____________

I understand that my taking part is voluntary; I can withdraw from the study at any time and I do not have to give any reasons for why I no longer want to take part: _________________

Use of the information I provide

I understand that the information I provide will be used only for research and educational purposes: _____________

I understand that my personal details such as name, email address and phone number will not be revealed to people outside the project: ___________________

I understand that my words may be quoted in publications, reports, web pages, and other research outputs: __________________

I understand that audio-recordings of me may be used in publications, reports, web pages, and other research outputs: __________________

I understand that other genuine researchers may have access to this data only if they agree to

preserve the confidentiality of the information as requested in this form: _______________

Name/address and contact details of participant

Name:

Address:

E-mail:

Telephone:

Signature Date

<u>Participant's Background Information</u>

Participant No: ………………	**Age:** ……………	**Profession:** ……………………..

How long have you been living in an English speaking country?

………
…………

How many hours do you speak English daily?

………
…………

Appendix B:
Stimuli (1) (English)

Please read the following words from left to right

Peak	Speech	Teeth	Steel	League	Key	Ski	Beak	Deal	Geese
Park	Spark	Tart	Star	Fun	Card	Scarf	Bar	Dark	Guard
Neat									
Pool	Spoon	Tool	Stool	Go	Cool	School	Boot	Do	Goose
Peak	Speech	Teeth	Steel	Now	Key	Ski	Beak	Deal	Geese
Have									
Park	Spark	Tart	Star	Far	Card	Scarf	Bar	Dark	Guard
Pool	Spoon	Tool	Stool	Race	Cool	School	Boot	Do	Goose
May									
Peak	Speech	Teeth	Steel	Need	Key	Ski	Beak	Deal	Geese
Park	Spark	Tart	Star	No	Card	Scarf	Bar	Dark	Guard
Pool	Spoon	Tool	Stool	Look	Cool	School	Boot	Do	Goose

First, each of the above words was read by each participant in isolation as they appear in the table.

Second, each of these words was also read in a carrier sentence which was I say ....again.

Appendix C:
Stimuli (2) (Arabic)

قيل	كير	بير	ديك	تين
قال	كاف	باب	دار	تاب
قول	كوب	بوك	دور	توت
قيل	كير	بير	ديك	تين
قال	كاف	باب	دار	تاب
قول	كوب	بوك	دور	توت
قيل	كير	بير	ديك	تين
قال	كاف	باب	دار	تاب
قول	كوب	بوك	دور	توت

من فضلك إقراء الكلمات التالية وباللهجة العامية التي تتحدثها يومياً، وليس بالعربية الفصحى:

Appendix D:

Detailed descriptive statistics for ENS

VOT values of initial voiceless plosives for English monolinguals in three vowel contexts (word in isolation)

English stop	Vowel	Manner	Maximum	Minimum	Mean	Standard Deviation
Labial	/i:/	Unaspirated	23.39	8.64	14.41	3.83
		Aspirated	111.87	40.68	66.95	15.98
	/ɑ:/	Unaspirated	24.00	4.32	14.81	4.40
		Aspirated	107.80	44.75	72.86	17.93
	/u:/	Unaspirated	40.68	4.67	16.47	7.74
		Aspirated	111.33	39.67	74.26	14.91
Coronal	/i:/	Unaspirated	40.33	12.71	23.47	6.27
		Aspirated	164.76	61.02	87.27	20.07
	/ɑ:/	Unaspirated	31.00	9.66	21.95	5.17
		Aspirated	127.38	51.87	87.86	19.05
	/u:/	Unaspirated	43.73	14.24	24.45	6.25
		Aspirated	128.65	68.65	91.74	13.50
Dorsal	/i:/	Unaspirated	39.15	18.05	29.05	5.99
		Aspirated	135.77	68.65	101.47	16.01
	/ɑ:/	Unaspirated	45.26	12.71	26.19	6.70
		Aspirated	127.84	62.55	96.97	16.47
	/u:/	Unaspirated	38.90	20.85	29.05	5.37
		Aspirated	130.69	64.07	89.47	14.97

VOT values of initial voiceless plosives for English monolinguals in three vowel contexts (word in sentence)

Stop	Vowel	Manner	Maximum	Minimum	Mean	Standard Deviation
Labial	/iː/	Unaspirated	23.39	1.78	11.89	4.02
		Aspirated	81.87	38.39	58.08	10.05
	/ɑː/	Unaspirated	24.41	3.31	11.32	4.19
		Aspirated	89.50	36.61	66.30	12.76
	/uː/	Unaspirated	27.46	1.78	15.66	4.51
		Aspirated	93.57	38.65	68.61	12.41
Coronal	/iː/	Unaspirated	35.09	10.17	20.22	5.39
		Aspirated	118.48	48.82	75.49	13.54
	/ɑː/	Unaspirated	24.92	9.34	16.60	4.00
		Aspirated	114.17	53.97	81.33	15.88
	/uː/	Unaspirated	33.31	12.71	20.95	5.61
		Aspirated	122.81	58.48	81.50	14.67
Dorsal	/iː/	Unaspirated	37.25	19.20	29.82	4.48
		Aspirated	114.92	47.29	86.86	15.56
	/ɑː/	Unaspirated	31.81	12.46	22.94	5.06
		Aspirated	112.38	44.24	78.72	17.59
	/uː/	Unaspirated	38.65	13.73	24.59	7.18
		Aspirated	108.06	52.67	74.70	13.57

Appendix E:

Overall ENS inferential analysis

Source		F	Sig.	Partial Eta Squared
context	Sphericity Assumed	28.997	.000	.569
	Greenhouse-Geisser	28.997	.000	.569
	Huynh-Feldt	28.997	.000	.569
	Lower-bound	28.997	.000	.569
place	Sphericity Assumed	153.410	.000	.875
	Greenhouse-Geisser	153.410	.000	.875
	Huynh-Feldt	153.410	.000	.875
	Lower-bound	153.410	.000	.875
vowel	Sphericity Assumed	.297	.745	.013
	Greenhouse-Geisser	.297	.625	.013
	Huynh-Feldt	.297	.629	.013
	Lower-bound	.297	.592	.013
vasp	Sphericity Assumed	1116.733	.000	.981
	Greenhouse-Geisser	1116.733	.000	.981
	Huynh-Feldt	1116.733	.000	.981
	Lower-bound	1116.733	.000	.981
context * place	Sphericity Assumed	3.027	.059	.121
	Greenhouse-Geisser	3.027	.072	.121
	Huynh-Feldt	3.027	.069	.121
	Lower-bound	3.027	.096	.121
context * vowel	Sphericity Assumed	1.010	.373	.044
	Greenhouse-Geisser	1.010	.351	.044
	Huynh-Feldt	1.010	.354	.044
	Lower-bound	1.010	.326	.044
place * vowel	Sphericity Assumed	16.255	.000	.425
	Greenhouse-Geisser	16.255	.000	.425
	Huynh-Feldt	16.255	.000	.425
	Lower-bound	16.255	.001	.425
context * place * vowel	Sphericity Assumed	.961	.433	.042
	Greenhouse-Geisser	.961	.406	.042
	Huynh-Feldt	.961	.415	.042
	Lower-bound	.961	.338	.042

context * vasp	Sphericity Assumed	8.584	.001	.281
	Greenhouse-Geisser	8.584	.004	.281
	Huynh-Feldt	8.584	.003	.281
	Lower-bound	8.584	.008	.281
place * vasp	Sphericity Assumed	9.346	.000	.298
	Greenhouse-Geisser	9.346	.000	.298
	Huynh-Feldt	9.346	.000	.298
	Lower-bound	9.346	.006	.298
context * place * vasp	Sphericity Assumed	6.190	.000	.220
	Greenhouse-Geisser	6.190	.001	.220
	Huynh-Feldt	6.190	.000	.220
	Lower-bound	6.190	.021	.220
vowel * vasp	Sphericity Assumed	1.052	.385	.046
	Greenhouse-Geisser	1.052	.346	.046
	Huynh-Feldt	1.052	.349	.046
	Lower-bound	1.052	.316	.046
context * vowel * vasp	Sphericity Assumed	.893	.472	.039
	Greenhouse-Geisser	.893	.405	.039
	Huynh-Feldt	.893	.412	.039
	Lower-bound	.893	.355	.039
place * vowel * vasp	Sphericity Assumed	4.789	.000	.179
	Greenhouse-Geisser	4.789	.003	.179
	Huynh-Feldt	4.789	.001	.179
	Lower-bound	4.789	.040	.179
context * place * vowel * vasp	Sphericity Assumed	.542	.824	.024
	Greenhouse-Geisser	.542	.707	.024
	Huynh-Feldt	.542	.746	.024
	Lower-bound	.542	.469	.024

Appendix F:

Overall ANS inferential analyses

Context x vowel x voicing x 2 places

Source		F	Sig.	Partial Eta Squared
Context	Sphericity Assumed	.592	.448	.020
	Greenhouse-Geisser	.592	.448	.020
	Huynh-Feldt	.592	.448	.020
	Lower-bound	.592	.448	.020
Place	Sphericity Assumed	23.055	.000	.443
	Greenhouse-Geisser	23.055	.000	.443
	Huynh-Feldt	23.055	.000	.443
	Lower-bound	23.055	.000	.443
Voice	Sphericity Assumed	2529.171	.000	.989
	Greenhouse-Geisser	2529.171	.000	.989
	Huynh-Feldt	2529.171	.000	.989
	Lower-bound	2529.171	.000	.989
Vowel	Sphericity Assumed	1.878	.162	.061
	Greenhouse-Geisser	1.878	.162	.061
	Huynh-Feldt	1.878	.162	.061
	Lower-bound	1.878	.181	.061
context * place	Sphericity Assumed	.056	.815	.002
	Greenhouse-Geisser	.056	.815	.002
	Huynh-Feldt	.056	.815	.002
	Lower-bound	.056	.815	.002
context * voice	Sphericity Assumed	3.382	.076	.104
	Greenhouse-Geisser	3.382	.076	.104
	Huynh-Feldt	3.382	.076	.104
	Lower-bound	3.382	.076	.104
place * voice	Sphericity Assumed	25.451	.000	.467
	Greenhouse-Geisser	25.451	.000	.467
	Huynh-Feldt	25.451	.000	.467
	Lower-bound	25.451	.000	.467
context * place * voice	Sphericity Assumed	2.242	.145	.072
	Greenhouse-Geisser	2.242	.145	.072
	Huynh-Feldt	2.242	.145	.072

	Lower-bound	2.242	.145	.072
context * vowel	Sphericity Assumed	7.577	.001	.207
	Greenhouse-Geisser	7.577	.002	.207
	Huynh-Feldt	7.577	.002	.207
	Lower-bound	7.577	.010	.207
place * vowel	Sphericity Assumed	4.963	.010	.146
	Greenhouse-Geisser	4.963	.010	.146
	Huynh-Feldt	4.963	.010	.146
	Lower-bound	4.963	.034	.146
context * place * vowel	Sphericity Assumed	4.568	.014	.136
	Greenhouse-Geisser	4.568	.017	.136
	Huynh-Feldt	4.568	.015	.136
	Lower-bound	4.568	.041	.136
voice * vowel	Sphericity Assumed	24.770	.000	.461
	Greenhouse-Geisser	24.770	.000	.461
	Huynh-Feldt	24.770	.000	.461
	Lower-bound	24.770	.000	.461
context * voice * vowel	Sphericity Assumed	2.196	.120	.070
	Greenhouse-Geisser	2.196	.121	.070
	Huynh-Feldt	2.196	.120	.070
	Lower-bound	2.196	.149	.070
place * voice * vowel	Sphericity Assumed	.997	.375	.033
	Greenhouse-Geisser	.997	.375	.033
	Huynh-Feldt	.997	.375	.033
	Lower-bound	.997	.326	.033
context * place * voice * vowel	Sphericity Assumed	.531	.591	.018
	Greenhouse-Geisser	.531	.587	.018
	Huynh-Feldt	.531	.591	.018
	Lower-bound	.531	.472	.018

Context x vowel x 3 places (voiced only)

Tests of Within-Subjects Effects
Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
context	Sphericity Assumed	3358.320	1	3358.320	1.721	.200	.056
	Greenhouse-Geisser	3358.320	1.000	3358.320	1.721	.200	.056

	Huynh-Feldt	3358.320	1.000	3358.320	1.721	.200	.056
	Lower-bound	3358.320	1.000	3358.320	1.721	.200	.056
vowel	Sphericity Assumed	6214.219	2	3107.109	7.616	.001	.208
	Greenhouse-Geisser	6214.219	1.844	3370.089	7.616	.002	.208
	Huynh-Feldt	6214.219	1.963	3165.043	7.616	.001	.208
	Lower-bound	6214.219	1.000	6214.219	7.616	.010	.208
place	Sphericity Assumed	1847.618	2	923.809	1.838	.168	.060
	Greenhouse-Geisser	1847.618	1.890	977.482	1.838	.171	.060
	Huynh-Feldt	1847.618	2.000	923.809	1.838	.168	.060
	Lower-bound	1847.618	1.000	1847.618	1.838	.186	.060
context * vowel	Sphericity Assumed	4390.129	2	2195.064	5.755	.005	.166
	Greenhouse-Geisser	4390.129	1.583	2772.697	5.755	.010	.166
	Huynh-Feldt	4390.129	1.660	2645.317	5.755	.009	.166
	Lower-bound	4390.129	1.000	4390.129	5.755	.023	.166
context * place	Sphericity Assumed	1297.149	2	648.575	1.123	.332	.037
	Greenhouse-Geisser	1297.149	1.921	675.318	1.123	.331	.037
	Huynh-Feldt	1297.149	2.000	648.575	1.123	.332	.037
	Lower-bound	1297.149	1.000	1297.149	1.123	.298	.037
vowel * place	Sphericity Assumed	6503.770	4	1625.942	4.923	.001	.145
	Greenhouse-Geisser	6503.770	3.594	1809.557	4.923	.002	.145

	Huynh-Feldt	6503.770	4.000	1625.942	4.923	.001	.145
	Lower-bound	6503.770	1.000	6503.770	4.923	.034	.145
context * vowel * place	Sphericity Assumed	2057.099	4	514.275	1.565	.188	.051
	Greenhouse-Geisser	2057.099	3.265	630.008	1.565	.199	.051
	Huynh-Feldt	2057.099	3.729	551.702	1.565	.192	.051
	Lower-bound	2057.099	1.000	2057.099	1.565	.221	.051

Appendix G:

Overall learner inferential analysis

Tests of Within-Subjects Effects
Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
context	Sphericity Assumed	17517.017	1	17517.017	1.618	.213	.051
	Greenhouse-Geisser	17517.017	1.000	17517.017	1.618	.213	.051
	Huynh-Feldt	17517.017	1.000	17517.017	1.618	.213	.051
	Lower-bound	17517.017	1.000	17517.017	1.618	.213	.051
vowel	Sphericity Assumed	21039.488	2	10519.744	14.186	.000	.321
	Greenhouse-Geisser	21039.488	2.000	10521.381	14.186	.000	.321
	Huynh-Feldt	21039.488	2.000	10519.744	14.186	.000	.321
	Lower-bound	21039.488	1.000	21039.488	14.186	.001	.321
place	Sphericity Assumed	68742.866	2	34371.433	15.417	.000	.339
	Greenhouse-Geisser	68742.866	1.337	51410.167	15.417	.000	.339
	Huynh-Feldt	68742.866	1.376	49944.020	15.417	.000	.339
	Lower-bound	68742.866	1.000	68742.866	15.417	.000	.339
	Lower-bound	133764.747	30.000	4458.825			
voice_asp	Sphericity Assumed	4879335.693	2	2439667.847	499.422	.000	.943

	Greenhouse-Geisser	4879335.693	1.156	4222103.526	499.422	.000	.943
	Huynh-Feldt	4879335.693	1.173	4160788.898	499.422	.000	.943
	Lower-bound	4879335.693	1.000	4879335.693	499.422	.000	.943
context * vowel	Sphericity Assumed	197.172	2	98.586	.148	.863	.005
	Greenhouse-Geisser	197.172	1.944	101.418	.148	.857	.005
	Huynh-Feldt	197.172	2.000	98.586	.148	.863	.005
	Lower-bound	197.172	1.000	197.172	.148	.703	.005
context * place	Sphericity Assumed	12155.386	2	6077.693	3.937	.025	.116
	Greenhouse-Geisser	12155.386	1.548	7854.401	3.937	.036	.116
	Huynh-Feldt	12155.386	1.616	7522.524	3.937	.034	.116
	Lower-bound	12155.386	1.000	12155.386	3.937	.056	.116
vowel * place	Sphericity Assumed	6108.334	4	1527.083	3.370	.012	.101
	Greenhouse-Geisser	6108.334	3.287	1858.322	3.370	.018	.101
	Huynh-Feldt	6108.334	3.740	1633.392	3.370	.014	.101
	Lower-bound	6108.334	1.000	6108.334	3.370	.076	.101
context * vowel * place	Sphericity Assumed	4725.964	4	1181.491	2.695	.034	.082
	Greenhouse-Geisser	4725.964	3.483	1356.716	2.695	.042	.082
	Huynh-Feldt	4725.964	3.997	1182.400	2.695	.034	.082
	Lower-bound	4725.964	1.000	4725.964	2.695	.111	.082
context * voice_asp	Sphericity Assumed	2635.407	2	1317.703	.188	.829	.006
	Greenhouse-Geisser	2635.407	1.070	2462.358	.188	.684	.006

	Huynh-Feldt	2635.407	1.078	2445.317	.188	.686	.006
	Lower-bound	2635.407	1.000	2635.407	.188	.667	.006
vowel * voice_asp	Sphericity Assumed	32064.361	4	8016.090	13.984	.000	.318
	Greenhouse-Geisser	32064.361	2.249	14258.696	13.984	.000	.318
	Huynh-Feldt	32064.361	2.440	13141.427	13.984	.000	.318
	Lower-bound	32064.361	1.000	32064.361	13.984	.001	.318
context * vowel * voice_asp	Sphericity Assumed	3573.933	4	893.483	1.465	.217	.047
	Greenhouse-Geisser	3573.933	2.325	1537.237	1.465	.237	.047
	Huynh-Feldt	3573.933	2.532	1411.531	1.465	.235	.047
	Lower-bound	3573.933	1.000	3573.933	1.465	.236	.047
place * voice_asp	Sphericity Assumed	277253.443	4	69313.361	49.802	.000	.624
	Greenhouse-Geisser	277253.443	2.230	124324.989	49.802	.000	.624
	Huynh-Feldt	277253.443	2.417	114688.803	49.802	.000	.624
	Lower-bound	277253.443	1.000	277253.443	49.802	.000	.624
context * place * voice_asp	Sphericity Assumed	1702.004	4	425.501	.644	.632	.021
	Greenhouse-Geisser	1702.004	2.824	602.753	.644	.580	.021
	Huynh-Feldt	1702.004	3.147	540.761	.644	.596	.021
	Lower-bound	1702.004	1.000	1702.004	.644	.429	.021
vowel * place * voice_asp	Sphericity Assumed	31723.505	8	3965.438	11.502	.000	.277
	Greenhouse-Geisser	31723.505	4.390	7226.482	11.502	.000	.277
	Huynh-Feldt	31723.505	5.236	6059.078	11.502	.000	.277

	Lower-bound	31723.505	1.000	31723.505	11.502	.002	.277
context * vowel * place * voice_asp	Sphericity Assumed	5605.675	8	700.709	2.010	.046	.063
	Greenhouse-Geisser	5605.675	4.312	1299.974	2.010	.092	.063
	Huynh-Feldt	5605.675	5.126	1093.572	2.010	.078	.063
	Lower-bound	5605.675	1.000	5605.675	2.010	.167	.063

Appendix H:

Group differences between VOT means: by place, voice/aspiration and vowel, with contexts combined

V	Place	Voicing/ Aspiration	CV:	Learner - ENS		Learner -ANS		ENS - ANS	
				Mean diff.	p	Mean diff.	p	Mean diff.	p
/i:/	labial	v-less asp.	/pi:/	-30.10	<.001				
		v-less unasp	/spi:/	13.92	<.001				
	coronal	v-less asp.	/ti:/	-20.46	<.001	-2.37	.435	18.10	<.001
		v-less unasp	/sti:/	41.20	<.001				
	dorsal	v-less asp.	/ki:/	-11.39	<.001	1.67	.582	13.10	<.001
		v-less unasp	/ski:/	41.41	<.001				
	labial	voiced	/bi:/	-48.43	<.001	47.42	<.001	95.85	<.001
	coronal		/di:/	-94.96	<.001	11.26	.102	106.2	<.001
	dorsal		/gi:/	-92.43	<.001	17.94	.005	110.4	<.001
/a:/	labial	v-less asp.	/pa:/	-26.41	<.001				
		v-less unasp	/spa:/	14.20	<.001				
	coronal	v-less asp.	/ta:/	-29.43	<.001	-0.925	.768	28.51	<.001
		v-less unasp	/sta:/	28.24	<.001				
	dorsal	v-less asp.	/ka:/	-22.57	<.001	2.45	.293	25.02	<.001
		v-less unasp	/ska:/	28.81	<.001				
	labial	voiced	/ba:/	-57.78	<.001	32.11	<.001	89.89	<.001
	coronal		/da:/	-87.52	<.001	6.61	.282	94.13	<.001
	dorsal		/ga:/	-78.36	<.001	22.63	<.001	101.0	<.001
/u:/	labial	v-less asp.	/pu:/	-21.72	.012				
		v-less unasp	/spu:/	26.23	<.001				
	coronal	v-less asp.	/tu:/	-11.56	.002	19.59	<.001	31.15	<.001
		v-less unasp	/stu:/	41.02	<.001				
	dorsal	v-less asp.	/ku:/	-0.012	.997	9.06	.002	9.08	<.001
		v-less unasp	/sku:/	42.07	<.001				
	labial	voiced	/bu:/	-55.25	<.001	29.85	.001	85.10	<.001
	coronal		/du:/	-106.00	<.001	-3.50	.657	102.5	<.001
	dorsal		/gu:/	-87.36	<.001	17.06	.013	104.4	<.001

Appendix I:

The perception study of English voiced stops

This was a separate study not designed to fit in fully with the production study, so in the end we felt it not cohesive to include it in the main text. It shared only that it concerned learners of English perceiving voiced stops before three vowels. It did not match the VOT production study in that it: did not concern perception of VOT of stops, but of stops as wholes, did not look at voiceless stops nor at stops in words in sentence contexts which were in the production study; on the other hand, it included non-words and stops in word medial position, which were not considered in the production study. Furthermore, it was not the kind of perception study which the SLM envisages as relevant since it did not measure how far learners perceived L2 stops as being similar to L1 stops, nor indeed did it ask learners to discriminate between different L2 stops: it rather asked them to say what written stop they heard (b d g).

We conceptualized it as two sub-studies:

A phonetic identification perception test was conducted to test the learners' perception of voiced stops in nonsense words in isolation (word initial and medial intervocalic).

A phonological identification perception test was used to test the learners' perception of voiced stops in real meaningful words in isolation (word initial).

These are briefly summarized next.

Method

Participants

Participants were the same learners as in the production study. One English native speaker also contributed by making the recordings to serve as stimuli, and five more by taking the test and obtaining 100% correct scores, showing that the stimuli were valid.

Procedure

Stimulus words wee each heard by participants through headphones spoken three times from a PC in a Linguistics lab. Learners were asked to listen to each numbered set of three repeated words and then write on an answer sheet against the corresponding number what sound they had heard (e.g. b d or g or something else). They were not told in advance what set of sounds they were going to hear. Nativelikeness of learners was scored directly from how close their perception scores approximated to 100% correct.

Phonological perception test stimuli

The list of target words contained 9 different real English words, with voiced stops in initial position followed by the three quantum vowels i.e. high front, high back and low. The words were common monosyllabic words of English, known to the participants, but did not include any of the words used in the production task: *beak deem geese bark dart guard boo do goose*.

A total of 27 (3 consonants * 3 vowels * 3 repetitions) target tokens of English were heard. The list also included some other words as distracters, all read out preceded by a number.

Phonetic perception test stimuli

A phonetic identification test of nonce words was also conducted. For that purpose, and following the common practice for such tests (Johnson & Babel 2010, Iverson et al 2008 etc.), a set of VCV and CV nonsense word stimuli was recorded in the voice of the English native speaker. The VCV stimuli were recorded in structures with C as the target sound inserted between repetitions of one of the three long quantum vowels, i.e. [i:], [ɑ:] and [u:], yielding words such as /ɑ:bɑ:, i:bi:, u:bu:/ stressed on the second syllable.

Data analysis

Scores out of three were calculated for each voiced stop in each condition (3 vowels, real-unreal, initial-medial). With respect to reliability, participants scored either 0 or 3 in 84% of the sets of three repetitions to unreal words. On real words they did this 82.6% of the time. Hence, participants were highly self consistent on the sets of three sounds.

The perception data, totally failed to be suitable for ANOVA because there were conditions where every participant scored perfect 3 correct. Such data does not have the normal distribution. Therefore, for perception test data the ordinal option in Generalized Linear Model - Generalized Estimating Equations (GZLM) was used instead. This performs exactly the analyses which ANOVA does, but has an ordinal option which allows data to be analysed that is unsuitable for ANOVA. For post hoc comparisons we also used GZLM, with Bonferroni adjustment for multiple comparisons.

Results

Three questions were answered.

'Does learner ability to correctly identify the voiced stops differ depending on whether the stops occur word initially or medially, or in real words versus unreal ones, and is either POA, or accompanying vowel a factor?'

In perception of voiced stops, learners consistently identified /b/ less well than /d/ and /g/, where they were near nativelike in ability.

Where real words versus unreal phonetic words were tested, perception was better on unreal words than real words, due primarily to more accurate performance on /b/, and to a small extent on /g/. Where initial versus medial word position was tested (only in unreal words), there was no overall difference in accuracy between positions. Vowel difference had no overall effects on perception.

'Does length of residence in the UK or mean daily use of English relate meaningfully to learner accuracy scores for perception of voiced stops?'

There were almost no correlations of LOR and mean daily use with perception scores, even for /b/ where there was some scope for improvement. On /d g/ ceiling effect made it hard in any case for any correlations to appear: perception of these stops had effectively already been acquired. Thus, there was no real sign of ongoing acquisition of stop perception of the type which we measured.

'Is there any relationship between perception and production in learner performance on English voiced stops?'

Learners who were more or less nativelike on accurate perception of voiced stops, relative to other learners, were not systematically closer in their VOT production to that of ENS. Thus, there was no evidence of the two abilities being related. However, of course, the perception ability being measured was not perception just of VOT but of voiced stops as wholes, so learners may well have been relying on phonetic cues other than VOT to identify them.

A reverse pattern was found for production and reception in the voiced stop POAs: /b/ was the most nativelike in production but the least in perception. Furthermore, /d g/ were nativelike in perception but some way from that in production (VOT). Superficially that might seem to run counter to the claims of the SLM which expects perception success to lead acquisition, prompting new phonetic categories to be created for target sounds, and then production based on that to later match perception. However, our perception test was not of the kind of perception which the SLM refers to (comparing perception if sounds in L1 and L2), and was not of the same phonetic feature that was measured from production (i.e. just VOT).

Appendix J:

How ProsodyPro Script was used to extract VOT values

A step by step guide to how the script was used to obtain the VOT of the participants' stops in this study is presented as follows:

1- Put the script in a folder which contains the sound file to be analyzed, and then open Praat.

2- Select Open Praat Script... from the top menu.

3- Find ProsodyPro.praat (the script) in the dialogue window and select it. Sometimes the script file cannot open and the computer gives an error message. All that needs to be done in that case is just to rename the sound file trying a different name perhaps using numbers instead of letters or using very short words.

4- When the script window opens in Praat, click on Run from the Run menu (or using the keyboard, type command-r or control-r).

5- In the startup window, we have to make sure that the task in the run window is set on: "interactive labelling" and leave the default values and settings for everything else.

6- Click OK and three windows will appear. The first window (PointProcess) shows the waveform together with vocal cycle marks: it is not needed in getting the VOT durations, so it can be hidden or even closed.

7- The second window (TextGrid) displays the waveform and spectrogram of the current sound and at the bottom of this window there are two TextGrid tiers, where you can insert interval boundaries (Tier 1) and add comments (Tier 2). In order to obtain the VOT durations, mark the beginning and the end of the duration with the mouse cursor on the waveform and press Enter. Then start labeling the interval boundaries this way for all the words and the files. The labels or word boundaries could be for example, the first letter or the whole word being analyzed.

8- The third window (Pause) allows you to control the progression of the analysis. When you finish labeling all the boundaries, end the progression of the current analysis session by clicking on "Finish" in the Pause window.

9- Find the script window and click or Run again. The same window as in step (5) will appear but the task this time should be changed to “Get ensemble files” and click OK.

10- The last step is to launch Microsoft Excel, click on open, then locate the folder containing the sound file and the script. A number of additional folders would have been created after the previous step (step 8). Locate the folder named “duration” to find all the VOT values for the ranges that you have marked in the sound file.

Only how to extract VOT values with this script has been explained here, however, the script can help with various other analyses. An interested reader is referred to the script’s developer’s website for a more detailed understanding of how the script works to get other values. (http://www.homepages.ucl.ac.uk/~uclyyix/)

The point of voice onset for the initial stop of each word was taken as the first cycle of the wave showing periodical vibration of the vocal folds, as advised by Cho & Ladefoged (1999:215). VOT was measured as the interval between that point and the instant of release of the stop after closure, which might of course be either before or after the release of the stop. Great efforts were made to obtain the most accurate measurement of the VOT productions of the participants. Waveforms and spectrograms were presented on the computer's screen and the VOT limits were manually marked with the use of the mouse cursor so that the duration could be calculated between the onset of the release and the onset of the first complete vibration of the vocal folds, as shown in Figure 0.1.

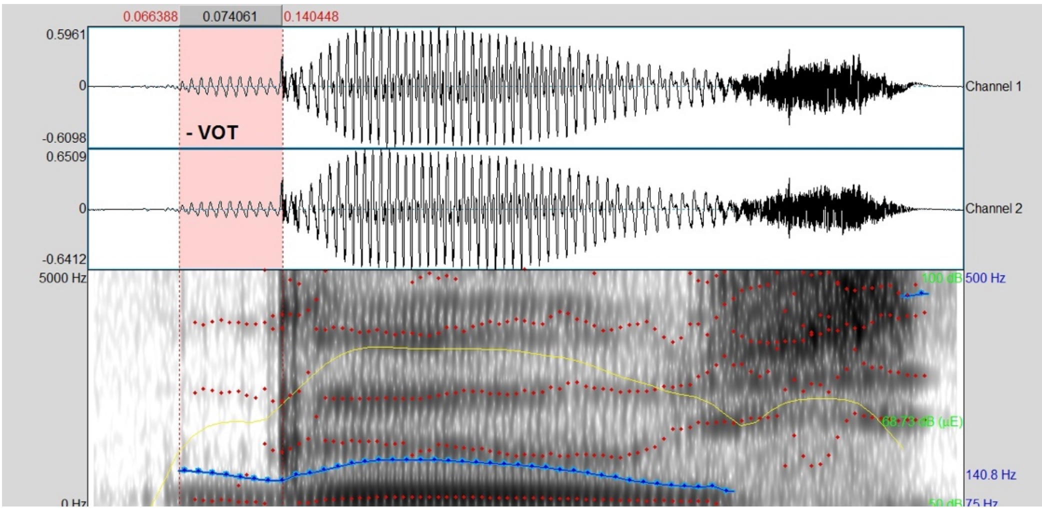


Figure 0.1: Waveform and spectrogram of a pre-voiced stop produced by a second language learner for the English 'g'.

REFERENCES

Al-Ani, S. H. (1970). *Arabic phonology: An acoustical and physiological investigation* (Vol. 61): Walter de Gruyter.

AlDahri, S. S. (2013). A study for the effect of the emphaticness and language and dialect for voice onset time (VOT) in Modern Standard Arabic (MSA).

Alghamdi, M. (1990). *Analysis, synthesis and perception of voicing in Arabic.* (PhD), University of Reading.

Alghamdi, M. M. (2006). Voice Print : Voice Onset Time as a Model. *Arab Journal for Security Studies and Training, 21(42)*.

Ali, A. M. A., Van der Spiegel, J., & Mueller, P. (2001). Acoustic-phonetic features for the automatic classification of stop consonants. *IEEE Transactions on Speech and Audio Processing, 9*(8), 833-841.

Alwasel, T. A. (2017). *The influence of diglossia on learning Standard Arabic*. (PhD), King's College London.

Ambrazaitis, G., & Frid, J. (2012). The prosody of contrastive topics in Southern Swedish. *Proceedings of FONETIK 2012*.

Archibald, J. (1998). *Second language phonology* (Vol. 17): John Benjamins Publishing.

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics, 39*(1), 39-49.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words, 167*(224), 233-277.

Best, C. T. (1995). A Direct Realist View of Cross-Language Speech Perception. *Speech perception and linguistic experience: Issues in cross-language research*, 171-204.

Best, C. T., M. (2006). Nonnative and second-language speech perception: Commonalities and complementarities. *Language experience in second language speech learning, 10389*.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001a). 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*, 775. doi:10.1121/1.1332378

Best, C. T., McRoberts, G. W., & Goodell, E. (2001b). 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. doi:10.1121/1.1332378

Black, M. P., Bone, D., Skordilis, Z. I., Gupta, R., Xia, W., Papadopoulos, P., . . . Narayanan, S. S. (2015). Automated evaluation of non-native english pronunciation quality: Combining knowledge- and data-driven features at multiple time scales. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Boersma, P., & Weenink, D. (2012). *Praat: doing phonetics by computer*: http://www.praat.org/.

Boersma, P. P. G. (2002). Praat, a system for doing phonetics by computer. *Glot international, 5*.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. *Second language acquisition and linguistic theory, 1*, 4-64.

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

Chao, K.-y., Khattab, G., & Chen, L.-m. (2006). *Comparison of VOT patterns in Mandarin Chinese and in English.* Paper presented at the Proceedings of the 4th Annual Hawaii International Conference on Arts and Humanities.

Chen, L.-m. (2007). VOT productions of word-initial stops in Mandarin and English : A cross-language study. *English*, 303-317.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics, 27*, 207-229. doi:10.1006/jpho.1999.0094

Choudhury, A., & Kaiser, E. (2012). *Prosodic focus in Bangla: A psycholinguistic investigation of production and perception.* Paper presented at the LSA Annual Meeting Extended Abstracts.

Cook, V. J. (1992). Evidence for multicompetence. *Language learning, 42*(4), 557-591.

Davenport, M., & Hannahs, S. (2010). *Introducing phonetics and phonology*: Routledge.

Deuchar, M., & Clark, A. (1996). Early bilingual acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics, 24*(3), 351-365.

Docherty, G. J. (1992). The Timing of Voicing in British English Obstruents. 289.

Fant, G. (1973). *Speech sounds and features*. Cambridge, MA, US: The MIT Press.

Fischer-Jørgensen, E. (1954). Acoustic analysis of stop consonants. *Miscellanea*.

Flege, J., & Port, R. (1981). CROSS-LANGUAGE PHONETIC INTERFERENCE : ARABIC TO ENGLISH. *Language, 24*, 125-146.

Flege, J. E. (1980). Phonetic Approximation in Second Language Acquisition. *Language learning, 30*, 117-134.

Flege, J. E. (1987a). Effects of Equivalence Classification on the Production of Foreign Language Speech Sounds in Sound Patterns in Second Language Acquisition. *Studies on Language Acquisition (SOLA)*(5), 9-39.

Flege, J. E. (1987b). The instrumental study of a L2 speech production: some methodological considerations. *Language learning, 37*, 285-296.

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

Flege, J. E. (1992). Speech learning in a second language. *Phonological development: Models, research, implications, 565*, 604.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America, 93*(3), 1589-1608.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 233-277.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. *Phonetics and phonology in language comprehension and production: Differences and similarities, 6*, 319-355.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. *Laboratory phonology, 9*, 353-382.

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *The Journal of the Acoustical Society of America, 76*(3), 708-721.

Flege, J. E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech communication, 40*(4), 467-491.

Giannini, A., & Pettorino, M. (1982). The emphatic consonants in Arabic (Speech Laboratory Report). *Naples: Istituto Universitario Orientale*.

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. *The Journal of the Acoustical Society of America, 107*, 2711-2724.

Hardcastle, W. J. (1973). Some observations on the tense-lax distinction in initial stops in Korean. *Journal of Phonetics, 1*(3), 263-272.

Honeybone, P. (2005). Diachronic evidence in segmental phonology : the case of obstruent laryngeal specifications. Linguistics, 1-36.

Ingham, B. (1982). North east Arabian dialects: Kegan Paul.

Iverson, G. K., & Salmons, J. C. (1995). Aspiration and laryngeal representation in Germanic. *Phonology, 12*, 369-396.

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.

Jesry, M. (1996). *Some cognitively controlled coarticulatory effects in Arabic and English, with particular reference to voice onset time.* University of Essex.

Johnson, K., & Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics, 38*, 127-136.

Kager, R., Feest, S. V. D., Fikkert, P., Kerkhoff, A., & Zamuner, T. (2007). Representations of [voice]: Evidence from Acquisition. In J. V. D. Weijer & E. J. V. D. Torre (Eds.), *Voicing in Dutch* (pp. 41-80). Amsterdam John Benjamins.

Keating, P., Linker, W., & Huffman, M. (1983). Closure duration of stop consonants. *Journal of Phonetics, 11*, 277-290.

Kent, R. D., & Read, C. (1992). *The Acoustic Analysis of Speech*: Singular Publishing Group.

Khattab, G. (2000). VOT production in English and Arabic bilingual and monolingual children. *AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4, 8*, 95-122.

Khattab, G. (2002). *Sociolinguistic competence and the bilingual's adoption of phonetic variants: auditory and instrumental data from English-Arabic bilinguals.* (Doctoral dissertation), University of Leeds.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal Of Speech And Hearing Research, 18*, 686-706.

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics*. Boston, MA: Wadsforth/Cengage Learning.

Ladefoged, P., & Maddieson, I. (1998). The sounds of the world's languages. *Language, 74*(2), 374-376.

Lado, R. (1957). Linguistics Across Cultures: Applied Linguistics for Language Teachers.

Larson-hall, J. (2010). A Guide to Doing Statistics in Second Language Research Using SPSS. *Language, 20*, 438.

Laver, J. (1994). *Principles of phonetics*: Cambridge University Press.

Lisker, L. (1975). Is it VOT or a first-formant transition detector? *The Journal of the Acoustical Society of America, 57*(6), 1547-1551.

Lisker, L. (2002). The voiceless unaspirated stops of English. *AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4*, 233-240.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word Journal Of The International Linguistic Association, 20*, 384-422.

Lisker, L., & Abramson, A. S. (1967). Some Effects of Context On Voice Onset Time in English Stops. *Language and Speech, 10*, 1-28.

Lisker, L., & Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language*, 767-785.

Liu, F. (2010). Single vs. double focus in English statements and yes/no questions. *Proceedings of Speech Prosody. ISCA Press, Chicago*.

MacKay, I. R., Flege, J. E., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *The Journal of the Acoustical Society of America, 110*(1), 516-528.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants. *Journal of Child Language, 7*, 41-74.

Mitleb, F. M. (2009). Voice Onset Time of Jordanian Arabic Stops. *3rd International Conference on Arabic Language Processing (CITALA'09), May 4-5, 2009, Rabat, Morocco*.

Nasukawa, K. (2010). Place-Dependent VOT in L2 Acquisition. *English*, 197-210.

Nearey, T. M., & Rochet, B. L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association, 24*(01), 1-18.

Newman, D. (2002). The phonetic status of Arabic within the world's languages: the uniqueness of the lughat al-daad. *Antwerp papers in linguistics., 100*, 65-75.

Newman, D. (2006). The Phonetics of Arabic. *Journal of the American Oriental Society, 46*, 1-6.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic research, 5*(3), 261-283.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 32*(6), 693-703.

Port, R. F., & Mitleb, F. M. (1983). Segmental features and implementation in acquisition of English by Arabic speakers. *Journal of Phonetics, 11*, 219-229.

Port, R. F., & Rotunno, R. (1979). Relation between voice-onset time and vowel duration. *The Journal of the Acoustical Society of America, 66*(3), 654-662.

Radwan, M. (1996). *An experimental investigation of the acoustical temporal correlates of voicing contrast in stop consonants (with reference to Arabic).* University of Essex.

Reis, M., & Nobre-Oliveira, D. (2007). Effects of perceptual training on the identification and production of English voiceless plosives aspiration by Brazilian EFL learners. *New Sounds: International Symposium on the Acquisition of Second Language Speech*, 398-407.

Roca, I., & Johnson, W. (1999). *A course in phonology*: Wiley-Blackwell.

Rochet, B. L., & Yanmei, F. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception. *Canadian Acoustics, 19*(4), 105-106.

Rogers, H. (2000). *The sounds of language : an introduction to phonetics*. Harrow, England; New York: Longman.

Schmidt, A. M. (1996). Cross-language identification of consonants. Part 1. Korean perception of English. *The Journal of the Acoustical Society of America, 99*(5), 3201-3211.

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

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*: Newbury House Publishers.

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics, 20*, 213-223.

Shimizu, K. (2011). A study on VOT of initial stops in English produced by Korean, Thai and Chinese speakers as L2 learners. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011)*, 1818-1821.

Simon, E. (2008). Laryngeal stop systems in contact: Connecting present-day acquisition findings and historical contact hypotheses * Ellen Simon Ghent University. 1-24.

Simon, E. (2009). Acquiring a new second language contrast: an analysis of the English laryngeal system of native speakers of Dutch. *Second Language Research, 25*, 377-408.

Smith, B. L. (1979). A phonetic analysis of consonantal devoicing in children's speech. *Journal of Child Language, 6*(01), 19-28.

Stevens, K. N., Keyser, S. J., & Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. *Invariance and variability in speech processes*, 426-449.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America, 55*(3), 653-659.

Syed, N. A. R. (2013). *The acquisition of English consonants by Pakistani learners.* (Unpublished PhD), University of Essex.

Treiman, R. (1985). Spelling of stop consonants after/s/by children and adults. *Applied Psycholinguistics, 6*(03), 261-282.

Wang, B., & Xu, Y. (2011). Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics, 39*(4), 595-611.

Weismer, G. (1979). Sensitivity of voice-onset time (VOT) measures to certain segmental features in speech production. *Journal of Phonetics, 7*(2), 197-204.

Weismer, G. (1980). Control of the voicing distinction for intervocalic stops and fricatives: Some data and theoretical considerations. *Journal of Phonetics*.

Whiteside, S. P., Henry, L., & Dobbin, R. (2004). Sex differences in voice onset time: a developmental study of phonetic context effects in British English. *Journal of the Acoustical Society of America, 116*, 1179-1183.

Wu, W. L., & Xu, Y. (2010). *Prosodic focus in Hong Kong Cantonese without post-focus compression.* Paper presented at the Proceedings of speech prosody.

Xu, Y. (2013). ProsodyPro - A tool for large-scale systematic prosody analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody*, 7-10.

Yeni-Komshian, G. H., Caramazza, A., & Preston, M. S. (1977). A study of voicing in Lebanese Arabic. *Journal of Phonetics, 5*(1), 35-48.

Zampini, M. L. (2008). L2 speech production research: Findings, issues and advances. In J. G. Hansen & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition*: John Benjamins.