

## The Effect of Context on Voice Onset Time

Ziyad Rakan Kasim <sup>(1)</sup>

### المخلص

عندما تكون الأصوات الانكليزية /p,t,k/ مسبوقة بـ /s/ ومتبوعة بصوت علة فإن زمن بداية الجهر (VOT) يقل ينتج عن ذلك اختفاء نَفَسِيَّة (aspiration) هذه الأصوات. يهدف هذا البحث إلى اختبار هذه الظاهرة في اللهجة العربية الموصلية. درست الأصوات النَفَسِيَّة الثلاثة /p,t,k/ ضمن موقعين: في بداية الكلمة وعندما تكون مسبوقة بـ /s/. وتم تسجيل كلمات تحتوي على هذه الأصوات وتحليلها تحليلًا صوتيًا طيفيًا (spectrographically) وقياس زمن بداية الجهر (VOT) لكل من هذه الأصوات في الموقعين كلاهما. أظهرت نتائج البحث أن قيم زمن بداية الجهر للأصوات الانفجارية النَفَسِيَّة الثلاثة لم تختلف، على العكس من اللغة الانكليزية، إذ لم يكن لاختلاف الموقع تأثير إحصائي دال على هذه القيم.

### Abstract

Voice onset time (VOT) is reduced when English /p,t,k/ are preceded by initial /s/ before a vowel, which results in a loss of the aspiration of /p,t,k/. The present paper attempts to test this phenomenon in Mosuli Arabic. The three voiceless stops /p,t,k/ are examined in two contexts: initially and when preceded by /s/. Words containing these stops were recorded and analyzed spectrographically and the VOT was measured for each stop initially and when preceded by /s/. The results of the study show that the VOT values of the voiceless stops do not show the difference noted in English; the context did not have a statistically significant effect on the VOT values.

### 1. Introduction

<sup>(1)</sup> Asst. Lecturer, Department of English, College of Education, University of Mosul.

Voice onset time (VOT) refers to the beginning of vocal-fold vibration relative to the release of a stop (Ladefoged, 1993:142). It has been found that VOT has an important role, and may be the most salient acoustic cue in discriminating the voiceless stops from the voiced ones in English (see Lisker and Abramson, 1964:399). The voiced stops usually have a negative value of VOT, i.e. voicing starts before the release of the stop, while the voiceless stops have a positive value of VOT, i.e. voicing starts after the release of the stop.

Aspiration is defined as "a period of voicelessness after the stop articulation and before the start of the voicing for the vowel" (Ladefoged, 1993:50)<sup>1</sup>, (for similar view see Gimson, 1976:152; Fry, 1979:122; and Roach, 2000:34, among others). This "period of voicelessness" represents the positive value of VOT. The phenomenon of aspiration has received a great deal of attention in English since it plays an important role in distinguishing /p,t,k/ from /b,d,g/ in initial position. Roach (2000:34), for example, emphasizes the importance of aspiration stating that "the most noticeable and important difference [...] between initial p,t,k and b,d,g is the aspiration of the voiceless plosives p,t,k".

Aspiration in English is affected by context. The aspiration of initial /p,t,k/ is said to disappear when these stops are preceded by /s/ (see Jones, 1973:68-74; Gimson, 1976:151-2; Ladefoged, 1993:50,

---

(<sup>1</sup>) For a detailed account of aspiration, see Ladefoged and Maddieson (1996).

and Catford, 1994:57 among others). Gimson (1976:151) states that “when /s/ precedes /p,t,k/ initially in a syllable, there is practically no aspiration, even when the syllable carries a strong accent”<sup>1</sup>.

What is of interest in this study is the effect of the loss of aspiration. When aspiration is lost in this context the contrast between /p,t,k/ and /b,d,g/ disappears. Roach (2000:34) shows that “the unaspirated p,t,k of the initial combinations sp, st, sk have the sound quality that makes English speakers perceive a plosive as one of b,d,g”. Experimental evidence supports this observation (see Lotz *et al.*, 1960). This means that aspirated /p,t,k/ have greater VOT values than their unaspirated counterparts; i.e. initial /p,t,k/ have greater VOT values than /sp, st, sk/, respectively.

This study aims at investigating the effect of context on VOT in Mosuli Arabic. The question that this study tries to answer is: Does /s/ affect the VOT of /p,t,k/ in Mosuli Arabic?

## **2. Methodology**

### **2.1 Subjects**

Twelve native speakers of Mosuli Arabic served as subjects of this study. The subjects were 20 to 45 years old, eight males and four females. They were all educated and had no language defects.

### **2.2 Material**

---

(<sup>1</sup>) It is worth mentioning that Gimson (1976:152) extends the effect of /s/ on /p,t,k/ even when the /s/ is in another syllable, as in “discussed”.

A list of six monosyllabic words was arranged in pairs of initial stop and a stop after initial /s/ as shown in Table 1 below.

**Table 1:** The Mosuli Arabic words used in the test.

initial stop	meaning	stop after initial /s/	meaning
/pe:r/	pair	/spe:r/	spare
/tu:l/	kind of cloth	/stu:l/	stool
/ka:fi/	enough	/ska:fi/	shoe maker

This procedure allows making a comparison between the two stops in relation to VOT and thus any difference between the two contexts would be attributed to the presence (or absence) of /s/. The words were printed on 7 cm x 14 cm flash cards.

### 2.3 Procedure

The subjects' performance was recorded individually in the language laboratory of the Department of English, College of Education, University of Mosul. A mobile phone (type Nokia 6120c) with a connected lapel microphone<sup>1</sup>(type HS-47) was used as a recorder. The sound format used for recording was "wav"<sup>2</sup> with a sampling rate of 8000 Hz.<sup>3</sup>

---

(<sup>1</sup>) This type of microphone has many advantages for recording, e.g. the ability to "mount it in precisely the right place" Tatham (1984).

(<sup>2</sup>) The advantage of this sound format is that it is the same format that is used by the computer software utilized for analyzing the data. As such, no converting software was used when the recordings had been transformed from the mobile phone to the computer and thus there was no signal loss of any type.

(<sup>3</sup>) This sampling rate is found to be the best rate for analyzing VOT spectrographically since it clearly shows the separate vocal-fold==vibration in the spectrogram. This has been tested by the researcher

Each subject was presented with the flash cards one at a time and was asked to read the words in normal everyday speed. A brief session preceded the recording in which each subject familiarized him/herself with the words. The subjects read each word three times. The flash cards were shuffled for each round. Thus, a total of 18 tokens was obtained for each subject. When analyzing the data it was found that one of the male subjects did not perform well; his speech seemed unnatural. Thus, this subject's performance was eliminated from analysis. Only the performance of the other 11 subjects was considered. Therefore, the total number of tokens obtained was 198 (see Appendix for the subjects' performance).

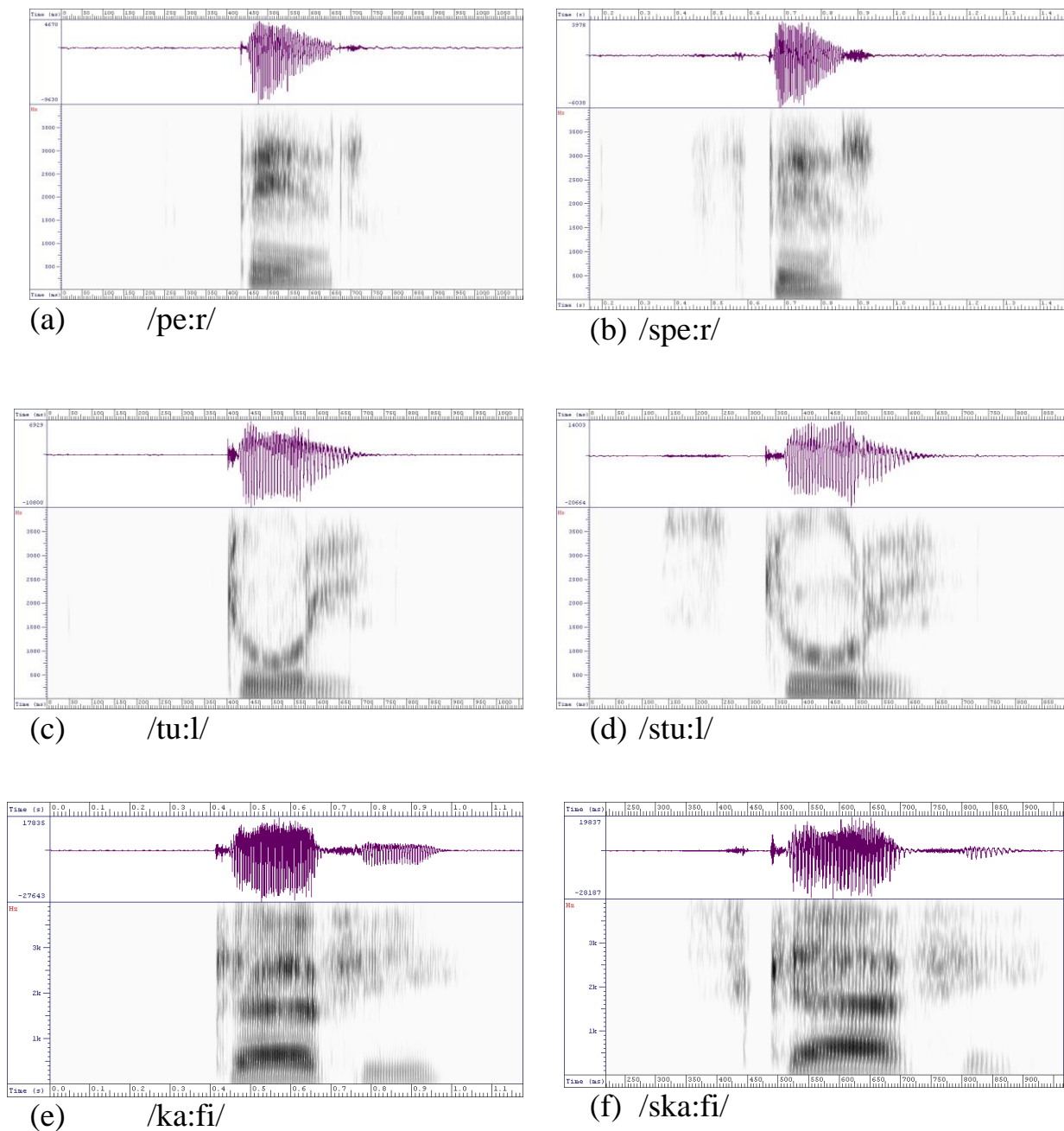
### **3. Data analysis**

A computer software was used for data analysis<sup>1</sup>. This software generated a wide-band spectrogram with waveform for each word. VOT was measured in the following way: VOT represents the period from the beginning of the release (represented by the presence of noise energy in the spectrogram, specially at higher frequencies) to the beginning of vocal-fold vibration (represented by the vertical striations present, specially in F1 in the spectrogram as shown in Figure 1 below).

---

using different sampling rates, a utility provided by the software WASP that is used for data analysis (see note 6 below).

(<sup>1</sup>) The computer software is called WASP (Waveform Annotations Spectrograms and Pitch).

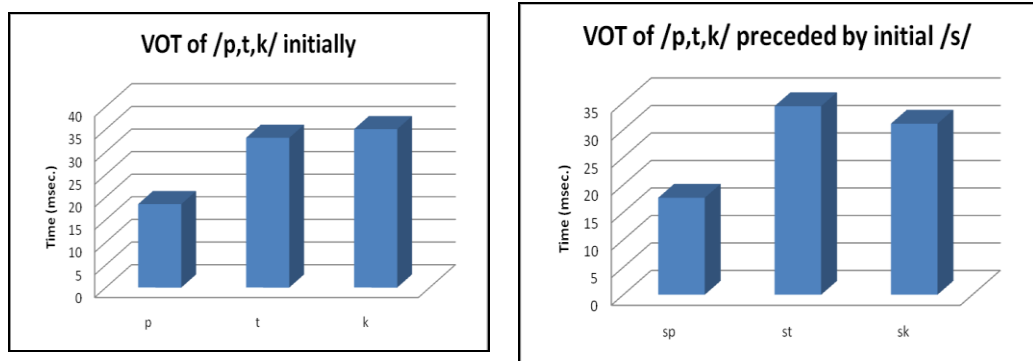


**Figure 1:** Waveform (upper part) and wide-band spectrograms (lower part) of the words studied.

The only difference in the spectrograms of /p,t,k/ and /sp, st, sk/ is that /sp, st, sk/ have a noise energy present at higher frequencies (2000 Hz and higher) preceding the stop closure representing the /s/ sound (see Figure 1, b, d, f). The data were fed into Microsoft Excel to obtain mean values, which were then transformed into graph bars (see Figures 2 and 3).

#### 4. Results

The results of the data analysis are presented in Figure 2 below.

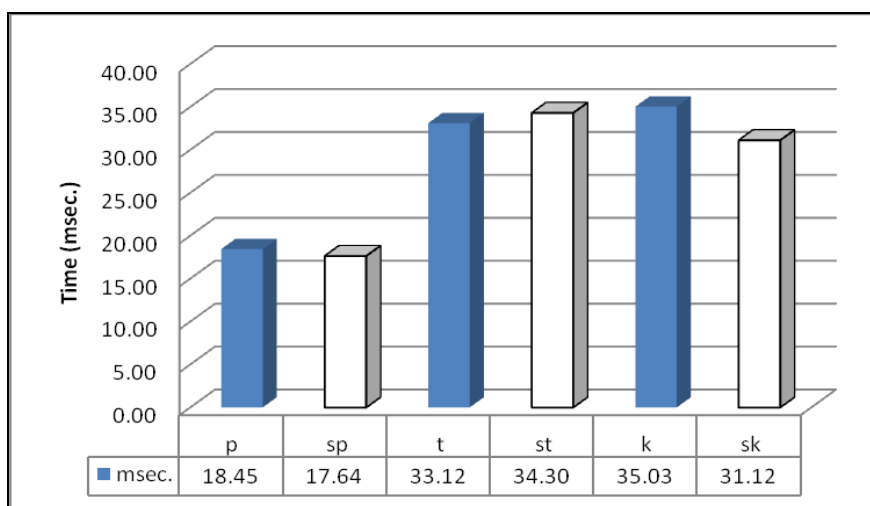


**Figure 2:** (a) VOT of /p,t,k/ initially and (b) when preceded by /s/

Figure 2a shows the VOT values of initial /p,t,k/. The figure shows a pattern in which the VOT values increase from /p/ to /k/. Figure 2b shows the VOT values of /p,t,k/ when preceded by initial /s/ (henceforth /sp, st, sk/). The increase of VOT values noted in Figure

2a is absent in Figure 2b; VOT increases from /p/ to /t/ but decreases from /t/ to /k/<sup>1</sup>.

Figure 3 below shows a comparison between VOT values of /p,t,k/ and /sp, st, sk/.



**Figure 3:** VOT comparison between /p,t,k/ and /sp, st, sk/.

A look at this figure reveals that there is little difference between the two contexts, i.e. between initial stops and stops after initial /s/. In order to test the validity of this, statistical analysis was used<sup>2</sup>; the tests

(<sup>1</sup>) It would be of interest to investigate this difference between the different places of articulation. However, the main interest of this study is the “inter” rather than the “intra” difference between the members of each pair of stops.

(<sup>2</sup>) The statistical analysis was performed by Mr. Marwan Abdul-Aziz Dabdoub (Assistant Professor at the Department of Statistics and Informatics, College of Mathematics and Computer Sciences, University of Mosul), whom the researcher is much indebted to.



used were F-test and Levene's test. The results of the statistical analysis indicate that there is no significant difference between the VOT values of p, sp; t, st; and k, sk respectively<sup>1</sup>.

## 5. Discussion

This study has shown that there is no significant VOT difference between initial /p,t,k/ and when these voiceless stops are preceded by /s/ in Mosuli Arabic; i.e. there is no context effect. This indicates that even if we were to cut the /s/ part of the sequences /sp, st, sk/ still native speakers of Mosuli Arabic would perceive the remaining stops as /p,t,k/, respectively.

In English, however, removal of the /s/ part of the sequences /sp, st, sk/ would yield /b,d,g/, respectively; i.e. the native speakers of English would perceive /p,t,k/ as /b,d,g/ (see Catford, 1994:57). This has proved to be the case in an experiment that tested the effect of removing the /s/ part of the sequences /sp, st, sk/. Lotz *et al.* (1960:72) have found that the /p,t,k/ of /sp, st, sk/ “were identified with voiced lenis stops” when the /s/ part was removed. This result can be explained with reference to data on the VOT of English stops reported by Klatt, 1975 (cited in Bond and Wilson, 1980). The data reported show that the VOT of /p,t,k/ when preceded by /s/ ranged between 12 and 30 msec., and the VOT of /b,d,g/ ranged between 11 and 27 msec.

---

(<sup>1</sup>) Although the VOT value of /st/ is a little higher than that of /t/, yet this difference is not significant.

There is an obvious overlap between the two ranges. In this case, the stop following /s/ could equally be identified with either the voiceless stop or the voiced stop.

In Mosuli Arabic the picture is different. In an earlier study by Rahim and Kasim (forthcoming)<sup>1</sup> the VOT values of /b,d,g/ ranged between -56.3 and -78.1 msec. A comparison between /b,d,g/ VOTs and /p,t,k/ VOTs when preceded by /s/ (which ranged between 17.6 and 34.3, see Figure 3) shows that there is no overlap between the two ranges, i.e. the native speakers of Mosuli Arabic did not confuse the voiceless stops with their voiced counterparts. For this reason (i.e. that the voiced stops have negative VOT values), the VOT of the voiceless stops (in /sp, st, sk/) is not affected by a difference in context, in our case the presence of /s/, simply because they would not be confused with their voiced counterparts.

## 6. Conclusion

This research has attempted an investigation of the effect of context on VOT in Mosuli Arabic; more specifically the effect of /s/ on /p,t,k/. Unlike English in which VOT has been found to be affected by a preceding /s/, VOT in Mosuli Arabic is not affected by this context. Aspiration is said to play an important role in the identification of /p,t,k/ from /b,d,g/ in initial position in English. However, it does not seem to play the same role in Mosuli Arabic.

---

(<sup>1</sup>) This paper is to appear in the Journal of Education and Science.

Obviously, cues other than aspiration seem to play such a role in the distinction between the voiceless stops and their voiced counterparts; namely, the different VOT ranges of the voiceless stops and the voiced stops.

### **Acknowledgments**

The researcher would like to express his sincere gratitude to the subjects that participated voluntarily in this study, and to the teaching staff of the Department of Psychology, College of Education, specially to Dr. Yaser Mahfouth, and to Mr. Mahmoud Abdul-Hadi, of the Centre of Developing Teaching Methods and Staff Training, for their advice on statistical matters. Many thanks are also due to Prof. Abbas J. Rahim, of the Department of English, College of Education, for reading an earlier draft of this paper.

## References

- Bond, Z. S. and Wilson, Howard F. (1980). /s/ plus stop clusters in children's speech. *Phonetica*, 37, 149-158.
- Catford, J. C. (1994). *A Practical Introduction to Phonetics*. Oxford: OUP.
- Fry, D. B. (1979). *The Physics of Speech*. Cambridge: CUP.
- Gimson, A. C. (1976). *An Introduction to the Pronunciation of English*. (2<sup>nd</sup> ed.). London: Edward Arnold.
- Jones, D. (1973). *The Pronunciation of English*. (4<sup>th</sup> ed.). London: CUP.
- Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word initial clusters. *Journal of Speech and Hearing Researches*, 3, 71-77.
- Ladefoged, P. (1993). *A Course in Phonetics*. (3<sup>rd</sup> ed.). New York: Harcourt Brace Jovanovich.
- Ladefoged, P. and Maddieson, I. (1996). *Sounds of the World's Languages*. Oxford: Blackwell.
- Lisker, L. & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384-422.
- Lotz, J., Abramson, A. S., Gertsman, L. J., Ingemann, F., and Nemser, W. J. (1960). The perception of English stops by speakers of English, Spanish, Hungarian, and Thai: a tape cutting experiment. *Language and Speech*, 3, 71-77.
- Rahim, A. J. and Kasim, Z. R. (forthcoming). A spectrographic study of voice onset time in Arabic. *Journal of Education and Science*.
- Roach, P. (2000). *English Phonetics and Phonology: A practical course*. (3<sup>rd</sup> ed.). London: CUP.
- Tatham, Marcel A. A. (1984). Recording and displaying speech. In Chris Code and Martin Ball (eds.), *Experimental Clinical Phonetics: Investigatory techniques in speech pathology and therapeutics*. London: Croom Helm.

**Appendix:** The six words that contain /p,t,k/ initially and preceded by /s/ with the subjects' performance;

S= subject, T=token, and the values represent time in msec.

	S1			S2			S3			S4			S5			S6			S7			S8			S9			S10			S11			Total	Mean value
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3		
/perr/	14	11	17	11	9.8	18	28	24	43	10	6.1	12	12	17	17	26	36	29	18	16	29	26	33	22	12	11	15	22	11	13	17	12	11	608.9	18.45152
/kper/	16	18	17	13	12	17	20	18	38	10	11	12	18	15	17	26	34	42	14	18	10	16	23	16	14	12	12	13	16	11	15	14	24	582	17.63636
/tper/	27	44	61	37	37	24	24	17	33	18	16	26	36	39	33	38	37	29	33	28	39	34	27	32	16	22	50	37	32	26	55	34	52	1093	33.12121
/stet/	36	64	43	26	33	32	48	41	41	21	19	17	42	39	48	30	25	38	37	31	27	30	33	38	29	31	26	26	28	46	40	39	1132	34.30303	
/kset/	46	56	45	29	29	28	14	16	47	22	26	25	29	32	34	37	51	45	31	39	27	31	32	31	40	36	28	29	29	45	48	55	44	1156	35.0303
/stet/	37	55	37	22	23	20	19	31	41	18	19	17	35	37	36	43	40	46	29	30	23	28	29	24	26	32	26	31	28	30	38	39	38	1027	31.12121