

Supraglottal cavity shape, linguistic register, and other phonetic features of Somali

Jerold A. Edmondson*, John H. Esling§, and Jimmy G. Harris§
University of Texas at Arlington* and University of Victoria§
Arlington, TX 76019-0559 USA and Victoria BC V8W 3P4, Canada

1.0. Introduction. The *aryepiglottic folds* (a-e folds) have, for the most part, not been accorded a place of importance in the phonetic literature on states of the glottis, the production of phonation types, voice qualities, glottalization, and initiation (airstream mechanisms). One early counterexample of this neglect, however, was Lindqvist 1969.¹ We have, in a series of recent papers, assembled direct visual evidence in support of the importance of the supraglottal area in sounds made in the larynx, lower pharynx, and aditus laryngis. This paper is a continuation of that effort.

We simultaneously intend to expand on the phonetic work on Somali of L. E. Armstrong 1934, B. W. Andrzejewski 1955, J. I. Saeed 1999, and M. Orwin 1994, but we will not strive to be comprehensive, but focus on the involvement of the a-e sphincter in phonation types, details of the consonants, vowels, and tones utilizing laryngoscopic, acoustic, and airflow data to support our observations.

Our data is taken from the speech of one *Somali* native speaker, Mr. Mohamed H. Mohamoud, about 40 years of age. Somali is a Cushitic language spoken in Somalia, Ethiopia, Djibouti, and Kenya with a population of about 8 -10 million. Standard Somali is based on northern forms of the language, which is the variety spoken by Mr. Mohamoud.

The most noteworthy linguistic features of Somali are: (a) its *cavity-induced voice quality/vowel quality system* overlaid by a set of vowel harmony rules, (b) its *tone-accent system*, which is mostly active at the grammar and discourse levels to indicate grammatical category and information structure, and (c) its very complex *topic-focus grammatical system*, which encodes the information structure of sentences. In regard to the last named feature, Saeed (1999:230) has described it, focus is: (a) found in declarative and interrogative sentences, (b) found in main clauses only, (c) more restricted in interrogative and negative sentences, and (d) found to affect more than just NPs.

Aside from its typological uniqueness in regard to its grammar, Somali is also a language of great phonetic and phonological complexity extending to many subsystems. The key feature, however, we believe to be the system of its voice registers that result from its differences of *supraglottal cavity shapes*. The previous work on Somali has called the phenomenon by many

¹The authors' names appear in alphabetic order. Esling wishes to acknowledge the assistance of the Social Sciences and Humanities Research Council of Canada for the research equipment and the support of STR-Speech Tech Ltd for the use of their facilities. All authors wish to express their utmost appreciation to Mr. Mohamed Hassan Mohamoud for his willingness to undergo the experiments reported on here as well as for his exceptional insight into the workings of the Somali language. He is the perfect example of why the active assistance of linguistic informants in no way resembles the passive subjects of medical experiments, and he certainly deserves to be recognized for his contribution to this work. He has also provided us with a letter of informed consent allowing us to give him that recognition and formal acknowledgement. We wish also to express our appreciation for helpful comments on this paper provided by Professors A.C.L. Remijsen of the University of Leiden and Professor Martin Orwin of the School of Oriental and African Studies, London; the usual disclaimers of responsibility, of course, apply.

different terms. For example, Berchem (1992:29) describes two sets of vowels, dividing into a “...neutrale und eine zentralisierte Form”. Orwin (1994) says the system of Somali vowels is divisible into a “front series” and a “back series”, the former marked by a plus sign and the latter by a cedilla, following Andrzejewski 1955. Orwin’s 1993 article entitled “Phonation in Somali phonology” argues for the need of a feature [\pm spread glottis], but this term is not used to distinguish vowels but the voiced and voiceless consonants. Saeed (1999:11-6) speaks about plus or minus Advanced Tongue Root [ATR] to describe these two sets of vowels. He emphasizes that the two series: (1) are not allophonic variants of one another, (2) are involved in independent harmony sets so that all vowels in a polysyllabic root belong the same [ATR] series.

The common feature of all these previous accounts rests on changes of the vowel quality. However, it is our contention that Somali in addition to its vowel quality contrast has a voice quality contrast. Languages which possess such linked contrasts of vowel quality and voice quality are best treated as *register languages* following Eugenie Henderson in her famous description of Khmer vowels (Henderson 1952). In many register languages the root causes of changes of vowel quality and differences of voice quality are not to be sought in adjustments of oral tongue position or glottal aperture. In fact, they are engendered by sphincteric constriction of the a-e folds. Thus in Somali there are two basic registers:

Tense supraglottal cavity (Set 1)

1. Sphincteric compacting of the arytenoid-epiglottal aperture in the posterior-anterior dimension.
2. Vowel quality that is more retracted.
3. Voice quality that is tense.

Lax supraglottal cavity (Set 2)

1. Expansion of the arytenoid-epiglottal aperture in the anterior-posterior dimension.
2. Vowel quality that is more fronted and/or raised.
3. Voice quality that is lax.

These supraglottal cavities do have real effects on vowel quality and voice quality. Indeed, as all previous investigators, we have confirmed the existence of systematic difference between the two sets of vowels. As the acoustic plot of formant frequencies in Figure 1 demonstrates one set is always both higher and more front than the second. The meanings displayed (Set 1) vs. (Set 2) are: (1) *cad* ‘white’ vs. ‘flesh’, (2) *maag* ‘to provoke’ vs. ‘to shy away from’, (3) *hel* ‘to find’ vs ‘he found’, (4) *hees* ‘a song’ vs. ‘he sang’, (5) *bid* ‘to fly up’ vs. ‘corpse’, (6) *diid* ‘to faint’ vs. ‘to refuse’, (7) *hogo* ‘the last days of the dry season’ vs. ‘to cower’, (8) *doon* ‘to want’ vs. ‘canoe’, (9) *sug* ‘to ascertain’ vs. ‘to wait’, and (10) *duul* ‘to fly’ vs. ‘to wage war’.

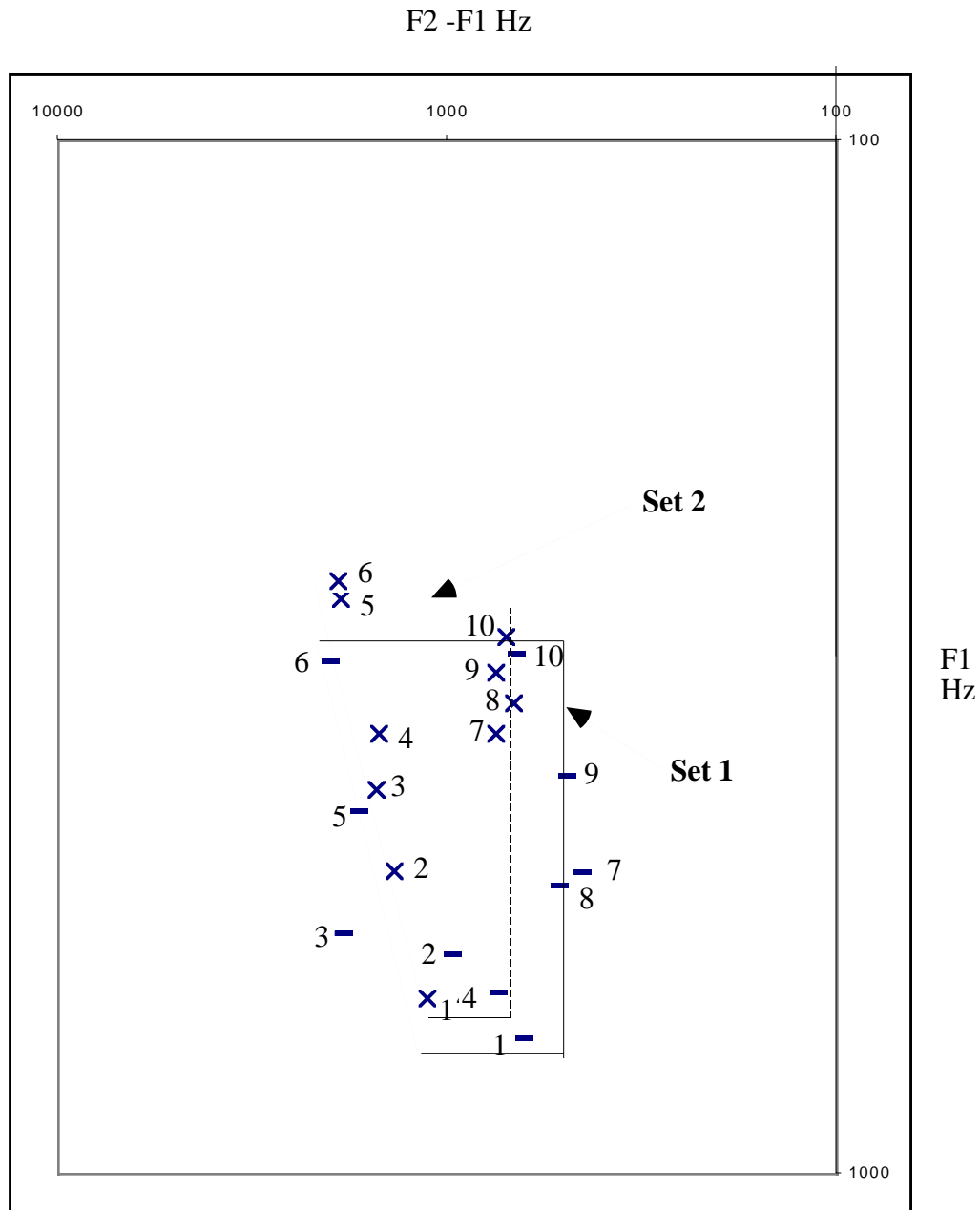


Figure 1: Vowel positions for the two Somali vowel sets as determined by the mean value of F2-F1 (Hz) vs. F1 (Hz) indicated by the arrow heads; labels for lexical items are given in Somali script, which does not distinguish the two.

But a contrast of vowel quality is not sufficient to describe the situation. In Figure 2 below we present a comparison of the states of the glottis for the two sets using a minimal pair *dûd* ‘to

faint' (Set 1) vs. *dîd* 'to refuse' (Set 2).²

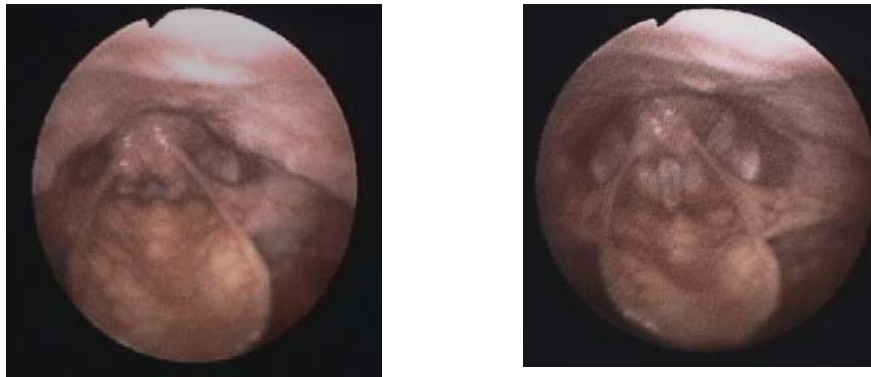


Figure 2: Laryngoscopic images of the Somali *dîd* 'to faint' and *dîd* 'to refuse'

As Figure 2 shows, *dîd* 'to faint' demonstrates sphinctered a-e and ventricular folds covering the sides of the glottal folds with obvious compacting of the aryepiglottic folds. This state contrasts with the non-constricted example, *dîd* 'to refuse'. Figure 2 presents direct visual evidence for a distinction of laryngeal/pharyngeal posture in addition to a distinction of vowel quality shown in Figure 1. We will return to these themes in discussion below, but for the moment we sketch some of the other dramatic features of Somali sounds.

2.0. Somali consonants. Somali has a complex and varied consonant system, especially the pharyngeals. We have just suggested that the two voice qualities *tense* (constricted supraglottal cavity or Set 1 register) and *lax* (expanded supraglottal cavity or Set 2 register) are responsible for changes of vowels. There are, however, also interactions with consonants. The same complexity extends to the tones and stress of Somali. In the transcription below we indicate tense voice with an underscore of the vowel, for example [ạ], stress with a macron [ˈ] before the syllable that is stressed, and tones will be marked over vowels as follows: HIGH is marked with [˥]; LOW is marked with [˩], and MID is marked with [˨˨]. For the moment we withhold further comment on these features as they will be treated in sections 4.0 and 5.0 below. We now turn to the variant manifestations of Somali consonants, as many of the consonant sounds have different positional variants within the word, thus increasing the number of consonant sounds. All nouns are given in the absolutive form, which is the citation form and is used when the nouns are regarded as outside of any syntactic or discourse context. Previous studies have not reached the level of description that is presented below.

The orthographic **b** initially represents a partially voiced bilabial oral stop [b̥]. We agree with Armstrong (1934), who states that a number of kymograph tracings showed no voicing at all in the speech of one informant. Her description of the sound was a “weak unaspirated p”. Intervocalic single **-b-**, especially in stressed syllables is a voiced bilabial fricative [β]. The word final variant in careful, overly correct speech is a voiced bilabial stop with a schwa offglide [-b̥^ə]. In conversational style speech the variant is a voiceless bilabial glottalized oral stop [-p̥]. When the letter is doubled intervocalically the first part is a voiceless glottalized bilabial oral stop followed by a partially voiced bilabial stop. Examples in all positions are: *bir* [ˈb̥ịr̥] ‘iron’, *boog* [ˈb̥ọọk̥] ‘wound’, *gabadh* [g̥ā ˈβạ̄r̥] ‘girl’, *dibbir* [ˈd̥ịp̥ b̥ịr̥] ‘to bloat’, and *xeeb*

²This slides and subsequent ones were taken from digital video files, which were made using a Kay Elemetrics 9100 Rhino-Laryngeal-Stroboscope with an Olympus ENF-P3 flexible fiberoptic bundle nasally inserted to a position behind the tongue and uvula. The video camera was a three chip CCD model by Sony interfaced to a digital camcorder.

[¹Hĕĕŕp̄] ‘coast’.³

The orthographic **t** is initially an apico-dental lamino-alveolar voiceless oral stop [¹t_ᵛ^h], which could be described as an apico-dental oral stop with a wide posterior wipe-off. It is aspirated and does not occur finally. Examples are: *tus* [¹t^hú̄s] ‘to show’ and *gaatir* [¹ḡǣə̄ t^hirr̄] ‘callus’. The orthographic **t** and orthographic **k** do not appear in word-final or syllable-final position.

The orthographic **d** represents initially a partially voiced apico-dental lamino-alveolar oral stop [¹d_ᵛ], which could be described as a apico-dental oral stop with a wide posterior wipe-off [¹d̄]. Intervocally, this sound is manifested as a voiced lamino-dental fricative, [ð]. Final **d** has two variants. In careful, overly correct speech it is a voiced denti-alveolar stop with a voiced schwa offglide, i. e. [-d^ᵛ] and in normal conversational style it is a voiceless glottalized apico-dental lamino-alveolar oral stop, i.e. [-¹t̄]. Examples in all positions are *dal* [¹d̄á̄l] ‘country’, *daan* [¹d̄ǣə̄n] ‘jaw’, *hadal* [h̄ā̄ð̄á̄l] ‘to speak, speech’, *xiddig* [H̄īŕ̄t̄ d̄īŕ̄k̄] ‘star’, and *cad* [¹ʔ̄á̄ŕ̄t̄] ‘white’.

The orthographic **dh** represents initially a partially voiced apico-postalveolar oral implosive [d̄ʔ̄]. This can be called a retroflex implosive. In the speech of our Somali speaker the tongue tip is up touching the posterior side of the alveolar ridge. The tip of the tongue is not curved back very much. Finally, in careful, overly correct speech it represents a voiced apico-postalveolar retroflex implosive with a schwa offglide [d̄ʔ̄̄]. In normal conversational style finally it is voiced retroflex flap followed by a voiceless apico-postalveolar retroflex glottalized stop that can sometimes be released as an ejective. Intervocally, it represents a voiced retroflex flap [ɾ] followed by a partially voiced retroflex oral implosive. Examples are: *dhiig* [¹d̄h̄īŕ̄k̄] ‘blood’, *badh* [¹b̄āɾ̄ŕ̄t̄] ‘to halve, cut in half’, *gabadh* [ḡā̄ β̄á̄ɾ̄ŕ̄t̄] ‘girl’, *qudhun* [¹q̄ū̄ d̄ū̄n] ‘stench’ and *dhudhun* [d̄h̄ū̄ɾ̄ŕ̄t̄ d̄ū̄n] ‘elbow to the end of the middle finger, the lower arm’.

The orthographic **g** represents initially a partially voiced dorso-velar oral stop [ḡ]. Intervocally, when it occurs singularly, it represents a voiced dorso-velar fricative [ɣ]. Finally, it is a voiceless glottalized dorso-velar oral stop [ŕ̄k̄]. The word final variant in careful, overly correct speech is a voiced dorso-velar stop with a schwa offglide [-g^ᵛ]. When it occurs intervocally as a geminate, then it is complex sound. First it is a glottalized dorso-velar stop, followed by a partially voiced dorso-velar stop. Examples in all positions are: *geel* [¹ḡè̄l] ‘camel’, *sagaal* [s̄ā̄ ʔ̄ā̄q̄á̄l] ‘nine’, *oggol* [ŕ̄ð̄ŕ̄k̄ ḡó̄l] ‘to be willing’, and *ilig* [¹ŕ̄ī ŕ̄īŕ̄k̄] ‘tooth’.

The orthographic **k** represents initially and intervocally an aspirated dorso-velar oral stop. Examples are: *kac* [¹k^hā̄ŕ̄] ‘to stand up’, *bakayle* [b̄ā̄ k^hā̄l̄l̄ē̄] ‘rabbit’, and *cukur* [¹ʔ̄ú̄ k^hū̄r̄] ‘wrist’.

The orthographic **q** represents a voiceless dorso-uvular stop, accompanied by epiglottal components depending upon the word position. The peculiar thing about this stop is that it might be called a voiceless epiglottalized uvular stop [q̄ŕ̄], because it is accompanied by an epiglottal stop. Finally, in careful, overly correct speech it represents a voiceless dorso-uvular stop with a schwa offglide [q^ᵛ]. In more normal conversational style it will be an epiglottalized dorso-uvular stop that can be released as an ejective. Examples are: *god* [¹q̄ŕ̄ŕ̄ŕ̄t̄] ‘to dig’, *maqal* [¹m̄á̄ q̄ŕ̄à̄l] ‘to hear’ *aqal* [¹ʔ̄á̄ q̄ŕ̄à̄l] ‘a house’, *liq* [¹l̄īq̄ŕ̄] ‘to swallow’.

The orthographic ‘ represents a glottal stop. An orthographic glottal stop does not occur initially but does occur medially and finally. Examples in all positions are: *iib* [¹ʔ̄īŕ̄ŕ̄p̄] ‘sale’,

³The orthography allows the doubling of <b n d r m g l>, but there is reason to believe that there are geminate forms for <s k f> and perhaps other stop and fricative sounds.

ba [ʰbáʔ] ‘calamity’ *ba’a* [ʰbáʔà] ‘the calamity’. It should be contrasted with *bah* [ʰbāh] ‘a family group, household’ and *bax* [ʰbāH] ‘to go out’.

The orthographic **j** represents a voiceless aspirated palato-alveolar affricate [tʃ^h] with a wide-grooved fricative release. In southern varieties of Somali this sound often appears as a voiced palato-alveolar affricate [dʒ]. Examples are: *jid* [ʰtʃ^hɪʔt] ‘way’ and *gujuf* [ʰgū tʃ^hɔf:] ‘dumpster’.

The orthographic **sh** represents a voiceless palato-alveolar wide grooved fricative [ʃ]. In final position it can be phonetically long. Examples are *shaah* [ʰʃáʔh:] or [ʰʃæ̀h:] ‘tea’,⁴ *shaash* [ʰʃáʔ:] ‘head scarf’ *shabeel* [ʃā βéé] ‘leopard’. *tasheel* [t^hā ʰʃéè] ‘to economize’

The orthographic **s** represents a voiceless lamino-alveolar narrow grooved fricative [s]. Word finally it is frequently phonetically long. Examples are *sal* [ʰsāl] ‘base’, *naas* [ʰnás:] ‘breast’, *qosol* [ʰqə sə] ‘to laugh, laughter’, and *bisil* [bīs ʰsíl] ‘ripe’.

The orthographic **f** represents initially, intervocalically, and finally a voiceless labio-dental fricative. Examples in all positions are: *far* [ʰfárr] ‘finger’, *dafir* [ʰdā fírr] ‘to deny’, *afar* [ʰf áfárr] ‘four’, and *af* [ʰʔáf:] ‘mouth’. In final position it is frequently phonetically long.

The orthographic **kh** represents a voiceless dorso-uvular fricative [χ]. Many lexical items with this sound are loan words from Arabic, though there are some Somali words possessing this sound as well. Examples are *khad* [χāʔt] ‘ink’ *khuuro* [χūúràrr] ‘snoring’, *dukhar* [dūχàrr] ‘to smash one’s head’.

The orthographic **h** is a voiceless glottal fricative [h]. It occurs initially and finally. Intervocalically and before lax (Set 2) register, it becomes a voiced glottal fricative [ɦ]. Examples in all positions are: *hees* [ʰhéēs:] ‘to sing a song’, *bohol* [bə ʰhól] ‘a cave’, and *baah* [ʰbæ̀h:] ‘to spread, be diffused’, *haad* [ʰhæ̀h:] ‘generic term for big birds of prey’ and *haad* [ʰhāʔt] ‘vulnerability’

The orthographic **c** is a voiceless epiglottal stop [ʔ] with a voiced pharyngeal approximant offglide [ʔʕ] when it occurs initially. It is sometimes accompanied by weak trilling of the epiglottal surface. Examples are *cad* [ʔʔáʔt] ‘white’, *jacayl* [tʃ^hā ʔʔáil] ‘love’ *sacab* [ʰsā ʔʔáip] ‘palm of the hand’ and *sac* [ʰsāʔ] ‘cow’.

The orthographic **x** represents a voiceless epiglottal fricative with trilling at the laminal surface of the epiglottis [H]. Examples are: *xaas* [ʰHás:] ‘family’, *waan naaxaa* [nāá ʰHāá] ‘I become fat’ *waxar* [wà Hár] ‘a female kid goat’, *naax* [nāáH] ‘to become fat’

The orthographic **m** represents initially a voiced bilabial nasal. Intervocalically, it is a voiced bilabial nasalized fricative [-β̥-] or approximant [-v̥-] and it rarely occurs finally. Examples are: *maag* [māáʔk] ‘to provoke’, *waan dhaamaa* [ʰdāá ʰβ̥āá] ‘I am better than’.

The orthographic **n** represents initially a voiced apico-alveolar nasal [n] and finally it is a long voiced apico-alveolar nasal with or without an offglide, especially after short vowels. The intervocalic orthographic n is short and with less strength of articulation. The tongue is curled only slightly up toward the gumridge but does not touch. Examples are: *naas* [ʰnás:] ‘breast’, *waan faanaa* [ʰfāá ʰnāá] ‘I boast’, and *waan faannaa* [ʰfāán ʰnāá] ‘we boast’.

The orthographic **l** represents initially and finally a voiced apico-alveolar lateral [l-].

⁴The form with lax voice is, according to Mr. Mohamoud, often used in southern varieties of Somali.

Like orthographic **n** the intervocalic orthographic **l** is short and with less strength of articulation. Examples are *lis* [lɪs] ‘to milk’, *galab* [gā ʼl.áʼp] ‘afternoon’ but also *galab* [ʼgē ʼl.èʼp] (very lax) ‘to overturn (onomatopoeia)⁵, *kallah* [ʼkʰāl ʼlāh], ‘to leave early in the morning’, and *dil* [ʼdɪl] ‘to kill’ *laal* [ʼlāʼl] ‘to suspend’, *waan daalaa* [ʼd̪æ̪æ ʼl.æ̪æ] ‘I get tired’, *waan daallaa* [ʼd̪æ̪æ ʼl.æ̪æ] ‘we get tired’, and *daala* [ʼd̪æ̪æ ʼl.æ̪] ‘the one/man that gets tired’.

The orthographic **r** represents initially a partially voiced apico-alveolar trill [r-]; southern Somali often demonstrates a flap r pronunciation instead of the trill. Initially, the trill often begins voiceless [r̥r]. Finally, it is apico-alveolar that starts voiced and ends voiceless [r̥r]. When it is singular intervocalically, the r is an apico-alveolar tap [r]. Examples are: *rab* [r̥rāʼp] ‘to desire’, *garab* [ʼgārāʼp] ‘shoulder’, *xeer* [ʼhɛ̪ɛr̥r] ‘tribal law’, *aar* [ʼʔāʼr̥r] ‘male lion’, *waan saaraa* [ʼsāʼr̥r] ‘I am taking out right now’, *waan saarraa* [ʼsāʼr̥r] ‘we are taking out right now’ and even with three stresses *saaraayaa* [ʼsāʼr̥r] ‘I am taking out right now’ (this forms is restricted to speakers of the eastern regions of Somalia).

The orthographic **w** represents a voiced labio-velar apoproximant [w]. Examples are *wan* [wān] ‘a ram’, *haween* [hā ʼwē̪ɛ̪n] ‘women’.

The orthographic **y** represents a fronto-palatal approximant. Examples are *yaab* [ʼjāʼp] ‘to be surprised’, *yeeɗh* [ʼjē̪ɛ̪ʔt] ‘to call’⁶, *ayeef* [d̪ā ʼjē̪ɛ̪f:] ‘monkey’ and *hooyo* [ʼhōó ʼjōʔ] ‘to go home’⁶. When it is geminate (even though the geminate form is not recorded in the orthography) -yy- intervocalically, it becomes a fronto-palatal fricative [j̠]. *hooyo* [ʼhōó ʼj̠] ‘mother’.

3.0. Somali vowels. As Berchem describes them, there are five basic vowels, which can be long or short in duration. Each of the ten resulting forms can, in turn, be “neutral or centralized” in his terms, or *sphinctered* versus *expanded supraglottal cavity* in ours, which gives a grand total of 20 vowel contrasts. The Somali orthography does not notate the register difference of vowels.⁷

The orthographic **a** represents two vowel sounds, a low back unrounded vowel when it is long [ɑ:] that is centralized or a low central unrounded vowel when it is short [ɑ] (sphinctered register), which contrasts with a low front unrounded vowel that is centralized [æ:] when long and a near-open central vowel [ɐ] when short (expanded register). Examples are: *maag* [ʼmāʼk] vs. [ʼmæ̪æʼk] ‘to provoke’ vs. ‘to shy away from’ and *cad* [ʼʔāʼt] vs. [ʼʔɛ̪ɛʼt] ‘white’ vs. ‘flesh’.

The orthographic **e** represents two vowel sounds, a open mid front unrounded vowel [ɛ] that can be long or short (sphinctered register) or an open mid unrounded vowel that is slightly raised and fronted [e] that can be long or short (expanded register). Acoustically, the long and short variants are basically quite close in quality. Examples are: *hel* [ʼhē̪l] ‘to sing’ vs. [ʼhē̪l] ‘he sang’ and *hees* [ʼhē̪s] ‘to find’ vs. [ʼhē̪s] ‘he found’

The orthographic **i** represents two vowel sounds, a close front unrounded vowel [i] in both long and short variants (sphinctered register) or a close front unrounded vowel that is

⁵There appear occasionally to be examples of onomatopoeic or ideophonic speech that may occur in an even laxer voice quality.

⁶This item seems to have level tone on the first syllable despite the long vowel. An alternative pronunciation is [hò ʼjōʔ] ‘to go home’.

⁷Somali has has a Final Vowel Assimilation phenomenon, cf. Saeed (1999:30-31), in which determiner suffixes are added to noun stems and cause the final vowel of the noun stem to assimilate to the suffix.

slightly raised and fronted [i] in long and short variants (expanded register). Acoustically the quality of the long and short variants is very similar. Examples are: *diid* [ˈd̪iːiʔt̪] ‘to faint’ vs. [ˈd̪iːiʔt̪] ‘to refuse’ and *bid* [ˈb̪iːiʔt̪] ‘to fly up (as of dust)’ vs. [ˈb̪iːiʔt̪] ‘corpse (archaic)’.

The orthographic **o** represents two vowel sounds, a open mid rounded back vowel [ɔ] (sphinctered register) or a open mid rounded back vowel [o] that is raised and slightly centralized (expanded register). Long and short variants are similar in quality. Examples are: *doon* [ˈd̪oːɔn] ‘to want’ vs. [ˈd̪oːɔn] ‘canoe’ and *tog* [ˈt̪ʰɔʔk̪] ‘to tie up’ vs. [ˈt̪ʰoʔk̪] ‘wadi’.

The orthographic **u** represents two vowel sounds, a close back rounded vowel [u] (sphinctered register) or a a close back rounded vowel that is raised and centralized [u] (expanded register). Long and short variants are similar in quality. Examples are for the long vowels: *duul* [ˈd̪uːuʔr] ‘to fly’ vs. *duul* [ˈd̪uːuʔr] ‘to wage war’. The short vowel contrast is *dun* [ˈd̪uːn] ‘thread’ vs. [ˈd̪uːn] ‘to collapse’.

While we have sometime transcribed the vowels of the two registers with the same or with differing IPA notations, for all examples, though, the sphinctered set always represents a position in acoustic vowel space that is more retracted than the corresponding value for the expanded set. This difference can be seen clearly in Figures 3 and 4 below.

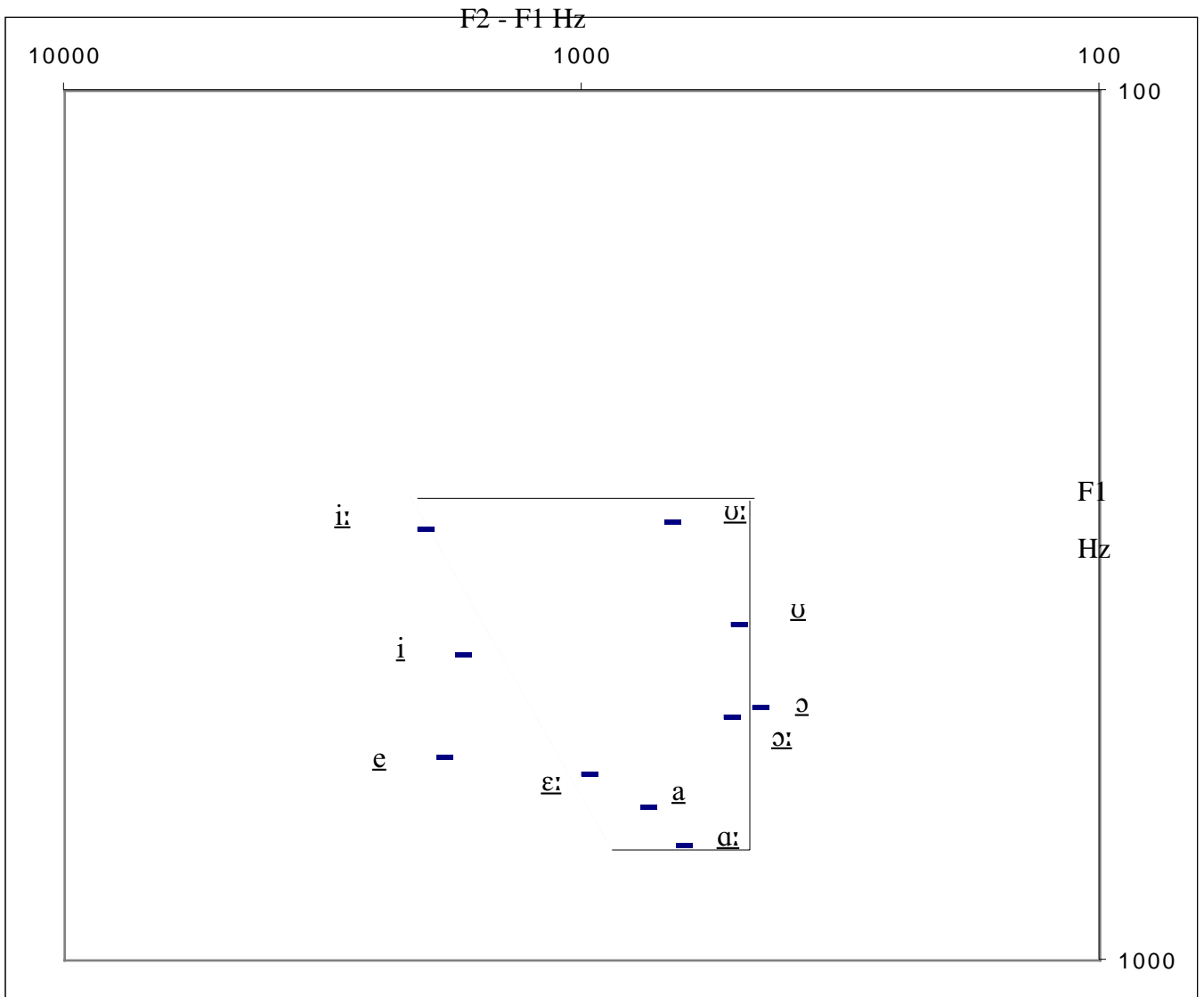


Figure 3: Acoustic plot of formants of the ten (five long/five short) vowels (sphinctered register)

F2-F1 Hz

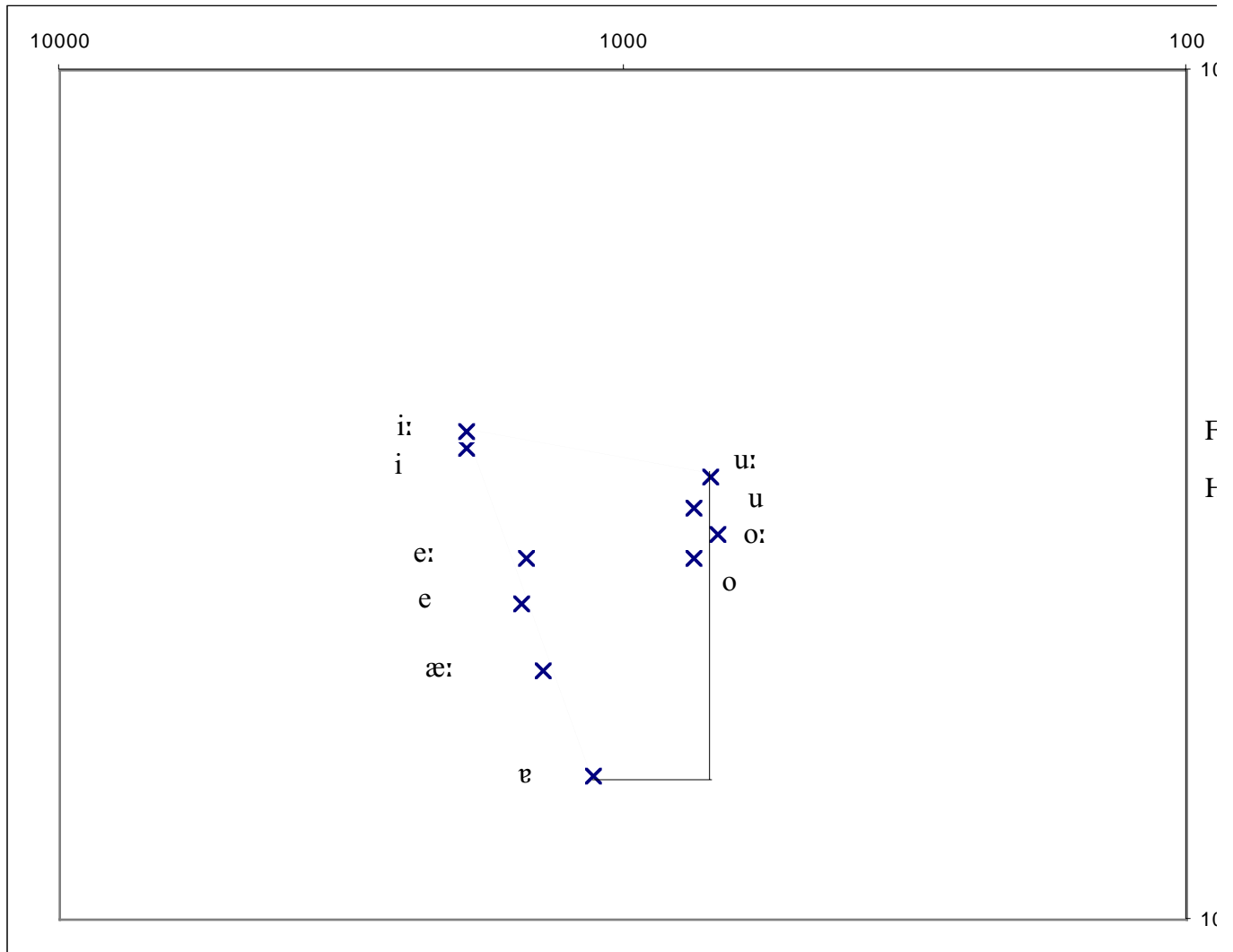


Figure 4: Acoustic plot of formants of the ten (five long/five short) vowels (expanded register)

It is further to be noted that there seem to be many more lexical items in the sphinctered register than in the expanded register.

4.0. Tone-accent. Somali has been treated as a *tone-accent language*. The main investigators of the phenomena have been Hyman 1981, Berchem 1992, Saeed 1999, and Le Gac 2001, 2002. The special feature of Somali tone-accent is that for some word categories, e.g. nouns, both syntactic and discourse factors co-determine the tone-accent of the language, cf. the discussion about the focus-topic typology of Somali above. Because discourse and information structure issues are involved, accounts have been a subject of some interest and of much controversy. In our paper we will not consider any cases but the very simplest of lexical states in regard to tone-accent and ignore completely the wider syntagmatic or discursal context and its effects on pitch trajectory. That said, it is necessary to establish a few generalizations about the tonal system of Somali before continuing.

Tonal Features

1. Somali is a tone-accent language with two or three lexical tones (Berchem 1993 assumes three, whereas Saeed 1999 posits two). The tones are called HIGH, MID, and LOW.

HIGH is often rising; LOW is often slightly falling.⁸ In long syllables there can additionally be the tone combinations HIGH-LOW and MID-HIGH. The stress or accent and tone co-determine each other, but see Saeed (1999:41-47) for a thorough discussion of this complex phenomenon. If a syllable has HIGH tone, then an accent always falls on that syllable. If there is no HIGH, then the accent falls on a MID. Accent or stress never falls on a LOW. In some instances there may be disyllabic words with stress on both the stem and inflectional endings. In such cases both syllables with the HIGH-LOW or MID-HIGH will receive full stress, e.g. *waan naaxaa* [ˈnāá ˈHāá] ‘I become fat’ and *maroodi* [mà ˈròð ˈðì] ‘elephant’. It is the property that every HIGH tone must have stress that make Somali unlike other stress or accent languages such as English, where only one syllable of a word, whether stem or affixes, has main stress.

2. Tone is not a part of the *harmony sets* of Somali vowels. If the last vowel has the *expanded pharynx voice quality*, then all preceding vowels have this feature as well. But tone is independent of this system.

3. Tone is a primary marker of grammatical gender as well as an indicator of grammatical category. Somali has grammatical gender in nouns.

4. Somali speakers can pre-empt some change rules from happening, e.g. when they want to indicate that they wish to continue speaking, Saeed (18).

First consider Figure 5, which shows the three tones in short syllables.⁹ Since LOW tone is never stressed, it is likely to appear only as an unstressed syllable in a compound.

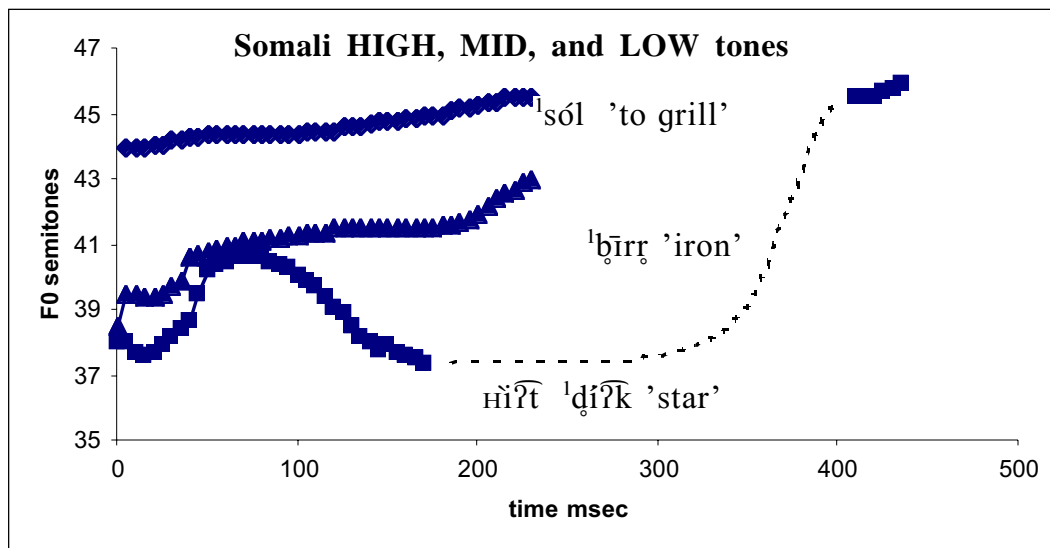


Figure 5: Comparative plot of the three Somali tones on short syllables: ˈsól ‘to grill’, ˈb̥írr̥ ‘iron’, and H̥íṛ̥t̥ ˈd̥íṛ̥k̥ ‘star’

⁸Saeed (1999:17-24) has a slightly more phonologically-focused analysis of the tones in which only HIGH and LOW are distinguished. He notes, though, that a tone becomes lower, whenever it is followed by a pause. We found that HIGH and MID may, at times, be difficult to distinguish. We, nevertheless, could distinguish pitch differences between the first syllable of *gābádh* ‘girl’ and *x̥íḏdíḡ* ‘star’ and many other examples. He notes as well that the Tone Bearing Unit in Somali is the mora, and long vowels and diphthongs possess two moras, which we would certainly agree with. In our more phonetically-oriented treatment we have kept HIGH and MID distinct even though they might be regarded as two members of the same phonological class.

⁹The tone values plotted here have been constructed from computer-generated composites of three repetitions of a syllable.

In Figure 5 we can note that HIGH tone is slightly rising with a pitch trajectory that begins about 44 semitones and rises to more than 45 semitones;¹⁰ the MID tone begins at 41 semitones and rising to about 43 at the end of the syllable. The LOW tone has an onset frequency of about 40 semitones and falls below 37 semitones. The LOW tone is illustrated here with a two syllable word.

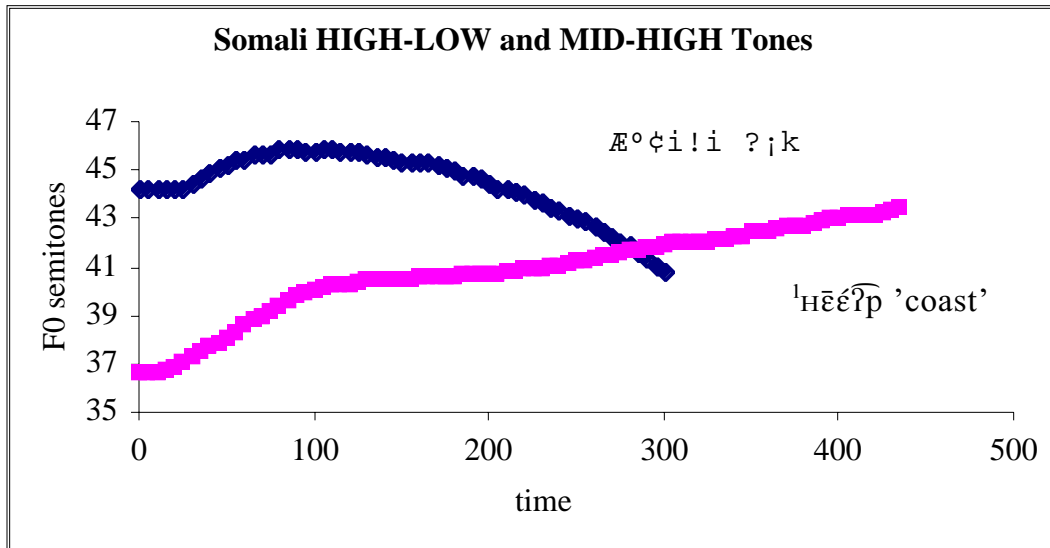


Figure 6: Comparative plot of the two Somali tones on long syllables: ^ldʒiŦk 'blood' and ^lHĒÉŦp 'coast'

In Figure 6 we see a plot for the long syllables corresponding to the Figure 5 above for the short syllables. In Somali there appear to be two such combinations of basic tones in long syllables. The HIGH-LOW after some lowering at onset begins its fall from 46 semitones and drops to 40 semitones. The MID-HIGH tone shape again after some influences of the initial starts its rise from 41 semitones and achieves its peak value of 44 semitones at the end of the syllable.

We have also found support for the claim in Berchem 1992 that Somali is a tone-accent language in the sense that accent and tone are not free. The following plot shows this connection of intensity and pitch in the minimal pair ^lnàn 'boy' (MID plus stress followed by LOW tone) and ^lnán 'girl' (MID tone followed by stressed HIGH tone).

¹⁰The semitone scale used here is based on the well-tempered Klavier scale of the piano with a reference point of 16.35 Hz.

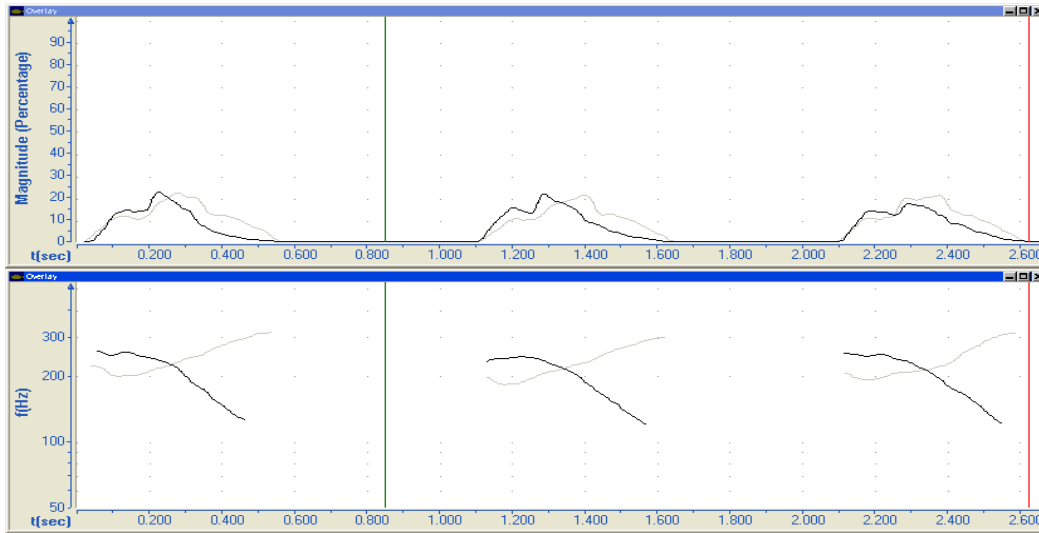


Figure 7: Comparison of three repetitions of $\bar{t}i\ n\grave{a}n$ ‘boy’ with dark lines and $\bar{t}i\ n\acute{a}n$ ‘girl’ with lighter lines, showing intensity and pitch differences

As Figure 7 shows, in this disyllabic word-form the first syllable gets the stress in $\bar{t}i\ n\grave{a}n$ ‘boy’ and the second syllable gets it in $\bar{t}i\ n\acute{a}n$ ‘girl’. This stress difference can be seen in the top window, as the word-form for ‘girl’ has low intensity on the first syllable and a strong second syllable. The word-form for ‘boy’ does have a strong onset of the second syllable but falls off rapidly across the pitch trajectory. Stress and tone also co-distinguish *grammatical categories*; verbs are LOW and adjectives HIGH, and *grammatical case*; absolutive case has tone according to its gender, as just described; nominative and objective have tones other than these, and *discoursal focus*; rules for this phenomenon are provided by Le Gac and depend upon noun class and other factors such as $[\pm\text{focus}]$, $[\pm\text{subject}]$, and $[\pm\text{final}]$. There are very few examples of tonal difference as the only lexical difference. An example of this rare difference the contrast between $k^h\acute{a}?$ ‘from’ and $k^h\bar{a}$ ‘the one’. As one can see $k^h\acute{a}?$ ‘from’ begins at the 42 semitones level and rises to about 44 semitones. It is quite a lot shorter than $k^h\bar{a}$ ‘the one’ begins at 39 semitones and rises to about 40 semitones.

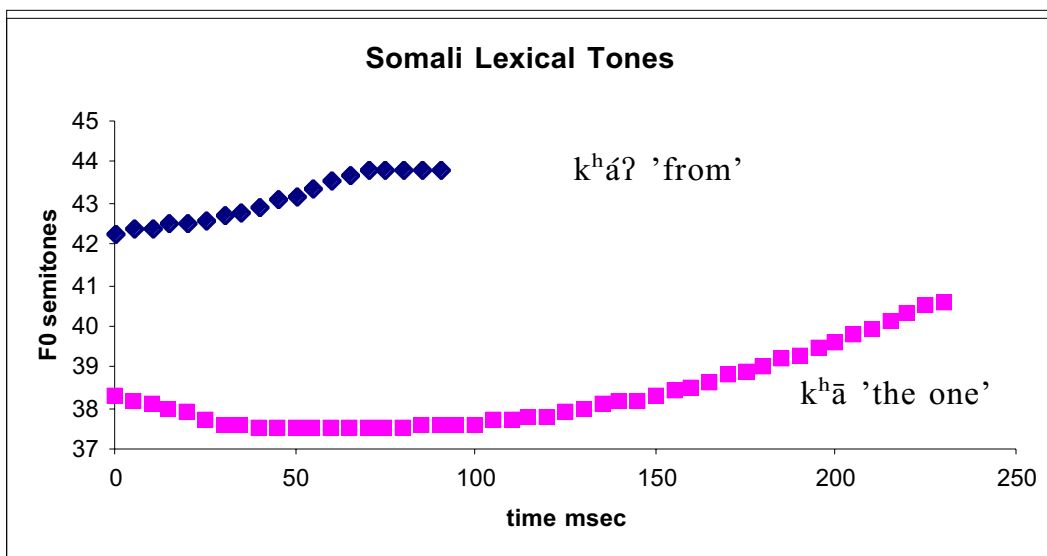


Figure 7: Lexical contrast between HIGH tone $k^h\acute{a}?$ 'from' and MID tone $k^h\bar{a}$ 'the one'

One interesting confirmation of the role of stress-tone interdependence is that the final item in a series of three repetitions in Mr. Mohamoud's speech tended to develop into low tone and was unstressed because of its in the series of repetitions.

5.0. Laryngoscopic observations. The following fiberoptic photo will help to familiarize the reader with the components of the larynx involved in this study.

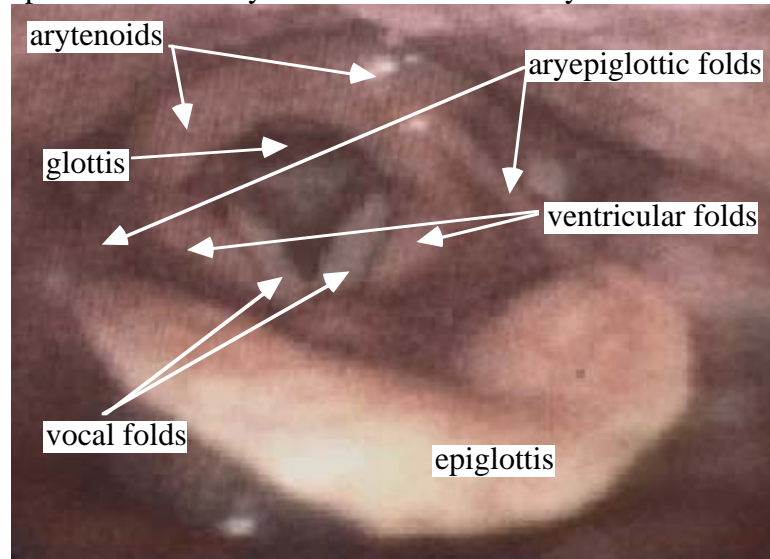


Figure 8: A view of the larynx and its components in the breath position

In Figure 2 above we demonstrated the contrast between sphincteric supraglottal cavity and expanded supraglottal cavity. In general, our video images suggest that for the Somali registers there are two aerodynamic valves with an intermediate valve, shaping the egressive airstream, which we show schematically in the following diagram.

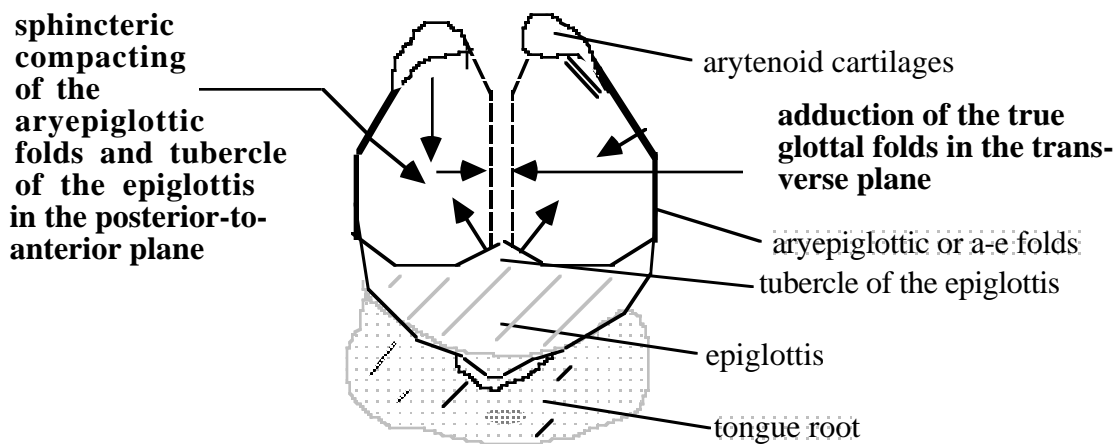


Figure 9: The two aerodynamic valves shaping the Somali airstream

There is first of all the adduction or approximation of the glottal folds and ventricular

folds in the medial direction. This movement is followed in time by a posterior-anterior collapse of the base of the epiglottis and arytenoids toward the center as well as a sphincteric compacting of the a-e folds, moving forward and upward toward the epiglottis and the root of the tongue, which in turn buckles antero-posteriorly as the larynx rises and the tongue retracts. At the same time, the tip of the epiglottis often rises in an upward direction, curling over and even hiding the arytenoids and glottal folds from view. Its opposite, the expansion of the arytenoid-epiglottal aperture, is effected by increasing the distance between these two structures via relaxation of the aryepiglottic sphincter and advancing the root of the tongue and the base of the epiglottis, away from the posterior pharyngeal wall. The impression from the video images is that both of these settings are deviations from the neutral breath position settings.

As an illustration of the interaction of consonants and phonation types, consider the lexical item *qiiq* [q̣ʰīīq̣ʰ] ‘smoke, to give off smoke’ (with H-L tone on a long vowel).



frame 1



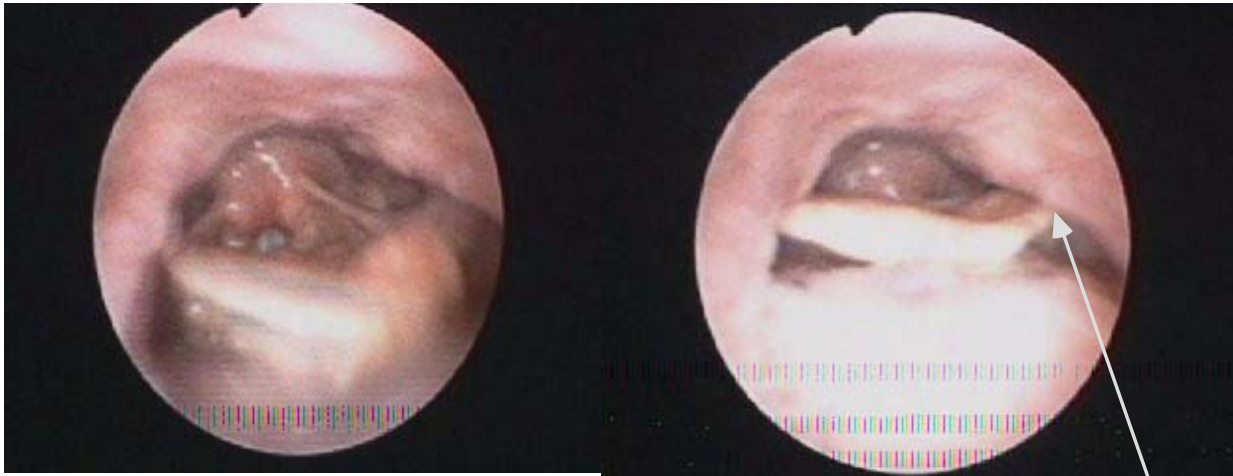
frame 6



frame 15



frame 23



frame 32

frame 40 Note the contact with back of throat



frame 44

Figure 10: Stages in the articulation of [q̤i̤q̤] ‘smoke, to give off smoke’, including laryngoscopic images and a spectrogram. Photographic frames are indicated at the bottom of the spectrograms

In Figure 10 we present a sound spectrogram and a sequence of selected frames taken from the original laryngoscopic video; camera speed 30 frames/sec. Before articulation of *qiiq* ‘smoke’ begins (frame 1), the glottis assumes the state of the glottis for breath. At onset of activity, the root of the tongue begins a backing gesture in synchrony with approximation of the vocal folds (frame 6). As the sequences comes to a close, the tongue pushes the underside of the epiglottis below the tubercle even further toward the central airway. The tip of the epiglottis curls up and back obscuring the glottal area until it almost touches the back of the throat (frame 15). The epiglottal tip descends again until the glottal area just becomes visible. At that time when the epiglottis is positioned directly over the center and very near to the closed vocal folds, the aryepiglottic schincter is much more tightly compressed beneath the epiglottis than in the typical configuration for a plain glottal stop. As the epiglottal stop is released, second formant activity on the spectrogram begins (frame 23) in a pattern consistent with pharyngealization. At a point about two-thirds of the way through the syllable, the tongue begins retracting towards the

back of the throat again, and the aryepiglottic sphincter closes and presses up anteriorly beneath the tongue. Once more the aryepiglottic closure is below the tongue root and the epiglottis, covering the glottal area. (frame 32). In frame 40 there is tight contact of the a-e folds up and under the base of the epiglottis with elevation of the larynx as a whole (notice that it is closer to the camera and brighter in the image); and there is also contact of the sides of the epiglottis with the back pharyngeal wall. At full release (frame 44) the a-e folds have opened fully, the tongue and epiglottis have advanced, and the vocal folds are fully abducted and visible at the glottis with the arytenoid cartilages fully parted.

As the data above and that in Figure 10 show, Somali is a language that makes use of plain glottal stops, with constriction of the epilaryngeal tube above the glottis. But Somali phonetic output also shows a set of phonetic gestures in phonological contrast to the plain glottal stop in which there is full epiglottal stop closure, that is maximum possible closure of the laryngeal mechanism rendering the entire pharyngeal vocal tract shut, cf. Esling et 2003. An example of this contrast is in initial position within the pair: 'sale' vs. 'a kind of fruit' (glottal stop vs. epiglottal stop).

Figure 11 for 'sale' and 'a kind of fruit' should go here

In separate file

Figure 12: Sequence of stages in the articulation of *dîid* 'to refuse' including laryngoscopic images and a spectrogram. Photographic frames are indicated at the bottom of the spectrograms

In Figure 12 are found a similar collection of frames that depict the stages in the production of *diid* [ᵀᵀᵀᵀᵀ] 'to refuse', which demonstrates expanded pharynx. Before articulation of [ᵀᵀᵀᵀᵀ] 'to refuse' begins (frame 1), the glottis assumes the state of the glottis for breath. At onset of activity, the root of the tongue begins to move in the opposite direction shown in Figure 9, as the tongue follows a vector away from the center of the airstream; this gesture transpires in synchrony with the approximation of the vocal folds. This movement continues until the tongue root almost disappears from view (frame 7). Voicing begins about at frame 15; the glottis is fully adducted from its open position in frame 1, but the supraglottic cavity is wide open as the arytenoids and the epiglottis have moved away from one another to the back and front, respectively, allowing the widest possible aperture for airflow above the glottis. This expanded pharynx position is maintained for about two-thirds of the duration of the syllable when the aperture of the glottal opening reverses again and begins to compact (frame 15). The arytenoids and tubercle of the epiglottis approach one another and the a-e sphincter and the tongue root begin to compact as the epiglottis obscures the view of the glottis (frame 38) preparing for the syllable ending glottalized [t]. After the glottal component of the t is completed, the a-e sphincter opens from its constricted position, and the vocal folds open and come back into view (frame 44).

One obvious difference between *diid* [ᵀᵀᵀᵀᵀ] 'to faint' and [ᵀᵀᵀᵀᵀ] 'to refuse' is the volume of air generated by the two articulations. In Figure 13 below we present the results of comparative airflow measures of Mr. Mohamoud during the production of this minimal contrast.

The airflow waves were gathered with the Rothenberg Mask, (Glottal Enterprises, Inc), a circumferentially-vented hard plastic mask for low distortion measurement of oral and nasal airflow, cf. Rothenberg (1977). It possesses two chambers with an air pressure transducer in each that converts dynamically changing air pressure into an electrical signal. In this study we connected the output of the oral airflow transducer to a small battery operated amplifier of our own construction to boost the signal to line level. The output of the amplifier, in turn, is connected to a Cecil Box (SIL International, Inc.) and laptop computer running the Cecil software for capturing the signal in volatile memory. Because the Cecil Box is a mono-channel device with no multiplexing capacity, only the oral waveform was captured.

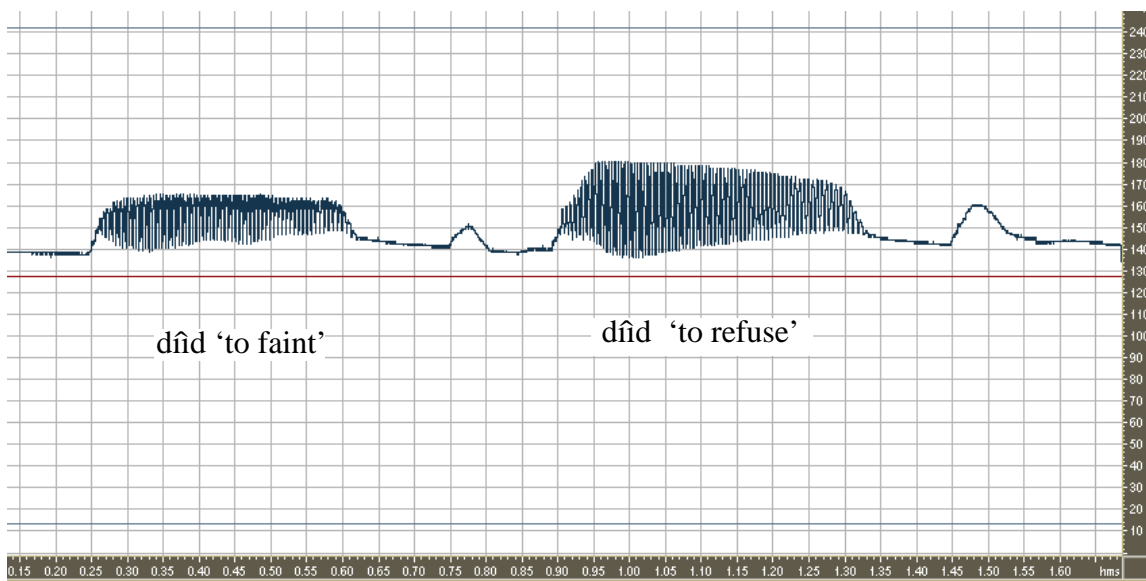


Figure 13: Comparison of airflow in the articulation of *dîid* ‘to faint’ and *dîid* ‘to refuse’

As this figure makes clear, *diid* [ˈd̥iːt̪] ‘to faint’ produced an airflow wave with an amplitude equal to about half that produced by *diid* [ˈd̥iːt̪] ‘to refuse’. This result shows the throttling and muting influence of the sphinctering. In this figure the release of the glottalized t is also very easy to discern.

One final example in Figure 14 of the supraglottal area in the articulation of Somali sounds is the famous case of Somali <x>, a voiceless epiglottal fricative and in word final position a trill. It was the subject of special emphasis in Armstrong 1934, who found trilling in the glottal area on the x-ray motion pictures taken at that time. We believe our data can fix the position of this trilling as in the supraglottal cavity. In *riix* [r̥iːx̥] ‘to push’ the sequence of frames begins in the breath position. The glottal folds approximate without change of the tongue root position (frame 10). After about 150 msec (frame 15) the sphinctering gesture begins without any retraction of the tongue. By the 21st frame there is strong sphinctering so that the glottal aperture is less than one-half that of the earlier maximum. When the 23rd frame is reached there is still voicing, the arytenoids are touching the tubercle of the epiglottis, and the tongue root and epiglottis retract rapidly to compact the supraglottic cavity. Voicelessness then begins (frame 25) tongue root is strongly backed with the tubercle not the highest point of the epiglottis. (frame 26). In frame 28 the laminal area just above the sphinctered a-e folds and anterior to the epiglottic tubercle trills in the high velocity airstream.

place about here

Figure 14: Sequence of stages in the articulation of *riix* ‘to push’ including laryngoscopic

images and a spectrogram. Photographic frames are indicated at the bottom of the spectrograms

6.0 Conclusions. Somali is an excellent example of a language that uses the sphinctering mechanism to mute, to modify, and to modulate acoustic sources coming through the glottis. Sphinctering and expanded pharynx are implemented phonetically as *constricted supraglottal cavity shape* versus *expanded supraglottal cavity shape* as well as by *trilling between the a-e structures and the epiglottal surface*. Vowels are also impacted by these changes as the posterior structures of the tongue retract into the lower pharynx, but are not to be taken as the primary exponent in this language the tongue root follows and enhances these cavity changes. At the same time there are strong connections between consonants and vocal register.

We know from our previous study (Edmondson, Esling, Harris, Li, and Lama 1999) that the Bai language of Yunnan Province, China has voice quality contrasts without vowel quality changes.

Our research findings have broad theoretical implications. Theoretically, we view the larynx and the supraglottal cavity as a tube containing three valves that can constrict in an incremental progression or completely close off the epilaryngeal tube through which the airstream passes. The inner most valve consists of the arytenoids and the true vocal folds; the middle valve is defined as the ventricular folds; the uppermost valve consists of the aryepiglottic folds and the epiglottis, including the sphinctering mechanism that controls their movements. The direct visual evidence from Somali and other languages in our research project supports our contention that pharyngeals are initially and primarily a function of the laryngeal sphincter, in which the aryepiglottic folds are the active articulator and the epiglottis is the passive articulator, rather than being simply the backmost lingual articulation. Evidence from Somali and other languages in our research supports our redefinition of glottal stop. A moderate glottal stop has completely adducted vocal folds at the vocal processes of the arytenoids, partially adducted ventricular folds, and a moderate sphinctering of the aryepiglottic folds. We have also observed that pharyngeals are not produced initially by the retraction of the tongue into the lower pharynx but primarily by constriction of the aryepiglottic laryngeal sphincter mechanism. Complete closure of the aryepiglottic laryngeal sphincter mechanism at the epiglottal place of articulation is an epiglottal stop.

In Somali a voiceless epiglottal stop is a complex sound containing a secondary glottal stop articulation with adduction of the ventricular folds. The primary articulation is the complete adduction of the aryepiglottic folds that rise up, move forward, and cover up the glottal opening making firm contact with the epiglottis as the tongue retracts. There is sometimes a voiced pharyngeal approximant offglide component which carries into the following vowel. It is also sometimes accompanied by weak trilling of the laminar surface of the epiglottis. Theoretically we view the glottal stop component of a voiceless epiglottal stop as of secondary rank order of stricture with the primary rank order of stricture being the complete closure between the aryepiglottic folds and the epiglottis. This, of course, has theoretical implications to the rank orders of strictures as described by Pike (1943) and Catford (1977). In Pike's rank order of strictures both degree of opening of the stricture and location of the stricture are important and the locations glottal, epiglottal, and pharyngeal are never primary locations. Catford's rank order of strictures is based solely on the degree of openness of the strictures. We follow most closely Catford's rank order of strictures, but with the inclusion of location of stricture, adapted from Pike, so that oral stops are primary over simultaneous glottal stop. And for us the glottal closure for glottal stop beneath the epiglottal stop is stricture which we view as secondarily contained within a primary stricture. In a Somali consonant like [q?], a epiglottalized uvular stop, we consider the oral stop primary, the glottal stop tertiary, and epiglottal stop secondary. In a plain epiglottal stop the glottal stop is secondary and the epiglottal stop is primary. More work needs to be done on rank order of strictures in general.

In Somali single intervocalic consonants tend to have a stricture difference and/or an articulatory strength difference from initial variants of the same phoneme. The articulatory strength differences are obvious in the differences between initial n- and l- and medial -n- and -l-. Stricture differences are obvious when the single (oral) stops b,d,g, and the single nasal (stop) m become fricatives intervocalically. Articulatory strength differences in consonants are not common in the literature but have been occasionally reported in such articles as Maddieson (1999:1965-1968). Articulatory strength contrasts, although rare in the literature, vary in different languages. The distinctions include variation in articulatory muscular tension, different degrees of place of articulation constriction, and different aerodynamic features such as degrees of aspiration. Palatograms and linguograms clearly show areas of strong articulation as more constricted with wider wipe-off areas. Acoustically duration appears to be longer in strong articulations than in weak articulations. More experimental studies of articulatory strength need to be done.

References

- Abercrombie, David. 1967. *Elements of General phonetics*. Edinburgh University Press.
- Andrzejewski, B. W. 1955. The problem of vowel representation in the Issaq dialect of Somali. *BSOAS* 17.3.567-80.
- Appleyard, David. 1990. The rôle of tone in some Cushitic languages", *York Papers in Linguistics*, 15.5-32.
- Armstrong, Liliás E. 1934. The phonetic structure of Somali. *Mitteilungen des Seminars für Orientalische Sprachen zu Berlin* No. 37 and reprinted 1964 Gregg Press Limited: Westmead, Franborough, Hants, England.
- Banti, G. 1988. Two Cushitic systems: Somali and Oromo nouns. In H. van der Hulst and N. Smith (eds.) *Autosegmental studies on pitch accent systems*. Dordrecht: Foris, pp 11-49.
- Berchem, Jôrg. 1992. *Referenzgrammatik des Somali*. Köln: Omimee Intercultural Publishers.
- Catford, J. C. 1977. *Fundamental problems in phonetics*. Edinburgh: University of Edinburgh Press
- Edmondson, Jerold A., John Esling, Jimmy G. Harris, Li Shaoni, and Lama Ziwo. 1999. The aryepiglottic folds and voice quality in the Yi and Bai language of Yunnan Province: laryngoscopic case studies. In *Chinese in Minzu Yuwen* 6.47-53 and in English in *Mon Khmer Studies* 31.1-21.
- Esling, John H, Barry F. Carlson, Jimmy G. Harris, Jerold A. Edmondson. 2003. The phonetic status of stops in the larynx. Submitted to *Language*.
- Handbook of the International Phonetic Association*. 1999. Cambridge University Press.
- Henderson, Eugenie. 1952. The main features of Cambodian pronunciation. *BSOAS* 19.1.1549-55.
- Hyman, Larry. 1981. Tonal accent in Somali". *Studies in African Linguistics* 12. 169–203.
- Le Gac, David. 2001. *Structure prosodique de la focalisation: le cas du somali et du français*. PhD dissertation, Université 7, Paris.
- Le Gac, David. 2002. Tonal alternations and prosodic structure in Somali. Paris 7.
- Lindau, Mona. 1975. Features for vowels. *UCLA Working Papers in Phonetics* 30.16-21.
- Lindau, Mona. 1976. Larynx height in Kwa. *UCLA Working Papers in Phonetics* 31.53-61.
- Lindqvist, J. 1969. Laryngeal mechanisms in speech. *QPST, STL, KRH*, Oct. 26-32.
- Orwin, Martin. 1994. *Aspects of Somali Phonology*. PhD Dissertation School of Oriental and African Studies, University of London.
- Orwin, Martin. 1993. Phonation in Somali phonology. In Mohamed Abdi Mohamed (ed.) *Anthropologie Somalienne (Proceedings of the 2ème Colloque des Etudes Somaliennes)*, pp.251-257
- Painter, Colin. 1978. Implosives, inherent pitch, tonogenesis and laryngeal mechanism. *Journal*

of Phonetics 6.249-74.

Pike, Kenneth L. 1943. Phonetics: a critical analysis of phonetic theory and a technique for the practical description of sounds. Ann Arbor: University of Michigan Press.

Saeed, John I. 1993. Somali reference grammar. Dunwoody Press. MD.

Saeed, John I. 1999. Somali (London Oriental and African Language 10).
Amsterdam/Philadelphia: John Benjamins Publishing Company.