

Realisation of Voice Onset Time (VOT) and Its Implication on Voicing of English (RP) Stops in Ghanaian English (GhE)

Lena Awoonor-Aziaku

Institute for Phonetics Johann Wolfgang, Goethe University, Frankfurt am Main, Germany
Email: awooleen@gmail.com

How to cite this paper: Awoonor-Aziaku, L. (2021). Realisation of Voice Onset Time (VOT) and Its Implication on Voicing of English (RP) Stops in Ghanaian English (GhE). *Open Journal of Modern Linguistics*, 11, 448-460.
<https://doi.org/10.4236/ojml.2021.113034>

Received: October 23, 2020

Accepted: June 22, 2021

Published: June 25, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

There has been a general perception that the English voiced stops /b, d, g/ before and after a pause or silence are more often confused with the initial voiceless stops /p, t, k/. Thus, in distinguishing the two, listeners most often depend on the onset of voicing of the following segments. Voice onset time, therefore, serves as a significant cue in distinguishing English voiced stops from their voiceless counterparts. Interestingly, voiceless stops in many Ghanaian indigenous languages appear to be weakly aspirated; they are perceived to be made with a very short voice onset time. However, listeners are able to distinguish the voiceless stops from their voiced counterparts. What this study sought to do was to investigate the production of English initial voiceless stops, /p, t, k/ in Ghanaian English. Specifically, it sought to examine the voice onset time of the stops, and whether listeners are able to distinguish English stops made with weak aspiration from the voiced stops before and after a pause. The study used 112 Ghanaian speakers, 76 for a production test; they produced 21 tokens of the stops in a /p/-, /t/-, and /k/- initial words, and 36 for a perception test; they listened to words produced with voiced stops in a /b-b/, /d-d/ and /g-g/ environments. The study has revealed that for all the voiceless stops, voicing for the following segments began after the stops were released, and there was no pre-voicing or voice lead. However, for some of the stops, voicing coincided with the release, while for some voicing began after an appreciable amount of time after the release burst. The length of the VOT was found to have been affected by place of articulation and vowel contexts. However, speaker sex seems to have no effect on the length of the VOT. It is clear that Ghanaian speakers of English would have no difficulty distinguishing weakly aspirated stops from voiced stops before and after a pause. Also, understanding words having the initial voiced stops before and after a pause and weakly aspirated stops produced by Ghanaians will not be a problem to listeners.

Keywords

Voice, Onset, Time, Stops, Burst, Articulation, Lag, Pre-Voice, Voiceless, Bilabial, Alveolar, Velar

1. Introduction

In phonetics, voice onset time (VOT) is a very important feature in the production of stop consonants. It is defined as the period between the release of the articulators for a stop and the onset of vocal fold vibration of the following segment. According to (Kaur, 2015), it is a measure of the time between the burst of a plosive and the onset of voicing of a subsequently voiced phoneme. In English, we can distinguish two types of contrastive stops: voiced /b, d, g/ and voiceless /p, t, k/ stops according to the state of the glottis. That is, voiced stops in English contrast phonemically with the voiceless ones. But the difference is not just one of voicing during the consonant closure (see Ladefoge & Johnson, 2011). As they noted, most people have very little voicing when the lips are closed during the production of for example, /p/ in “pie” or /b/ in “buy”. But it is clear that in “pie”, there is a moment of aspiration (a period of voicelessness) after the release of the lip closure for /p/ and before the start of voicing for the vowel. It is this interval that shows that the stop is aspirated. Interestingly, the amount of voicing in each of the so-called voiced stops [b, d, g] depends on their context. They are only fully voiced (i.e. voicing usually occurs throughout the stop closure) when they occur in between two voiced sounds. But according to Ladefoge & Johnson (2011), most speakers of English have no voicing during the closure of the so-called voiced stops in silence initial position, or when they occur after a voiceless sound as in for example, “bag” /bæg/ or “brass band” /bæs bænd/ respectively.

Meaning that we cannot use the state of the glottis alone to determine the difference between voiced and voiceless stops in English. We can, therefore, distinguish two types of non-contrastive stops which depend significantly on the release of the stop and the onset of vocal fold vibration of the following segment. For instance, we can have zero, negative or positive voice onset times, and the timing is always relative to the stop release burst (Kaur, 2015). Thus, English stops are typically placed into one of three distinct types of voice onset time. We say a voice onset time is negative when voicing starts before the release of the stop; when voicing starts during the hold/closure phase for the stop. This is also described as “voice lead” or “pre-voice”. Negative VOT is typical with voiced stops, which normally have their voice onset times noticeably less than zero (Kaur, 2015). The voiced phonemes /b, d, g/ are therefore categorised as pre-voiced or “voicing lead”. The diagrams below show the production of the English stops.

Both **Figure 1(a)** and **Figure 1(b)** show the stages of plosive production. In **Figure 1(a)** we see the closure of the two articulators. In English, this closure

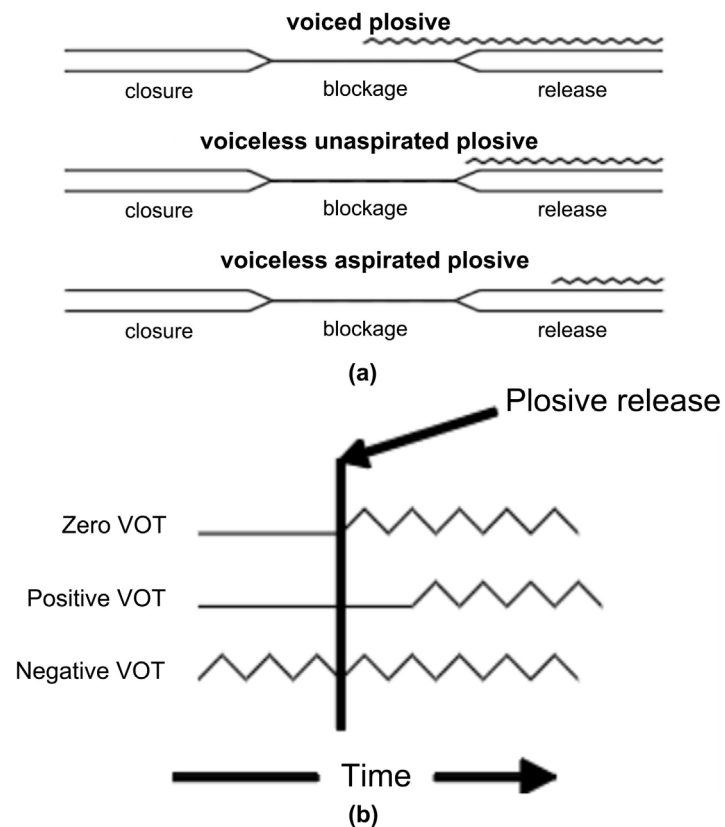


Figure 1. (a & b) Articulation of English stops.
https://en.wikipedia.org/wiki/Voice_onset_time

could be between the lips, or between the tongue tip and the alveolar ridge, or between the back of the tongue and the velum. After the closure, there is the hold phase (where supraglottal pressure builds, and also the release stage/phase). In 1b we see two phases; the hold phrase where subglottal activity takes place, and the release phase indicating the various voice onset times. Smith (1978), McCrea and Morris (2005) have observed that in the production of voice lead phonemes, voicing occurs before the burst which results in a negative VOT value shown in the figures.

VOT is however termed “positive” when voicing starts after the release of the stop (i.e. after the burst), resulting to what is called “voice lag”. The voice lag could be short or long depending on whether the stop is made with aspiration or without aspiration. The short lag is termed zero; if voicing occurs simultaneously or just after the burst. This results in a VOT value very close to zero or a small positive value approximately 20 milliseconds (Smith, 1978; Swartz, 1992). Stops with very short voice lag tend to have values ranging from 0 to 25 ms with a median value of 10 ms (Auzou et al., 2000). These are normally voiceless unaspirated stops in English; they turn to be made with a very short voice onset time at or near zero (voicing is simultaneous with the stop release). Auzou et al. (2000) believe that an offset of 15 ms or less for [t] and 30 ms or less for [k] is inaudible and counts as unaspirated. Aspirated stops are, however, made with long lag.

Nonetheless, they turn to have a longer VOT when followed by sonorants. The length of the VOT in such a case is usually a measure of aspiration; the longer the VOT, the stronger the aspiration. In Navajo, stops which are strongly aspirated, the aspiration (i.e. the VOT) lasts twice as long as it does in English: 160 ms vs. 80 ms for [k^h], and 45 ms for [k]. Some languages are, however, weak aspirated languages. The voiceless unaspirated velar stop [k], for instance, tends to have a VOT of about 20 - 30 ms; a weakly aspirated [k] of about 50 - 60 ms; a moderately aspirated [k^h] averaged 80 - 90 ms, and anything much over 100 ms would be considered strong aspiration (see [Smith, 1978](#); [McCrea & Morris, 2005](#)). [Smith \(1978\)](#) found that speakers produced all the English voiced stops as voice lead consistently (i.e. where voicing coincides with the burst); e.g. bilabial /b/ 56%, the alveolar /d/ approximately 50%, and the velar /g/ about 39%. The voiceless stops /p/, /t/, and /k/ are, however, classified as long voice lag phonemes, during whose production voicing occurs significantly after the burst, which normally results in positive VOT values ([Lisker & Abramson, 1964](#); [Smith, 1978](#)). The length of the values can vary from one speaker to the other.

[Lisker and Abramson \(1964\)](#) had a similar result. They found that English voiced stops /b, d, g/ are produced more frequently as short voicing lag phonemes. What they mean is that individual speakers most often use one voicing mode for voiced stop consonants. However, the study made by [Smith \(1978\)](#) has shown that speakers do not exclusively use one voicing mode for the stops, they tend to switch between voicing lead and short voicing lag modes depending on contexts. The lead or pre-voicing and short lag stops are, therefore, allophones or allophonic variants of /b, d, g/. They are in free variation and do not change meaning. The speakers are, thus, free to use either lead or short lag for /b, d, g/ at will. Many listeners treat the lead and short lag stops as equivalent. However, in the early native-English speech development, it was noticed that listeners tended to have some difficulty hearing the difference between lead and short lag allophones of /b, d, g/. Recent studies on VOT of English stops have shown that native English speakers use either short voicing lag or long voicing lag mode. And that for long voicing lag phonemes, VOT values tend to range between 60 and 100 milliseconds. In English, voicing can therefore successfully separate /b, d, g/ from /p, t, k/, but only when stops are at word-medial positions, meaning that this is not always true for word-initial stops. Strictly speaking, word-initial voiced stops /b, d, g/ are partially voiced, and sometimes are even voiceless. The concept of VOT finally acquired its name in the famous study of ([Lisker & Abramson, 1964](#)). One of their earliest works on voice onset time for instance, indicates that:

“... a difference of voicing not only separates voiced from voiceless stops, but that it equally well distinguishes aspirated from unaspirated stops, where the latter are both commonly called voiceless. The noise feature of aspiration, instead of being considered coordinate with voicing, is then regarded simply as the automatic concomitant of a large delay in voice onset. In English, at least, this seems reasonable: /b, d, g/ and /p, t, k/ probably differ everywhere in the time of

voice onset relative to release, but in certain positions the presence of aspiration noise tells us something about the absolute magnitude of delay in the onset time following /p, t, k/ releases ... ” (p. 387).

Voice onset time is affected by several factors. Among these are: place of articulation, vowel height, and stress (Kaur, 2015). The lag has been found to be greater for voiceless stops than it was for voiced stops. But it was found to be more profound as the place of articulation moved from bilabial to alveolar and then to velar stop. The difference in the lag, as was observed, could be as a result of differences in the pressure drops upon the release of the stop. For instance, the more abrupt the pressure drop is, the sooner the voicing of the next segment starts. This could result to a reduction in aspiration, that is, shorter lag. A test conducted by Kaur (2015) on the three places of articulation of stops shows that the tongue dorsum separates more slowly, thus less abrupt pressure drop from the velum for /k/ than the tip from the alveolar ridge for /t/, or from the lips for /p/. Their observation means that voice onset time increases as the place of articulation moves from front to back. It had also been noticed that the lag of the stops was greater when stops were followed by high vowels such as /i/ with greater vocal tract constriction, than when they were followed by low vowels which naturally have less vocal tract constriction. The reason for this effect has also been attributed to the abruptness of the pressure drop. High vowels have a more obstructed cavity than low vowels. Consequently, the pressure drops for stops produced in the environment of a high vowel would have less abrupt pressure drop, and a stop produced in such environment will have a longer lag than the one produced before a low vowel.

Stress has also been observed to have a significant effect on the timing of voice onset time. For instance, studies have shown that English voiceless stops in stressed syllables tend to have greater voicing lag than those in unstressed syllables. Lisker and Abramson (1964) examined the release burst of /p, t, k/ and revealed that the onset of the burst relative to the onset of voice onset time affected the formant transition structure. They observed that an increase in the stop burst could lead to an increase in the voice onset time. If the VOT is increased, transition into the following vowel may be largely complete at the beginning of voicing for the next vowel, so that the duration of devoiced transitions relative to voiced transitions is increased. Since released burst duration typically increases from labial to alveolar to velar points of articulation (Lisker & Abramson, 1964), it is possible to predict corresponding increases in the perceptual weight attached to devoiced transitions. Bursts tend to show a strength hierarchy: voiceless aspirated to voiceless unaspirated and then to voiced stop. That is, burst spectrum of voiceless aspirated stops have more energy than that of unaspirated and voiced stops, but the source spectra for aspiration is nearly the same as that for fricatives (Hilary, 2005).

Gay (1978) investigated three male American speakers and showed that F2 onset in the aspirated bilabial stop [p^h] in “pap” [p^hap] and [p^hup] was about 180 Hz higher than in the bilabial [b] in the syllables /bap/ and /bup/. It was however

about 125 Hz lower in bilabial stop /b/ before the high front vowel /i/ in [p^hɪp] than in /bɪp/. Fant (1973) therefore explains that in the articulation of voiced bilabial stop /b/, the tongue is more nearly in the position for the vowel before the release of the stop closure than it is for the [p^h]; hence a shorter VOT. As the articulators begin to move towards the vowel, the release of aspirated stops may occur earlier in time than that of unaspirated stops so that energy begins while the articulators are still farther away from the vowel target (Ohman, 1965; Fant, 1973). Also, the higher F2 onsets for aspirated stops may arise from the open glottis during aspiration. The aspirated stops therefore tend to have stronger bursts than the unaspirated stops (Zue, 1976). According to Lisker & Abramson (1967), although aspirated alveolar stops have longer VOTs than those of labial stops; VOT serves as a weak cue for stop place identification since VOT varies from one speaker to the other and so can be misleading; listeners will only perceive labials when VOT is relatively short and alveolars when VOT is relatively long and nothing else. This then implies that several acoustic cues are needed for stop identification.

Study Objective

The general perception is that the so-called English voiced stops tend to be confused with the voiceless stops unaspirated stops. But this is not the same in Ewe language, which appears to be a weakly aspirated language (my own observation). The purpose of this research was to examine the production of English voiceless stops among Ghanaian speakers of English. Also, to find out if listeners are able to distinguish voiced stops from weakly aspirated stops produced by Ghanaian speakers.

2. Methodology

Seventy-six (76) Ghanaian speakers of English from Peki and Keta were given task to perform. The first task was a reading task made of a list of words produced in isolation. This was presented visually on a monitor. The speakers were asked to read the words in a quiet room, which were recorded using Handy4 Next audio recorder with a sampling rate of 44.1 kHz (i.e. 44,100 times per second) with a 16 bit resolution. The words were English words with the voiceless stops at word initial position in different vowel environments. They were presented in random order for each participant. The stimuli consisted of 35 words: 21 test items, 7 each representing the three voiceless stops at word-initial position, and 14 distracter items with initial sonorants, fricatives, etc. All the test items were selected using Wells' (1982) lexical framework, and were all monosyllabic words. No words with onset and coda clusters or initial voiced stops were included except the distracter items.

The second task involved a perception test conducted with 36 Ghanaian speakers of both genders between 30 and 50 years of age across the 21 plus an additional nine words of initial /b, d, g/. The listeners for the test were selected

according to their demographic characteristics. This information was collected based on their medical records, which was retrieved from the university's hospital under the permission of the participants. All the participants provided written informed consent forms to participate in this study, which was then approved by the ethics committee. None of the subjects had a hearing impairment. All the listeners wore a headphone. All the speech perception tests were applied in a soundproof booth using an audiometer connected to an amplifier and an acoustic box. The subjects were positioned at a distance of one meter from the loudspeaker.

The signals were then played to the listeners for the purpose of identifying and labelling the variables as voiced or voiceless stops. The focus, therefore, was on whether the subjects would be able to hear any contrasts between the stops with short VOT (unaspiration) and the voiced initial stops. In this test, the stops with very short VOTs were selected and paired with the voiced stops. The subjects were taught the difference between the voiceless stops and the voiced stops. They were then asked to tick whether what they heard was voiceless or voiced. The researcher believes that once the listener is able to identify and label the stops with a short VOT as voiceless stops and the initial voiced stops as voiced, it is highly likely that they would be able to distinguish voiced stops from voiceless stops spoken in every context.

Data Analysis

Measuring the voice onset time (VOT)

Voice onset time is usually measured in milliseconds (ms). To measure this, wide-band spectrograms of the recordings were made using PRAAT software. From these, the voice onset times of each of the voiceless stops were measured for each speaker. The measurement was done by marking off the intervals between the release of the stop and the onset of glottal vibration (voicing). For instance, the point of onset of voicing was determined by locating the first of the regularly spaced vertical striations, which indicate glottal pulsing (on the PRAAT window). What this means is that the beginning of the lag was identified by a sharp spike where the waveform changes from quiescent to transient; the end point; the onset of vocal fold vibration was determined from where the waveform becomes periodic. Thus, for the spectrographic readings, the voice onset time intervals from the beginning of the release burst to the onset of voicing were analysed. The VOT values of the target stops were obtained from the waveform and verified with the spectrogram. **Figures 2(a)-(c)** show the production of the three English stops /p, t, k/ by the speakers.

The areas marked in square in **Figures 2(a)-(c)**, show both the start and the release of the stops. The gaps between the release and voicing of the next segment are marked in square. Here, the phoneme /t/ has the longest VOT of 88 ms, /p/ has 19 ms and /k/ has 48 ms.

The areas marked in square show both the start and the release of the stops.

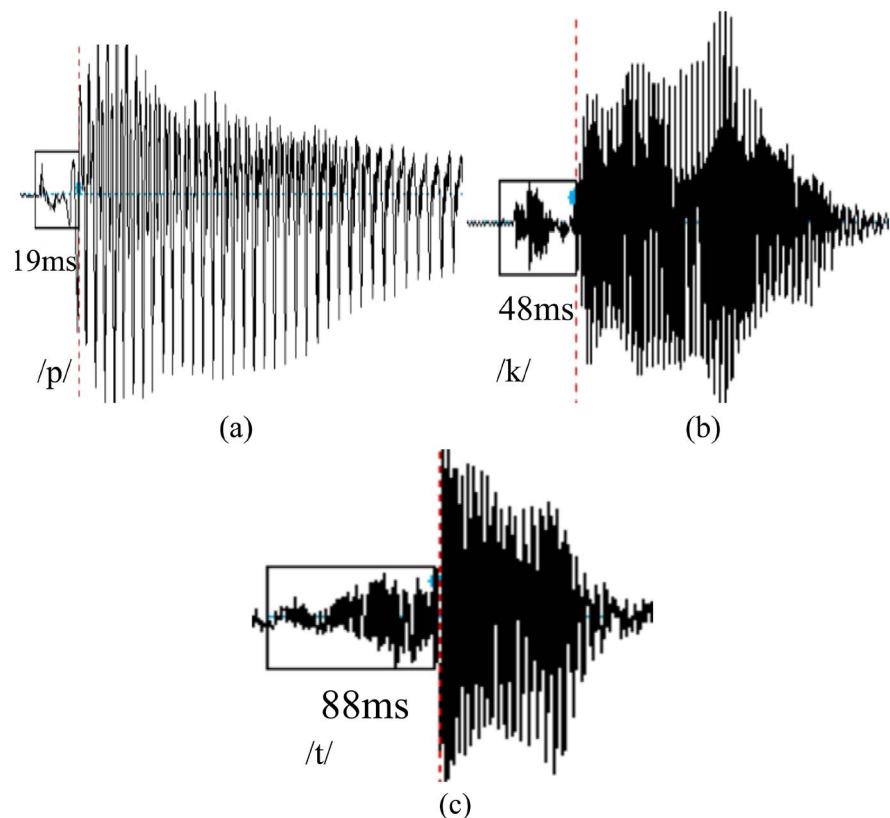


Figure 2. (a, b, c) Voice Onset Times of the English voiceless stops /p, t, k/.

The gaps between the release and voicing of the next segment are marked in square. Here, the phoneme /t/ has the longest VOT of 88 ms, /p/ has 19 ms and /k/ has 48 ms.

3. Results and Discussion

The study has shown that the three English voiceless stops, /p, t, k/ were produced through the three stages of stop production; approach, hold and release stages. It has been noticed that for all the stops, voicing began for the following voiced sounds after they were released. This means that for the production of all the stops, vocal fold vibration began after the release burst (i.e. they were all released with a positive VOT). Interestingly, two forms of the voice onset times were noticed; a very short voice onset time where voicing began immediately the stops were released, and a long voice onset time where voicing began after an appreciable amount of time after the release of the stops. Voicing therefore coincided with the release burst for some of them, while with others voicing began after an appreciable amount of time after they were released. But we need to note that there was no instance of pre-voicing or voice lead. Nevertheless, the voice onset time among all the speakers was found to be generally short (32 ms on average). The length of the timing, however, varied according to different factors, for example, place of articulation.

This variation came, for example, on average of 17 ms for bilabial, 60 ms for

alveolar, and 23 ms for velar. Surprisingly, the alveolar stop had recorded the highest value of about an average of 60 ms in all the contexts. This is contrary to Kaur's (2015) investigation which, for instance, showed the voiceless velar stop recording the longest voice onset time among the three English voiceless stops he had examined. Kaur (2015) attributed this to the abruptness of pressure drop across the glottis. This study, nonetheless, partially supports the general perception that the English voiceless alveolar stop /t/ is generally released with frication (i.e. it is released as if it were a fricative) by many Ghanaians. Releasing a stop with frication can subsequently lead to a long VOT. Unfortunately, this present study used Ewe speakers only. Also, there has not been any study on VOT in Ghanaian English, and any of the Ghanaian languages. To be able to establish this fact we would need more data from other language groups.

Apart from the place of articulation effect, I also noticed that variation in the voice onset time occurred within each of the stops. This could be attributed to different factors including vowel height; there has been an effect of vowel contexts on the voice onset time of the stops. A measure of the timing of all the initial stops, for example, has shown a relatively short VOT before high vowels such as /i/ and /u/ (33 ms) than it was for the low vowel /a:/ (50 ms). This means that the VOT was shorter when there was a following high vowel, and longer when there was a following low vowel. This was inconsistent with previous results which recorded longer VOTs for high vowels than for following low vowels. But one interesting thing I noticed was that all the high vowels were preceded by the alveolar stop which we have already seen has long voice onset time. This can subsequently contribute to the long VOTs that we observed in this environment. For example, if we compare the voice onset time of /t/ in "tall" in Figure 3(a) with that of "tool" in Figure 3(b), we will realise that /t/ before /ɔ:/ in "tall" has a higher voice onset time than the one before /u:/ in "tool".

This was the same with the /p/ of "peel" before the high vowel /i/, and of "park" before the low vowel /a/. In each of the words, the /p/ before the high vowel /i:/ had a longer voice onset time than the one before the low vowel /a/ in Figure 4(a) and Figure 4(b).

In Figure 4, three things happened. In the first place, /p/ in park in Figure 4(a)

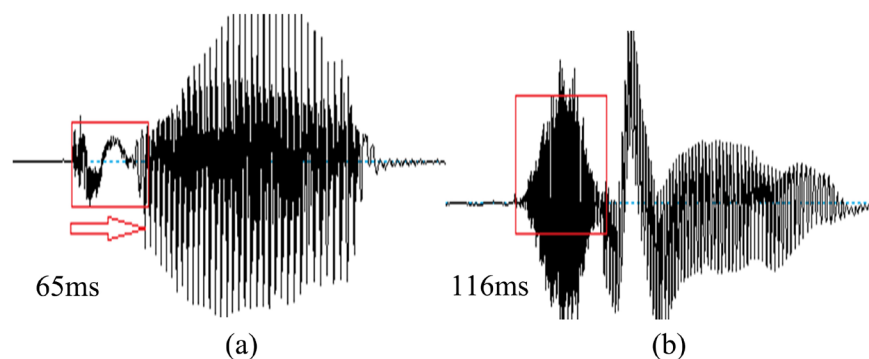


Figure 3. (a & b) Voice onset time of /t/ in "tall" and "tool". (a) /t/ in tall; (b) /t/ in tool.

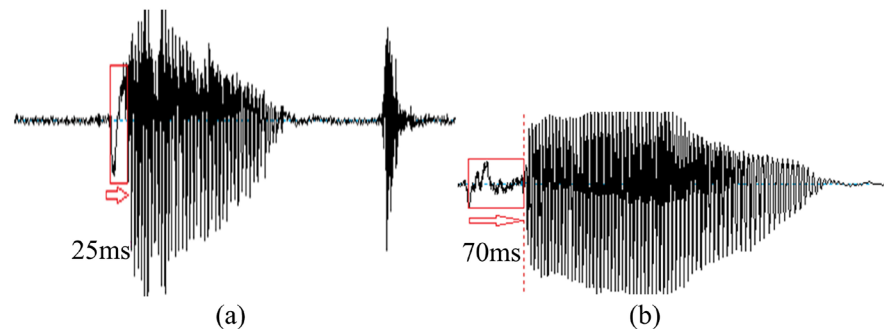


Figure 4. (a & b) Voice onset time of /p/ of “park” and “peel”. (a) /p/ of park; (b) /p/ of peel.

is followed by a low vowel /a/, and a voiceless coda /k/. /p/ in “peel” in **Figure 4(b)** is however followed by a high vowel /i/, enclosed with a liquid /l/. Note that it had also been observed in this study that stops enclosed with liquid codas generally have long VOTs while those enclosed with voiceless stops have short VOTs. A pattern that was also apparent here is that the words with longer VOTs tended to have longer vowels immediately following the /p/, /t/ or /k/, and the words with the shortest VOTs tended to have back lax vowels and voiceless coda consonants. At this juncture, the connection between vowel length and voice onset time is consistent with the results of previous studies which indicate longer VOTs coming at the expense of vowel duration. Nonetheless, longer VOT, therefore, meant more of the interval between the stop release and the end of the vowel becoming voiceless. Thus, since words with longer vowels have initial stops with longer VOTs in general, tokens with shorter vowels are more likely to have initial stops with shorter VOTs.

Another thing that was apparent in this study was that the length of the VOTs was interestingly large for some words. The words with the most environment effects include “peel”, “tool”, and “tall”, which had closed codas with liquids and long vowels. For example, these words which were followed by long vowels relatively tend to have long VOTs than those that were followed by short vowels. Similarly, words with initial stops closed with codas with the liquid /l/ had relatively long VOTs. The longer VOT of some of these words could be explained by the presence of the postvocalic liquid as there was an increase of VOT in these contexts. It is, therefore, difficult however, to attribute all the effects of variation to vowel contexts since not all vowel environments here were affected. For instance, /k/ in the word “course” has a shorter VOT than the /k/ in “curve.” We can see that although /k/ in both words are followed by long vowels, they have different lengths of VOTs. It would be difficult, therefore, to improve our understanding of the vowel effect without analysing data from more words. There is, therefore, the need to have more studies to establish this fact.

The estimated effect size for the liquid factor was greater than the effect of following vowel height/duration. With the presence of the liquid accounting for observed long VOTs, it has also been noticed that the VOT is shorter when the

next syllable starts with phonetically voiceless consonants, and even shorter when that voiceless obstruent is a stop. The effects of following voiceless obstruents and liquids help us to account for the reason why /p/ in the word “peel” which is closed by a postvocalic liquid /l/ has a longer VOT than the /k/’s in words like “cock”, “course” and “kept” with voiceless obstruent codas. It is also clear why /t/ of “top” which is enclosed with a stop has a VOT shorter than that of /k/ enclosed with /s/ in “course” even though both codas are voiceless. Furthermore, syllabic liquids are associated with a longer VOT than would be expected based on their duration. This could be because liquids, being produced with relatively closed vocal tract, similarly affect the tense vowels in delaying the onset of voicing by reducing airflow.

Another thing examined in this work was variation between speakers, that is, an interaction between VOT and the social variable, sex, of the speakers. It was noticed that the voice onset time of the stops did not show cross-speaker variability. For example, the effect of sex on other factors such as vowel duration and height were relatively consistent across speakers. The differences in the values of the voice onset time for both sex groups were not very huge for all the three stops. This showed that the variation in the VOT of the speakers is likely not as a result of their social differences. Reduction in VOT in stops that are followed by high vowels and voiceless obstruents is similar to Grassmann’s Law in which aspirated consonants in Greek and Sanskrit were de-aspirated when followed by an aspirated consonant in the next syllable (Mielke & Nelson, 2018). In addition to their occurrence in Greek and Sanskrit, similar anticipatory dissimilation of aspiration had been observed in Ofo, an extinct Siouan language (De Reuse, 1981), and four Salish languages (interior Salish languages Kalispel, Okanagan, Shuswap), and the other related studies.

The fact that anticipatory dissimilation of aspiration had been observed in various unrelated languages suggested a phonetic basis for this pattern. Simons (2009), for example, argues that Grassmann’s Law could have been allophonic at the time that the dissimilation pattern developed in the Indo-European languages.

Perception Test

A perception test was conducted with 36 Ghanaian speakers of both sexes between 30 and 50 years of age. The focus was, therefore, on whether the subjects would be able to hear any contrasts between the stops with short VOT (unaspiration) and the voiced stops. In this test, the listeners were given a recording of a list of words (nine in all) of the English stops, /b, d, g/ in /b-b/, /d-d/ and /g-g/ environments. These were compared with the voiceless stops produced with very short VOTs. Thus, they were tested using words with initial voiced stops before and after a pause or silence and those that were made with very short VOTs. The signals were then played to the listeners for the purpose of identifying and labeling the variables. They were to identify and label them according to what they

heard; whether what they heard was voiced or voiceless stops. But before the task, the subjects were taught the difference between the voiceless and voiced stops. Thus after this, the listeners were asked to tick whether what they heard was the voiceless stops, /p, t, k/ or the voiced /b, d, g/ ones.

Interestingly, all the listeners were able to identify and label all the stops made with short VOTs as voiceless and the voiced stops in all the words as voiced. They were also able to identify and label the initial voiced stops before and after a pause or silence as voiced and those that were made with very short VOTs as voiceless. That is, the listeners did not perceive any of the voiceless stops as voiced stops. They were also able to distinguish the stops with short voice onset times from those made with long VOTs. It is therefore obvious that the listeners had no difficulty distinguishing between the stops with short voice onset times and voiced stops before and after a pause. Only four (4) out of the 36 subjects perceived the stops with short VOTs as voiced. However, this effect appears to depend almost entirely on variation in the duration of the vowels immediately following the stops. The duration of the silent stop closure, and the duration of the test syllable itself was noticed to have had some influence on the identification of stops as voiced or voiceless. It is clear that the distinction between initial voiced stops before and after a pause and unaspirated stops will not be a problem to listeners.

4. Conclusion

This study has shown that the English stops, /p, t, k/ went through the same stages of plosive articulation in Ghanaian English. In their productions, there was no pre-voicing; voicing began for the following voiced segments after they were released. But it has been noticed that for some of the stops, voicing coincided with the release, while for some, voicing began after an appreciable amount of time after the burst release. Voicing for the following segments varied according to place of articulation, vowel contexts, vowel duration, and word type. Again, speaker sex appears to have no effect on the realisation of the voice onset times of the stops. Interestingly, the voice onset time was found to be generally short among all the speakers. Nonetheless, listeners were able to label them as voiceless; that is, they were able to distinguish between the voiceless stops with short VOTs and the voiced stops before and after a pause.

We can, therefore, say that Ghanaian speakers of English realise English voiceless initial stops with and without aspiration. However, they were able to identify both as voiceless. There is, therefore, the possibility that Ghanaian listeners will be able to tell the difference between voiceless stops (whether said with a short or long VOT) and voiced stops before a pause or after a pause. They are highly unlikely to have any difficulty distinguishing voiced stops before and after a pause or silence from voiceless unaspirated stops. Finally, speaker sex is highly unlikely to have any influence on the realisation of the voice onset times of the stops in Ghanaian English.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Auzou, P., Ozsancak, C., Morris, R. J., Jan, M., Eustache, F., & Hannequin, D. (2000). Voice Onset Time in Aphasia, Apraxia of Speech and Dysarthria: A Review. *Clinical Linguistics & Phonetics*, *14*, 131-150. <https://doi.org/10.1080/026992000298878>
- De Reuse, W. J. (1981). Grassmann's Law in Ofo. *International Journal of American Linguistics*, *47*, 243-244. <https://doi.org/10.1086/465693>
- Fant, G. (1973). *Speech Sounds and Features*. Cambridge, MA: The MIT Press.
- Gay, L. R. (1978). *Educational Research: Competencies for Analysis and Application* (4th ed.). New York: Merrill/Macmillan.
- Hilary, L. (2005). Introduction to Acoustic Phonetics. *Principles of Experimental Phonetics*, *2*, 180-222.
- Kaur, J. (2015). Factors Influencing Voice Onset Time (VOT): Voice Recognition. *International Journal for Research in Applied Science & Engineering Technology*, *3*, 174.
- Ladefoge, P., & Johnson, K. (2011). *A Course in Phonetics* (6th ed.). Belmont CA: Wadsworth.
- Lisker, L., & Abramson, A. (1967). Some Effects of Context on Voice Onset Time in English Stops. *Language and Speech*, *10*, 1-28. <https://doi.org/10.1177/002383096701000101>
- Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word*, *20*, 384-422. <https://doi.org/10.1080/00437956.1964.11659830>
- McCrea, C. R., & Morris, R. J. (2005). The Effects of Fundamental Frequency Level on Voice Onset Time in Normal Adult Male Speakers. *Journal of Speech, Language, and Hearing Research*, *48*, 1013-1024. [https://doi.org/10.1044/1092-4388\(2005\)069](https://doi.org/10.1044/1092-4388(2005)069)
- Mielke, J., & Nelson, K. (2018). Voice Onset Time in English Voiceless Stops Is Affected by Following Postvocalic Liquids and Voiceless Onsets. *The Journal of the Acoustical Society of America*, *144*, 2166. <https://doi.org/10.1121/1.5059493>
- Ohman, S. (1965). Coarticulation in VCV Utterances: Spectrographic Measurements. *Journal of the Acoustical Society of America*, *39*, 151-168. <https://doi.org/10.1121/1.1909864>
- Simons, G. F. (2009). Linguistics as a Community Activity: The Paradox of Freedom through Standards. In W. D. Lewis, S. Karimi, H. Harley, & S. Farrar (Eds.), *Time and Again: Theoretical Perspectives on Formal Linguistics: In Honor of D. Terence Langendoen* (pp. 235-250). Amsterdam: John Benjamins. <https://doi.org/10.1075/la.135.18sim>
- Smith, (1978). Effect of the Place of Articulation and Vowel Environment on Voiced Stop Consonant Production. *Glossa*, *12*, 163-175.
- Swartz, B. L. (1992). Gender Differences in Voice Onset Time. *Perceptual and Motor Skills*, *75*, 983-992. <https://doi.org/10.2466/pms.1992.75.3.983>
- Wells, J. C. (1982). *Accents of English: The British Isles* (2nd ed.). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511611766>
- Zue, V. W. (1976). *Acoustic Characteristics of Stop Consonants: A Controlled Study. Technical Report No. 523*. Lexington, MA: Lincoln Laboratory.