Koffi, E. (2012). Intelligibility assessment and the acoustic vowel space: An instrumental phonetic account of the production of English lax vowels by Somali speakers. In. J. Levis & K. LeVelle (Eds.). *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference,* Sept. 2011. (pp. 216-232). Ames, IA: Iowa State University.

INTELLIGIBILITY ASSESSMENT AND THE ACOUSTIC VOWEL SPACE: AN INSTRUMENTAL PHONETIC ACCOUNT OF THE PRODUCTION OF ENGLISH LAX VOWELS BY SOMALI SPEAKERS

Ettien Koffi, Saint Cloud State University

Previous phonetic attempts to account for the intelligibility of L2 English vowels have relied exclusively on impressionistic acoustic approaches. In the impressionistic framework, native speakers (and sometimes nonnative speakers) are called upon to render comprehensibility and intelligibility judgments regarding speech samples that they hear and transcribe. In this paper a different approach is proposed whereby intelligibility assessments are based solely on the acoustic signals produced by nonnative speakers. This exploratory study deals with the acoustic vowel space of 10 Somali male participants. The first two formants of their lax vowels are measured and compared with those of General American English (GAE) in accordance with Peterson and Barney's landmark study of 1952. The comparisons make it possible to predict which GAE lax vowels are most likely to cause intelligibility challenges when these Somalis speak English.

INTRODUCTION

The propensity for vowels to contribute to regional variations is well documented in the sociolinguistic literature. Second language acquisition researchers also attribute aspects of unintelligibility to vowels, especially lax vowels (Flege et al., 1986: Munro & Derwing, 1995; 2008). In a comprehensive review of vowels systems, Crothers (1978) found that lax vowels are more infrequent (i.e., more marked) than tense vowels. These findings seem to validate the claims made by the proponents of the Contrastive Analysis Hypothesis and the Markedness Differential Hypothesis that lax vowels are troublesome for learners of English whose first language lacks them. Indeed, nearly all the "Note to Teachers" found in Celce-Murcia et al. (2010, pp. 117-22) have to do with the difficulties associated with lax vowels. Very little information exists in the second language phonology/phonetic literature as to whether or not speakers of languages that have a similar inventory of lax vowels as General American English (GAE) would have difficulties producing them. The goal of this paper is to provide some preliminary answers to this question. English and Somali are strikingly similar with respect to their phonemic inventory of simple vowels. It is commonly agreed that English has 11 phonemic monophthongs (Fromkin, Rodman & Hyams, 2011, p. 248) and Somali has 10 (Saeed, 1999, p. 11). Moreover, the two languages share almost the same number of lax vowels, namely, ɪ, ɛ, æ, ɔ, ʊ/ except that /ʌ/ is missing from the vocalic inventory of Somali. In this paper an instrumental phonetic analysis of F1 and F2 formants of lax vowels in Somali English (SoE) is done and plotted against those of GAE. This methodology serves as the basis for the inferences that are made in this paper about the intelligibility of lax vowels in SoE.

ARTICULATORY PHONETIC CLASSIFICATIONS OF VOWELS

First and foremost, it must be noted that linguists have had a hard time describing with accuracy the articulatory gestures involved in producing vowels. Fromkin and Rodman (1998, p. 235) recounts her frustration with vowels as follows:

> There have been many different schemes for describing vowel sounds. They may be described by articulatory features, as in classifying consonants. Many beginning students of phonetics find this method more difficult to apply to vowel articulations than to consonant articulations. In producing a [t] you can feel your tongue touch the alveolar ridge. When you make a [p] you can feel your two lips come together, or you can watch your lips move in a mirror. Because vowels are produced without any articulators touching or even coming close together, it is often difficult to figure out just what is happening. One of the authors of this book, at the beginning of her graduate work, almost gave up the idea of becoming a linguist because she could not understand what was meant by "front," "back," and "low" vowels.

The labels "front," "back," and "low" are not the only confusing terms that linguists use to describe vowels. Another vague descriptor that is often used is the label "tense." According to Fromkin and Rodman (1998), a tense vowel is one that is "produced with greater tension of the tongue muscles" (p. 239). This description appears to be simple and straightforward but Thomas (2011) doubts its usefulness. He contends that "the name tenseness is not especially helpful in understanding what's going on. It's based on the questionable notion that 'tense' vowels show more muscular tension rather than 'lax' vowels, but muscular tension won't help you a bit when you're trying to measure acoustic signals or conducting a perception experiment" (p. 147).

The vagueness of the labels tense vs. lax in describing some vowels has led linguists to look for a descriptively more accurate term. After trying various labels such as "narrow vs. wide," "primary vs. wide," "expanded vs. non-expanded," Ladefoged and Maddieson (1996, p. 300) note that over the past two decades, the term "ATR" (advanced tongue root) has gained wide acceptance among phoneticians and phonologists. The ATR label is deemed more acceptable than the traditional label "tense" because it depicts accurately the physiological mechanisms involved in producing vowels. Tiede's (1993, p. 114) MRI study of tongue advancement in English and Akan has shown that in producing tense vowels the tongue does indeed move forward a few more millimeters than in producing their lax counterparts, as shown in Table 1:

Table 1
Tongue Root Advancement Measurements

Vowel	Akan Root Advancement	English Root Advancement
[i]	22.97 mm	22.97 mm
[ɪ]	17.50 mm	21.88 mm
[e]	21.88 mm	21.88 mm
[ɛ]	19.69 mm	19.69 mm
[u]	32.81 mm	30.63 mm
[ʊ]	18.50 mm	18.59 mn

NOTE: Tiede (1993) did not provide measurements for [o], [ɔ], [æ], [ʌ], and [ɑ].

For the tense back vowel [u], the tongue advances twice as forward as in the back lax vowel [ʊ]. This forward thrust of the tongue root brings about secondary articulatory gestures such as enlargement of the pharyngeal area and the lowering of the larynx (Ladefoged & Maddieson, 1996). Even so, Ladefoged and Maddieson remark that the label [+ATR] vowel is not completely synonymous with the term tense vowel. They contend that "there is an overlap in the usage of these terms" but that "among back vowel pairs, there is no such parallel" (p. 300). The supporting evidence for this claim will be discussed in the section where GAE and SoE back vowels are compared. This caveat notwithstanding, I will follow contemporary usage and consider tense vowels to be the same as [+ATR] vowels; and lax vowels to be identical with [-ATR] vowels.

The Distribution of Lax Vowels Worldwide

The original study that led to the discovery of [±ATR] vowels was done on West African languages. For this reason, the literature has focused more extensively on [±ATR] vowels of languages from that region. However, Saeed (1999) notes that Somali, a language spoken in the Horn of Africa, also [± ATR] has vowels: "Somali has an interesting version of a five vowel system, involving two sets of five vowels. The two sets form five pairs of vowels where in each pair one vowel is pronounced with the tongue more forward than the other. Each pair of vowels can be differentiated by the phonetic feature advanced tongue root (ATR): thus we can label the Front series [+ATR] vowels and the Back series [-ATR]" (pp. 11-12). If we accept the proposition that tense vowels are synonymous with [+ATR] and lax vowels are the same as [-ATR] vowels, then according to Whitley S(2004), tense/[+ATR] and lax/[-ATR] have the following worldwide distribution. The frequency of occurrence of lax vowels worldwide listed in the third column provides additional support to the claim that lax vowels are troublesome to nonnative speakers of English. There seems to be a correlation between frequency of occurrence and learnability. The general impression is that lax vowels that occur infrequently are harder to acquire than those that occur more frequently.

Table 2

Worldwide Distribution of Lax Vowels

N0	Vowels	Feature	Worldwide Distribution
1.	[ɪ]	[-ATR] / Lax	17.4%
2.	[ɛ]	[-ATR] / Lax	38.5%
3.	[æ]	[-ATR] / Lax	13.6%
4.	[ɔ]	[-ATR] / Lax	32.5%
5.	[ʊ]	[-ATR] / Lax	15.5%
6.	[ʌ]	[-ATR] / Lax	1.3%

NOTE: There are disagreements among linguists regarding the status of some vowels as to whether they are tense or lax. The list of lax vowels presented here follows Ladefoged's (2006, p. 96) classification.

Now that some of the articulatory and distributional facts about tense and lax vowels are known, we will proceed with an acoustic phonetic investigation of GAE lax vowels as produced by Somali speakers.

PARTICIPANTS, TOKENS, AND EQUIPMENT

In summer 2011 I began an experimental study entitled "An Exploratory Study of Somalis' Pronunciation of English Vowels." It was approved by the Institutional Review Board (IRB) of Saint Cloud State University. The study replicates Peterson and Barney's classic 1952 study of GAE vowels. Twenty-two Somali participants were recruited to pronounce the words <heed>, <hid>, <hayed>, <head>, <had>, <hawed>, <hoed>, <hod>, <hood>, <who'd>, and <HUD> (an acronym for "Housing and Urban Development") and to read an expanded version of the George Mason University Speech Accent text (http://accent.gmu.edu/browse_language.php).

Seventeen Somali males and five females participated in the study. The participants were selected on the basis of their age of arrival (AOA) in the United States. AOA is important because in second language research, a correlation has been found to exist between the age of arrival/learning and accentedness in vowel production. For instance, Munro et al. (1996) found in their study of vowel production by 240 Italian immigrants to Canada that "the age of arrival in Canada of the Italian speakers examined in this study had an effect on the degree of perceived accent in all the English vowels studied here" (p. 326). The Somalis who participated in this study were all post-pubescent, that is, 15 or older prior to immigrating to the USA. Nearly all of them are in their late twenties or mid-thirties now. One participant is over fifty years old. Due to cultural and religious factors, availability of female Somali speakers of the same demographic profile was very limited. Also, the literacy level of the female participants was so low that some of them could not read the research tokens. In the end, the speech samples provided by the female participants were discarded in this analysis because some of the tokens were contaminated. The female research assistant who helped gather the data was heard in the background of one of the recordings coaching a participant. The speech samples from seven other Somali males could not be used in the present study because of technical difficulties.

All in all, the study is based on the speech samples provided by 10 participants because the data they provided is acceptable. The participants were all students at Saint Cloud State University at the time of the recording. Each person provided me with 33 tokens (11 words repeated three times). Collectively, they provided 330 tokens (33 x 10). Although the number of participants and tokens is far smaller than Peterson and Barney's data (76 participants and 1,520 recorded words), the number of participants and tokens for the present analysis is more than sufficient for an acoustic phonetic analysis. By way of comparison, Daniel Jones's classic study of British Cardinal Vowels was based on his own pronunciation (Thomas, 2011, p. 146). Furthermore, Thomas adds that "for studies in which speakers' entire vowel inventories are mapped, some authorities recommend measuring at least 20 tokens each. However, I've found that measuring as few as seven to ten is adequate if atypical or outlier tokens are excluded" (p. 159). Ladefoged and Maddieson's (1996, p. 283) analyses found in *The Sounds of the World's Languages* were often based on speech samples provided by no more than five people. Many phoneticians, including Ladefoged (2006), sometimes base their findings on their own speech samples or the speech sample provided by one or two family members (Thomas, 2011, p. 240; Yavaş, 2006, p. 103). Given the fairly large number of the tokens analyzed in this study, it can be claimed that the findings reported here are representative of SoE.

EQUIPMENT, DATA COLLECTION TECHNIQUES, AND PROCEDURES

All the recordings were done with an Olympus Ws-710 M Digital Voice Recorder. The recorded data were converted into .wav files with Switch Sound File Converter Plus, Version 4.09 by NCH Software. In some instances, all 11 words were recorded as a single sound file. The Wave Pad Sound Editor Masters Edition, Version 5.0 by NCH Software was used to create smaller files. The acoustic measurements and analyses were done with Praat, an online open source software designed for acoustic phonetic analyses.

The “word list speech style” also known as “citation-form speech style” was the elicitation technique used to collect the data (Thomas, 2011, p. 292). It is essentially the same style that Peterson and Barney used (1952, p. 175). This style has many advantages for an exploratory study such as this one. First, it helps eliminate the phenomenon of undershooting because when vowels are pronounced in a naturalistic fashion (casual speech style), they may not be enunciated fully. However, when vowels are read in a word list speech style, they are realized fully so that their onset and their offset are clearly visible on a spectrogram. An additional benefit associated with the word list speech style is that all the vowels are stressed. This makes it easier to measure their duration in various environments. Thomas (2011, pp. 138-139) summarizes the advantages of citation-form speech style as follows:

> This practice [casual speech] contrasts with that of most phoneticians, who usually favour citation-form speech. Citation-form speech, whether in word lists or in phrases, has two advantages of its own. For one thing, it yields more heavily stressed, longer tokens than most of those in conversational speech. As a result, the tokens in citation-form speech approximate their phonetic targets more closely than most tokens from conversational speech and hence show less coarticulation and undershoot. For another, the words elicited in citation-form speech can be controlled, which is why even sociolinguists often use reading passages and minimal pairs.

Finally, a word should be said about how the vowels occurring between /h/ and /d/ were measured. There are several methods for measuring vowel tokens (Thomas, 2011, pp. 41, 45, 138, 150-153). Ladefoged and Maddieson (1996, p. 287) and many phoneticians recommend collecting information about vowels at the midpoint so as to minimize the effects of the preceding or the following consonant on the vowel under investigation. However, in keeping with Peterson and Barney’s original methodology (1952, p. 176, Figure 2), F1, F2, and duration information were collected by measuring the whole vowel from the onset to the offset. Thomas (2011) does not see any serious harm in doing so for an exploratory study because, after all “any measurement of a formant is an estimate – it isn’t really possible to determine a formant value exactly” (p. 41). He only recommends that one sticks only with one method: “The key is to be consistent – choose the guidelines you want to follow and then stick with them” (p. 139). Figure 1 highlights the onset and offset areas of the vowels used in the analysis. They are indicated on the spectrogram by vertical lines.

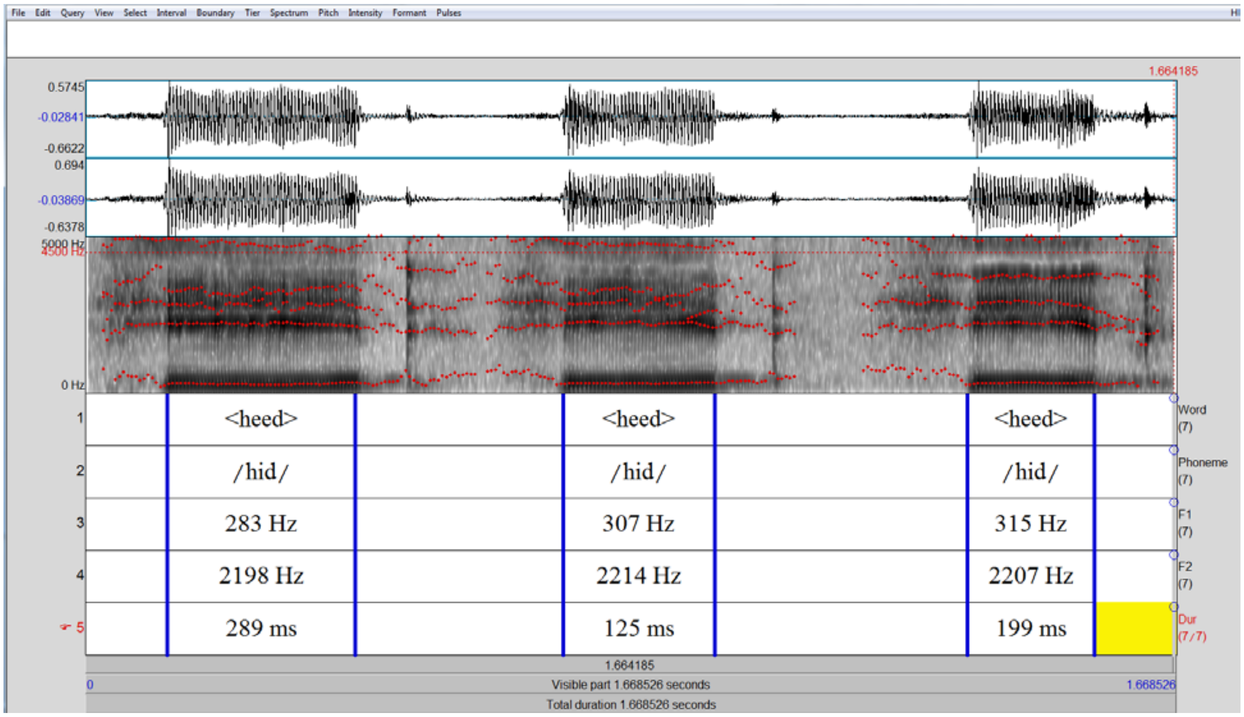

Figure 1. Sample spectrograph of vowels.

Data was collected from each repetition of the token. The numerical values were then averaged. This process was repeated for all 330 tokens. Only F1 and F2 formant information were gathered because most studies of English vowels concern themselves only with height and backness features. Ladefoged (2006, p. 272) and Ladefoged and Maddieson (1996, pp. 282-292) contend that height and backness parameters are acoustically the most relevant in describing vowels in world languages (see the section on feature hierarchy and intelligibility for the full rationale).

IMPRESSIONISTIC VS. INSTRUMENTAL ASSESSMENT OF INTELLIGIBILITY

For a little over two decades, Tracey Derwing and Murray Munro have devoted a considerable amount of professional energy to the intelligibility of the speech of nonnative speakers of English. The depth and breadth of their scholarship defies summarization lest their seminal work be oversimplified. Fortunately, my goal here is not to summarize their findings but to highlight the methodology that they use in assessing intelligibility. To do this, I will focus on two of their studies, one published in 1995 and the other in 2008. Determination of intelligibility and comprehensibility are made on the basis of how judges rate the speech of nonnative speakers. Munro and Derwing go to great lengths in many of their publications to note that the judges have passed a pure-tone hearing screening test. They also give ample sociometric information about the judges: native speakers versus nonnative speakers, linguistically trained versus linguistically untrained, their ages, etc. These judges (including Derwing and Munro) listened to speech samples provided by nonnative speakers (Mandarin Chinese speakers for the 1995 study, Mandarin and Slavic language speakers for the 2008 study). In the 1995 experiment, in addition to the selected tokens, the researchers had the participants produce short extemporaneous utterances. For the 2008 study, the participants embedded carefully selected 10 vowel tokens into the frame "Now, I say _________" where the target word had a CVC syllabic structure. In the 1995 study, a nine-point Likert scale (1995, p. 77) was used to assess the speech of nonnative speakers. The 2008 was a longitudinal study which sought to assess the intelligibility of vowels over six time frames. In the 2008 paper, the researchers reported on page 486 that "interjudge agreement was determined in terms of whether each production was assessed as the target vowel. Overall, three out of four listeners gave the same assessment 91% of the time. Complete agreement from all four listeners was reached on 67% of the time." In the 1995 paper, on pages 92-94, the authors discussed various issues that caused discrepancies among the judges in their assessment of intelligibility and its correlations (or lack of) with comprehensibility and accentedness. As one reads their other publications, (see Munro 2011, pp. 10-12 for example), one sees that they discuss linguistic and paralinguistic factors that may affect a judge's assessment of intelligibility. They often mention the following: rate of speech, speech clarity, voice quality, word choice, ambient noise, listeners' dispositions, among others.

The adjective "impressionistic" is an accurate description of the methodology used by Munro and Derwing to assess intelligibility and comprehensibility. I use the label "impressionistic" in relation to their methodology because they rely solely on the perception of selected judges to rate the intelligibility and comprehensibility of spoken tokens. There is nothing inherently pejorative about an impressionistic methodology.[1] Ladefoged (2003) confirms the usefulness of this methodology by saying that "early phoneticians did wonderful work relying simply on their ears" (p. 30). The methodology that I use in this paper to assess the intelligibility of SoE has been labeled "instrumental" because it does not rely on human agency to assess intelligibility but rather on acoustic devices and techniques. Ladefoged (2003) and many sociophoneticians recommend such a methodology because "instrumental phonetics has made it possible to document descriptions of languages more precisely" (p. 30). The use of this instrumental methodology presupposes that I view intelligibility first and foremost as an acoustic phonetic event. Thus, I redefine intelligibility in this paper as the acoustic phonetic cues that feed into word recognition and ultimately into semantic comprehension. This definition is in line with the distributed network approach discussed by Thomas (2011, p. 310).

This methodological stance does not in any way suggest that there is a simplistic correlation between the acoustic signals emitted by the talker and those perceived by the hearer. Any cursory look at the phonetic literature shows that the issue has been vigorously debated for fifty years or more. Johnson (2012) underscores this in a quote that he attributes to Cooper: "There are many questions about the relation between acoustic stimulus and auditory perception which cannot be answered merely by an inspection of spectrograms, no matter how numerous and varied these might be" (p. 123) To start with, the physiological structure of the ear makes a perfect correlation impossible (Johnson, 2012, pp. 83, 180-181). As a result, Johnson (2012) concludes that "acoustic analyses give only a rough approximation to the auditory

[1] There are two schools of thought in theoretical linguistic phonetics: the phonetics of speech production versus the phonetics of speech perception. Some of the leading theoreticians of the two schools are at odd with each other on methodological grounds (Johnson, 2012, p. 112). Production scholars focus on the acoustic vowel space to account for unintelligibility or confusion of speech sounds. Perception scholars emphasize the perceptual vowel space to do the same (Johnson, 2012, pp. 144-148). This paper highlights the insights that the acoustic vowel space brings to assessing the intelligibility of vowels. However, I do not discount the contributions made by perception scholars. My preference is simply a matter of my training. I'm more familiar with the instruments used to measure production than those used to measure perception. Ideally, production practitioners and perception practitioners should work collaboratively to fine tune intelligibility assessment in L2 acquisition.

representations that listeners use in identifying speech sounds" (p. 94). These caveats notwithstanding, phoneticians have also known for a very long time that sounds that are phonetically and acoustically closer are often very easily confused. Johnson (2012, p. 119-123) and others are now using "multidimensional scaling" computational techniques to visually represent the distance between sounds. So, in spite of the unresolved issues that keep theoreticians awake at night, comparing acoustic vowel spaces, as in Figure 4, can lead to important insights for teaching pronunciation to nonnative speakers of English or any language. An additional rationale for assessing intelligibility instrumentally is provided by Flege et al. 1986, p. 362):

> Perhaps some of these misidentifications [of vowels] were due to *talker* rather than listener variations. Vowels might be misidentified as the result of overlapping tongue positions for vowels adjacent in the phonetic space. Given the relatively small volume of the oral cavity and the large size of the tongue, the need to produce all 15 vowels of English with non-overlapping tongue configurations (or movement patterns) seems to represent an enormous control problem.

The "enormous control problem" mentioned in the quote increases tremendously for multilingual speakers because, somehow, they have to use the same tongue in the same restricted vowel space for all the languages that they speak. Somehow, they have to remember (subconsciously) that when they are speaking this or that language, they should allocate slightly different spaces to the vowels of their second, third, or fourth language. Failure to do so may result in unintelligible utterances.

Findings and Analyses

We will now investigate whether or not Somali speakers allocate the same vowel space as speakers of GAE do when producing English lax vowels. To find answers, a straightforward comparison is done between the formant values of GAE lax vowels and those found in SoE. Thomas (2011, p. 162) notes that this is a commonly used method:

> One common method that is sometimes used in variationist studies is to compare the vowel or vowels that exhibit the variation that's being studied against another vowel or vowels that are thought to be stable in the dialect or community in question. For example, a study might compare F1 and F2 values of some vowel against those of an [i] vowel, such as the FLEECE vowel in English. This method is easy to use and requires no complicated mathematical transformations: all you need is a ratio of the formant value of the stable vowel against that of the varying vowel.

For this study, I take the F1 and F2 formant values given by Peterson and Barney (1952, p. 183) to be representative of GAE vowels. It is against these formant values that SoE vowels are compared and contrasted, as shown in Table 3.

Table 3
Mean Formant Values in GAE and SoE

N0	Vowels	US Male F1	Somali Male F1	US Male F2	Somali Male F2
1.	<hid> [ɪ]	390	549	1,990	1,886
2.	<head> [ɛ]	530	570	1,840	1,810
3.	<had> [æ]	660	678	1,720	1,674
4.	<hawd> [ɔ]	570	609	840	1,339
5.	<hood> [ʊ]	440	436	1,020	1,396
6.	<hud> [ʌ]	640	629	1,190	1,532

The methodology adopted in this study consists in comparing lax vowels of the same types in GAE and SoE. In comparing F1 and F2 formant values among different dialects/languages, it is good to keep Baart's (2010, p. 67) interpretive framework in mind:

> A frequency difference of, say, 200 Hz is much more noticeable for people (and perceived as a much greater difference if lower frequencies are involved (as in the difference between 200 and 400 Hz) than if higher frequencies are involved (as in the difference between 2000 and 2200 Hz).

The 200 Hz frequency that Baart uses is just for illustrative purposes. Basic frequency calculations must be performed for each language under investigation to gauge frequency differences that matter. For this study, I have decided to use Peterson and Barney's (1952) benchmark acoustic measurements of GAE. The median frequency range is 135 Hz for F1, and 170 Hz for F2, as displayed in Table 4:

Table 4
Frequency Distance between GAE Vowels

N0	Vowel Pairs by Natural Class	F1 Difference	F2 Difference
1.	[ɪ] vs. [ɛ]	140 Hz	150 Hz
2.	[ɪ] vs. [æ]	270 Hz	270 Hz
3.	[ɛ] vs. [æ]	130 Hz	120 Hz
4.	[ʊ] vs. [ɔ]	130 Hz	180 Hz
5.	[ʊ] vs. [ʌ]	200 Hz	170 Hz
6.	[ɔ] vs. [ʌ]	70 Hz	170 Hz

Median frequency ranges are often displayed in vowel charts with ellipses drawn with solid lines to show areas on the chart where a specific vowel is pronounced slightly differently by different people. Sometimes, no ellipse is drawn but the various realizations of the vowel under consideration are scattered on the chart, see Ladefoged (2003, p. 129-30, figures 5.17, 5.18) for an example of both methods. In evaluating vowel intelligibility, it is assumed that if the F1 and F2 frequencies between GAE and SoE vowel of the same type are lower or equal to 135 Hz and 170 Hz respectively, then the SoE vowel is intelligible. The reason for this is because the difference in frequency falls within the median range. However, if F1 and F2 frequencies are in excess of 135 Hz or 170 Hz, then the SoE vowel under consideration is moderately to strongly accented.[2] It is hard to state conclusively that a vowel is unintelligible just by looking at frequency differences. However, when frequency differences are plotted in the same vowel quadrant, a clearer picture of which vowel(s) may or may not be intelligible emerges.

Height Comparison between GAE and SoE Vowels

Comparisons of GAE and SoE vowel pairs of the same type yield the following frequency differences:

Table 5
F1 Distance between GAE and SoE Vowels

N0	Minimal Pairs	F1 Frequency	F1 Difference
1.	English [ɪ] vs. Somali [ɪ]	390-549	159 Hz
2.	English [ɛ] vs. Somali [ɛ]	530-570	40 Hz
3.	English [æ] vs. Somali [æ]	660-678	18 Hz
4.	English [ɔ] vs. Somali [ɔ]	570-609	39 Hz
5.	English [ʊ] vs. Somali [ʊ]	440-436	4 Hz
6.	English [ʌ] vs. Somali [ʌ]	640-629	11 Hz

[2] According to F1 formant data provided by Koon (2006, pp. 150, 152), in Central Minnesota /æ/ is strongly accented because it raises from 644 Hz to 836 Hz when produced by men and women before the voiced velar /g/. When compared with Peterson and Barney's data, /æ/ before /g/ is raised by 176 Hz.

Given the information in Table 5, it appears that the only GAE lax vowel that is produced in an accented fashion is [ɪ] because the frequency difference between GAE and SoE exceeds 135 Hz. The Somali pronunciation of [æ], [ʊ], and [ʌ] do not appear to be accented because the frequency differences between them are subsonic, that is, they are below 20 Hz. Ferrand (2007, p.34) and Ladefoged (1996, p. 21) note that human ears cannot detect frequencies below 20 Hz. It is therefore doubtful that a GAE hearer can detect any difference if a Somali speaker repeats the words /hæd/, /hʊd/, and /hʌd/ directly after a GAE speaker. With the exception of /ɪ/, SoE and GAE and lax vowels are similar with regard to F1. The Somali vowel [ɪ] is much lower than its GAE counterpart because, as is well known in acoustic phonetic studies, there is an inverse relationship between vowel height and F1 frequency.

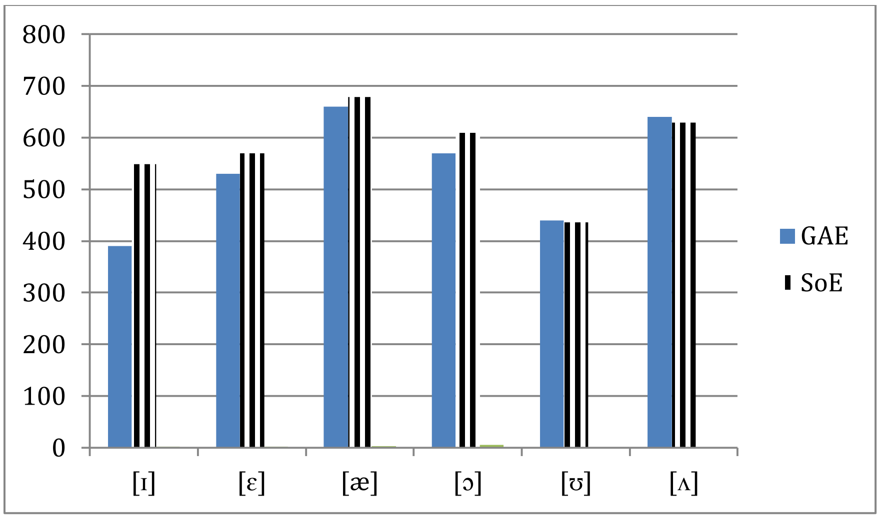

Figure 2. F1 Frequency Graphs.

Backness Comparison between GAE and SoE Vowels

The median F2 frequency range for English lax vowels is 170 Hz. This means that if a vowel is produced with an acoustic energy equal or less than 170 Hz, that vowel is not considered accented. Accordingly, the vowels [ɔ], [ʊ], and [ʌ] may tentatively be thought of as being accented in SoE but the vowels [ɪ], [æ] and [ɛ] are not, as shown in Table 6. However, the word "tentatively" is the key word here because later discussions will show that F2 has only a marginal effect on vowel perception.

Table 6
F2 Distance between GAE and SoE Vowels

N0	Minimal Pairs	F2 Frequency	Difference
1.	English [ɪ] vs. Somali [ɪ]	1,990-1,886	104 Hz
2.	English [ɛ] vs. Somali [ɛ]	1,840-1,810	30 Hz
3.	English [æ] vs. Somali [æ]	1,720-1,674	46 Hz
4.	English [ɔ] vs. Somali [ɔ]	840-1,339	499 Hz
5.	English [ʊ] vs. Somali [ʊ]	1,020-1,396	376 Hz
6.	English [ʌ] vs. Somali [ʌ]	1,190-1,532	432 Hz

It is worth noting that the F2 frequencies in Table 6 confirm what Ladefoged and Maddieson (1996, p. 300) have said about the difference in acoustic behavior of back and front vowels. They found that [-ATR] front vowels across a wide range of languages behave similarly acoustically but [-ATR] back vowels do not. We see this clearly in the bar graphs in Figure 3. All three front vowels in GAE and SoE are similar but there is a notable difference in the acoustic behavior of back vowels.

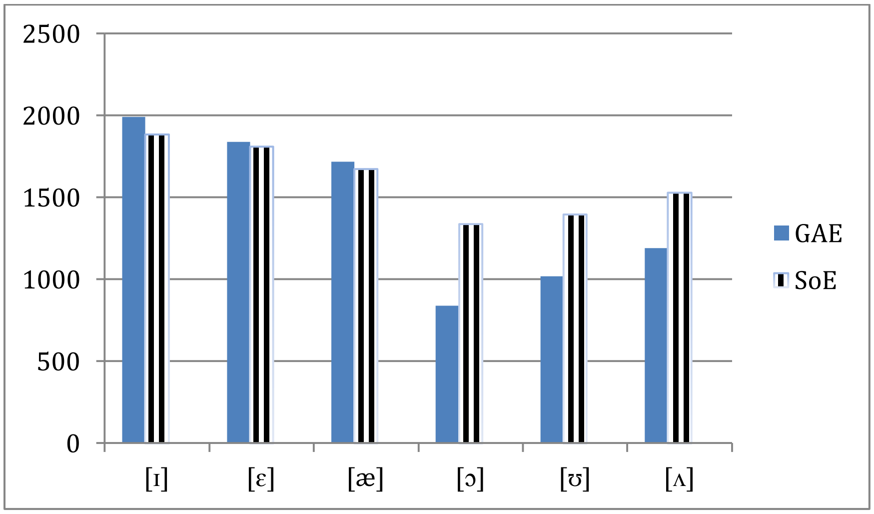

Figure 3. F2 Frequency Graphs.

The comparison between the lax back vowels of GAE and SoE show that English back vowels are fronted, that is, they are pronounced towards the front of the mouth by Somali speakers. Additional insights between the two vocalic systems can be gained by plotting GAE and SoE vowels in the same vowel quadrant (Figure 4).

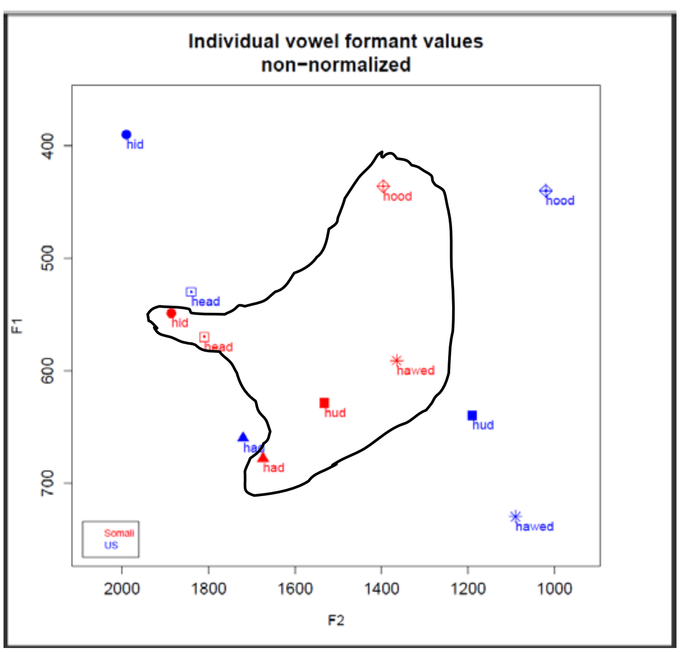

Figure 4. Comparative Vowel Quadrant.

A line has been drawn around SoE vowels to make them stand out. Figure 4 reveals two essential differences between SoE and GAE vowels. First, the space occupied by SoE vowels is far more restricted than that of GAE. Secondly, all GAE back vowels are fronted in SoE. These observations can shed some light on SoE vowels that are more or less likely to be intelligible. Ladefoged (2001) recommends plotting the vowels of different dialects of the same language in the same vowel quadrant so as to highlight dialectal similarities and differences. He contends that "vowel charts provide an excellent way of comparing different dialects of a language" (p. 43). He provides an additional rationale for his preference by saying that "this kind of plot arranges vowels in a similar way to the vowels in the IPA vowel chart. The formant frequencies are spaced in accordance with the Bark scale, a measure of auditory similarity, so that the distance between any two sounds reflects how far apart they sound" (Ladefoged 2003, p. 130). This last piece of information is important because it shows that a display such as the one in Figure 4 simulates as closely as possible how a human being who listens to SoE and GAE vowels perceives them acoustically.

Feature Hierarchy, Formant Hierarchy, and Intelligibility

The comparisons done in the previous sections suggest that the Somali pronunciation of vowels [ɪ], [ɔ], [ʊ], [ʌ] may pose intelligibility challenges to GAE hearers but [ɛ] and [æ] would not. However, this extrapolation is unwarranted because it fails to take into account feature hierarchy. Phonologists claim that in processing vowel sounds, some features are more prominent than others. Ladefoged (2006, pp. 271-272) lists five main features for vowels and places them in the following hierarchical scale: Height > Backness > Rounding > Tongue Root > Rhotic. Ladefoged and Maddieson (1996) rank height as the most important feature because, they argue, "all languages have some variations in vowel quality that indicate contrast in the vowel height dimension" (p. 286). With respect to the features backness and rounding, they note that "the languages of the world make much more limited use of the front-back and rounded-unrounded dimensions" to distinguish between utterances (p. 290). Furthermore, Fischer-Jørgensen (1985, p. 93) adds that the dimension front-back is more complicated and has given problems to phonologists because very few languages rely on this feature in formulating phonological rules. So for phonologists, the most salient feature for vowels is height.

Phonetic experimentations have confirmed that the intuition of phonologists about feature hierarchy is acoustically verified. Even though vowel quality encompasses many formants, only the first three (F1 for height, F2 for backness, and F3 for lip rounding) are acoustically relevant in many instances. Moreover, in many acoustic phonetic studies, very little mention is made of F3 because "[it] has very little function in distinguishing the vowels shown" (Ladefoged 2001, p. 46). As for F2, nobody doubts its relevance in the acoustics of vowels. Yet, its role in accounting for the perception of intelligibility is rather marginal and limited. Sociophonetic studies suggest that dialectal changes involving F1 are more noticeable and more pervasive than those involving F2. For instance, nearly all the examples of the Northern Cities dialect shift have more to do with variations in F1 than variations in F2. The examples of /bæd/ pronounced as [bɛd] is a case of vowel raising (Ladefoged 2001, p. 45, Figure 7). The phenomenon dubbed "Canadian Raising" by Fromkin et al. (2011, p. 312) also involves variations in F1. In any event, Ladefoged (2006) provides acoustic data to explain why F1 plays a more dominant role than F2. He writes, "As a further refinement, because the second formant is not as prominent as the first formant (which, on average has 80% of the energy in a vowel), the second formant scale is not as expanded as the first formant scale" (p. 188). Since F1 has 80% of the energy in the vowel, it is clearly the most important formant. Consequently, it plays the most salient role in the perception and assessment of the intelligibility of vowels.

Using Formant Hierarchy to Account of Intelligibility in SoE

It has already been noted that [ɪ] in SoE is considerably lower than [ɪ] in GAE. Figure 4 reinforces this drastic difference visually. In fact, [ɪ] in SoE is pronounced so low that it virtually occupies the same articulation space as the [ɛ] of GAE. There is only a 19 Hz difference between the Somali [ɪ] and the GAE [ɛ]. Acoustically speaking, this difference is subsonic, that is, it is so infinitesimal that most listeners are unable to distinguish between the two. As a result, the SoE pronunciation of a word such as [hɪd] may sound like [hɛd] to a GAE listener. GAE speakers who have informally[3] listened to the audio files have not been able to tell <hid> apart from <head>, and vice versa. It is not unusual for people to confuse these two vowels. Peterson and Barney (1952, p. 179) reports that about 7% of the occurrences of these two vowels in their 1,520 tokens were misidentified by native speakers listening to other native speakers. However, the occasional confusions between [ɪ] and [ɛ] among GAE speakers are nothing compared with the situation that occurs in listening to SoE. Since both vowels have overlapping spaces, the likelihood of unintelligibility is fairly high. If [ɪ] and [ɛ] are not clearly differentiated in pronunciation, unintelligibility increases tremendously because, according to Faircloth and Faircloth (1973, p. 18), [ɪ] and [ɛ] occur very frequently in GAE. The vowel [ɪ] ranks first and [ɛ] ranks sixth in GAE. Moreover, the pair [ɪ] and [ɛ], has according to Celce-Murcia et al. (2010, p. 160) has a functional load of 43%.

The spacing of SoE lax vowels also shows that the Somali pronunciation of [ʌ] and [ɔ] can lead to unintelligibility in some instances. There is only a 21 Hz difference between the two sounds. Even though a frequency of 21 Hz can be detected by human beings, these two sounds are acoustically so close that unless a person pays very close attention, they might mistake one sound for another. This means that a GAE speaker hearing a Somali person say the words <cut> and <caught> could very easily mistake them for <caught> and <cut> and vice versa.[4] In fact, Munro et al. (1996, p. 326) found that [ɝ], in all likelihood the Canadian counterpart of [ʌr], was the most poorly produced vowel by Italian immigrants.

Pedagogical Implications

Researchers who work within L2 phonology frameworks such as the Contrastive Analysis Hypothesis and the Markedness Differential Theory may be misled in believing that if two languages have the same sets of vocalic segments, they may be transferred positively. This assumption has led to erroneous conclusions such as the following:

> Somali and English share a number of the same vowel phonemes and diphthongs. Because of this, problems with pronunciation will not likely come because a student can't produce the vowel in question (Lindsay, 2006, p. 47).

---

[3] The word "informally" is used here because I have not done a formal perception study of Somali vowels.

[4] When I was writing this paper, I attended a meeting of the Immigrant Research Group at Saint Cloud State University. An African faculty member who is involved in community planning used the phrase "housing bust." Another faculty member, a native speaker of Spanish, heard "housing bossed" and asked aloud "who is this housing boss?" The production difficulties associated with [ʌ] and [ɔ] are widespread in many forms of African Englishes.

However, the instrumental acoustic analysis done in this study does not support such a conclusion. Given the intelligibility challenges posed by the lax vowels [ɪ, ɛ, ɔ, ʌ], what pedagogical strategies can teachers use to improve the production of vowels by post-pubescent Somali students? Fischer-Jøgensen (1985, pp. 93, 95) suggests an articulatory phonetic regimen based on manipulating jaw movements. He describes it as follows:

> Moreover, a number of phonetic and phonological rules and developments involving height are better understood when described in articulatory terms. … I suppose production plays a role in this connection. As demonstrated by Lindblom and Sunberg, the simplest way to produce differences in vowel height is by raising and lowering the mandible, and it may be relatively easy to control this movement. First, the proprioceptive sensitivity seems to be more developed for jaw movements than for advancing or retracting the tongue. This may have something to do with the fact that jaw opening and closing is used for other biological purposes, for example, eating. Second, jaw movement is visible. (It may happen that a student starting a phonetics course believes that he produces an [e] by retracting his tongue, but he will not maintain that he produces [a] by closing his mouth.

The preeminence of the first formant over the second is reaffirmed again here. According to Fischer-Jøgensen, it is easier to translate height information into pedagogical practice than backness movement because most people cannot feel their tongue advancing or retracting a few millimeters from its position at rest (see Table 1). In their longitudinal study of L2 vowel acquisition, Munro and Derwing (2008) report that intelligibility of English vowels can improve over time. It is my contention that targeted instruction based on findings such as the ones in this study can hasten and improve intelligibility.

SUMMARY

The findings presented in this paper are not intended to predict in absolute terms that every occurrence of [ɪ] and [ɛ], or [ɔ] and [ʌ] will lead to unintelligibility. There are a number of factors that mitigate an erroneous perception of these vowels even if they are acoustically mispronounced. For instance, if the discourse context is sufficiently rich, sentential clues will alleviate the phonetic processing load on the hearer. Byrd and Mintz (2010) explain how this is possible even in the face of a heavily accented speech: "A speech sound will be perceived as intended by the speaker, even if the speaker introduces some articulatory, and hence acoustic variation from production to production" (p. 143). A rich syntactic context enables reliance on top-down processing strategies rather than bottom-up strategies. The differences in the acoustic of GAE and SoE vowels provide us with quantifiable data as to why when most post-pubescent Somali immigrants speak English, they may be misunderstood more often than not if other discursive strategies fail. Last but not least, it should be borne in mind that the data in this study deal with averages. As a result, it does not take into account intraspeaker variations. However, this caveat is not an admission of methodological weakness but rather an acknowledgment that Somali speakers, just like other human beings, have their own particular speaking styles and speech idiosyncrasies that fall outside the reach of instrumental acoustic phonetics.

ABOUT THE AUTHOR

Ettien Koffi is Professor of Linguistics at Saint Cloud State University, Minnesota, USA. He teaches the linguistic courses in the TESOL/Applied Linguistics M.A. program in the English Department. He has written three linguistic books and is the author of many peer-reviewed articles on various topics in linguistics. His primary area of specialization is at the interface between acoustic phonetics and phonology. He also has extensive experience in emergent orthographies and in the acoustic phonetic and phonological description of dialect variation.

REFERENCES

Baart, J. (2010). *A field manual of acoustic phonetics.* Dallas, Texas: SIL International.

Byrd, D. & Mintz, T. (2010). *Discovering speech, words, and mind*. Malden, MA: Wiley-Blackwell.

Celce-Murcia, M., Brinton, D. & Goodwin, J., with Griner, B. (2010). *Teaching pronunciation: A reference for teachers of English to speakers of other languages, 2nd Ed.*, New York: Cambridge University Press.

Crothers, J. (1978). Typology and universals of vowel systems. In J. Greenberg (Ed.) *The universals of human language: Volume 2, phonology,* pp. 93-152. Stanford, CA: Stanford University Press.

Faircloth, S. & Faircloth, M. (1973). *Phonetic science: A program of instruction*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Ferrand, C. (2007). *Speech science: An integrated approach to theory and clinical practice, 2nd ed*. New York: Pearson Allyn And Bacon.

Fisher-Jørgensen E. (1985). Some basic vowel features, their articulatory correlates, and their explanatory power in phonology. In V. Fromkin (ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, pp. 79-99. New York: Academic Press.

Flege, J., Fletcher, S., McCutcheon, M. & Smith, S. (1986). The physiological specification of American English. *Language and Speech, 29* (4), 361-88

Fromkin, V. & Rodman R. (1998). *An introduction to language, 6th ed.*, Orlando, Florida: Harcourt Brace College Publishers.

Fromkin, V., Rodman R. & Hyams, N. (2011). *An introduction to language*, 9th edition. New York, NY: Wadsworth Cengage Learning.

Johnson, K. (2012). *Acoustic and auditory phonetics, 3rd ed.*, Malden, MA:Wiley-Blackwell.

Koon, D. (2006). *Acoustic analysis of NAE vowels in central Minnesota*. MA Thesis. Saint Cloud State University: St. Cloud, MN.

Ladefoged, P. (1996). *Elements of Acoustic Phonetics*, 2nd edition. Chicago: University of Chicago Press.

Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of language.* Malden, MA: Blackwell Publishers.

Ladefoged, P. (2003). *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Malden, MA: Blackwell Publishers.

Ladefoged, P. (2006). *A course in phonetics*, 5th edition. New York: Thomson-Wadsworth.

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages*. Malden, MA: Blackwell Publishers.

Lindsey, M. (2006). *An ELL teacher's guide to Somali-English interlingual phonology*. MA Thesis. Saint Cloud State University: St. Cloud, MN.

Munro, M. (2011). Intelligibility: Buzzword or buzzworthy. In J. Levis & K. LeVelle (Eds). *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*, Sept. 2010, 7-16, Ames, IA: Iowa State University.

Munro, M. & Derwing, T. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning 45*(1), 73-97.

Munro, M.J. & Derwing, T.M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning, 58*, 479-502.

Munro, M., Flege J. & Mackay, I. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics, 17, 313-334*.

Peterson, G. & Barney, L. (1952). Control methods in a study of the vowels. *The Journal of the Acoustical Society of America, 24* (2), 176-84.

Saeed, J. (1999). *Somali*. Philadelphia, PA: John Benjamins Publishing Company.

Tiede, M. (1993). An MRI-based study of the pharyngeal volume contrast in Akan. *Haskins Laboratories status report on speech research*, SR-113, 107-30.

Thomas, E. (2010). *Sociophonetics: An introduction*. New York, NY: Palgrave-Macmillan.

Whitley, S. (2004). *The Phonicon: A Multimedia Digital Encyclopedia of the Sounds of the World's Languages*. CD Rom. Dallas, Texas: SIL International.

Yavaş, M. (2006). *Applied English phonology*. Malden, MA: Blackwell Publishing.