DETECTION OF AUTOMATIC THE VOT VALUE FOR VOICED STOP SOUNDS IN MODERN STANDARD ARABIC (MSA)

Sulaiman S. AlDahri

King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia
saldahri@kacst.edu.sa

ABSTRACT

Signal processing in current days is under studying. One of these studies focuses on speech processing. Speech signal have many important features. One of them is Voice Onset Time (VOT). This feature only appears in stop sounds. The human auditory system can utilize the VOT to differentiate between voiced and unvoiced stops like /p/ and /b/ in the English language. By VOT feature we can classify and detect languages and dialects. The main reason behind choosing this subject is that the researches in analyzing Arabic language in this field are not enough and automatic detection of VOT value in Modern Standard Arabic (MSA) is a new idea. In this paper, we will focus on designing an algorithm that will be used to detect the VOT value in MSA language automatically depending on the power signal. We apply this algorithm only on the voiced stop sounds /b/, /d/ and /dˁ/, and compare that VOT values automatically generated by the algorithm with the manual values calculated by reading the spectrogram. We created the corpus, and used CV-CV-CV format for each word, the target stop consonant is in the middle of word. The algorithm resulted in a high accuracy, and the error rate was 0.80%, 26.62% and 11.71% for the three stop voiced sounds /d/, /dˁ/ and /b/ respectively . The standard deviation was low in /d/ sound because it is easy to pronounce, and high in /dˁ/ sound because it is unique and difficult to pronounce.

KEYWORDS

Arabic, VOT, MSA, POA, TEO

1. INTRODUCTION ABOUT VOICE ONSET TIME

VOT (Voice Onset Time) feature appears important in distinguishing between voiced and unvoiced stops in various languages. Phonation onset or VOT is defined as the interval (period) between the release burst of the stop and the onset of glottal vibration [1]. This period is measured in msec. This makes the VOT detection automatically is difficult. VOT as we have just described is relevant only for stop consonants [2]. Stop consonants are produced with a closure of the vocal tract at a specific place which is known as the place of articulation (POA) [3]. So, VOT is affected by the stop consonant's POA, which is different from one language to another. Also, VOT is affected by the speaker linguistic knowledge and vowel duration [3][4].

David C. Wyld et al. (Eds) : DBDM, CICS, CSIP, AI&FL, SCOM, CSE, CCNET-2016
pp. 45–56, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.60505

VOT value can be divided to three types, zero VOT, positive VOT and negative VOT. Zero VOT means where the onset of vocal fold vibration coincides very close to the release of the stop closure. Positive VOT means that there is a delay in the onset of vocal-fold vibration after the release of the stop closure; in this case, the voicing lag. In Negative VOT, the onset of vocal fold vibration precedes the release of the stop closure; in this case, the voicing lead [1][5]. Categories range from two-four depending on the certain language. For example, English and Spanish have two voicing categories, whereas Eastern Armenian and Thai have three voicing categories [6].

The VOT is an important characteristic of stop consonants that play a great role in perceptual discrimination of phonemes of the same POA [7]. Also, VOT is an important feature in stress related phenomena, word segmentation, and dialectal and accented variations in speech patterns [2]. Moreover, the previous researches, found that VOT values are not affected by the change of the age in both male and female [8]. It is well known that VOT varies to some extent with place of articulation. The principal findings are that: (1) the further back the closure, the longer the VOT (Fischer Jorgensen, 1954; Peterson & Lehiste, 1960); (2) the more extended the contact area, the longer the VOT (Stevens, Keyser & Kawasaki, 1986); and (3) the faster the movement of the articulator, the shorter the VOT (Hardcastle, 1973). These patterns have been known for many years [3].

VOT values are generally unobserved in fixed length frame based speech investigation. On the other hand, it is known that automatic speech recognition and performances can be understood by the help of VOT. Among the various applications of the use of VOT is the difficulty of accent detection. Non-native language can affect both the length and the quality of the VOT of English stops [9]. Depending on a research effort [9], VOT values can be used to discriminate Mandarin, Turkish, German, Spanish and English accents.

In languages which process two categories of voicing. Voiced and voiceless, the voicing onset usually starts before the release of the stop closure for a voiced stop. However, languages differ in the manifestation of these two categories. Depend on VOT, Lisker et al. [1] have a distinct number of voicing categories that are used contrastively, thereby forming different phonological entities of the respective system. Such categories range from two-to-four depending on the particular language. For example, English, Spanish, Tamil, Hungarian, and Dutch have two voicing categories, whereas Eastern Armenian and Thai have three voicing categories [1][6].

2. LITERATURE REVIEW

2.1. Automatic detection the VOT

There have been a number of previous researches proposing algorithms for automatic VOT measurement. Previous studies have used automatic measurements for speech recognition tasks (Niyogi and Ramesh, 1998, 2003; Ali, 1999; Stouten and van Hamme, 2009), phonetic measurement (Fowler et al., 2008; Tauberer, 2010), and accented speech detection (Kazemzadeh et al., 2006; Hansen et al., 2010). Some studies, focus on the problem of VOT measurement itself, and evaluate the proposed algorithm by comparing automatic and manual measurements (Stouten and van Hamme, 2009; Yao, 2009; Hansen et al., 2010; Lin and Wang, 2011). Moreover, one study used machine learning technique to design an algorithm to automatically measure VOT (Morgan and Joseph, 2012) [4].

On the other side, English was studied by Lisker et al. [1] using American and Britain dialects among more than nine other languages and dialects under different environments. In general, languages like English, Japanese, and German, were investigated for more than forty years for VOT stops [1]. Several studies that have been conducted in English, showed similar results to those of Lisker et al [1]. Peterson et al. [10] present their VOT results of /p/, /t/, /k/ as 58 msec, 69 msec and 75 msec respectively. Flege et al. [11] found the VOT of /p/ is 46 msec, /t/ is 56 msec and /k/ is 67 msec. In addition, close relations were found between the 18 languages investigated by Cho T. and ladefoged P. [3] which represented 12 different language families.

In Spanish and French languages that treat long lead or pre-voiced stops (i.e., long, negative VOTs) as voiced, while short-lag stops are classified as voiceless (Caramazza & Yeni-Komshian, 1974; Deuchar & Clark, 1996; Zampini & Green, 2001). In other words, in English (and Swedish) voiced stops are characterized by an articulator timing similar to Spanish voiceless stops [12]. In French language, the VOT is sufficient phonological cue for the distinction of the homorganic stop consonant pairs in French [13].

However, to replace manual measurement, we believe that an automatic VOT measurement algorithm should meet three criteria. Both the burst and voicing onsets are often highly transient, and because the effects of interest (e.g., VOT difference between two conditions) in studies using VOT measurements are often very small, the algorithm should have high accuracy by the chosen measure of performance. The cues to the burst and voicing onset locations vary depending on many factors like speaking style, speaker's native language, and different labs have slightly different VOT measurement criteria [4].

In addition, there was a general age effect on the L2 learners' categorical-perception behavior mirrored by negative correlations and also the overall differences between listener groups were significant for all three voicing continua [12].

Das and et al. was Automatic detected of VOT for unvoiced stops (/t/, /k/ and /p/) used the Teager Energy Operator (TEO). This algorithm is applied to accent classification using English, Chinese, and Indian accented speakers. Using the 546 tokens and consisting of 3 words from 12 speakers. The VOT is detected with less than 10% error when compared to the manual detected VOT. Also, pairwise English accent classification are 87% for Chinese accent, 80% for English accent, and 47% for Indian accent [9]. The TEO is a nonlinear energy tracking signal operator which has been used in speech and signal processing. It has been shown that TEO can be useful for detecting voiced/unvoiced speech, speech under stress, speech under vocal fold pathology and an automated sub-band frequency analysis is performed to detect VOT value [9].

In another effort, Okalidou A. et al. found that the developmental patterns Standard-Greek (SG) and Cypriot-Greek (CG) were different due to the number of contrasting voicing categories in each language/dialect. They found the VOT value three-way voicing contrast in CG is acquired later than the two-way voicing contrast in SG [6].

In several methods of VOT detection are considered, with the most accurate method based on tracking the laryngographic signal. This method is not possible unless a laryngograph is used while recording speech from the speaker. The other methods are based on tracking formant frequencies (F1, F2 or F3), performing spectrographic analysis, or tracking the onset of speech (f0) periodicity in the acoustic waveform. Manual involvement is required in all of these methods

to calculate the VOT value [9]. In study Parkash C. et al., they found approach for detection of VOT based on Bessel expansion and amplitude modulation component of the TEO [14].

2.2. VOT in Arabic language

The researches and resources in the speech processing field for Arabic language is not enough, especially in the subject related to VOT. Studies based on Arabic language varies in results, Al-Ani [15] and Mitleb [16] suggested that Arabic is a member of Group A, but Yeni-Komshian et al. [17], who based his research on Lisker al et. [1], found it to be from Group B. On the other hand, Flege [11] considered it neither belongs to Group A nor B.

In a study by Alghamdi [2], he analyzed the role of VOT in speaker identification and the effect of acquiring a second language in Ghamdi dialect for Saudi speaker. This research showed that a phonetic diversity between the first and second language is maximized when the speakers are more fluent in the second language. Moreover, Alghamdi [2] investigated the Saudi dialect of Arabic language, and the results of the average VOT for /t/, /k/ and /t$^{?}$/ were 39 msec, 42 msec and 21 msec, respectively.

Another effort, Mitleb [16] analyzed VOT of Jordanian Arabic stops. One of his results is that the VOT value depends on the vowel length in long vowel environment more than its dependency on the length in short vowel environment. Also, he realized that VOT distinguishes Arabic unvoiced and voiced stops as the case in English.

In Aldahri research, he concluded that VOT values of these stops 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 English language, where voiced stops have negative VOT values, but it is positive for unvoiced (e.g., /t/) [19]. Another research, Aldahri and Alotiabi, they found the emphaticness property decreases VOT values if compared with VOTs of nonemphatic [20]. In another effort, MSA Arabic language is found to have both long and short VOT for unvoiced and voiced sounds respectively [18]. However, the researches to detect the VOT in Arabic language are not applied until now. So, that appears the important this research in this field.

The aim of this study is to design an algorithm to automatically detect the VOT for voiced stop sounds in MSA Arabic. The rest of this paper is organized as following: Section 3 describes the used corpus and the methodology. Section 4 gives the results of the research in addition to some discussions. Before the final section, Section 5 summaries the results of the research. Finally, Section 6 is to list our references.

3. EXPERIMENTAL FRAMEWORKS

The set of stop phonemes in MSA language consists of eight phonemes that can be classified into: emphatic and non-emphatic or voiced and unvoiced [21]. These sets are shown in Table 1 with a description of their place of articulation, voicing, and emphaticness property [19].

Table 1. Stop phonemes in MSA Arabic language.

			Bilabial	Alveo-dental	Velar	Uvular	Glottal
Stop	Voiced	Emphatic		/d$^{?}$/			
		Non-emphatic	/b/	/d/			
	Unvoiced	Emphatic		/t$^{?}$/			
		Non-emphatic		/t/	/k/	/q/	/?/

3.1. Data set

The used corpus is based on the previous work explained in [18][19], with each word containing one of the seven targeted sounds. The recorder speakers were chosen carefully to insure achieving the utterance quality required for the work of this search.

Those speakers include native and non-native Arabic speakers. The speakers are between thirteen and forty years old. The words are chosen to make sure that the targeted sounds are in the middle of the word while the preceding and the succeeding phonemes with respect to the targeted phonemes are always the same (/a/). The word structure is CV-CV-CV. The speaker repeats this set of words for 5 trials.

The total number of the recorded utterances is 2800 (80 speaker's × 7 words × 5 trials) recorded words. As we know, the voiced stop sounds in MSA Arabic are /d/, /b/ and /d$^{?}$/. We managed to record three words (each one containing one of the voiced stop sounds) for 1200 recorded words (80 speaker's × 3 words × 5 trials). The sampling rate was set at 16000 sample/seconds (16 kHz) and resolution at 16 bit using one channel (mono).

Each record file is named according to the following naming pattern: SxxCyEzTw.wav. In this string S, C, E and T stand for speaker, consonant, emphatic, and trial, respectively. The xx (two digits number) displays the speaker number. The one digit y, 1 refers to /d/ or /d$^{?}$/, 3 refers to /t/or /t$^{?}$/, 5 refers to /k/, 6 refers to /q/ and finally 7 refers to /b/. The fourth digit z, is a binary flag set to 0 for non emphatic and 1 for emphatic. One digit z is the emphatic/non-empathic sound identifier as following: 1 refers to the pair /d$^{?}$/ or /d/ , 3 refers to the pair /t$^{?}$/ and /t/. The last digit, w, is a one-digit number representing the trial number.

For the goal of this paper, we used carrier words that only carry the stop voiced consonants /d/, /d$^{?}$/ and /b/. The Table 2 is lists the part of the corpus that is targeted in this research work.

Table 2. Stop sounds and carrier words which used in this research

Arabic Alphabet Carrier	IPA Symbol	Carrier words (CV-CV-CV)	Transcription	Code
daal د	/d/	نَدَرَ	/nadara/	C1e0
Dhaad ض	/dˀ/	نَضَرَ	/nadˀara/	C1e1
Baa ب	/b/	نَبَرَ	/nabara/	C7e0

3.2. VOT Detection Algorithm

The methodology used to extract the VOT value is written using Java. The Power values of the signal file are used as input to the program in form of an array, each line or each value represents 1 msec in time. Then, the array is scanned line by line to create queues of increasing values. Each queue ends when the next value in the array is less than the current value (each queue represents a part of the signal from a local min to the next local max). However, In order to handle noise, the following condition is performed:

The queue contains a series of increasing values, but to ignore slight decreases that happened due to signal noise, we check if the decrease from the current power value to the next power value is less than or equal to a given value (0.1 dB by default), ignore it and continue building the queue, but if it's larger than that, stop and consider the current value the end of this queue.

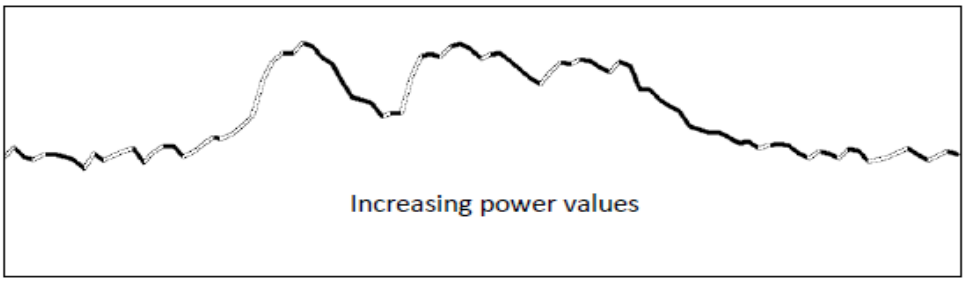

Figure 1. All created queues after scanning

To remove useless queues (that can't be a VOT), each queue is then checked for the following:

1. Power Delta: which is the difference between the highest and the lowest power values in the queue (the start and end). If it is less than a given value then ignore the queue. The value depending on our experience is between 2 to 4 dB.

2. Signal Length (Queue Length): If the length in msec is less than a given value then ignore the queue. Depending on our experience the VOT value is always more than 5 msec in MSA language.

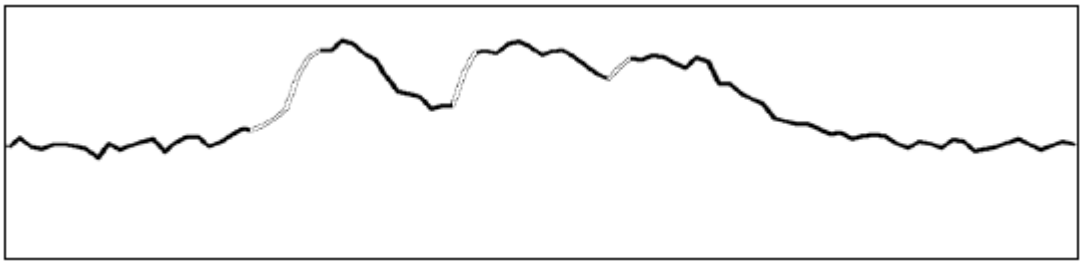

Figure 2. Remaining queues after the cleaning step

We're always looking for the VOT in the second letter (the middle letter) of three letter words. The algorithm benefits from the Pitch Contour data (if available) to skip the parts of the signal with a Pitch Contour value of 0 (unvoiced sound, which is usually the beginning and end of the file) as show in Figure 3. It also has an option to only read a window of a given length (usually between 200 to 40 mesc) that represents the center of the signal, and ignores the rest of the file. The equation to calculate the center is:

$$X1 = \frac{\text{Pitch counter}}{2} - \text{Given length}$$

$$X2 = \frac{\text{Pitch counter}}{2} + \text{Given length}$$

The VOT value should be between X1 and X2 as show in Figure 3.

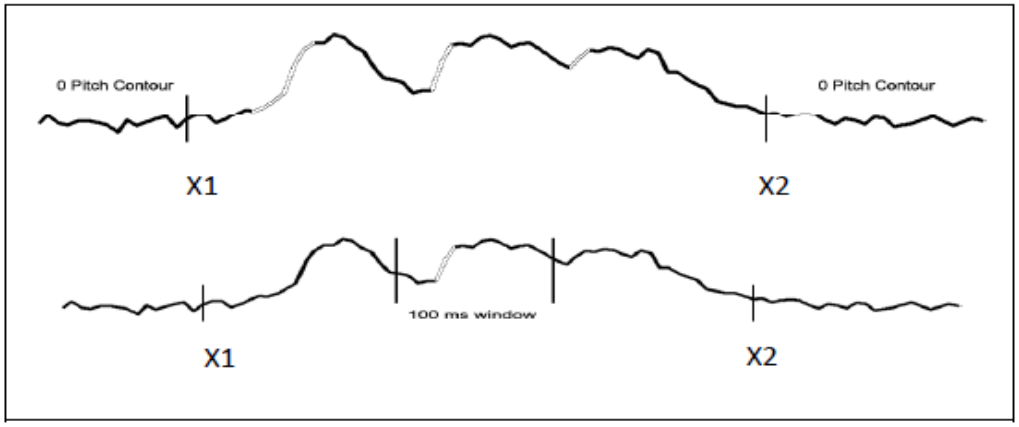

Figure 3. Remaining VOT candidates

4. DISCUSSION AND RESULTS

We apply and investigation the algorithm on voiced stop sounds in MSA Arabic which are /d/, /dˀ/ and /b/. We examined many audio files in our corpus, but the tables lists partial subset of VOT values, specifically sixteen audio files for each of the three investigated stops. Twenty audio files are for each sound as shown in Tables 1, 2 and 3. VOT values were obtained by measuring the distance between the onset of energy in the formant frequency range representing the release of air pressure and the first vertical striations representing glottal pulsation from wide-

band spectrograms of recorded words [13]. Our algorithm depends on the power signal to determine VOT value. So, the start release of vocal tract is represented by the point where the signal start increasing and the start vibration vocal cord is represented by the point where the signal stops increasing as shown in Figure 3.

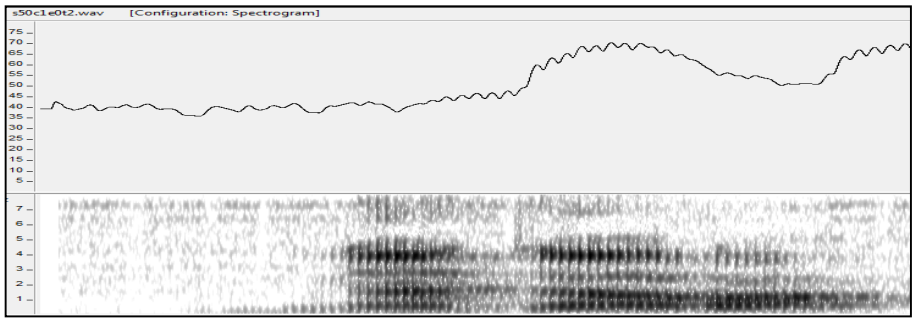

Figure 4. Spectrogram and power signal for stop voiced sound. Down arrow represent start release and up arrow represent start vibration

In any language, stop sounds are divided to voiced sounds and voiceless sounds. Lisker and et al.[1] divided languages based on VOT value. From our previous researches [18] we found that the VOT value in Arabic language is positive. Also, we found that the VOT value in voiced sounds is short VOT, and long VOT in unvoiced sounds.

Each Table contains three spectrogram readings for VOT values, the average for it, VOT value detected by our algorithm, standard division between algorithm VOT value and the average VOT value by reading the spectrogram, and the error rate. The error rate is calculated by the following equation:

$$\text{Error Rate} = \frac{|x-y|}{x} * 100$$

where x represents VOT value by reading the spectrogram and y represents VOT value by the algorithm.

Table 3. Algorithm statistics for /d/ sound.

Name File	By Read Spectrogram 1	By Read Spectrogram 2	By Read Spectrogram 3	Mean	By algorithm	Standard Deviation	Error Rate
s50c1e0t2	16	12	12	13.33333333	15	1.178511302	12.5
s51c1e0t2	13	11	10	11.33333333	7	3.064129385	38.24
s52c1e0t2	19	12	13	14.66666667	14	0.471404521	4.55
s53c1e0t2	12	11	12	11.66666667	12	0.23570226	2.86
s54c1e0t1	18	15	18	17	19	1.414213562	11.76
S55c1e0t2	9	9	10	9.333333333	10	0.471404521	7.14
s56c1e0t2	13	8	14	11.66666667	11	0.471404521	5.71
s59c1e0t4	10	10	12	10.66666667	14	2.357022604	31.25
s60c1e0t2	11	8	9	9.333333333	7	1.649915823	25.00
S61c1e0t2	19	18	11	16	15	0.707106781	6.25
Average	14	11.4	12.1	12.5	12.4	0.070710678	0.80

In Table 3, we analyze the algorithm for sound /d/. The average VOT value by the algorithm is 12.4 msec which are inside the VOT value range of three spectrograms readings. Eight files have VOT value inside the range of the three spectrogram readings and two files were outside this range. The pronunciation of this sound is easy, because it is found in most languages. So, the error rate is 0.80%. Also, the standard division for VOT value between the algorithm and manual spectrogram reading is 0.07 msec.

Table 4. Algorithm statistics for /d?/ sound.

Name File	By Read Spectrogram 1	By Read Spectrogram 2	By Read Spectrogram 3	Mean	By algorithm	Standard Deviation	Error Rate
s51c1e1t2	11	11	12	11.3333333	15	2.592724864	32.35
s53c1e1t2	10	8	13	10.3333333	18	5.421151989	74.19
s54c1e1t1	15	15	10	13.3333333	19	4.006938427	42.50
s55c1e1t3	12	8	13	11	15	2.828427125	36.36
s57c1e1t2	7	8	8	7.66666667	8	0.23570226	4.35
s58c1e1t2	10	8	12	10	10	0	0.00
s59c1e1t2	12	12	6	10	14	2.828427125	40.00
s60c1e1t3	11	10	9	10	10	0	0.00
s61c1e1t2	7	11	8	8.66666667	6	1.885618083	30.77
s62c1e1t2	10	11	10	10.3333333	15	3.299831646	45.16
Average	10.5	10.2	10.1	10.26666667	13	2.309882152	26.62

The outcomes of VOT value by algorithm and spectrogram readings for /d$^{?}$/ sound is represented in Table 4, which contains 10 speakers. This sound is unique to MSA Arabic language and difficult to pronunciation. So, it is difficult to measure VOT value by reading the spectrogram hence it is difficult to detect it automatically. By algorithm, the VOT values in 6 files are outside the range of three spectrogram readings. Also, the maximum VOT value outside the spectrogram range by 5 msec and the minimum is outside by 2 msec. In addition, the average VOT value detected by algorithm is 13 msec, which is outside the three spectrogram readings. The error rate is 26.62% for this sound.

Table5. Algorithm statistics for /b/ sounds.

Name File	By Read Spectrogram 1	By Read Spectrogram 2	By Read Spectrogram 3	Mean	By algorithm	Standard Deviation	Error Rate
s50c7e0t2	8	7	9	8	9	0.7071068	12.50
s51c7e0t2	13	13	12	12.6666667	15	1.6499158	18.42
s52c7e0t2	10	9	8	9	7	1.4142136	22.22
s53c7e0t2	10	13	9	10.6666667	11	0.2357023	3.13
s54c7e0t3	11	11	12	11.3333333	7	3.0641294	38.24
s55c7e0t2	10	12	13	11.6666667	14	1.6499158	20.00
s58c7e0t2	8	8	7	7.66666667	5	1.8856181	34.78
s59c7e0t2	12	14	8	11.3333333	5	4.4783429	55.88
s60c7e0t2	8	12	7	9	6	2.1213203	33.33
s61c7e0t2	15	15	12	14	14	0	0.00
Average	10.5	11.4	9.7	10.53333333	9.3	1.7206265	11.71

In Table 5, we investigate and analyze 10 speakers for /b/ sounds. We found by algorithm four files with a VOT value far from the spectrogram range by 3 msec or less, and another two files

far by 5 msec. We ignore the different between the algorithm and spectrogram reading when it equals 1 msec. The accuracy or error rate for this sound is 11.71%. In addition, the average VOT value by algorithm is inside the range of spectrogram readings for this sound.

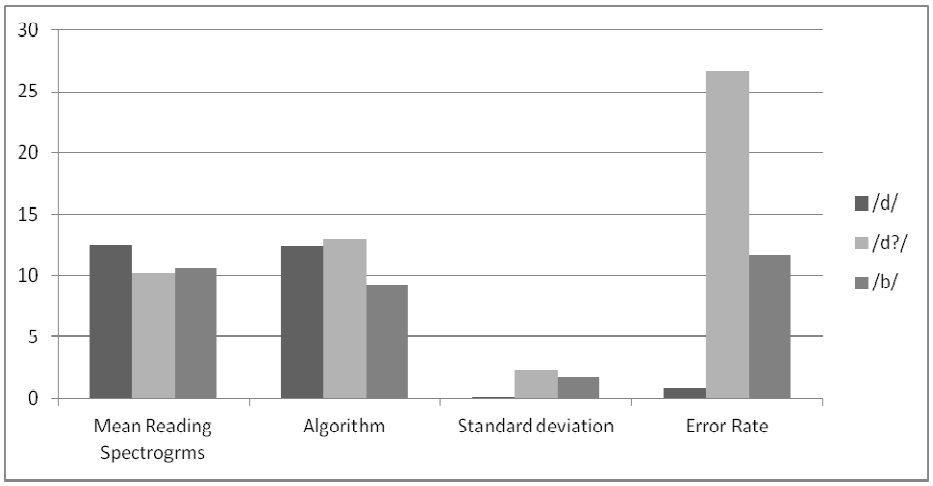

Figure 5. Statistic of VOT value for three voiced stop sounds.

In another side, Figure 5 shows the statistic to compare the VOT value between the automatic VOT estimated by the algorithm and the manual VOT by spectrogram reading for three stop voiced sounds. We found that the VOT values are positive and short VOT. In addition, the smallest standard deviation for these three phonemes is 0.07 msec in sound /d/. So, the error rate of the algorithm for this sound is low, which represents 0.80%. Also, sound /d$^{?}$/ standard deviation is high with 2.3 msec, because this sound is unique, and difficult to pronounce and detect manually by spectrogram. Therefore, this sound is difficult to detect automatically, and the error rate is 26.62% which is high comparing to other voiced stop sounds.

5. CONCLUSION AND FUTURE WORK

In this paper, we detected the VOT value automatically for voiced stop sounds in MSA Arabic namely /d/, /d$^{?}$/ and /b/ and compared the accuracy algorithm with spectrogram reading. Automatically detecting VOT in speech signal is a challenging problem because it combines temporal and frequency structure over short duration. Our algorithm depends on reading the power signal. We create our database and the structure CV-CV-CV. We ended to a conclusion that the error rates were 0.80%, 26.62% and 11.71% for /d/, /d$^{?}$/ and /b/ respectively. Moreover, we found that the standard deviation between the algorithm and spectrogram reading is low in /d/ sound because this sound is easy to pronounce, and high in the unique /d$^{?}$/ sound, the main reason is that it is difficult to pronounce.

As a future and continuing research of this effort, we will continue to study automatic VOT for unvoiced stop sounds in MSA Arabic and compare our algorithm with manual spectrogram reading. Also, we will extend our database. On the other hand, we will study the environmental effect such as gender, age, effect of neighboring phonemes and place of autocorrelation to the VOT value.

ACKNOWLEDGEMENTS

Grateful acknowledgment to King Abdulaziz City for Science and Technology for its support of this work. Also, acknowledgment to Eng. Hazzam Alhakami.

REFERENCES

[1] L. Lisker and A. S. Abramson, "A Crosslanguage Study of Voicing in Initial Stops: Acoustical Measurements", Word, Volume 20, No. 3, , pp. 384-442, December 1964.

[2] M. Alghamdi, "Voice Print": Voice Onset Time as a Model. Arab Journal for Security Studies and Training. 21, pp. 42: 89-118, 2006 (in Arabic).

[3] Cho, T. and Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. Journal of Phonetics, 27:207–229.

[4] Sonderegger, M. and Keshet, K., "Automatic measurement of voice onset time using discriminative structured prediction", Journal Acoustical Society of America, Soc. Am., 132, Dec 2012.

[5] A. S. Abramsont and L. Lisker, "A Cross-language Experiments in identification and Discrimination", 21-24 May 1968.

[6] Okalidou, A., Petinou, K., Theodorou, E. & Karasimou, E. (2010). Development of voice onset time in Standard Greek and Cypriot Greek speaking preschoolers. Clinical Linguistics and Phonetics, 24(7), 503–519.

[7] J. Jiang, M. Chen, and A. Alwan,"On the perception of voicing in syllable-initial plosives in noise", Journal of the Acoustical Society of America, vol. 119, no. 2, Febuary 2006.

[8] P. Sweeting, and R. Baken, "Voice onset time in normal-aged population", Journal of Speech and Hearing Research. 25, 129-134, 1982.

[9] S. Das, and J. Hansen, "Detection of voice onset time (VOT) for unvoived stops (/p/, /t/, /k/) using the Teager energy operator (TEO) for automatic detection of accented English", Proceedings of the 6th NORDIC Signal Processing Symposium (NORSIG'4), pp. 344 – 347, June 9-11 2004.

[10] G. E. Peterson and I. Lehiste, "Duration of syllable nuclei in English", Journal of the Acoustical Society of America 32, PP. 693-703, 1960.

[11] J. E. Flege and R. Port, "Cross-language phonetic interference: Arabic to English", Language and Speech 24, PP. 125-146, 1981.

[12] Stölten, Katrin; Abrahamsson, Niclas; Hyltenstam, Kenneth, Effects of Age of Learning on Voice Onset Time: Categorical Perception of Swedish Stops by Near-native L2 Speakers, December 2014, Language & Speech;Dec2014, Vol. 57 Issue 4, p425.

[13] Caramazza A. and Yeni-Komshian G. H,(1974). Voice onset time in two French dialects, Journal of Phonetics, 2, 239–245.

[14] C. Prakash, N. Dhananjaya, S. Gangashetty, Bessel features for detection of voice onset timeusing AM-FM signal, in Proc. of Int. Conf. on the Systems, Signals and Image Processing (IWSSIP), (IEEE, Sarajevo, Bosnia and Herzegovina, 2011), pp. 1–4.

[15] S. Al-Ani, “Arabic Phonology”, The Hague, 1970.

[16] F. Mitleb, "Voice onset time of Jordanian Arabic stops", The 3rd International Conference on Arabic Language Processing (CITALA'09), Rabat, Morocco, pp. 133-135, May 4-5 2009.

[17] G. H. Yeni-Komshian, A. Caramaza and M. S. Preston, “A study of voicing in Lebanese Arabic”, Journal of Phonetics 5:1, PP. 35-48, 1977.

[18] AlDahri S., Studying VOT of unique Arabic stop and designing system to identify stop sounds in Modern Standard Arabic (MSA), Intelligent Signal Processing and Communication Systems (ISPACS), November 12-15, 2013, Naha, Okinawa, Japan.

[19] Alotiabi Y. and Aldahri S., Phobnnetic Investigation of MSA Arabic Stops (/t, d/)”, Yantai, China, 2010.

[20] Aldahri S., The Effect of Arabic Emphaticness on Voice Time Onset (VOT), This conference (the 3rd International Conference on Audio, Language and Image Processing (ICALIP 2012), China Shanghai, 2012.

[21] J. Deller, J. Proakis and J. H. Hansen, “Discrete-Time Processing of Speech Signal”, MacmillAnn, 1993, Autor Engineering in computer science, researcher in field speech processing.