
A Contrastive Study of the Voice Onset Time (VOT) in English and Arabic Languages

Dr. Rafida Mansoor Mahmood

Mustansiriyah University - College of Arts

Department of English Language and Literature

Ph. D student at the Lebanese University

Abstract

The signal sound contains many different features. One of these features is voice onset time henceforth (VOT) and this feature refers to the ways different people of different languages have been distinguished by the way they articulate stop consonants of their own language.

This feature (VOT) can be utilized by the human auditory system to distinguish between voiced and devoiced stops such as /p/ and /b/ in English, /t/ and /tʔ/ Arabic.

The study is contributed into five sections:-

Section One is introductory, which contains the introduction, the problem, the hypothesis, the aim, the limitation and the value of the study. Section Two shows the definitions and types of VOT: positive, negative, zero VOT and role of VOT. Section Three deals with the measurement and categorical perception of VOT, these ways of measurements are spectrograms, waveform and lagtime.

Section Four investigates the VOT of two languages, Arabic and English in details with a comparison between these two languages. It ends with a number of conclusions. One of these conclusions is that Arabic VOT is different from English VOT and this approved the hypothesis.

List of Abbreviations

CA: Classical Arabic.

CV: Consonants vowel.

MSA: Modern Standard Arabic.

MS: Milliseconds. VOT: Voice Onset Time

Section One

Voice Onset Time (VOT) is a temporal parameter associated with stop consonants in general. It is the main feature used to differentiate between voiced stops and devoiced stops. It is used to refer to the difference of articulation between the users of Arabic and English languages and the way they produce them.

1.1 The problem

VOT causes problem to the users of Arabic and English languages, due to the difference of duration of VOT in these two languages.

1.2 The hypothesis

It is hypothesized that Arabic stops /p, t, d, tʔ, dʔ/ and English stops /p, t, d, k, g, b/ have different VOT for each language.

1.3 The aim

The aim of this study is to investigate the VOT in both Arabic and English languages to find out the differences of VOT between English and Arabic.

1.4 The limitation

It is limited to discuss the VOT of Arabic stops /p, t, d, tʔ, dʔ/ and English stops /p, t, d, k, g, b/ consonants only.

1.5 The value

This study is pedagogically very important for the users(learners) of these two languages to distinguish between the Arabic and English stops.

Section Two

Definitions, Types and the role of VOT

2.1 Definitions

Lisker and Abramson(1964:384), Rogers(2000:243), and Davenport and Hannahs (2010:69) all agree upon the definition that phonation onset or VOT is defined as the length time (period) between the onset of voicing pulses and the release of the primary occlusion of the vocal tract. VOT is relevant only for stop consonants. This period is usually measured in milliseconds.

Crystal (1989:329) defines VOT as "the point in time at which vocal-cord vibration starts, following the release of a closure in a voiceless unaspirated plosive, there is a delay before the voicing starts; in a voiceless aspirated plosive, the delay is much longer, depending on the amount of aspiration .The amount of delay, in relation to the types of plosive, varies from language to language."

2.2 Types of VOT

There are three types of VOT; zero VOT, positive VOT, negative VOT (Rogers, 2000:243).

2.2.1 Zero VOT

Lisker and Abramson (1968:21) point out that zero VOT means where the onset of vocal fold in vibration coincides approximately with the plosive release. Morris et al (2000:136) explain that zero VOT is the case that the vibration of the vocal fold (their onset) coincides with the release of the oral closure. Ashby (2011:124) adds that what happens through the process of producing zero VOT is that the vocal folds only start to vibrate simultaneously with the removal of the oral constriction, as shown in the following figure

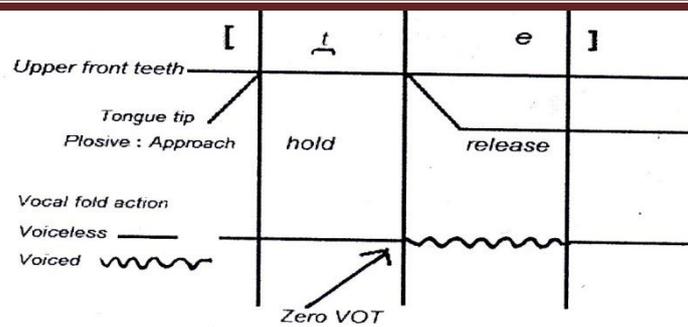


Fig () : Zero VOT

Figure () adapted from Ashby (2011 : 124

If someone produces a quite unaspirated /p/and that the glottis is already "throttled down" to about the size of the whisper orifice during the closed phase of the stop consequently the vocal folds can start vibrating immediately upon the release of the oral of the closure, i.e. the VOT is virtually zero and we can represent the utterance as /pa/. The unaspirated stop /p/ in /pa/ is still voiceless, since the vocal folds are not vibrating at all during the closed phase of the stop (Catford, 1988:193).

2.2.2 Positive VOT

Lisker and Abramson (1964:390) mention that positive VOT means that there is a delay in the onset of vocal fold vibration after the plosive release. Ashby (2011:125) adds that voiceless continues into the beginning of the following sonorant, after the removal of the consonantal constriction there is along or positive VOT, as shown in the following figure

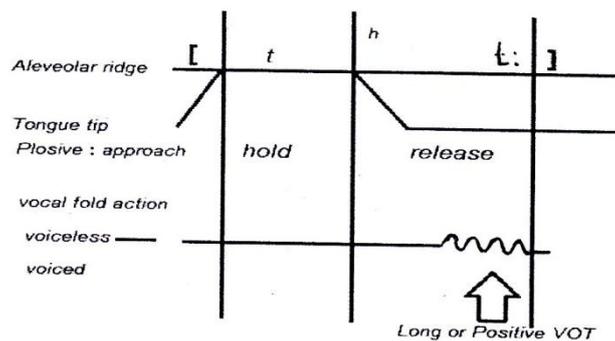


Fig f) : Long or Positive VOT

Figure () adapted from Ashby (2011: 126)

What happens here is that, the vocal folds remains fully open, effectively devoicing the onset to any following sonorant.

2.4 The Role of VOT in Speech

VOT is an important characteristic of stop consonants. It plays great role in perceptual discrimination of phonemes of the same place of articulation (Jiangetal, 2006:5).

It is also used in word segmentation, stress, related phenomena, and accented variation in speech patterns (Lisker and Abramson ,1964:399).

It is found, according to O'shaughnessy (1981:406), and Kessinger & Blumstein (1990:127) that VOT values are not affected by the change of gender of the addressee.

Section Three

Measurement & Categorical Perception of VOT

3.1 Measurement of VOT

There are three ways of measuring VOT. These ways are: spectrogram, waveform, lag time.

3.2 Spectrograms

"Phoneticians use a machine called a spectrograph which allows measurement and analysis of frequency durations, transitions between speech sounds and the like" (Davenport & Hannahs, 2010:59).

Rogers (2000:244), Morris et al (2000:133) and Ladefoged (2006:146) agree upon the idea that through this process of pronouncing VOT corresponds to the interval between the onset of energy "burst" representing the release of an articulatory constriction, and the first of the regularly spaced vertical striations representing the vocal fold vibrations.

3.3 Waveforms

Rogers (2000:131-8) and Ashby (2011:20-55) mention that the waveform is the easiest tool of analyzing and measuring VOT. It shows the pulses corresponding to each vibration of the vocal folds. The vibration in air pressure which associated with speech sound is recorded by the waveform. Thus, voiced sounds show up on the waveform as larger patterns than voiceless sounds. Consonants are also distinct from one another. Waveforms allow us to see differences in voicing and in manner of articulation and can be useful when used in conjunction with spectrograms.

3.4 Lag time

The third way of measuring VOT is called Lag time, which is obtained from the burst to the onset of the first formant of the following vowel. In some studies by Morris et al (2000:134) Lag time measurements are obtained from the closure releases at the beginning of the burst to the beginning of the second formant of the following vowel. The choice of the second formant as a determiner of VOT is particularly functional for voiced /b,d,g/.

3.5 Categorized perception of VOT in The English language:

Clark & Yallop (1995:313) and Radford et al. (2009:113) agree upon the definition of VOT as a type of perception in which listeners can discriminate the phonetic categories, (voiced, voiceless), very well but they can't hear differences within these categories because listeners perceive in terms of categories (voiced and voiceless) rather than in terms of minutes gradation of sound.

It's possible to create a set of syllables, containing series of /p-b/ in which VOT is systematically varied. Radford et al.(2009:112) believe that with a short VOT one to expect subjects to perceive /b/, whereas with a long VOT, one predicts that they will perceive /p/; but with intermediate VOT value the case is different, Figures (1) and (2) below.

According to Clark & Yallop (1995:313) and Radford et al. (2009:112), if the delay is increased in small steps (say 10 ms.) from around zero to about 100 ms. after the release of the occlusion, English-speaking listeners continue to hear the stop as voiced to 20 and 30 ms. ,always depending on the particular stimulus properties. If, however, the stimuli fall on the same side of the boundary, the subjects' responses indicate that they are guessing, i.e. they cannot perceive the difference between a stimulus with a VOT of, say, 40ms. and another with a VOT of, say, 60 ms.

Radford et al. (2009:112) show the result of an identification experiment on the perception of /p/ and /b/ with VOT varying along the X-axis.

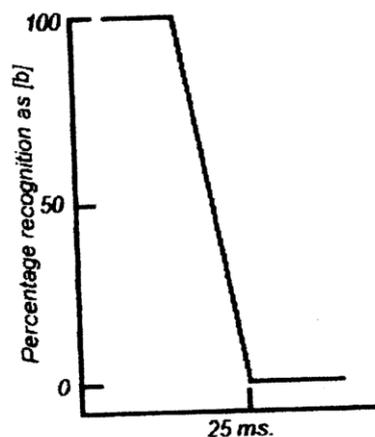


Figure (1) :Results of an identification experiment for a [b_p]-series

Figure (1) Adopted from Radford et al. (2009 : 112)

Radford et al. (2009:113) show the result of a discrimination experiment for a /b-p/ series, as follows:-

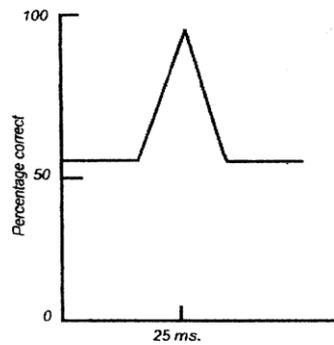


Figure (9) : Results of a discrimination experiment for a [b _ p]-series.

Figure (9) adapted from Radford (2009 : 113)

Section Four

Arabic VOT vs. English VOT

4.1 The Arabic language Overview

Arabic is a Semitic language, and it is one of the oldest languages in the world. Currently, it is the second language in the number of speakers. (<http://encrta.msn.com/media>)

Arabic is the first language in the Arab world, i.e., Saudi Arabia, Jordan, Yemen, Egypt, Syria, Lebanon, etc., Arabic alphabets are used in several languages such as Persian and Urdu. The Modern Standard Arabic (MSA henceforth) consists of (34) sounds: (28) consonants and (6) vowels (Alkhouli, 1990:9).

MSA has three long vowels (/i:/,/a:/,/u:/) and three short vowels (/i/,/a/,/u/), while English has twelve vowels. Arabic has fewer vowels than the English language.

Arabic sounds contain two distinct classes. They are pharyngeal and emphatic sounds. These two classes can be found in only Semitic languages like Hebrew (Elshafei, 1991:80).

The allowed syllables in the Arabic language are:

CV, CVC, and CVCC where (V) indicates a (long and short) vowel, while (C) indicates a consonant. Arabic utterances can only start with a consonant (Elshafei, 1991:80).(Alkhouli, 1990:10)All Arabic syllables must contain at least one vowel. Also Arabic vowels can't be initialed and they can occur either between two that exist in MSA also exist in classical Arabic (CA henceforth). The CV type is short while all the others are long syllable can also be classified as open or closed, an open syllable ends with a vowel while a closed syllable ends with a consonant(Furui,2001:32).

House (1961:1176) illustrates the set of stop sounds by classifying them into groups, saying that the set of stop sounds in the Arabic language consist of eight sounds and they can be classified into

emphatic and non-emphatic or voiced and devoiced. These sets are best illustrated in the following table.

Table(1)stop sounds in the Arabic language adopted from (House, 1961:1176)

			Bilabial	Alveo-dental	Velar	Uvular	Glottal
Stop	Voiced	Emphatic		/d/			
		Non-emphatic	/b/	/d/			
	Devoiced	Emphatic		/t/			
		Non-emphatic		/t/	/k/	/q/	/ʔ/

4.2The Arabic language VOT

There is a huge shortage of modern research in the Arabic language in the fields of references and resources regarding digital speech and language processing(Alghamdi,2004:65).

Alghamdi (2006:99) investigates the values for Saudi dialect in the Arabic language. The results of average VOT:-

- 1- For /t/ is 39 milliseconds.
- 2- For /k/ is 42 milliseconds.
- 3- For /t/ is 21 milliseconds.

Mitleb's (2009:133) findings about an Arabic Jordanian accent VOT values are as following for neighboring short vowel /i/ :-

- 1- For /d/ is 10 milliseconds.
- 2- For /t/ is 37 milliseconds.
- 3- For /k/ is 39 milliseconds.
- 4- For /g/ is 15 milliseconds.

For neighboring long vowel /i:/ :-

- 1- For /d/ is 23 milliseconds.
- 2- For /t/ is 64 milliseconds.
- 3- For /k/ is 60 milliseconds.
- 4- For /g/ is 20 milliseconds.

According to Alghamdi's experiment (2006:100), he finds out that for Ghamidi dialect of the Arabic language, the results of average VOT for vowel /a/:-

- 1- For /t/ is 25 milliseconds in the initial position of the word.
- 2- For /k/ is 30.3 milliseconds in the initial position of the word.

In Aldahri & Alotaibi (2010:29-31) experiment, they find out that the /d/ VOT value in (MSA) range:-

- 1- For /d/ between 12 and 22 milliseconds.
- 2- For /t/ between 38 and 93 milliseconds.

They (ibid) conclude that the VOT values of these stops /t/, and /d/ are positive regardless of the voicing, where /d/ is a voiced sound, but /t/ is not.

This is not the case for the same sounds in the English language where the voiced stops have negative VOT values, but devoiced ones (e.g., /t/) have positive VOT values. (ALDahri, 2012:17)

According to ALDahiri (ibid), there are four MSA Arabic stops namely /d/, /d?/, /t/ and /t?/ by analyzing their VOT values. He reaches the conclusion that VOT values for these stops are positive, he realized also the fact that VOT values of /d/ and /t/ phonemes are always more than VOT values of /d?/ and /t?/ phonemes.

There are two main standards in the Arabic language which are MSA Arabic and CA Arabic by computing, analyzing and comparing the VOT. It is found that for the MSA and CA Arabic, voiced sounds have short VOT while the devoiced sounds have long VOT. In addition, it is found that VOT values vary from one Arabic dialect to another. This shows that VOT can be used for dialect classification or detection <http://encrta.msn.com/media>

4.3 VOT values of English stops

Lisker and Abramson (1964:400) show that the VOT values for the English stops are the following:-

- 1- For /d/ values are ranging between 0 milliseconds to -155 milliseconds.
- 2- For /t/ values are ranging between 30 milliseconds to 105 milliseconds.
- 3- For /b/ values are ranging between 0 milliseconds to -130 milliseconds.
- 4- For /p/ values are ranging between 20 milliseconds to 120 milliseconds.
- 5- For /g/ values are ranging between 0 milliseconds to -150 milliseconds.
- 6- For /k/ values are ranging between 50 milliseconds to 135 milliseconds.

Peterson & Lehiste (1960:700) present the results for their studies of VOT English stops:-

1- For /p/ it is 58 milliseconds.

2- For /t/ it is 69 milliseconds.

3- For /k/ it is 75 milliseconds.

Flege et al (1981:127) shed light on the result of their studies of VOT English stops as follows:-

1- For /p/ it is 46 milliseconds.

2- For /t/ it is 56 milliseconds.

3- For /k/ it is 67 milliseconds.

Lisker & Abramson (1964:400) investigate the VOT values showing that the voiced sounds have negative VOT, whereas the unvoiced sounds have positive VOT and this means that the vibration of the vocal cords in voiced sounds start before devoiced sounds.

4.4 Comparing VOT between English and Arabic

After detecting the references Lisker & Abramson (1964:440), Alghamdi (2006:115), and Mitleb(2009,131) the researcher put the average of VOT values of /d/, /t/ for the two languages under study (English and Arabic).

Table (2) the average of VOT values of /d/ and /t/ for the two languages (English and Arabic)

Source	/d/ VOT (msec) Voiced	/t/ VOT (msec) Devoiced	Language & Dialect
Mitleb(2009) (short vowel)	10	37	Jordanian Dialect
Mitleb(2009) (long vowel)	24	64	Jordanian Dialect
Algamdi (2006)	-----	30	Saudi Dialect
Lisker & Abramson (1964)	-102	70	English Language

From the previous table it is clear that the VOT for the voiced and devoiced Arabic MSA stops are positive. This is not the case in English where voiced stops have negative VOT values as reported in Lisker & Abramson (1964:440). English voiced stops have negative VOT values whereas devoiced ones have positive VOT values as reported by Lisker & Abramson (1964:440).

We can also note that there are wide variations of /t/ VOT values among different Arabic Dialects. This means that Arabic speakers vocalize /d/ stops in all three dialects in the same way. Arabic speakers of these three dialects have big variations of vocalizing the /t/ stop which is a devoiced stop.

This comparison shows that the hypothesis of this study is approved, since that English VOT differs from the Arabic VOT in that each language has different ranges of VOT.

The voiced stops have negative VOT in the English language and voiceless stops have positive VOT in English.

This is not the case in the Arabic language, in which the voiced stops have positive VOT and voiceless stops have negative VOT, according to the measures of VOT for voice and voiceless for the two languages, as what we have investigated previously and this is the point of the study.

4.5 Conclusions

In the light of the literature being surveyed it can be concluded as the following:

1- The Arabic stops /p,t,d,tʔ,dʔ/ have different VOT from the English stops /p,t,d,b,k,g/ .

This conclusion approved the hypothesis of the study.

2- It is found by <http://encrta.msn.com/media> that voiced sounds have short VOT while the devoiced sounds have long (VOT) in the two main standards in Arabic language which are MSA Arabic and CA Arabic.(Ibid) shows that VOT values vary from one Arabic dialect to another.

3- English language has short VOT for voiceless, but negative VOT for voiced sound.

4- Values of the English stops, the voiceless sounds have positive VOT, while the voiced sound has negative VOT. (Lisker & Abramson,1964:400)

5- Arabic stops /t/ and /d/ have positive VOT regardless of voicing, in which /d/ is voiced whereas /t/ is devoiced.

6- There are wide variations of /t/ VOT values among different Arabic dialects, the Arabic speakers vocalize /d/.

7- Languages can be divided into two groups according to the length of VOT. Arabic languages belong to group A because it has long VOT for voiceless but short voice (Alghamdi,2004:25). English language belong to group B which has short VOT for voiceless, (Das et al,2004:34).

8- Values of Arabic stops /d/ and /t/ phonemes are always more than VOT values of /tʔ/ and /dʔ/ phonemes.

References

- Abramson, A & Lisker, L. (1968), A cross language experiment in identification and discrimination, London: Longman.
- Al-Ani, S. (1970) Arabic phonology, The Hague: Mouton.
- Al-ghamdi, M. (2004), Analysis Synthesis and perception of voicing in Arabic, Riyadh: Al-toubah.
- Al-ghamdi, M. (2006), "Voice Print: Voice Onset Time as a Model." Arab Journal for security studies and Training, Riyadh: Al-Toubah.
- Ashby, P. (2011), Understanding phonetics, London: Hodder education.
- Clark, J. & Yallop, C. (1995) An introduction to phonetics and phonology, Oxford: Blackwell publisher Ltd.
- Crystal, D. (1989) A Dictionary of Linguistics and phonetics, London: Oxford. Basil Blackwell Ltd.
- Darenport, M. & Hannahs, S. (2010), introducing phonetics and phonology, London: Hodder education.
- Das, S. & Hansen, J. (2004) Detection of voice Onset Time (VOT) for unvoiced stops (/p/,/t/,/k/) using the Teager energy operator (TEO) for automatic detection of accented English, proceedings of the 6th Nordic signal processing Symposium.
- Al-Dhahri,S. and Alotaibi(2010) A Crosslanguage survey of VOT value for stops(/d/ ,/t/) ,The 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS 2010), Xiamen, China.
- Al-Dhahri, S. (2012) The Effect of Arabic Emphaticness on Voice Onset Time(VOT),2012 International Conference on audio, Language and Image processing(ICALIP 2012),Shanghai, China
- Flege, J. & Port, R. (1981) "Cross- language phonetic interference: Arabic to English", Language and speech, no. 24.
- Flege, J. (1979) phonetic interference in second language acquisition, India, Bloomington.
- Furui, S. (2001) Digital speech processing, Synthesis, and Recognition, UK: Newcastle.

- House, A. (1961) "On Vowel duration in English", Journal of The Acoustical Society of America, no.33.
- Jiang, J. & Chen, M. & Alwan, A. (2006) " On the perception of voicing in syllable-initial plosives in noise", Journal of The Acoustical Society of America, Vol.119, no. 2 .
- Al-khouli, M. (1990) Alaswatt Alaghawaiyah, Jordan: Amman, Darl Alfalah.
- Kessinger, R. & Blumstein, S. (1990) "Effects of speaking rate on voice-Onset time and vowel production:some implications for perception studies". Journal of phonetics, no. 26.
- Ladefoged, P. (2006) A Course in phonetics, Boston: Thomson wordsworth.
- Lisker, L. & Abramson, A. (1964), A cross language study of voicing in initial stops: Acoustical measurements, word, volume, No.3.
- Mitleb, F.(2009) Voice Onset of Jordanian Arabic Stops, The 3rd, International conference on Arabic Language Processing, Morocco: Rabat.
- Morris, J & Francis, E., pascal, A.(2000), Clinical linguistics and phonetics, Florida: Taylor and Francis Ltd.
- MSN Encarta. (2009) Languages spoken by more than 10 million people. <http://encrta.msn.com/media>
- O'shaughnessy,D. (1981)"A study of French Vowel and Consonants Durations", Journal of phonetics. No.9.
- Peterson, G. & Lehiste, I. (1960) "Duration of Syllable nuclei in English", Journal The Acoustical Society of America, no.32.
- Radford, A. & Mortin, A. & David, B. & Harold, C. Andrew, S. (2009) Linguistic: An Inroduction, United kingdom: Cambridge university press.
- Rogers, H.(2000), The sounds of language, London: Longman.
- Yeni-Komshian, G & Caramaza, A. & Peterson, M. (1977) " A study of voicing in Labanese Arabic ", Journal of phonetics.
- Al-shafei, M. (1991) "Toward an Arabic Text-To-Speech System", The Arabian Journal for Science and Engineering, vol.16, no.4b.

المخلص

تحتوي الأصوات في اللغة على مختلف الخواص من هذه الخواص هي وقت استهلال الصوت (VOT) وهذه الصفة تشير الى الأختلاف بين اللفظ للأصوات الأنفجارية في اللغتين العربية والأنكليزية ويمكن ان نميز بين الأصوات الأنفجارية المهموسة والمهموزة عن طريق (VOT) .

قسمت الدراسة الى خمسة اجزاء:-

الجزء الأول يتضمن المقدمة، المشكلة، الفرضية، الهدف، حدود البحث وقيمة البحث.

الجزء الثاني يوضح التعاريف وانواع VOT ودورها في اللغة.

الجزء الثالث يتعامل مع قياس (VOT) وطريقة ادراكه.

الجزء الرابع يبحث ال (VOT) في اللغتين العربية والأنكليزية والمقارنة بينهما ونتائج البحث التي توصل لها الباحث.

ينتهي البحث اخيرا في المصادر التي استخدمت في البحث