An Exploratory VOT Analysis of Somali Stop Consonants

Mohamed Abdiraman  
*St. Cloud State University*

Ettien Koffi  
*St. Cloud State University*

Follow this and additional works at: [http://repository.stcloudstate.edu/stcloud_ling](http://repository.stcloudstate.edu/stcloud_ling)

Part of the [Applied Linguistics Commons](http://repository.stcloudstate.edu/stcloud_ling)

Recommended Citation
Available at: [http://repository.stcloudstate.edu/stcloud_ling/vol1/iss1/2](http://repository.stcloudstate.edu/stcloud_ling/vol1/iss1/2)

This Article is brought to you for free and open access by The Repository at St. Cloud State University. It has been accepted for inclusion in Linguistic Portfolios by an authorized administrator of The Repository at St. Cloud State University. For more information, please contact kewing@stcloudstate.edu.
AN EXPLORATORY VOT ANALYSIS OF SOMALI STOP CONSONANTS

MOHAMED ABDIRAMAN AND ETTIEN KOFFI

1.0 Introduction

This study focuses on the production of group of consonants that linguists classify as stops. They are classified this way because the pulmonic air is momentarily stopped before being released. This happens when two articulators come together to block the airstream. The consonants that are produced this way are \([p, b, t, d, k, g]\). Phoneticians study some characteristics of these consonants by measuring the time lag between closure and the release of closure. The acoustic term for this is voice onset time, or VOT for short. The goal of this study is to examine how six Somali speakers produce these consonants when speaking English. Though this is only an exploratory study, important insights are gathered in support of two important issues of great importance in acoustics and in second/foreign language phonetics. The first preliminary finding validates a universal tendency that shows that the VOT of consonants that have the feature \([+\text{anterior}]\) is progressively shorter than those that have the feature \([-\text{anterior}]\). The second significant insight is that age of arrival seems to matter in the acquisition of \([p]\) for Somali speakers of English.

2.0 VOT and Stops Consonants

The VOT is defined aerodynamically and articulatorily by Yavas (2011) as “the duration of the period of time between the release of a plosive and the beginning of vocal fold vibration.” The nature and amount vibration is directly proportionate to the configurations of the glottis during the articulation of a particular sound. Voiced sounds are produced when the vocal cords come closer together and the air molecules that pass through the glottis causes them to vibrate. For voiceless sounds, the vocal cords are further apart, and so the air molecules pass through freely without causing any vibration. This can be illustrated by focusing on how the segments in <cat> are pronounced. In producing the consonant \([k]\), the vocal folds are further apart, so air molecules pass through the glottis freely. However, this is not the same for \([a]\) during the articulation of which the vocal folds come closer together and the air molecules that pass through cause vibration. For the pronunciation of \([t]\), the vocal folds move further apart again, and the air molecules pass through without causing any vibration. The VOT for a sound such \([k]\) is taken to be the time lag between when the articulators come apart and the vocal cords start vibrating. This acoustic behavior is generally easily seen on spectrograms. The VOT is calculated in milliseconds, abbreviated as “ms”. In an earlier analysis of the VOT of Somali speakers of English, Conway (2008, p. 48) made the following observations:
Looking at the VOT, five aspects of Somali pronunciation of stops emerge. First, VOT lengthened as the phoneme is articulated farther back in the oral cavity ... Secondly, word-initial VOTs are shorter than word-final VOT; medial stops tend to have the shortest VOT of all three positions within a word. Thirdly, voiced stops are devoiced in sometimes initial and always final positions of words. Fourth, word-medial voiced stops are articulated as fricatives in over 50% of all samples. Finally, the distinction between /t, d/ in Somali seems to rely on factors other than VOT and voicing.”

This paper is also a VOT study of Somali-accented English. However, it is not only a descriptive study. It also seeks to answer important questions about the effect of age arrival on the acquisition of English stops consonants in general. It also examines Somali VOT in the light of universal acoustic tendencies. Before answering these research questions, let’s first acquaint ourselves with the consonants of the Somali language.

3.0 Somali Stop Consonants

Koffi (2010)\(^1\) proposes the following table for Somali consonants:

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Alveolar</th>
<th>Palatoalveolar</th>
<th>Retroflex</th>
<th>Velar</th>
<th>Uvular</th>
<th>Pharyngeal</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stops</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>t</td>
<td></td>
<td></td>
<td></td>
<td>k</td>
<td></td>
<td></td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Voiced</td>
<td>b</td>
<td>d</td>
<td></td>
<td></td>
<td>q</td>
<td>g</td>
<td>g</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fricatives</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>f</td>
<td>s</td>
<td></td>
<td></td>
<td>χ</td>
<td>h</td>
<td>h</td>
<td></td>
<td>ξ</td>
</tr>
<tr>
<td>Voiced</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Affricates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>tʃ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nasals</strong></td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Trill</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>Glides</strong></td>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>j</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Somali Consonant Chart

A causal look at the chart shows that the consonant [p] is missing from the Somali inventory. Consequently, according to contemporary theories of second/foreign language acquisition, [p] is

\(^1\) The consonant chart in Saeed (1999:7) omits the places of articulation. This piece of information has been supplied by comparing Saeed’s chart with the IPA chart at http://weston.ruter.net/projects/ipa-chart/view/keyboard/. Moreover, Saeed does not follow the IPA convention of listing voiceless consonants before listing their voiced counterparts. My adaptation of Saeed’s chart complies with the IPA.
a prime candidate for a negative phonetic transfer. In this paper, we examine the acoustic behavior of all stops consonants, not only [p] to see if this prediction is borne out. Furthermore, the data allows us to see if the Somali speakers’ pronunciation of these stops consonant concord with universal phonetic tendencies about VOT. A small-scale study was conducted with six participants with the expectation that their pronunciation would answer help the research questions. Table 2 provides us with some basic sociometric information about the speakers. For the purpose of this study generic Somali names are used to conceal the identity of the participants:

<table>
<thead>
<tr>
<th>No</th>
<th>Name</th>
<th>Gender</th>
<th>Age of Arrival</th>
<th>Current Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jamal</td>
<td>Male</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>Aisha</td>
<td>Female</td>
<td>8</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>Salah</td>
<td>Male</td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>Zamia</td>
<td>Female</td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>Amino</td>
<td>Female</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>Ahmed</td>
<td>Male</td>
<td>10</td>
<td>26</td>
</tr>
</tbody>
</table>

Table 2: Participants

All the participants report that Somali is their primary language when interacting with other Somali speakers. They also describe themselves as bilingual English-Somali speakers. At the time of the study, the participants were enrolled in courses at three institutions of higher education in Minnesota: Augsburg College, Saint Cloud State University, and the University of Minnesota.

The participants were recorded producing words in their citation forms which contain the segments [p, b, t, d, k, g] in initial position. Each person pronounced a word containing the segment under consideration 18 times. The list of words appears in Table 3:

<table>
<thead>
<tr>
<th>No</th>
<th>Selected Target Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>/p/ the phoneme spelled &lt;p&gt; in &lt;pat&gt;: voiceless bilabial stop</td>
</tr>
<tr>
<td>2</td>
<td>/b/ the phoneme spelled &lt;b&gt; in &lt;bat&gt;: voiced bilabial stop</td>
</tr>
<tr>
<td>3</td>
<td>/t/ the phoneme spelled &lt;t&gt; in &lt;tot&gt;: voiceless alveolar stop</td>
</tr>
<tr>
<td>4</td>
<td>/d/ the phoneme spelled d in &lt;dot&gt;: voiced alveolar stop</td>
</tr>
<tr>
<td>5</td>
<td>/k/ the phoneme spelled &lt;c&gt; in &lt;cap&gt;: voiceless velar stop</td>
</tr>
<tr>
<td>6</td>
<td>/g/ the phoneme spelled &lt;g&gt; in gap: voiced velar stop</td>
</tr>
</tbody>
</table>

Table 3: Words and Segments
The total amount of tokens produced for this study is 108, that is, \((6 \times 3 \times 6)\). All six participants pronounced each word three times. All the participants produced all the target words. Figure 1 gives an example of the three repetitions of the word \(<\text{pat}\>\):

![Sample Data](image)

Figure 1: Sample Data

The recordings were done on a Macbook computer with a built in microphone in a quiet room. The participants were asked to pronounce each word as naturally as possible. The recorded data was analyzed using Praat, an online acoustic analysis software.

The VOT is extremely important in the study of second/foreign language accent because, according to Ferrand (2007, p. 267) speakers in all languages rely on this acoustic feature to categorize a segment either as voiced or voiceless. Ladefoged (2003, p. 94) underscores the importance of the VOT by stating that “any description of the phonetic structure of a language should include an account of VOT.” Impressionistic evidence indicates that Somali speakers have difficulties with some stop consonants, but not with others. However, this anecdotal evidence does not tell us a whole lot about what is going on, nor does it give us any information about the type of speaker who may have a difficulty pronouncing these sounds. This exploratory acoustic study is aimed at providing us with some answers.

4.0 The VOT of \([p]\) and \([b]\)

The analyses in all the sections follow the same logic and pattern. The VOT of the contrasting stop segments are analyzed according to place of articulation. The segments \([p]\) and
[b] are discussed first, followed by [t] and [d], and finally by [k] and [g]. The spectrograms in Figures 2 and 3 represent Amino’s pronunciation of [p] and [b]:

![Figure 2: VOT of [b]](image1)

![Figure 3: VOT of [p]](image2)

Figure 4 provides us with a graph summary of the VOT measurements for all the 36 tokens containing [p] and [b]:

![Figure 4 graph summary](image3)
The mean VOT of [p] across all participants is 61.5 ms whereas that of [b] is 20.33 ms. According to the mean VOT, the Somali speakers in the study do not have any difficulty differentiating [p] and [b] because the VOT of these two segments fall within expected norms. Byrd and Mintz (2010, p. 131) write that for the bilabial stops, the VOT difference between [p] and [b] is +25 ms. This measurement corresponds to what Byrd and Mintz refer to as “category boundary.” They describe it as follows:

The point at which perception switches from one sound to another is called category boundary. The VOT boundary for voicing differs slightly depending on the place of articulation of articulation. For bilabials, it’s +25 ms, for alveolars +34 ms, and for velars +42.

The mean VOT difference for all six Somali speakers is indeed +41.17 ms. This might lead the unsuspecting researcher to conclude rather erroneously that Somalis do not have a problem with [p] and [b]. The mean categorical difference between [p] and [b] masks a very important insight about Somali English. To shed some light on this hidden insight, let’s turn our attention to Amino’s pronunciation.

The data shows that the VOT of Amino’s realizations of [p] and [b] are respectively 66 ms and 39 ms. The VOT of [b] is particularly significant in this regard. Ladefoged (2006, pp. 147-8) notes that “a normal value for the VOT of English stressed initial /p/ would be between +30 ms distinguish voiceless stops from voiced stops.”

Baart (2010, pp. 90-91) and Yavas (2010, p. 108) state generally that a category boundary of +30 ms distinguish voiceless stops from voiced stops.
50 and 60 ms. English initial /b/ … may have a VOT of about 10 ms.” If we go strictly by how speakers of General American English (GAE) perceptually contrast [p] and [b], they need at least 40 ms of category boundary difference between the two sounds. However, in Amino’s pronunciation, the difference is only 25 ms long. This is significantly below the threshold that they need to tell [p] apart from [b].

What does this mean for the perception of the speech of Amino and other Somali people who speak like her? GAE interlocutors are likely to misperceive <pat> for <bat>. Amundson-Cisse (2009, pp. 65-6) provides spelling and perception errors to show that this is indeed the case. She recounts a story of how she was misheard once in class. She used the word <pastor> but a Somali student misunderstood it to be <bastard> and told her that she has used a “bad word.” The confusion came from the fact that the student could not hear the difference between [p] and [b]. On pages 94-98 she also lists many spelling mistakes where her students regularly confuse initial [p] and [b].

The sociometric information in Table 3 indicates that only Amino has a problem with [p] and [b]. Coincidentally, she is also the only participant who immigrated to the USA when she was 15. All the others arrived before the onset of puberty. Amino’s data confirms Koffi (2010) and (2011)’s observations that the speech of Somalis who were born outside of America or emigrated after puberty is generally more accented than those who immigrated before puberty.

5.0 The VOT of [t] and [d]

The spectrograms in Figures 5 and 6 are those of the initial consonants in <tip> and <dip> respectively. The pronunciations are those of Zamia:
The VOT measurements for [t] and [d] do not reveal any pronunciation problem. The mean length of [t] across all the speakers is 74.5 ms, and that of [d] is 12.5 ms. These measurements indicate that Somali speakers of English make a clear distinction between voiceless [t] and voiced [d]. Furthermore, GAE interlocutors are not likely to misunderstand a Somali person on account of his/her pronunciation of [t] and [d] because the category boundary for all the speakers
exceed the limit of +34 ms reported by Byrd and Mintz. Furthermore, the acquisition of [t] and [d] is not affected by the age of arrival of the participants.

6.0 The VOT of [k] and [g]

The consonant chart in 3.0 shows that the Somali language has the velar consonants [k] and [g]. Consequently, from the point of view of articulatory phonetics no pronunciation difficulties are expected for these two sounds. This observation is borne out by the spectrographic evidence provided by Ahmed’s pronunciations of the initial consonants in the words <cap> and <gap>:

![Figure 8: VOT of [k]](image)
The mean duration of VOT of [k] is 89.5 ms whereas that of [g] is 21.66 ms. Since the category boundary of these two sounds exceeds the limit of +42 ms, GAE listeners can perceive the voicing difference between [k] and [g] clearly. Consequently, it can be postulated that Somali speakers would not any difficulty producing the English velars [k] and [g].
7.0 Summary

All in all, the six participants produced 108 tokens containing the segments [p, b, t, d, k, g]. The sample size is too small for any strong conclusions to be drawn from it. However, this exploratory investigation has yielded data that accords with the impressionistic assessments of Somali English as described by Lindsey (2006), Amundson-Cisse (2009), Koffi (2010) and (2011), namely that some speakers such as Amino do not differentiate between [p] and [b] acoustically. A large scale study involving many speakers who fit Amino’s linguistic profile is needed to validate whether or not age of arrival has an impact on the pronunciation of [p] and [b].

The Somali data also lends support to the universal acoustic tendency which shows that the VOT of voiceless stops increases progressively as we move from the front of the mouth towards the pharyngeal wall. This is borne out by comparing the mean VOT of [p], [t], and [k] which are respectively 61.5 ms, 74.5 ms, and 89.5 ms. Furthermore, the Somali data lends credibility to the concept of “category boundary.” The fact that the category boundary of Amino’s pronunciation lies exactly at the threshold of +25 ms explains why her [p] may be confused with her [b].

ABOUT THE AUTHORS

Mohamed Abdirahman is a graduate student in the MA TESL program at the university. He was born in Muqdisho, Somalia and immigrated to the U.S.A. when he was ten years old, settling with his family in Virginia and living there for five and half years. His family later moved to Minnesota, where he has lived for the past fourteen and half years. His research interests include vocabulary acquisition and second language accents. He plans to devote his master’s thesis to Somali VOT patterns and bilabial stop consonants. He hopes to pursue a Ph.D. in linguistics.

Ettien Koffi is Professor of Linguistics. He teaches the linguistic courses in the MA TESL/Applied Linguistics at Saint Cloud State University. His research interest is at the interface between acoustic phonetics and phonology.

References


