Linguistic Portfolios

Volume 7 Article 6

2018

Sibling Acoustics: Phonetic Measurements of the Vowels Produced by Three Brothers Who "Sound Alike"

Ettien Koffi
St. Cloud State University, enkoffi@stcloudstate.edu

James Lyons
jlyonsesl@gmail.com

SIBLING ACOUSTICS: PHONETIC MEASURMENTS OF THE VOWELS PRODUCED BY THREE BROTHERS WHO "SOUND ALIKE"

ETTIEN KOFFI AND JAMES LYONS[1]

ABSTRACT

This paper seeks to apply an acoustic phonetic methodology to measure and account for the impressionistic assessment of the speech of three brothers whose relatives, friends, and acquaintances say "they sound exactly alike." The current investigation is limited to the pronunciation of the 11 phonemic monophthong vowels in American English. The psychoacoustic instruments of masking and Just Noticeable Difference (JND) are used to verify if the brothers "sound alike" acoustically. The correlates of the vowels that are investigated in this study are F0, F1, F2, F3, intensity, and duration.

1.0 Introduction

For the most part, acoustic phonetic research has focused on the variability of acoustic correlates between speakers. This is highlighted in a recent article by Hazan (2017:36) who observes correctly that "the acoustic patterns that cue phonetic distinctions vary from talker to talker." A few lines later, she states that variability is found in the speech of the same talker, "variability in this complex code arises because a given talker will not produce the acoustic patterns in an identical fashion even when uttering the same word on different occasions." A reasonable person cannot deny that between-talker and within-talker variability is an acoustic phonetic norm. Yet, any reasonable person cannot also deny that on occasion people, especially siblings of the same gender, "do sound exactly like each other." Acoustic phonetics has spent a lot of resources investigating variability. In this paper, we give some consideration to "sounding alike." We investigate it in four steps. First, we give some background information on how this topic came about. Secondly, we discuss the methodology. Thirdly, we describe the interpretive framework. Finally, we compare and contrast the acoustic measurements of the vowels produced by the three brothers to determine if they really "sound alike" as has been claimed by their parents, relatives, friends, and acquaintances.

2.0 Experiential Classroom Acoustic Phonetic Research

The phrase "experiential acoustics" is borrowed from Heller (2013) and is contrasted with "experimental acoustics" to describe the kind of acoustic phonetic research the first author engages in with his students. He has designed a unique methodology to teach acoustic phonetics using the students' own speech data. At the beginning of the semester, well before students embark on any acoustic measurements, he has them record themselves reading five different texts. With these five recordings, students are able to analyze and measure all the segments and suprasegments of English and their various acoustic correlates. Throughout the semester, the students complete 13 graded assignments that range from vowel initial glottalization to the acoustic correlates of lexical stress. The focus of this paper is on the vowel project. The students produce the 11 phonemic "simple" vowels of English in their citation forms. Each vowel is repeated three times. They

[1]**Authorship responsibilities:** This paper grew out of an acoustic phonetic course that the first author taught and that the second author took. James Lyons is recognized as the second author to the extent that he did all the recordings and measurements. The use of the data and the argumentations for verifying that the brothers sound acoustically alike are made by the first author and assumes full responsibility for any erroneous interpretation of acoustic data.

create their own acoustic vowel spaces which they compare and contrast with the one in Peterson and Barney (1952:183) or Hillenbrand et al. (1995:3103). The pedagogical assumptions behind this "experiential" approach is that students are more motivated to study acoustic phonetics if they can apply it to themselves directly. This approach has given opportunities to students to explore various topics such as the one described in this paper.[2]

2.1 Equipment and Methodology

The three brothers were recorded wearing a Logitech G230 Stereo Gaming Headset. The recordings were made on an HP EliteBook with an Intel Core i5 laptop computer. They recorded themselves reading the following words:

1. Heed
2. Hid
3. Hayed
4. Head
5. Had
6. Hod
7. Hawed
8. Hoed
9. Hood
10. Who'd
11. Hud

Measurements were taken of the nucleus. Tiers were created to extract only pertinent vowel information. The measurements do not include the consonants /h/ and /d/, as shown in the annotated figure below:

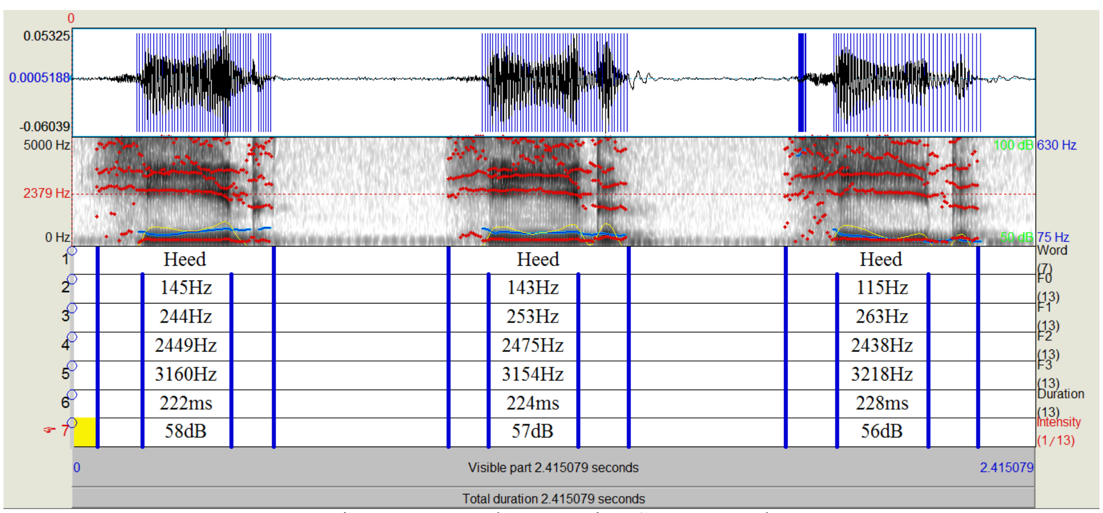

Figure 1: Vowel Annotation Spectrograph

[2] This "experiential" approach has been very successful judging by the employment and scholarship opportunities it has afforded my students. Some have been hired by software companies working on automatic speech recognition, some work for state and federal governments in forensics, other apply their acquired skills to teaching English as Second/Foreign language. Many have earned scholarships or assistantship for Ph.D. programs in linguistics with specializations in phonetics or phonology.

Vowels are chosen for this project because speakers of the same language differ more in their pronunciation of vowels than in their pronunciation of consonants. Moreover, both Peterson and Barney (1952) and Hillenbrand et al. (1995) have used the same list in their seminal works on General American English and US Midwest English. The acoustic correlates investigated are: F0, F1, F2, F3, Duration, and Intensity.[3]

3.0 Acoustic Phonetic Account of "Sounding Alike"

At its core, speech intelligibility research highlights the acoustic cues that are robust for intelligibility. For this reason, experts often look at the interplay of cues between the frequency intensity, and temporal domains. The correlates in these three domains are investigated in the paper to see if they can shed some light on why hearers perceive the three brothers as sounding alike. The correlates in frequency domain deserve the lion's share of attention because they are multifaceted. To understand better how they may be used to measure "sounding alike", we need to make a short detour into Critical Band Theory (CBT).

In the 1940s, Fletcher, a physicist, demonstrated mathematically that the human ear compartmentalizes the audibility range, 20 to 20,000 Hz, into frequency bandwidths known as "critical bands." Another physicist, von Békésy, demonstrated clinically that Fletcher's critical bands were grounded in physiological reality. This discovery earned the latter the Nobel Prize in 1961 in Physiology and Medicine. Their findings and subsequent refinements have shown that "the 1/3- octave frequency response is one of the best indicators of subjective frequency balance as it approximates the critical bandwidth accuracy of our hearing," (Everest and Pohlmann 2015:529). As a result, this system has been adopted nationally and internationally (Pope 1998:1346) for measuring and accounting for most auditory-perceptual phenomena. It is used in manufacturing audio products, sound level meters, and a wide variety of hearing devices. In the aural perception of vowels, the frequency bands ranging from 60 to 3,000 Hz are considered the most relevant. It is labeled "Speech Banana" in Kent and Read's (2002:187) and compartmentalized into four formant bands: F0 goes from 60 to 250 Hz, F1 from 250 to about 1000 Hz, F2 from 1000 to 2000 Hz, and F3 from 2000 to 3000 Hz. In each of these frequency domains, there are acoustic thresholds that can be used to determine if in fact the three brothers sound alike. Additionally, intensity and duration thresholds can be used to measure "sounding alike." In the technical literature, these thresholds are known as Just Noticeable Difference (JND). All in all, five JNDs (F0, F1, F2, F3, intensity, and duration[4]) are used verifying if the brothers sound acoustically alike.

3.1 Verification of "Sounding Alike" on the F0 Bandwidth

Psychoacoustic experiments have shown that in the F0 frequency band, people can detect variations as small as 0.30%. Young (2011:609) explains it as follows:

[3] Because our interest is in speech intelligibility, and not in speaker identification, no effort was made to take measurements of F4 and F5. Reetz and Jongman (2009:184) note that these two formants are good in revealing idiosyncratic characteristics about speakers.

[4] The JND thresholds discussed in this paper apply only to male speakers. Separate thresholds apply to female speakers. Since the participants in this study are all males, we focus only on JNDs that are relevant to males.

> The perception of frequency is called pitch. Most of us have excellent relative pitch, which means that we can tell whether one sound has a different pitch from another. Typically, we can discriminate between two sounds if their frequencies differ by 0.3% or more. For example, 500 and 501.5 Hz are noticeably different. Pitch perception is directly related to frequency and is not directly affected by other physical quantities such as intensity.

However, to make calculations easier, Stevens (2000:228), Lehiste (1970:64) and many phoneticians have rounded the 0.30% threshold to 1 Hz. In other words, if the mean F0 distance of the vowels produced by the three brothers is <1 Hz, this is proof that they sound acoustically similar.

3.2 Verification of "Sounding Alike" on the F1 Bandwidth

F1 correlates with vowel height, i.e., mouth aperture. The F1 value of high vowels is usually less than <300 Hz, but the F1 value of low vowels is usually >600 Hz. The frequency of mid vowels is between 400 to 500 Hz.[5] On the F1 frequency band, vowels produced by the speakers of the same language/dialect can vary as much as 50 Hz (Ladefoged 2003:128, Kent and Read 2002:110). For this reason, if the F1 distance between the three brothers is 50 Hz, it does not necessarily mean that they sound alike. However, if the distance between their vowels is ≤ 20 Hz, this is convincing proof that they sound alike because human ears cannot perceive frequency signals below this threshold on the F1 frequency band. F1 plays an important role in measuring "sounding alike" acoustically because, according to Kent and Read (2002:33):

> [It] is typically the most intense formant, largely because of the interaction with the amplitude of the other formants. One way of thinking about this is to say that F1 rides on the low-frequency tails of the other formant curves, so that F1 is boosted in amplitude relative the other formants. Loudness judgments of speech tend to be highly correlated with the amplitude of F1, which is not surprising given that this formant tends to be the strongest.

This view is also shared by Ladefoged and Johnson (2015:207) who state that F1 alone accounts for 80% of the acoustic energy found in vowels. In other words, the speech of the brothers will be described as sounding acoustically alike if the F1 distances between their vowels are between 20 and 30 Hz.

3.3 Verification of "Sounding Alike" on the F2 Bandwidth

For vowels, F2 correlates with the horizontal movements of the tongue. In male speech, an F2 value of ≥ 2000 Hz corresponds to front vowels, a value between 1,600 and 1,400 Hz corresponds to central vowels, while frequencies ≤1300 Hz correlate with back vowels. The JND for intelligibility on the F2 bandwidth is ≤ 200 Hz. This means that if the F2 of the vowels produced by the brothers is less than this threshold, they sound alike. In fact, Kent and Read (2002:111) indicate that "The F2 frequency is more sensitive to dialectal and idiolectal variation…" In other words, the smaller the F2 distance between the vowels produced by the three brothers, the more acoustically similar they sound.

---

[5] The values and all other formant values are based male speakers since the participants in this study are males.

3.4 Verification of "Sounding Alike" on the F3 Bandwidth

F3 measures lip rounding. Vowels produced with unrounded lips have greater F3 values than those produced with rounded lips. As a rule of thumb, the F3 value of unrounded vowels produced by male speakers is ≥ 2600 Hz corresponding to unrounded vowels, while those ≤ 2400 Hz correspond to rounded vowels. However, Kent and Read (2002:110) note that F3 measurements are not very robust because F3 does not vary much among speakers. It is, therefore, not surprising that many phoneticians do not even bother with F3 in their analyses of vowel intelligibility. Stevens (2000) does not provide any masking values for F3. However, it is estimated that masking occurs if the acoustic distance between segments is ≤ 400 Hz. Consequently, if the F3 distances between the brothers' vowels is below this threshold, it means that they sound acoustically alike.

3.5 Verification of "Sounding Alike" on the Intensity Scale

The JND threshold at which hearers can determine whether one signal is louder than the other is ≥3 dB (Stevens 2000:225-6). This threshold is even specified on the package of many audio devices. If the intensity of the vowels produced by the three brothers is ≤3 dB, we will conclude that they do sound alike. If it is ≥ 5dB, we may conclude that they do not sound alike.

3.6 Verification of "Sounding Alike" Using the Duration Correlate

Hirsh (1959:767) conducted several experiments and concluded that the minimal temporal threshold to determine that one segment is longer than the other is 10 ms. Stevens (2000228-9) makes a distinction between segments lasting less than 100 ms and those lasting 200 ms or longer. For the latter, the JND is 15-20 ms. However, Hirsh (1959:767) found that when the distance between two signals is 17 ms, the signals are perceived "correctly". If the distance between the vowels produced by the brothers is less than 17 ms, it means that they sound acoustically alike.

4.0 Controllable and Incontrollable Correlates

The JNDs discussed in the previous sections will be applied to the vowels produced by the three brothers to determine if the impressionistic claims that they sound alike can be verified acoustically. However, before embarking on such an analysis, we should first make an important distinction between the acoustic correlates that speakers can control/manipulate consciously and those that they cannot. Every speaker can intentionally or willfully raise his/her voice and speed up or slow down the tempo of his/her speech. This is especially true for a language such as English where segmental duration is not phonemic. Intensity and duration are, therefore, controllable correlates, but under normal modal phonation, F0, F1, F2, and F3 are not controllable. Under normal circumstances, talkers are not aware of very subtle vocal fold vibrations, tongue height movements, tongue advancement or retraction, or lip rounding and spreading that affect the overall quality of their vowels. The data collected for this study is fairly large. It can be examined from various angles. However, for the sake of brevity and maximum clarity, we will present the data of two brothers at a time. This makes it easy to highlight the similarities and differences between the acoustic correlates of their vowels. In so doing, we display a total of 15 tables. The data from Andrew and James are compared and contrasted first, then comes the data from Andrew and Donovan. Thereafter, we show James' and Donovan's measurements.

4.1 Comparing and Contrasting F0

The mean F0 measurements of the vowels produced by the three brothers are as follows: Andrew: 130 Hz, James: 142 Hz, and Donovan: 124 Hz. The F0 distances between their vowels are respectively 12 Hz between Andrew and James, 6 Hz between Andrew and Donovan, and 18 Hz between James and Donovan, as shown in the tables below:

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Andrew	F0	141	137	130	129	126	124	121	124	130	138	134
James	F0	153	160	143	137	137	140	139	138	146	145	131
Difference	F0	12	23	13	8	11	24	18	14	16	7	3

Table 1.1: Comparison between Andrew and James

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Andrew	F0	141	137	130	129	126	124	121	124	130	138	134
Donovan	F0	134	125	121	115	120	120	116	126	124	147	120
Difference	F0	7	12	9	14	6	4	5	2	6	9	14

Table 1.2: Comparison between Andrew and Donovan

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
James	F0	153	160	143	137	137	140	139	138	146	145	131
Donovan	F0	134	125	121	115	120	120	116	126	124	147	120
Difference		19	35	22	17	17	20	23	12	22	2	11

Table 1.3: Comparison between James and Donovan

Since the F0 distances are all higher than the JND of 1 Hz, we conclude that the three brothers do not sound alike. However, this conclusion must be nuanced because many experts are of the opinion that this correlate plays virtually no role in intelligibility. Miller (1989:2128) states this view as follows:

> It is well known that, under most conditions, the identity of a perceived vowel depends strongly on the formant values of the spectrum and is independent of voice pitch. However, under certain circumstances the voice pitch can influence a vowel's identity, even when the formants are fixed. Within the auditory-perceptual theory, two hypothetical mechanisms are included that would allow voice pitch, under certain circumstances, to control the perception of vowel category.

The "two hypothetical mechanisms" to which Miller alludes do not apply in the production of vowels in their citation forms. Consequently, the fact that the F0 distances between the brothers are higher than the required ≥1 Hz does not negate that they sound alike.

4.2 Comparing and Contrasting F1

As noted previously, F1 accounts for 80% of the acoustic energy in vowels. This means that F1 plays an important role in intelligibility. Therefore, F1 is useful for measuring "sounding alike" acoustically. The mean F1 of all the vowels produced by Andrew, James, and Donovan are respectively 538 Hz; 533 Hz, and 496 Hz. The tables below show the vowel-by-vowel measurements:

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Andrew	F1	296	448	410	587	628	803	749	480	498	347	626
James	F1	340	487	454	635	685	708	685	483	476	377	595
Difference	F1	44	39	44	48	57	95	64	3	22	30	31

Table 2.1: Comparison between Andrew and James

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Andrew	F1	296	448	410	587	628	803	749	480	498	347	626
Donovan	F1	253	400	372	567	693	763	722	438	419	289	542
Difference	F1	43	48	38	20	65	40	27	42	79	58	84

Table 2.2: Comparison between Andrew and Donovan

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
James	F1	340	487	454	635	685	708	685	483	476	377	595
Donovan	F1	253	400	372	567	693	763	722	438	419	289	542
Difference	F1	87	87	82	68	8	60	37	45	57	88	53

Table 2.3: Comparison between James and Donovan

The mean acoustic distances between the F1s of the three brothers are respectively 5 Hz between Andrew and James, 42 Hz between Andrew and Donovan, and 37 Hz between James and Donovan is 37 Hz. As noted in 3.2, an acoustic distance of ≤ 50 Hz cannot be used as prima facie evidence that the brothers sound acoustically alike. According to CBT, masking takes place when the acoustic distance between vowels is less than 20 Hz.[6] When we apply this criterion to the vowels produced by the brothers, we conclude that Andrew (538 Hz) and James (533 Hz) sound acoustically alike because the distance between their vowels is only 5 Hz. If we apply Stevens' (2000:228) threshold of 30 Hz to James' and Donovan's (496 Hz) vowels, we conclude that they also sound acoustically alike because the acoustic distance is 37 Hz. However, the JND in F1 shows that Andrew (538 Hz) and Donovan (496 Hz) do not sound as alike as Andrew and James, or James and Donovan do since the distance between their vowels is 42 Hz, that is, higher than the threshold of 30 Hz.

4.3 Comparing and Contrasting F2

Recall that from the discussions in 3.3 that Kent and Read (2002:111) consider F2 to be the correlate that is "more sensitive to dialectal and idiolectal variation." In sociolinguistics, an idiolect is defined as a set of speech mannerisms that are peculiar to an individual. The expected JND threshold on the F2 frequency band is ≤ 200 Hz. If the F2 acoustic distance between the brothers is considerably less than this threshold, this would be an indication that they sound acoustically alike. Let's see what the measurements in the tables tell us about their pronunciation.

[6] Stevens (2000:228) contends that in free fields (in normal listening conditions), the JND is estimated at 30 Hz.

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Andrew	F2	2296	1903	2122	1771	1809	1330	1329	991	1295	1157	1336
James	F2	2407	1676	2123	1786	1735	1268	1258	926	1275	1143	1365
Difference	F2	111	227	1	15	74	62	71	65	20	14	29

Table 3.1: Comparison between Andrew and James

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Andrew	F2	2296	1903	2122	1771	1809	1330	1329	991	1295	1157	1336
Donovan	F2	2454	2010	2306	1807	1766	1290	1172	914	1255	1037	1303
Difference	F2	158	107	184	36	43	40	148	77	40	120	33

Table 3.2: Comparison between Andrew and Donovan

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
James	F2	2407	1676	2123	1786	1735	1268	1258	926	1275	1143	1365
Donovan	F2	2454	2010	2306	1807	1766	1290	1172	914	1255	1037	1303
Difference	F2	47	334	183	21	31	22	86	12	20	106	62

Table 3.3: Comparison between James and Donovan

The mean F2 value of Andrew's vowels is 1576 Hz, that of James is 1542 Hz, and Donovan that of 1574 Hz. The distance between Andrew and James is 34 Hz, the one between James and Donovan is 32 Hz, while the one between Andrew and Donovan is only 2 Hz. Acoustically speaking, these three brothers sound exactly alike. Kent and Read (2002:110-111) note that F2 usually varies considerably between speakers of the same language/dialect. The fact that the F2 distances on this bandwidth are so insignificant is a conclusive piece of evidence that the brothers have very similar idiolects, i,e., they are acoustically identical.

4.4 Comparing and Contrasting F3

On the F3 frequency band, an acoustic distance of less than 400 Hz between segments means that they are aurally imperceptible. The data in the tables below underscores incredible similarities in how the brothers produce their vowels. The average F3 measurements for Andrew, James, and Donovan are respectively 2540 Hz, 2645 Hz, and 2651 Hz. The distance between Andrew and James is 105 Hz, the one between Andrew and Donovan is 111 Hz, and the one between James and Donovan is 6 Hz. Because frequency is perceived on a logarithmic scale, not on an arithmetic scale, this means that the F3 of three brothers is acoustically identical. F3 does not vary considerably between speakers of the same language/dialect. For this reason, the perceptual similarities between the brothers should not be overemphasized. Yet, the fact that James and Donovan produced F3 identically cannot be overlooked.

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Andrew	F3	2752	2601	2548	2507	2458	2568	2581	2562	2553	2274	2538
James	F3	2931	2641	2609	2784	2608	2622	2579	2550	2597	2361	2813
Difference	F3	179	40	61	277	140	54	2	12	44	87	275

Table 4.1: Comparison between Andrew and James

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Andrew	F3	2752	2601	2548	2507	2458	2568	2581	2562	2553	2274	2538
Donovan	F3	3177	2667	2754	2687	2603	2501	2366	2466	2598	2514	2836
Difference	F3	425	66	206	180	145	67	215	96	45	240	298

Table 4.2: Comparison between Andrew and Donovan

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
James	F3	2931	2641	2609	2784	2608	2622	2579	2550	2597	2361	2813
Donovan	F3	3177	2667	2754	2687	2603	2501	2366	2466	2598	2514	2836
Difference	F3	246	26	145	97	5	121	213	84	1	153	23

Table 4.3: Comparison between James and Donovan

4.4 Comparing and Contrasting Intensity

If the mean intensity distances between the vowels produced by the brothers is ≤ 3 dB, this would be a strong indication that the brothers sound alike acoustically alike. Let's see what we can learn from the data in Tables 5.1, 5.2, and 5.3:

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Andrew	INT	61	66	64	65	64	66	63	63	64	62	66
James	INT	73	77	76	76	74	75	74	76	77	75	75
Difference	INT	12	11	12	11	10	11	11	13	13	13	9

Table 5.1: Comparison between Andrew and James

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Andrew	INT	61	66	64	65	64	66	63	63	64	62	66
Donovan	INT	57	65	62	65	63	66	66	67	66	63	64
Difference	INT	4	1	2	0	1	0	3	4	2	1	2

Table 5.2: Comparison between Andrew and Donovan

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
James	INT	73	77	76	76	74	75	74	76	77	75	75
Donovan	INT	57	65	62	65	63	66	66	67	66	63	64
Difference	INT	16	12	14	11	11	9	8	9	11	12	11

Table 5.3: Comparison between James and Donovan

The mean intensity of the vowels produced by the brothers is respectively 64 dB for Andrew and Donovan and 75 dB for James. By this correlate, Andrew and Donovan sound identical. However, the intensity difference between James and his two brothers is 11 dB. This can be interpreted to mean that James does not sound like his brothers. The intensity level in a normal conversation is 60 dB (Schnitta Bonnie 2016:55). This means that Andrew and Donovan recorded themselves speaking at a normal conversational level. However, James recorded himself using an intensity level that Pearsons and Horonjeff (1982:14) describe as a "raised level." This speech level is commonly associated with public speaking or classroom lecturing. A comprehensive study conducted by Pearsons and Bennett (1977:19) shows most teachers use intensity levels between 67-78 dB. Since James is training to be a teacher, it is not surprising that he recorded himself using a "teacher's voice volume." The fact that James' intensity is higher than that of his brothers does not in any way indicate that they do not sound alike acoustically. Intensity, as was noted in 4.0, is a controllable acoustic correlate. This means that when James is interacting normally with his relatives, friends, and acquaintances, and he is not "wearing his teacher's hat", he sounds exactly like Andrew and Donovan.

4.5 Comparing and Contrasting Duration

The mean duration of the totality of vowels produced by the brothers is 211 ms. The speakers' individual mean vowel durations are as follows: Andrew (190 ms), James (245 ms), and Donovan (199 ms). Since the overall duration of their vowels is higher than the 200 ms threshold, we will use the JND for 17 ms proposed by Hirsh (1959:767) to determine whether or not the brothers sound acoustically alike in vowel duration.

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Andrew	DUR	187	133	214	167	239	214	225	222	162	192	142
James	DUR	286	173	252	184	293	292	326	299	172	272	147
Difference	DUR	99	40	38	17	54	78	101	77	10	80	5

Table 6.1: Comparison between Andrew and James

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Andrew	DUR	187	133	214	167	239	214	225	222	162	192	142
Donovan	DUR	224	141	228	170	288	211	233	211	165	205	118
Difference	DUR	37	8	14	3	49	3	8	11	3	13	24

Table 6.2: Comparison between Andrew and Donovan

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
James	DUR	286	173	252	184	293	292	326	299	172	272	147
Donovan	DUR	224	141	228	170	288	211	233	211	165	205	118
Difference	DUR	62	32	24	14	5	81	93	88	7	67	29

Table 6.3: Comparison between James and Donovan

The durational distance between and Andrew's vowels (190 ms) and James' vowels (245 ms) is 55 ms. The one between James' vowels and Donovan's vowels (199 ms) is 46 ms. In both instances, the temporal distance is higher than 17 ms threshold. Since the temporal distance is higher than 17 ms, it indicates that James does not sound like his two brothers in vowel duration. However, when we compare and contrast vowel duration in Andrew's and Donovan's data, we see that the difference between them is only 9 ms. Therefore, the two sound acoustically alike in duration. However, before concluding that James does not sound like his brothers, we need to take the following into account.

Recall that Andrew and Donovan are monolingual English speakers. Recall also that James is majoring in linguistics with the intent of teaching English to Japanese speakers. Finally, recall that duration is a controllable correlate, i.e., that one can slow down or speed up one's pronunciation at will. James' bilingualism may have something to do with the fact that his vowels are longer than those of his two siblings. James describes himself as being an imbalanced bilingual in Japanese. He spent a year abroad studying Japanese. Hazan (2017:36) states that speaking another language in addition to one's native tongue can introduce variability into one's speech. Vowel lengthening may be a feature from Japanese that James has incorporated into his English. We note in passing that vowel duration is phonemic in Japanese. It is also likely that James' training as a future teacher of English as Foreign Language (EFL) may be responsible for vowel lengthening. As a future ESL/EFL teacher, it has been drilled into him that he should enunciate clearly. This translates into speaking deliberately and slowly. Even though the recording instructions tell the participants to speak as naturally as possible, James may have gone into an "EFL teacher mode" when he was recording himself. However, when he is interacting normally with family members, friends, and acquaintances, chances are that he does not change his normal speech tempo, namely, he speaks just like Andrew and Donovan. Since duration is a controllable speech correlate, the aforementioned distance between him and his brothers cannot be used as prima facie evidence that all three do not sound acoustically similar.

5.0 Visualization of "Sounding Alike"

As the saying goes, "a picture is worth a thousand words." A comparative acoustic vowel Space of the brothers can give us a lot of information about what goes on in their mouths when they produce the 11 vowels used in our study. Figure 1 displays that data for us:

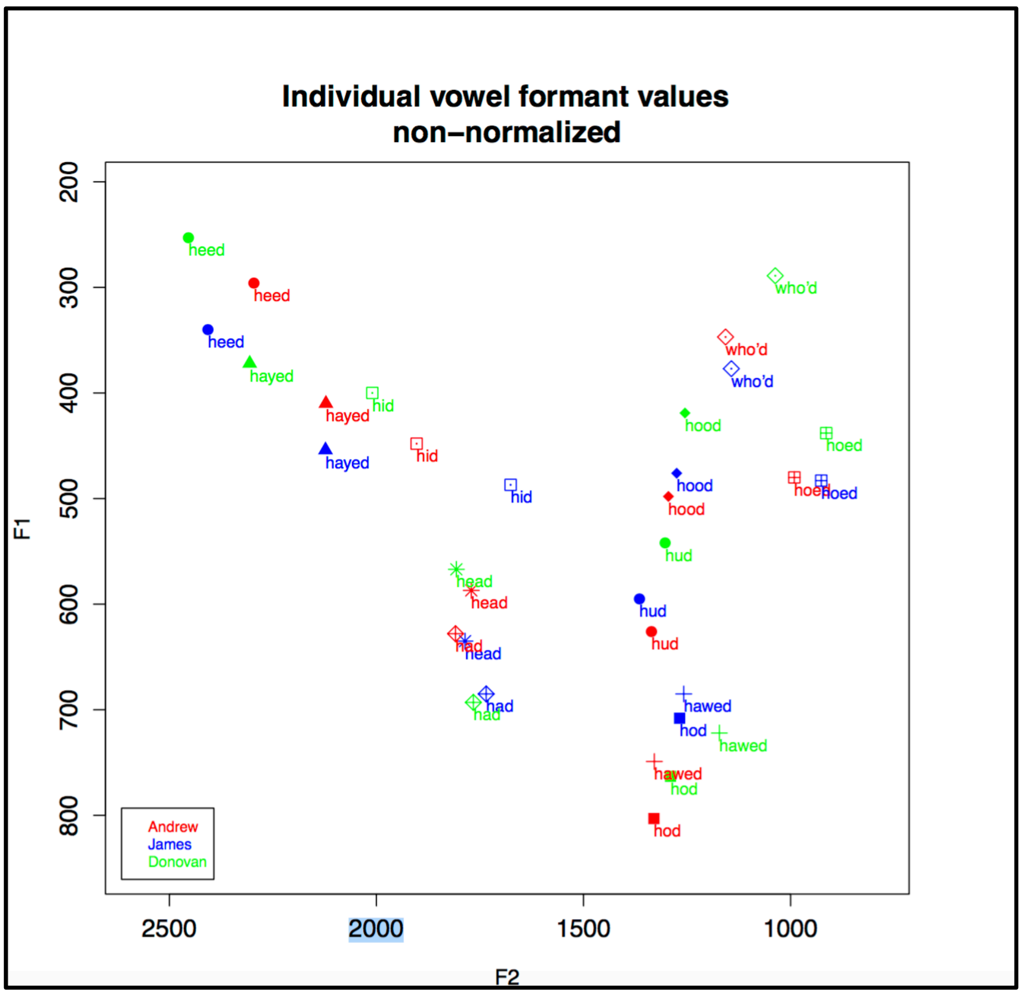

Figure 1: Comparative Acoustic Vowel Space

As is the custom in acoustic phonetic display of vowels, only F1 and F2 measurements are used in providing vowel diagrams. A lot can be said about the spacing of the vowels produced by the brothers. However, to avoid an unnecessary repetition of what been said in previous sections, we will only draw attention to one vowel that is produced markedly differently by the brothers. The vowel in question is [ɪ] in <heed>. It is strongly centralized by James, but it is less so in the speech of Andrew and Donovan. It is also worth noting that the vowel <e> in <hayed> has risen higher than [ɪ] in <heed> in the speech of the three brothers. Also, the <ʊ> in <hood> has lowered substantially in their speech. It is lower than the vowel <o> in <hoed> in the speech of Andrew and James. The migration downward has not yet reached its destination in Donovan's speech. According to Koffi (2016), these are the signature characteristics of Central Minnesota English. The presence of these two features attest that the Lyons brothers have grown up in this region. One major difference revealed by the data in Figure 1 is that the vowel [ɛ] (635 Hz) produced by James and the vowel [æ] (628 Hz) produced by Andrew mask each other because the acoustic distance between them is only 7 Hz. James produces this vowel so low that it encroaches on the space that [æ] occupies in Andrew's pronunciation.

6.0 Summary Table of "Sounding Alike"

The degree to which the three brothers sound alike acoustically is summarized in Table 7. The symbol [-] means that, according to the feature under consideration, the speakers do not sound alike. The symbol [+] means that they sound alike acoustically, while the symbol [±] indicates that the acoustic measurements are variable:

	F0	F1	F2	F3	Intensity	Duration
Andrew and James	-	+	+	+	±	±
James and Donovan	-	+	+	+	±	±
Andrew and Donovan	-	±	+	+	+	+

Table 7: Feature Summary of Acoustic Correlates

The F0 correlate has the symbol [-] in all three rows because this correlate fails to discriminate between speakers of the same biological gender. It is, therefore, not surprising that it cannot be used to verify that the brothers sound alike acoustically. According to the F1 correlate, Andrew and James, on one hand, and James and Donovan on the other, sound acoustically alike, but Andrew and Donovan do not because the F1 distance between their vowels is 42 Hz. However, since this distance is below the 50 Hz threshold, to the naked ears, they sound almost identical. The brothers' F2 measurements are stunningly identical. It even suggests that they have the same idiolect, i.e., speech mannerisms. F2 is by far the strongest acoustic phonetic piece of evidence that the brothers sound alike acoustically. The F3 correlate also supports the overall finding that the brothers sound exactly alike acoustically. The variable symbol [±] in the Intensity and Duration columns suggest otherwise because the duration and intensity of James' vowels differ from his brothers. However, as was also noted earlier, these two correlates can be controlled and manipulated at will. James does not sound like his brothers according to these two correlates because he recorded himself speaking like an EFL teacher. He raised his speech level as all teachers do. Furthermore, he spoke slowly and deliberately as teachers of English as a Second Language or teachers of English as a Foreign Language do. We anticipate that when James is talking to his relatives, friends, and acquaintances, he uses a conversational speech level and a normal speech tempo. Therefore, in such communicative contexts, he would sound acoustically like his two brothers.

In our considered opinion, the merit of this paper is that it has provided a principled methodology for measuring and explaining acoustically what is meant when we hear that two people "sound exactly alike." In the past, acoustic phonetics has discounted such impressionistic statements because experts have been fixated on the variability of acoustic signals and have ignored statements about invariance, that is, when people sound alike. Ladefoged (2003:27) contends that, "There is no doubt that the ultimate authority in all phonetic questions is the human hear. But nowadays instrumental aids can often illuminate particular points, acting like a magnifying glass when we need to distinguish between two similar sounds." If it is true that the ear is indeed the final authority in acoustic phonetics, as Ladefoged suggests, then we cannot dismiss when many people conclude that such and such persons "sound alike." It is the responsibility of acoustic phoneticians to assess the veracity of such impressionistic claims. This is what we have done in this paper by using the psychoacoustic instruments of Just Noticeable Difference and masking thresholds in F0, F1, F2, F3, intensity, and duration. Using these tools, we have successfully verified that the Lyons brothers sound alike acoustically as far as their pronunciation of vowels is concerned. A future study may investigate the production of their consonants and other prosodic features such as lexical stress and sentence intonation patterns.

Overall Data

Words		heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
Vowels		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Andrew	F0	141	137	130	129	126	124	121	124	130	138	134
James	F0	153	160	143	137	137	140	139	138	146	145	131
Donovan	F0	134	125	121	115	120	120	116	126	124	147	120
GAE	F1	270	390	*476	530	660	730	570	*497	440	300	640
Andrew	F1	296	448	410	587	628	803	749	480	498	347	626
James	F1	340	487	454	635	685	708	685	483	476	377	595
Donovan	F1	253	400	372	567	693	763	722	438	419	289	542
Andrew	F2	2296	1903	2122	1771	1809	1330	1329	991	1295	1157	1336
James	F2	2407	1676	2123	1786	1735	1268	1258	926	1275	1143	1365
Donovan	F2	2454	2010	2306	1807	1766	1290	1172	914	1255	1037	1303
Andrew	F3	2752	2601	2548	2507	2458	2568	2581	2562	2553	2274	2538
James	F3	2931	2641	2609	2784	2608	2622	2579	2550	2597	2361	2813
Donovan	F3	3177	2667	2754	2687	2603	2501	2366	2466	2598	2514	2836
Andrew	DUR	187	133	214	167	239	214	225	222	162	192	142
James	DUR	286	173	252	184	293	292	326	299	172	272	147
Donovan	DUR	224	141	228	170	288	211	233	211	165	205	118
Andrew	INT	61	66	64	65	64	66	63	63	64	62	66
James	INT	73	77	76	76	74	75	74	76	77	75	75
Donovan	INT	57	65	62	65	63	66	66	67	66	63	64

ABOUT THE AUTHORS

Ettien Koffi, Ph.D., is a professor of Linguistics at St. Cloud State University, Minnesota, USA, specializing in acoustic phonetics (Speech Intelligibility). His research interests center around sociophonetic variations in Central Minnesota English, acoustic phonetic accounts of intelligibility in L2 English, and acoustic phonetic and general description of Anyi, a West African Language spoken in Cote d'Ivoire. He is the author of four books and numerous papers covering topics as varied as syntax, translation, language planning and policy, orthography, and indigenous literacy training manuals. He can be reached at enkoffi@stcloudstate.edu.

James Lyons is a graduate of Saint Cloud State University (SCSU) with a Bachelor of Arts degree in Linguistics and a minor in Teaching English as a Second Language. He is currently teaching and managing an English conversation school in Kumamoto, Japan. He has served as a tutor in the Intensive English Center at SCSU as well as an ESL teaching assistant in Saint Cloud schools. He has spent a year abroad studying Japanese language and culture at Seinan Gakuin University in Fukuoka, Japan. James dreams of one day teaching English and Literature at a University in Japan. He can be reached at jlyonsesl@gmail.com.

References

Everest, Alton F. and Ken C. Pohlmann. 2015. *Master Handbook of Acoustics.* 6th edition. New York: McGraw Hill Education.

Hazan, Valerie. 2017. Speech Communication Across the Life Span. *Acoustics Today 13* (1):36-43.

Heller, Eric J. 2013. *Why You Hear What You Hear: An Experiential Approach to Sound, Music, and Psychoacoustics*. Princeton, NJ: Princeton University Press.

Hillenbrand, James, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler. 1995. Acoustic Characteristics of American English Vowels. *Journal of the American Acoustical Society*, 97 (5) 3099-3111.

Hirsh, Ira. 1959. Auditory Perception of Temporal Order. *The Journal of the Acoustical Society of America* 31 (6): 759-767.

Kent, Ra D. and Charles Read. 2002. *Acoustic Analysis of Speech.* 2nd edition. Clifton Park, NY: Delmar Cengage Learning.

Koffi, Ettien. 2017. *Relevant Acoustic Phonetics of L2 English: Focus on Intelligibility.* St. Cloud, MN: Unpublished Manuscript.

Koffi, Ettien. 2016. The Lowering of High Lax Vowels in Central Minnesota English: Does it Happen in Other Dialects? *Linguistic Portfolios* 5:2-14

Ladefoged, Peter and Keith Johnson. 2015. A Course in Phonetics. 7th edition. Stamford, CT: Cengage Learning.

Ladefoged, Peter. 2003. *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Malden, MA: Blackwell Publishing.

Lehiste, Ilse. 1970. Suprasegmentals. Cambridge, MA: The MIT Press.

Miller, James D. 1989. Auditory-Perceptual Interpretation of the Vowel. *The Journal of the Acoustical Society of America* 85 (5): 2114-2134.

Pearsons, Karl S. and Ricarda L. Bennett. 1977. Speech Levels in Various Noise Environments. *The Environmental Health Effects Research*. EPA 600/1-77-025.

Pearsons, Karl S. and Richard D. Horonjeff. 1982. Measurement of Speech Levels in the Presence of Time Varying Background Noise. *NASA Contractor Report 3547*. NAS1-16527.

Peterson, Gordon F and Harold L. Barney. 1952. Control Methods Used in a Study of the Vowels. *Journal of the American Acoustical Society*, 24 (2): 175-184.

Pope, J. 1998. Analyzers. In *Handbook of Acoustics*, ed. by Malcolm J. Crocker, pp. 1341-1353. New York: A Wiley-Interscience Publication.

Reetz, Henning and Allard Jongman 2009. *Phonetics: Transcription, Production, Acoustics, and Perception*. Malden, MA: Wiley-Blackwell.

Schnitta Bonnie. 2016. Residential Quietude, the Top Luxury Requirement. *Acoustics Toda*y 12 (3):49-56.

Stevens, Kenneth N. 2000. *Acoustic Phonetics.* Cambridge, MA: The MIT Press.

Young, H. D. 2011. *College Physics*, 9th edition. http://blogg.lnu.se/vagrorelse/files/2013/02/OpenStax_ch17.pdf.