Linguistic Portfolios

Volume 13 Article 2

2024

A COMPREHENSIVE REVIEW OF FORMANTS: LINGUISTIC AND SOME PARALINGUISTIC APPLICATIONS

Ettien Koffi
St. Cloud State University

Recommended Citation

Koffi, Ettien (2024) "A COMPREHENSIVE REVIEW OF FORMANTS: LINGUISTIC AND SOME PARALINGUISTIC APPLICATIONS," *Linguistic Portfolios*: Vol. 13, Article 2.
Available at: https://repository.stcloudstate.edu/stcloud_ling/vol13/iss1/2

A COMPREHENSIVE REVIEW OF FORMANTS: LINGUISTIC AND SOME PARALINGUISTIC APPLICATIONS

ETTIEN KOFFI

ABSTRACT

The ability to use Fast Fourier Transform (FFT) and spectrographs to disentangle, measure, and visualize formants has had a profound and beneficial impact on fields as varied as signal processing, acoustic-phonetics, speech pathology, speech synthesis, and voice biometrics. However, formants are relatively unknown to the average phonologist or linguist. This review is intended to explain the benefits of knowing formants and using them in the description of languages. The focus is on F1, F2, F3, and F4 because they are the most important formants. Yet, passing remarks are made about F5. Elements of the discussions include F0 (pitch) because, even though it is not a formant, it plays a supporting role to formants. The analyses and demonstrations provided in this paper are based on 2,904 formant tokens extracted from the speech of 22 speakers of American English, 17 females and 5 males.

Keywords: Formants, Frequency Spectrum, Just Noticeable Difference, Critical Bands, Critical Band Theory, One-third Octave Bandwidth, Speech Banana, Voice Biometrics, Relative Functional Load, Formant Combinatorics

1.0 Introduction

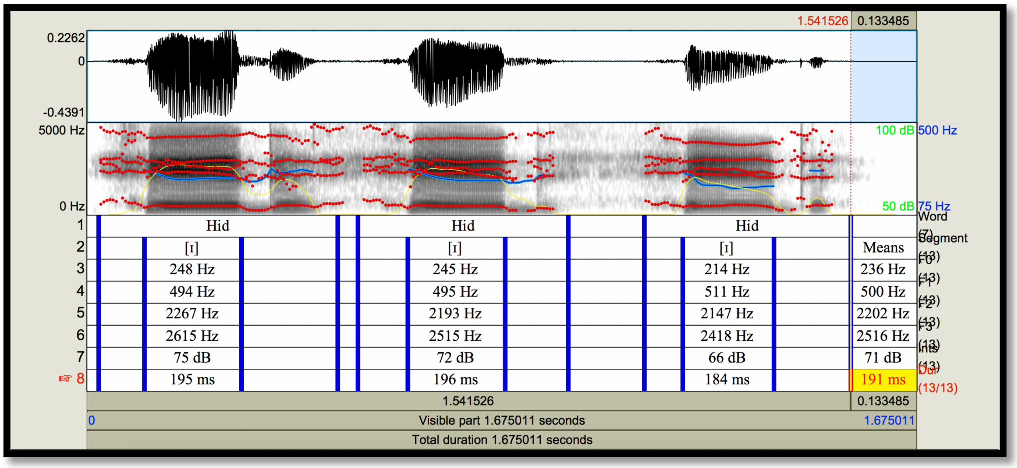
Formants are a staple in acoustic phonetics because they provide useful information about the production of speech segments in the supralaryngeal cavity (i.e., oral and nasal cavities). They are important in hearing sciences and speech intelligibility because they provide useful insights about how the Complex Auditory Nervous System (CANS) processes speech signals. They are also invaluable in voice biometrics, speech synthesis, and voice research. Tools such as the Fast Fourier Transform (FFT) and the spectrograph have given experts unparalleled insights into the measurements and behavior of formants. Nearly 80 years of accumulated knowledge and insights are reviewed in this paper to give phonologists and linguists in general the ability to use formant data to buttress their descriptions of languages. Because formants are of interest to physicists, audiologists, signal processing engineers, forensic scientists, and speech technologists, a dizzying depth and breadth of information is available. Even so, this paper is purposefully selective about the type of information that is reviewed. The reader should know at the outset that only aspects of formants that are most relevant to the linguistic description of languages is discussed, with only passing observations made about paralinguistic applications such as voice biometrics. The paper contains four broad installments. The first highlights foundational concepts. Here several definitions are provided to underscore that the formant is a multifaceted concept for which a single definition is woefully inadequate. The second introduces the participants from whom the data was collected. The third, the longest, focuses specifically on formants and underscores their relevance in phonological and acoustic phonetic analyses. Lastly, the fourth brings forth the usefulness of formants in some paralinguistic applications.

1.1 A General Definition of Formants

The concept of "formant" is surprisingly difficult to define. This is so because various academic disciplines appeal to it and define it to meet their disciplinary needs. Definitions of formants in physics and signal processing, such as the one in Steven (2000:131), are almost indecipherable to the average linguist. Simply put, physicists and acousticians define formants in relation to the mouth and the nose as resonance cavities. Here, the goal is to understand how air molecules are modified in different areas of the supralaryngeal cavities to produce speech sounds. Formants are also of interest to speech pathologists, articulatory phoneticians, voice instructors, and some musicologists. Lee et al. (2016:426) note that articulatory phoneticians define formants by examining the place of articulation and degrees of vocal tract constriction in the production of speech sounds. The oral cavity is understood broadly to include the lips, tongue, hard palate, pharyngeal wall, and sinuses. This is a first attempt at defining formants. Subsequent sections add to it by considering six more specific definitions.

1.2 A Spectrographic Definition of Formants

Acoustic phoneticians define formants by scrutinizing areas of concentrated bands of spectral energy. All speech segments have formants, but greater emphasis is placed on the formants of vowels and sonorants, i.e., nasals, liquids, and glides.[1] When these segments are produced, many formants are "nicely" staggered one on top of each other. Traditionally, investigations have focused on the first three formants: formant 1, formant 2, and formant 3, which are abbreviated as F1, F2, and F3. It is only recently that researchers have begun paying attention to F4. As for F5, it is hardly ever discussed in linguistic analyses because it is not very useful for intelligibility, as demonstrated in 7.1, 7.2, and 7.3. Formants appear in spectrographs in Praat as lines of red dots, as shown in Figure 1. This spectrograph displays the first five formants of the "kit" vowel [ɪ], repeated three times.



1	Hid	Hid	Hid		Word (7)
2	[ɪ]	[ɪ]	[ɪ]	Means	Segment (13)
3	248 Hz	245 Hz	214 Hz	236 Hz	F0 (13)
4	494 Hz	495 Hz	511 Hz	500 Hz	F1 (13)
5	2267 Hz	2193 Hz	2147 Hz	2202 Hz	F2 (13)
6	2615 Hz	2515 Hz	2418 Hz	2516 Hz	F3 (13)
7	75 dB	72 dB	66 dB	71 dB	Ints (13)
☞ 8	195 ms	196 ms	184 ms	191 ms	Dur (13/13)

1.541526 0.133485

0 Visible part 1.675011 seconds 1.675011

Total duration 1.675011 seconds

Figure 1: Formant Patterns [Formants are red Online]

---

[1] The formants of stops and fricatives are seldom extracted because they do not play a decisive role in speech intelligibility. Yet, Koffi (2023) devoted an entire paper to the formant analysis of stops in Anyi. Also, Koffi (2024) addresses the formants of fricatives. Extracting the formants of stops and fricatives becomes a necessity when formant-based speech synthesis is being considered for a language.

Even if the red lines are removed, formants can be singled out as areas of concentration of spectral energy, shown as darker bands in Figure 2A. Here, attention should be focused on the spectrographic behavior of the "dress" vowel [ɛ]. Observations can also be made about the formants of the sonorants [l, r, w, j]. The **spectrographic definition** of formants as bands of concentration of spectral energy still holds true even when the formant tracker (red online) is not activated.

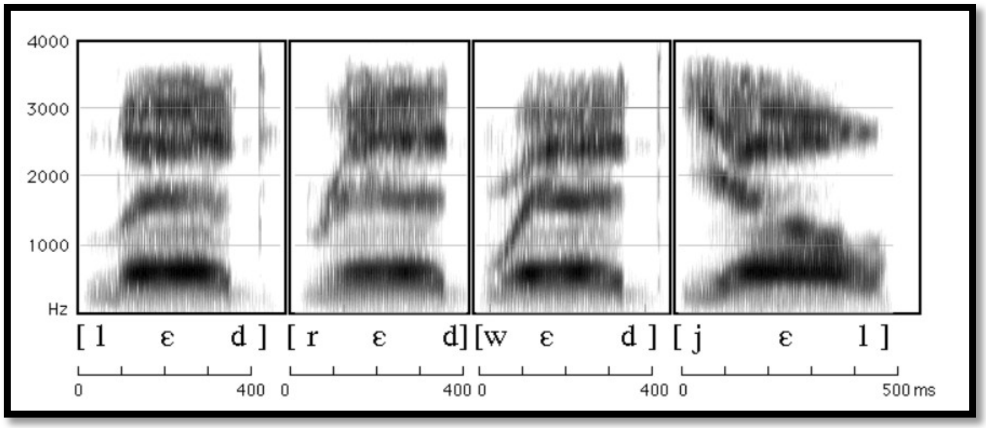


Figure 2A: Formant Patterns of Sonorants[2]

1.3 A Spectral Slice Definition of Formants

Formants can also be defined according to their spectral behavior, after one has taken a spectral slice of the segment under consideration. If a **spectral definition** is desired, one can appeal to Watt (2015:87) who defines formants as "amplitude peaks in the acoustic spectrum which result from the excitation of particular vocal-track resonances brought about by the vibration of the vocal folds setting the column of air inside the pharyngeal, oral and nasal cavities in motion." The spectral slice definition of formants can be illustrated with Figure 2B. This represents a spectral slice of the vowel [æ] in the word <that> produced by yours truly.

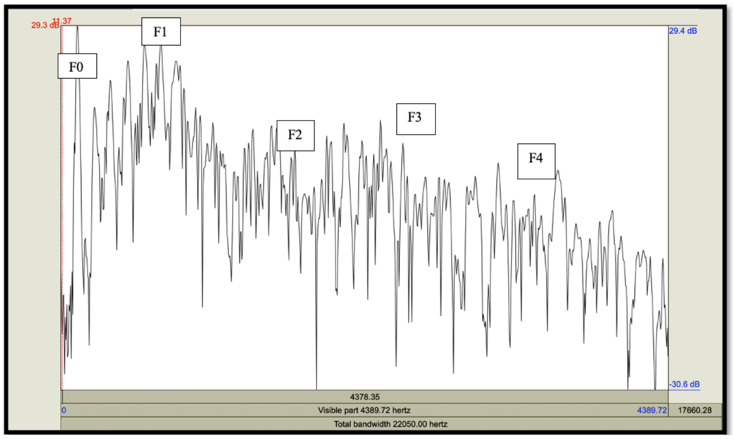


Figure 2B: Spectral Slice of Formants

---

[2] Source: https://corpus.eduhk.hk/english_pronunciation/index.php/3-2-acoustic-aspects-of-consonants/. Retrieved August 11, 2023.

1.4 An Articulatory Definition of Formants

We must first motivate an articulatory definition of formants by appealing to the Physiological Principle in accordance with Baken and Orlikoff (2000:3). It is stated as follows:

> **Physiological Principle**
>
> Measurements must have a known (or at least a very likely) and specific relationship to recognized aspects of the speech system physiology.

The most important publication that helps ground formants in articulatory phonetics is Lindblom and Sundberg (1971). They explain in great detail how before egressive air molecules exit the oral cavity, they are modified in specific ways by various articulators. The perfect combination, synchronization, and timing of articulatory gestures help to modify air molecules into various resonance frequencies. All this constitutes an **articulatory definition** of formants. Figure 3 taken from Lindblom and Sundberg (1971:1172) illustrates how formants are produced.

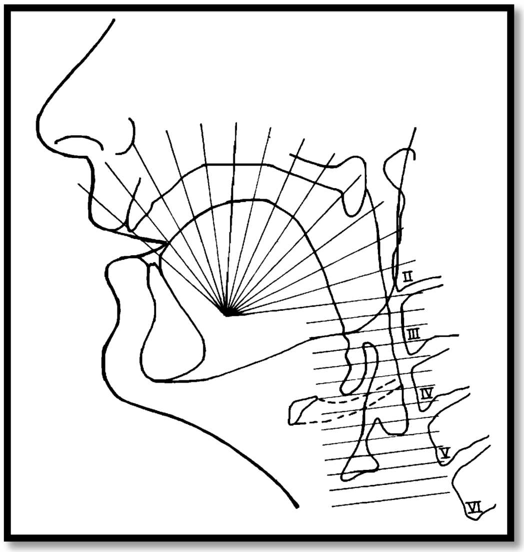


Figure 3: Illustration of the Physiological Definition
"Björn E. F. Lindblom, Johan E. F. Sundberg; Acoustical Consequences of Lip, Tongue, Jaw, and Larynx Movement. *J Acoust Soc Am* 1 October 1971; 50 (4B): 1166–1179. https://doi.org/10.1121/1.1912750; Figure 11, with the permission of the Acoustical Society of America."

The lines radiating from a single neutral tongue position are meant to underscore the prominent role that the tongue plays in speech production. The lines are also meant to highlight the extraordinary level of synchronization required by the tongue, lips, jaw, and various places of articulation when formants are produced. They note that the slightest change (i.e., ± 10 millimeters) from the neutral position of the tongue can greatly influence formant values. In other words, the muscular coordination and orchestration needed to produce the formants of a single speech sound defy a succinct description. Lee et al. (2016:426) are right in complaining that researchers have "simplified the relationships between tongue position and formant frequencies." Yet, the physiological dynamics undergirding formant production are so complex that researchers have no choice but to simplify them.

1.5 A Pattern Recognition Definition of Formants

From the 1940s up until the 1980s, pattern recognition was deemed the best way to describe and identify formants. Patterns such as those in Figure 4 were displayed prominently in numerous publications to help students master the prototypical behavior of formants. The patterns here focus mostly on the behaviors of F1 and F2 of vowels and the preceding stop consonants. Students are to notice that F1 and F2 are more maximally separated for front vowels versus back vowels. The subtle changes at the very beginning of formants are supposed to give away some information about the consonants. Additionally, the interval between F1 and F2 gives away information about mouth aperture, as depicted in Figure 4. The spacing is considerably wider for front vowels, whereas for back vowels, F1 and F2 are closer to each other.

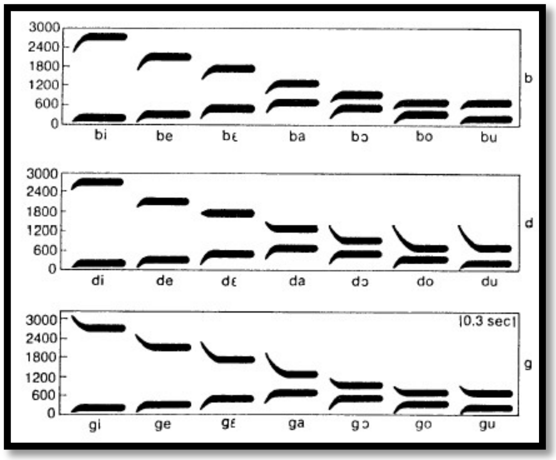


Figure 4: Formant Patterns of Vowels
"Pierre C. Delattre, Alvin M. Liberman, Franklin S. Cooper; Acoustic Loci and Transitional Cues for Consonants. *J Acoust Soc Am* 1 July 1955; 27 (4): 769–773. https://doi.org/10.1121/1.1908024; Figure 1, with the permission of the Acoustical Society of America."

Over the years, the pattern recognition approach to formants identification has lost much of its original appeal. Pattern recognition of formants is generally good when words are produced carefully in isolation. However, formant visualization proves problematic in running speech. Confounding factors such articulatory effort, speech tempo, interspeaker and intraspeaker variability, co-articulation, and the quality of the spectrograph, have led to some paradoxes, as noted in Johnson (2022:1963). At times the same sound would have wildly different formant patterns, while at other times two unrelated sounds would have similar formant patterns. As a result of these incongruities, there has been a gradual shift from formant "gazing" to formant measurements. It is hard to say exactly what was the catalyst for the change. I conjecture based on Gordon and Ladefoged (2001:396) that it had a lot to do with Kenneth Steven and Peter Ladefoged. In that paper, the authors acknowledged that Steven, the famed acoustic phonetician from the Massachusetts Institute of Technology (MIT), had suggested to Ladefoged and his co-author that "Phonation differences can be quantified through a number of phonetic measurements, even if certain physiological or auditory properties defining these phonation types are harder to define." Even though the specific context of this remark had to do with phonation types, the observation also applied to formants. Indeed, papers such as Peterson and Barney (1952) had already shown that formants could be reliably identified and categorized by extracting various measurements. I should hasten to add that formant visualization is not completely obsolete.

Johnson (2022) still teaches it to his speech signal processing engineering students at the University of California Los Angeles (UCLA). I also introduce it to my linguistics, engineering, and computer science students but do not emphasize it because it no longer holds the sway it once did.

1.6 An Auditory-Perceptual Definition of Formants

The formants emitted by the mouth of talkers play an important role in how hearers perceive and process the auditory input that they receive. For this reason, a **perceptual definition** of formants is needed. The process of auditory perception of formants is summarized by Russo (2020:38) as follows:

> Sound waves produced by voices or instruments are collected in the outer ear, mechanically amplified in the middle ear, and transduced by sensory hair cells in the cochlea within the inner ear to produce neuroelectric activity. This neuroelectric activity is then transmitted by the auditory nerve to the brainstem and onto the thalamus, which in turn projects to the auditory cortex within the temporal lobe of the brain.

The ears, the hair cells, the auditory brainstem, the thalamus, etc. are abbreviated as CANS (complex auditory nervous system). The preliminary research that elucidated the role of formants in speech perception can be attributed to the groundbreaking investigation that Harvey Fletcher and his team did at the Bell Telephone Laboratory. In a seminal publication in 1940, Fletcher, a physicist, calculated mathematically the frequency responses inside the basilar membrane. He posited that it is divided into some 29 **critical bands** where specific frequencies are perceived, processed, and transduced in neuro-electrical impulses, as displayed in Figure 5:

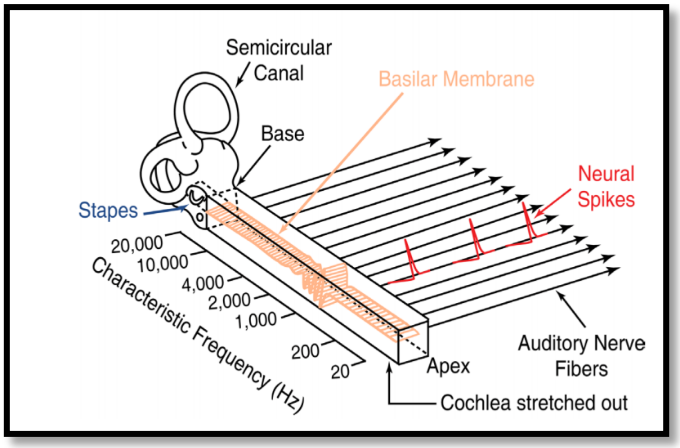


Figure 5: Audibility Range in the Frequency Domain [Color Online]
"Sachs, M. B., Bruce, I. C., Miller, R. L., and Young, E. D. (2002). Biological basis of hearing aid design. *Annals of Biomedical Engineering* 30, 157-168. doi:10.1114/1.1458592. Reprinted by permission of Biomedical Engineering Society."

Von Bekesy, another physicist, proved clinically that Fletcher's theory was grounded in physiological reality. For his tireless effort and for his ingenious pioneering ways of probing into the inner ear, von Bekesy was awarded a Nobel Prize in Medicine in 1961. The name of the theory that was developed to account for how the auditory system perceives and processes auditory input is called the **Critical Band Theory** (CBT).

The basilar membrane of mature adults is 32 to 35 mm long (Alkahby et al. 1999:116). Its length is subdivided into bands of 1.2 mm long, thereby yielding 25 to 29 critical bands. It processes and transduces frequencies as infinitesimal as 20 Hz and as gargantuan as 20,0000 Hz. This represents the **audibility range** of healthy humans, i.e., those who do not suffer from any hearing impairment. It has been posited that each critical band is subdivided into three parts: the lower, center, and upper frequency bands, as displayed in Table 1. This subdivision into **third octave bandwidths** mirrors how the human ear experiences frequency data (Everest and Pohlmann 2015:529).

N0	Lower Band Limits	Center Frequency	Upper Band Limits
1.	22	25	28
2.	35	40	44
3.	44	50	57
4.	57	63	71
5.	71	80	88
6.	88	100	113
7.	113	125	141
8.	141	160	176
9.	176	200	225
10.	225	250	283
11.	283	315	353
12.	353	400	440
13.	440	500	565
14.	565	630	707
15.	707	800	880
16.	880	1000	1130
17.	1130	1250	1414
18.	1414	1600	1760
19.	1760	2000	2250
20.	2250	2500	2825
21.	2825	3150	3530
22.	3530	4000	4400
23.	4400	5000	5650
24.	5650	6300	7070
25.	7070	8000	8800
26.	8800	10000	11300
27.	11300	12500	14140
28.	14140	160000	17600
29.	16000	20000	22500

Table 1: One Third Octave Critical Bands

The human ear performs a Fast Fourier Transform (FFT) type of analysis on incoming auditory inputs. In each critical band, the **center frequency** corresponds to the area of maximal excitation of the signals, that is, places where signals are boosted/magnified for maximum intelligibility. O'Shaughnessy (2023:33) notes that center frequencies are acoustic cues that hearers rely on to distinguish between phones. As will be shown below, they play an exceptionally crucial role in the intelligibility of formants. However, humans do not utilize the full extent of their audibility range in perceiving and processing sounds and noises. The upper frequencies ranges become inoperative as people age or if they take certain medicines that kill inner and outer hair cells. When I performed a simple hearing test[3], I found out that I could not perceive frequencies below 30 Hz nor frequencies higher than 12,000 Hz. This is not a cause for alarm for me or anybody else because, as we will see below, the frequencies needed for speech intelligibility are ≤ 5000 Hz.

1.7 A Speech Banana Definition of Formants

The concluding lines of the previous section are depicted pictorially in Figure 6 as a **"Speech Banana."** Experts use this metaphor to symbolize the range of formant frequencies that are relevant for speech production and perception. The nickname of "Speech Banana" is appropriate because the formant ranges have the shape of banana.

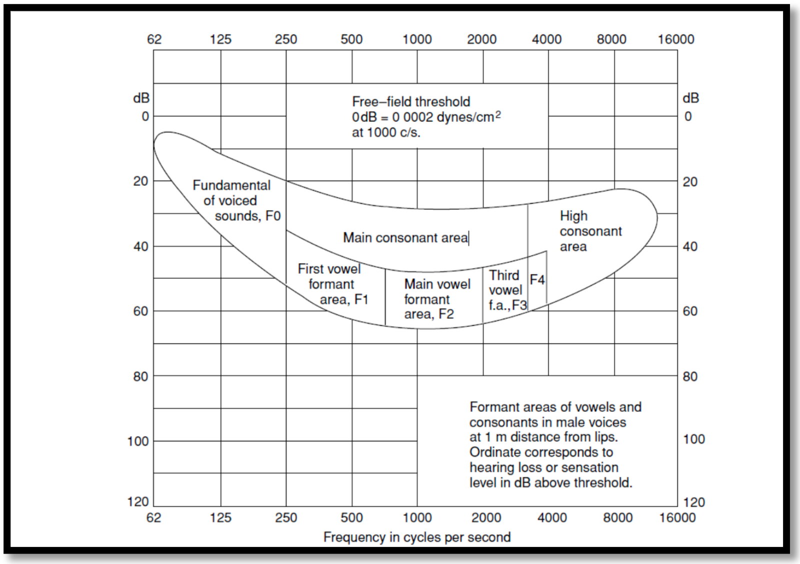


Figure 6: Formants and Speech Banana
"Nittayapa Klangpornkun, Chutamanee Onsuwan, Charturong Tantibundhit, Pittayapon Pitathawatchai; Predictions from "speech banana" and audiograms: Assessment of hearing deficits in Thai hearing loss patients. *Proc. Mtgs. Acoust* 2 December 2013; 20 (1): 060004. https://doi.org/10.1121/1.4879262; Figure 2, with the permission of the Acoustical Society of America."

The leftmost edge of the banana begins at 62 Hz. This is so because the lowest frequency that any human being can produce with their vocal folds when speaking is 60 Hz (Fry 1979:60). The area from 62 Hz to 250 Hz corresponds to F0/pitch in adult vocalization.[4] We will see in 2.2 that even though F0 is not a formant, it plays a supporting role to formants. The frequencies from 250 Hz to 4000 Hz are the actual formants deemed relevant for speech intelligibility. Figures 1 and 2A show how formants are staggered on each other. The lowest is called F1 and ranges 250 to 1000 Hz. Thereafter, all formants have a 1000 Hz interval between them. F2 covers the frequency zone of 1000-2000 Hz, F3 spans across 2000-3000 Hz, while F4 ranges from 3000 to 4000 Hz, and F5 spans from 4000 to 5000 Hz. Seldomly are F6 and higher formants mentioned because they do not play any role in speech intelligibility. F1, F2, and F3 are said to contain

---

[3] The test can be performed at: https://www.szynalski.com/tone-generator/ retrieved on January 17, 2023.

[4] Nobody produces an F0 as low as 60 Hz. Praat and other speech software set the lowest pitch detection algorithms at 75 Hz.

linguistically pertinent information, whereas F4 and F5 are seen as gold mines of paralinguistic information about speakers. More will be said about this in subsequent sections.

2.0 Participants, Data, and Methodology

Data is provided to illustrate aspects of the multifaced definitions of formant attempted earlier. The data that serves as the basis for the demonstrations and exemplifications come from 22 native speakers of American English from the state of Minnesota, 17 females and 5 males. They all signed informed consent forms that were approved by the Institutional Review Board (IRB) at St. Cloud State University. They read the same elicitation paragraph below from which the formants of vowels in red ink (red online) were extracted.

> Pl**ea**se call Stella. **A**sk her to bring these things w**i**th her from the store: Six g**oo**d spoons of fresh snow p**ea**s, five th**i**ck slabs of bl**ue** cheese, and m**ay**be a foot-long sandwich as a snack f**o**r her brother B**o**b. We als**o** need a small plastic snake, the little y**e**llow b**oo**k, a rubber d**u**ck, and a p**a**per I-p**a**d. She should not forget the d**o**g video game and the big toy fr**o**g for the k**i**ds. She must leave the faked gun at home but she may bring the ten sea turtles, the m**a**t that my mom b**ou**ght, and the black r**u**g. She can scoop these things into three r**e**d bags and two **o**ld backpacks. We will g**o** meet her, Sue, Jake, and Jenny Wednesday at the very l**a**st train station. The station is between the bus stop and the c**oo**kie store on Flag Str**ee**t. We m**u**st m**ee**t there 12 **o**'clock, for sure. The entrance **i**s at the **e**dge of the z**oo** in Zone 4 under the zebra sign. York's Treasure Bank is the tall building in the l**e**ft c**o**rner. She cannot miss it.

This text is an augmented version of the Speech Accent Archive elicitation paragraph.[5] Augmentation was necessary because the original text lacked the vowel [ʊ]. The 11 phonemic vowels from which formant measurements were extracted are the following:

Vowels	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Segments	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]

Table 2: Well's Lexical Set for Phonemic English Vowels

The formant data reported in this paper are based on 2,904 formant tokens. The females produced 2,244 formant tokens (17 speakers x 11 phonemic vowels x 3 repetitions x 4 formants), while the 5 males 660 tokens.[6] All male data, except for F0 measurements are placed in the appendix for consultation and will not be commented on any further. Also, instead of referring to individual vowels by their IPA symbol alone, Well's lexical set is used in tandem with IPA transcription, as modeled in Ladefoged and Johnson (2015:102-3). So, we will say the "fleece" vowel [i] instead of just simply displaying the IPA symbol [i]. Finally, it must be borne in mind that when formants are extracted from vowels in running speech, their values may be slightly different from the formants of vowels produced in citation form.

---

[5] Information retrieved from http://accent.gmu.edu/ on March 4th, 2024.

[6] The original data appeared in Koffi and Krause (2020).

2.1 Just Noticeable Difference (JND) Thresholds

Since the goal of this review is to familiarize phonologists with formants, we must underscore the role that they play in speech intelligibility. So, we appeal to **psychoacoustics** and to **Just Noticeable Difference (JND)** thresholds. Psychoacoustics is the branch of acoustics that describes how humans perceive and process auditory inputs. Fastl and Zwicker (2007:VII) define psychoacoustics simply as follows, "The correlation between acoustical stimuli and hearing sensations is investigated by acquiring sets of experimental data and by models which simulate the measured facts in an understandable way." The overarching goal of psychoacoustics is to discover when measurements are auditorily perceptible. The best way to achieve this goal is through JND thresholds at which speech stimuli are optimally perceived or not perceived. Stevens (2000:225) explains that JNDs are correct responses elicited from experimental subjects. To qualify as a valid JND, at the very least, 75% of the responses must agree. Everest and Pohlmann (2015:23, 515) report that JNDs are commonly used in physical and biomedical sciences to establish degrees of significance. The JNDs of formants that are used in the remainder of this paper have accumulated from nearly 80 years of psychoacoustic experimentations. When JNDs are used, they obviate the need of complex statistical analyses because they are de facto statistically significant.

2.2 F0 Measurements

F0/pitch is not a formant because it is produced in the laryngeal cavity whereas formants belong to the supralaryngeal cavities. Even so, it is discussed briefly in this paper because, according to Zahorian and Jagharghi (1993:1966-1967), it plays an important supporting role in the perception of F1 and F2. The human ear is exceptionally sensitive to the minutest variations in pitch because it perceives it on a linear (arithmetic) scale but formants are perceived on a logarithmic (non-linear) scale. Numerous authorities, including Stevens (2000:228), Lehiste (1970:64), Gandour (1978:57), and Rabiner and Juang (1993:152), to mention only the prominent ones, have reported that in ordinary listening conditions, people can detect a pitch difference of less than 1 Hz, hence the JND below:

JND of F0

> Of two speech signals A and B, the former is perceived as being auditorily different from the latter if and only if the acoustic distance between them is ≥ 1 Hz.[7]

When this JND is applied to the data in Tables 3A and 3B, we see that the participants produced a total of 506 tokens of F0 (46 speakers x 11 phonemic vowels). Pitch alone helps to discriminate between 453 of them. This means that the relative functional load (RFL) of F0 for English vowels is 89.52%. Pitch failed to discriminate between 53 vowels pairs, that is 10.47%. The take-away is that F0 plays a substantial role in the intelligibility of English vowels.

Words	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
F0 Males	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1M	114	130	114	122	114	113	114	117	124	126	120
Speaker 2M	146	153	137	144	139	128	140	145	154	162	130
Speaker 3M	102	102	92	98	96	93	92	91	95	93	92
Speaker 4M	148	141	135	117	127	133	126	126	134	130	128
Speaker 5M	105	105	101	98	98	98	99	102	105	106	100

---

[7] JND for the perception of F0 between syllable nuclei within the same word is 1 Hz. The JND for the perception of F0 between consecutive words in the same utterance is 5 Hz. These two JNDs should never be confused!

Speaker 6M	118	113	109	109	103	102	96	103	113	126	113
Speaker 7M	110	108	97	104	97	93	98	102	109	108	94
Speaker 8M	123	123	113	117	113	114	116	120	121	123	114
Speaker 9M	123	108	109	106	110	107	103	113	117	119	106
Speaker 10M	122	139	108	119	80	81	82	87	83	82	78
Speaker 11M	123	119	115	115	115	111	110	114	119	117	124
Speaker 12M	98	105	102	100	97	98	103	111	106	109	102
Speaker 13M	113	110	110	113	108	112	111	113	124	124	113
Speaker 14M	153	160	143	137	137	140	139	138	146	145	131
Speaker 15M	133	119	121	111	110	102	115	125	129	139	115
Speaker 16M	158	117	114	115	145	109	109	173	118	136	105
Speaker 17M	129	129	127	128	127	130	127	137	133	137	129
AVG	**124**	**122**	**114**	**114**	**112**	**109**	**110**	**118**	**119**	**122**	**111**
St. Deviation	17	17	14	12	17	15	15	21	17	19	15

Table 3A: F0 Measurements for Male Speakers

Words	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
F0 Females	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1F	186	176	171	179	169	173	169	172	176	187	172
Speaker 2F	224	220	212	222	205	220	211	231	234	269	229
Speaker 3F	277	276	253	245	240	233	226	234	236	253	252
Speaker 4F	248	246	237	234	219	236	227	246	241	255	243
Speaker 5F	191	189	180	180	177	165	172	188	188	218	181
Speaker 6F	199	192	155	190	178	172	170	169	201	199	182
Speaker 7F	243	248	192	251	198	216	209	244	303	301	218
Speaker 8F	217	210	206	204	205	203	200	207	210	224	208
Speaker 9F	225	229	216	225	212	211	209	211	219	221	217
Speaker 10F	241	236	240	230	230	227	231	232	241	248	236
Speaker 11F	163	191	192	192	190	172	190	185	198	174	181
Speaker 12F	235	223	233	205	201	208	211	235	232	248	203
Speaker 13F	207	196	200	192	194	189	190	197	196	219	191
Speaker 14F	207	191	189	190	186	190	191	193	193	203	173
Speaker 15F	223	211	227	221	219	220	221	215	229	224	227
Speaker 16F	219	226	227	230	238	231	228	225	220	213	234
Speaker 17F	217	224	208	211	205	239	187	199	216	233	196
Speaker 18F	266	282	270	244	237	254	245	259	253	269	249
Speaker 19F	207	214	211	349	216	205	211	214	209	203	225
Speaker 20F	245	232	236	222	207	232	220	213	217	224	217
Speaker 21F	242	246	248	234	243	233	254	225	234	245	240
Speaker 22F	236	218	210	205	209	190	211	208	216	225	202
Speaker 23F	223	258	207	244	238	220	234	236	241	249	229
Speaker 24F	215	222	221	209	212	212	205	206	212	237	214
Speaker 25F	210	214	188	191	169	173	182	179	185	167	167
Speaker 26F	226	234	223	223	201	208	217	212	215	220	204
Speaker 27F	177	160	154	157	169	153	161	159	178	184	182
Speaker 28F	224	236	224	229	212	204	211	221	229	238	170
Speaker 29F	207	215	212	216	203	186	195	206	207	222	202
AVG	**220**	**221**	**211**	**218**	**206**	**206**	**206**	**211**	**218**	**226**	**208**
St. Dev.	24	27	27	34	21	25	23	24	25	29	25

Table 3B: F0 Measurements for Female Speakers

3.0 General Acoustics of Formants

Now, we turn our attention to the role that formants play in the intelligibility of vowels. Formants are frequencies that are generated in the supralaryngeal cavities, i.e., the oral and nasal cavities after speech signals have first been modified in the larynx. Praat and other speech analysis software packages display only five formants. Yet, the settings can be changed to display more formants if one wishes to. However, it is not useful to do so because higher formants are not useful in the intelligibility of speech sounds. Even though spectrographic displays show that formants are in theory independent of each other, Bele (2005:567) notes that they are interdependent on each other. For example, a lowered F2 causes F3 and F4 to be also low. Yet, both the independence and interdependence of formants must be kept in mind as we proceed.

3.1 Relative Functional Load of Formants

Because our primary goal in this paper is to highlight the relevance of formants for linguistic analyses, we will calculate the RFL of each formant as we did for F0 in 2.2. When all the analyses are completed, we will rank formants to see their relative strength for speech intelligibility. The calculation of the RFL of formants is done by computing their discriminatory power vis à vis each other. If the acoustic distances between pairs of vowels meet or exceed the stated JND thresholds, we conclude that it plays a role in speech intelligibility. If the JND threshold is not met, then we conclude that it does not play a discriminatory role.

3.2 Vowel Articulation and the First Formant

The extraction of F1 measurements satisfies the Physiological Principle. Numerous authoritative studies have indicated that F1 correlates with mouth aperture. Speech segments that are produced with the mouth almost closed have smaller F1 values, whereas those produced with the mouth wide open have larger values (see Figure 4). It is often noted F1 measurements follow the law of inverse proportionality. This means that high vowels have smaller F1 values, while low vowels have larger F1 values. F1 measurements can help to classify vowels according to three degrees of height (aperture):

1. F1 measurements ≤ 400 Hz correspond to high vowels.
2. F1 measurements between 400-600 Hz correspond to mid vowels.
3. F1 measurements ≥ 700 Hz are indicative of low vowels.

In a nutshell, the traditional linguistic classification of vowels along the aperture continuum as **high**, **mid**, and **low** is supported by F1 formant measurements. The specifications above are based on the F1 formant of males. For females, these thresholds must be adjusted upward. Pavlovic (1987:415) suggests 16%, Stevens (2000:288) proposes 18%, while Heller (2013:370), Epps and Ambikairajah (2012:45) have 20%. I go with 20% because it makes calculations easier.

3.3 The JND of F1

Flanagan (1955:616), Mermelstein (1978:578), Rabiner and Juang (1993:152) have reported that the JND threshold at which F1 is optimally perceived by the naked ear is 60 Hz, as stated below:

JND of F1

> Of two speech signals A and B, the former is perceived as being auditorily different from the latter if and only if the acoustic distance between them is ≥ 60 Hz.

Rabiner and many of his co-researchers have appealed to this JND in their work on speech digitalization. Labov et al. (2006) have used it to measure the extent of dialect variation in *Atlas of North American English* and in other many other publications. A subset of this JND is the ≤ 20 Hz threshold. This JND indicates that humans cannot perceive any difference between speech sounds if they differ by less than 20 Hz on F1 and other higher formants (Flanagan 1955:616, Thomas 2011:56).

3.4 The RFL of F1

The JND above is applied to 187 vowel tokens (11x17) produced by 17 female speakers of American English. Only 7 vowel tokens are less than 60 Hz, which means that F1 is a robust correlate for 180 tokens. In other words, the RFL of F1 is 96.25%.

Vowels/F1	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Running Speech	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1F	405	544	489	734	924	845	692	614	648	417	790
Speaker 2F	406	529	487	722	837	797	557	551	573	421	721
Speaker 3F	361	376	421	604	892	878	505	525	603	365	657
Speaker 4F	412	547	464	689	937	843	547	526	571	468	783
Speaker 5F	415	493	436	673	827	742	526	473	525	487	657
Speaker 6F	379	503	465	732	885	738	637	414	551	508	759
Speaker 7F	393	575	570	760	952	897	640	637	632	476	804
Speaker 8F	398	351	443	523	610	768	506	457	572	398	691
Speaker 9F	417	489	424	394	846	776	614	562	527	421	604
Speaker 10F	396	467	447	643	791	669	625	599	512	417	622
Speaker 11F	369	450	427	695	771	788	631	596	489	401	640
Speaker 12F	465	496	484	583	767	741	594	510	494	391	575
Speaker 13F	420	425	415	600	753	743	507	508	495	411	562
Speaker 14F	410	495	482	676	824	779	618	531	550	461	747
Speaker 15F	396	489	467	597	762	758	495	456	512	372	588
Speaker 16F	426	535	501	606	856	769	587	532	531	430	637
Speaker 17F	411	493	435	587	732	733	539	467	527	430	630
Mean	404	485	462	636	821	780	577	526	547	427	674
St. Deviation	23	58	38	90	86	57	59	61	46	40	78

Table 4: F1of Vowels[8]

The high RFL of F1 underscores the fact that it plays an extremely important role in the intelligibility of vowels. Ladefoged and Johnson (2015:202) have singled out F1 as the most important of all formants because it alone accounts for 80% of the acoustic energy conveyed by vowels. Kent and Read (2002:33) explain why this is so:

> The first formant is typically the most intense formant, largely because of the interaction with the amplitudes 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 to the other formants. Loudness judgments of speech tend to be highly

---

[8] The measurements in Tables 3 through 5 are taken form Koffi and Krause (2020:60-85).

correlated with the amplitude of F1, which is not surprising given that this formant tends to be the strongest.

Researchers in all languages should expect to see a very high RFL for F1 since vowels usually carry a very high load in speech intelligibility.

4.0 Vowel Articulation and the Second Formant

The consensus in acoustic phonetics is that F2 correlates with the horizontal movement of the tongue. When it moves forward from its restful (i.e., neutral, see Figure 1) position, F2 values increase, when it retracts backward towards the pharyngeal wall, F2 values decrease (Thomas 2011:48). Lindblom and Sundberg (1971: 1175-1777) provide additional refinements. They note that a forward or backward movement of just ± 10 mm translates into significant F2 increases or decreases. The following thresholds help identify three articulatory landmarks for vowels:

1. F2 measurements ≥ 2000 Hz are indicative of front vowels.
2. F2 measurements are between 1800-1400 Hz are indicative of central vowels.
3. F2 measurements ≤ F2 1400 Hz are indicative of back vowels.

Again, it should be noted that females' formants are 20% higher than their male counterparts. These thresholds can form the basis for formulating phonological rules such as fronting, centralization, and retraction.

4.1 The JND of F2

There are slight differences with regard to the JND of F2. Mermelstein (1978:578) lists two JNDs, 151 Hz and 171 Hz, depending on the consonants that surround the vowel. Rabiner and Juang (1993:152) report a JND of 158 Hz, while Scharf (1961:215) has a JND of 200 Hz. Labov et al. (2006) used the JND of 200 Hz in *Atlas of North American English*. Since the 200 Hz JND lines up with the center frequency of critical bands (see Table 1), it should be the one used in the calculations of RFL in all languages. So, it is stated as follows:

JND of F2

Of two speech signals A and B, the former is perceived as being auditorily different from the latter if and only if the acoustic distance between them is ≥ 200 Hz.

4.2 The RFL of F2

When the JND of F2 is applied to the data in Table 5, we see that the 17 female participants produced a total of 187 F2 vowel tokens (11x17). In 50 instances (26.73%), masking occurs because the distances between tokens are less than 200 Hz, which means that if one were to listen to them with one's naked ear, one would not be able to detect any difference. In 137 other instances, the acoustic distances are greater than the JND of 200 Hz. This yields an RFL of 73.26%.

Vowels/F2	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Running Spch	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1F	2333	1886	2181	1590	1573	1406	1426	1388	1554	1718	1579
Speaker 2F	2473	1423	2391	1524	1397	1237	1013	1275	1392	1353	1461
Speaker 3F	2679	2008	2577	1883	1944	1575	1393	1295	1650	1698	1749

Speaker 4F	2501	2125	2612	1597	1776	1302	1332	1229	1362	1416	1606
Speaker 5F	2548	2025	2408	1369	1816	1480	1314	1168	1545	1042	1710
Speaker 6F	2566	1930	2552	1824	1721	1337	1531	1063	1478	1788	1575
Speaker 7F	2574	1996	2489	1734	1820	1511	1675	1617	1455	1856	1730
Speaker 8F	2260	1973	2386	1748	1726	1398	1294	1369	1571	1628	1612
Speaker 9F	2359	1819	2111	1866	1645	1340	1367	1358	1679	1620	1585
Speaker 10F	2322	2061	2309	1662	1626	1251	1340	1172	1562	1671	1491
Speaker 11F	2487	1990	2453	1733	1899	1459	1365	1235	1303	1699	1594
Speaker 12F	2160	1897	2074	1416	1757	1197	1102	1209	1358	1475	1473
Speaker 13F	2469	2134	2476	1688	1954	1435	1320	1310	1505	1633	1526
Speaker 14F	2451	1995	2333	1618	1684	1352	1265	1253	1319	1586	1487
Speaker 15F	2453	2045	2399	1625	1740	1492	1350	1195	1478	1315	1625
Speaker 16F	2324	1777	2290	1646	1665	1381	1414	1086	1329	1377	1473
Speaker 17F	2433	2047	2326	1715	1693	1355	1038	1366	1266	1598	1534
Mean	2434	1948	2374	1661	1731	1382	1325	1269	1459	1557	1577
St. Deviation	295	114	65	130	81	147	335	268	120	303	101

Table 5: F2 of Vowels in Running Speech

A quick glance at the "lot" vowel [ɑ] and the "cloth" vowel [ɔ] shows that 15 out 17 participants did not produce them intelligibly. This is not surprising because these two vowels have all but merged in the dialect of English spoken in Minnesota. The confusion analyses provided by Peterson and Barney (1952:182) and Hillenbrand et al. (1995:3108) support our findings, namely that these two vowels often mask each other in other dialects of American English.

4.2 Acoustic Vowel Spaces

A very practical way in which F1 and F2 measurements have contributed in enhancing linguistic analyses is that they have been used to generate acoustic vowel spaces such as the one in Figure 6 for many languages. By extracting F1 and F2 formant frequencies, one can import them into a vowel normalization suite such as Norm[9] and create an acoustic vowel space for any language. Ladefoged and Johnson (2015:103) explain the linguistic benefits of acoustic vowel spaces as follows:

> Vowel charts provide an excellent way of comparing different dialects of a language. 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.

---

[9] http://lingtools.uoregon.edu/norm/norm1.php retrieved on January 10, 2023.

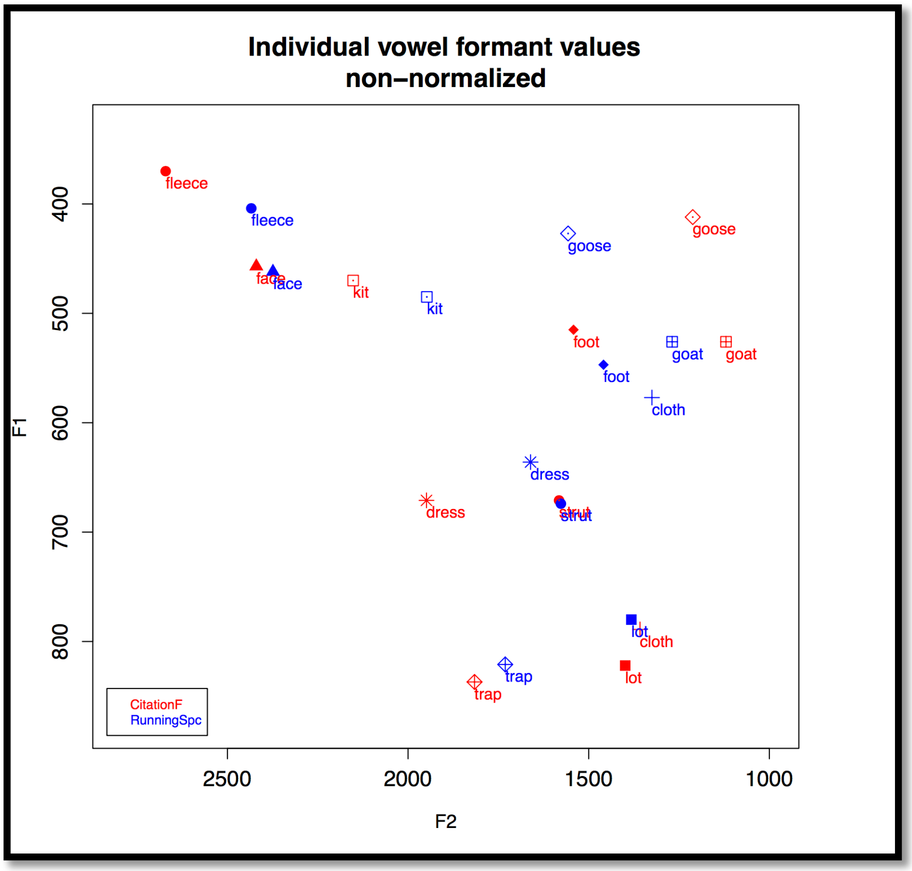


Figure 6: Acoustic Vowel Space for Females (Color Online)

Kent and Read (2002:111) note that F2 plays an important role in sociophonetics analyses because it is sensitive to dialectal variations. Figure 6 highlights one important way in which the dialect of English in Minnesota is different from other dialects of American English. We see for instance that the "kit" vowel [ɪ] (485 Hz), though classified as a high vowel in General American English is no longer so in the Minnesota dialect. It has lowered substantially to the point that it is lower than the "face" vowel [e] (462 Hz). This is true both in running speech and in words produced in citation form. Koffi (2016:2-14) also discusses the lowering of the "foot" vowel [ʊ]. This phenomenon also happens in California English (Ladefoged 1999:42).

5.0 Vowel Articulation and the Third Formant

There is still some uncertainty about the articulatory correlate of F3. The prevailing view is that F3 correlates with lip positions. Segments produced with unrounded lips have a greater F3 value than those produced with rounded lips. However, when we apply this received wisdom to the "trap" vowel [æ] and the "cloth" vowel [ɔ], we run into problems. The former is classified as unrounded, and the latter as a rounded. Yet, Peterson and Barney (1952:183) show that they have the same F3 value of 2410 Hz in male speech. Hillenbrand et al. (1995:3103) also show that the differences between these two vowels do not amount to much. Counterexamples such as these make the correlation between F3 and lip position tenuous, at best.

Delattre (1951:873) does not correlate F3 with lip position, but rather with the lowering of the velum. He contends that segments produced with a lowered velum have low F3. This interpretation is compatible with Lindblom and Sundberg (1971:1176). For example, when /k/ and /g/ are produced, F3 values are so low that that they often intersect with F2, which leads to the so-called "velar pinch." The lowering of the velum may be a better explanation for why [æ] and [ɔ] have similar low F3. Bradley (2018:382-3) makes the same observation that lowering the larynx decreases the value of F3. Choi (2012:39) offers another articulatory correlate of F3 that looks appealing. He points to the blade of the tongue. He contends that segments produced with the blade of the tongue up have a higher F3 value than those that do not involve the blade of the tongue. In other words, F3 correlates with the phonetic feature [±coronal]. This insight is powerful for explaining why laterals tend to have higher F3 values than rhotics.

5.1 The JND of F3

Scharf (1961:215) estimates the JND of F3 to be 400 Hz, while Rabiner and Juang (1993:152) report a JND of 355 Hz. Again, we appeal to the center frequency bandwidths in Table 1 in stating the JND of F3 as follows:

JND of F3

> Of two speech signals A and B, the former is perceived as being auditorily different from the latter if and only if the acoustic distance between them is ≥ 400 Hz.

5.3 The RFL of F3

When we use this JND to gauge the intelligibility of the vowels in Peterson and Barney (1952:183) and Hillenbrand et al. (1995:3103), we see that it is not a very robust correlate. The same is true for our data. The 17 participants produced 187 F3 vowel tokens (11x17). Of these, 158 (84.49%) mask each other, meaning that the acoustic difference between them is less than 400 Hz. In other words, there are only 29 instances where the acoustic distance between pairs of phonetically similar vowels is ≥ 400 Hz. It can be concluded that F3 plays a marginal role in the intelligibility of English vowels because its RFL is only 15.50%.

Vowels/F3	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Running Spch	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1F	2926	2895	2514	2532	2444	2504	2046	2886	3017	2734	2579
Speaker 2F	2864	2650	2845	2272	2371	2634	2353	2827	2766	2679	2486
Speaker 3F	3357	3103	3202	2817	2968	2818	2511	3259	3335	3200	2970
Speaker 4F	3148	2081	3024	2332	2386	2647	2536	2953	3098	2954	2333
Speaker 5F	3039	2544	2936	2452	2572	2538	2323	2710	2738	2893	2415
Speaker 6F	3172	2780	2999	2750	2713	2507	2552	3190	2876	3152	2373
Speaker 7F	3061	2971	3037	2603	2551	2818	2545	3009	2716	2986	2665
Speaker 8F	2801	2797	2804	2518	2416	2816	2535	3030	2950	1757	2749
Speaker 9F	2599	2516	2560	2627	2493	2382	1952	2696	2594	2678	2468
Speaker 10F	2824	2792	2746	2735	2569	2238	2398	2717	2566	2682	2364
Speaker 11F	3034	2781	2862	2423	2697	2240	2442	2848	2685	3004	2334
Speaker 12F	2815	2969	2653	2649	2809	2538	2553	2747	2718	2979	2657
Speaker 13F	2943	2958	3037	2931	2942	2718	2692	2866	2928	2855	2829
Speaker 14F	2869	2709	2666	2606	2719	2535	2198	2889	2842	2715	2555
Speaker 15F	2882	2777	2799	1914	2775	2582	2462	2976	2854	2676	2454
Speaker 16F	2907	2876	2786	2508	2597	2560	2307	2980	2838	2339	2566
Speaker 17F	3324	2648	2782	2462	2321	2705	2359	3042	2890	3230	2720
Mean	2974	2755	2838	2537	2608	2575	2397	2919	2847	2794	2559
St. Deviation	195	233	184	233	195	176	192	161	188	352	184

Table 6: F3 of Vowels in Running Speech

6.0 Vowel Articulation and the Fourth Formant

So far, there is no clear anatomical correlate for F4. In the following quote, Ladefoged (2006:205-6) hints that it may have something to do with the size of the speaker's head,

> No simple technique will enable one to average out the individual characteristics so that a formant plot will show only the phonetic qualities of the vowels. One way to deal with this problem is probably to regard the average frequency of the fourth formant as an indicator of the individual's head size and then express the values of the other formants as percentages of the mean fourth formant frequency. But this possibility is not open when the fourth formant frequencies have not been reported for the sets of the vowels being compared.

My best efforts to find supporting evidence for Ladefoged's claim have been fruitless. Instead, Lindblom and Sunberg (1971:1176) hint at correlating F4 with the pharynx, "The net effect of the pharynx lowering on F3 and F4 is to decrease the frequency distance between them." Bele (2005:567, 574) opines that "F4 frequency will be dependent both on vocal tract length and the larynx tube configuration." Takemoto et al. (2006:2228) also write that, "Although previous studies have reported that the fourth F4 or fifth F5 formant is sensitive to the laryngeal cavity shape e.g., Fant, 1960, it is still unclear how the laryngeal cavity generates such a formant." In other words, the anatomical bases of F4 remain elusive.

6.1 The JND of F4

Many experts, including Stevens (2000: 154, 300), Fastl and Zwicker (2007:235-5), Everest and Pohlmann (2015:13), and Scharf (1961:215) estimate the JND of F4 to be 600 Hz. Rabiner and Juang (1993:186) are the only ones who report a JND of 480 Hz. We side with the majority view and state the JND of F4 as follows:

JND of F4

Of two speech signals A and B, the former is perceived as being auditorily different from the latter if and only if the acoustic distance between them is ≥ 600 Hz.

6.2 The RFL of F4

Ladefoged (2006:205-6) laments the fact that very few studies report F4 values. Indeed, earlier studies such as Peterson and Barney (1952) did not measure F4. Lindblom and Sundberg's (1971) paper is one of the few early studies that have provided measurements for F4. We find F4 measurements also in Hillenbrand et al. (1995:3103). When the JND of F4 is applied to either Lindblom and Sundberg (1971:1176) or Hillenbrand et al. (1995:3103), we see that almost all vowel pairs do mask each other. Clearly, F4 does not contribute much to the intelligibility of English vowels. The same observation holds true in Table 7.

Vowels/F4	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Running Speech	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1F	3102	3050	3251	2942	3128	2972	3212	3381	3458	3394	3147
Speaker 2F	3991	3609	2730	3123	3305	3573	3768	3952	3938	4255	3482
Speaker 3F	4026	4076	3860	3701	3944	3983	3738	3845	3786	3887	3822
Speaker 4F	4081	2958	3046	3446	3462	3449	3214	3847	3862	4122	3646
Speaker 5F	3631	3380	3267	3229	3396	3431	3522	3762	3872	3527	3480
Speaker 6F	3583	3781	3619	3630	3660	3143	3463	3602	3845	3574	3286
Speaker 7F	4047	4057	4071	3629	3849	3557	3471	3976	3933	3996	3695
Speaker 8F	4007	4006	3974	4003	3975	3725	3857	3689	3810	3875	3652
Speaker 9F	3766	3414	3230	3288	3447	3481	3490	3757	3837	3757	3515
Speaker 10F	4137	4084	4158	4105	4329	4106	4178	4174	4198	4090	4151
Speaker 11F	4131	4200	4028	3872	4038	3857	3842	3908	3910	3845	4119
Speaker 12F	3850	3924	3613	3706	3496	3728	3554	3687	3860	3940	3732
Speaker 13F	4456	4241	4383	3886	3814	3548	3928	4179	4304	4259	3867
Speaker 14F	3652	3921	3533	3416	3351	3234	3474	3812	3455	3393	3802
Speaker 15F	4466	4424	4357	3935	3747	3764	3906	2972	4102	4129	3918
Speaker 16F	4091	3820	3925	3993	3851	3588	4010	3802	4019	4053	3770
Speaker 17F	4301	4353	4153	4090	3615	2762	4045	4328	4301	4249	4469
Speaker 18F	4240	4196	4332	3889	4209	3702	4043	4013	4361	4302	3860
Speaker 19F	3951	3770	3843	3861	3684	3551	3814	3483	3512	4305	3670
Mean	4022	3900	3784	3711	3731	3565	3739	3821	3939	3975	3774
St. Deviation	327	413	476	342	319	330	283	307	264	299	305

Table 7: F4 of Vowels

The 19 participants produced a total of 209 (19 x11) F4 vowel tokens.[10] Only six vowel tokens have acoustic distances greater than the JND of ≥ 600 Hz, which means that 203 tokens fall below the threshold of intelligibility. Therefore, the RFL of F4 is 2.87%. RFLs this low are deemed statistically insignificant by Catford (1987:89-90). This may explain why F4 measurements are often not extracted.

7.0 Vowel Articulation and the Fifth Formant

The anatomical correlate of F5 is even murkier than that of F4. Experts hedge and place heavy caveats on their pronouncements. For example, Takemoto et al. (2006:2228) hint at the larynx as a possible source of F5. Bele (2005:560, 567) notes that F4 is often indistinguishable from F5 because they appear like a cluster. Since this explanation seems unsatisfactory, I queried ChatGPT for an answer. It gave me the following result: "The fifth formant, or F5, is found at an even higher frequency than the fourth formant. It is associated with the shape of the oral cavity and the nasal cavity."[11] A similar statement is found in Ladefoged and Johnson (2015:223) who write that "The position of the fourth and higher formants in most vowels is indicative of the speaker's voice quality rather than the linguistic aspects of the sound. Similarly, the exact location of the higher formants in nasals depend to a great extent on individual physiological characteristics of the speaker." For the time being, the best answer regarding the anatomical correlate of F5 is "nobody knows."

---

[10] The data from the first 17 participants comes Koffi and Krause (2020). The measurements from the last two are from data that has been collected since then.

[11] This is in response to the following query: "Explain to me the formant frequencies of the fourth and fifth formants." Information retrieved on January 18th , 2024.

7.1 The JND of F5

F5 measurements are often unavailable because researchers do not bother extracting them. Even so, Hawk (1994:1074, 1081) indicates that they can be extrapolated from F4 values by increasing them by 13%, as shown in Table 8. The JND of F5 is stated as follows:

JND of F5

> Of two speech signals A and B, the former is perceived as being auditorily different from the latter if and only if the acoustic distance between them is ≥ 800 Hz.

7.2 The RFL of F5 in our Data

Since F5 measurements of English vowels are not available in the existing literature, those in Table 8 are extrapolated from the F4 measurements in Table 7.

Vowels/F5	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Running Speech	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1F	3505	3446	3673	3324	3534	3358	3629	3820	3911	3835	3556
Speaker 2F	4509	4078	3084	3528	3734	4037	4257	4465	4449	4808	3934
Speaker 3F	4549	4605	4361	4182	4456	4500	4223	4344	4278	4392	4318
Speaker 4F	4611	3342	3441	3893	3912	3897	3631	4347	4364	4657	4119
Speaker 5F	4103	3819	3691	3648	3837	3877	3979	4251	4375	3985	3932
Speaker 6F	4048	4272	4089	4101	4135	3551	3913	4070	4344	4038	3713
Speaker 7F	4573	4584	4500	4100	4349	4019	3922	4492	4444	4515	4175
Speaker 8F	4527	4526	4490	4523	4491	4209	4358	4168	4305	4378	4126
Speaker 9F	4255	3857	3649	3715	3895	3933	3943	4245	4388	4245	3971
Speaker 10F	4674	4614	4698	4638	4891	4639	4721	4716	4743	4621	4690
Speaker 11F	4668	4746	4551	4375	4038	4358	4341	4416	4418	4344	4654
Speaker 12F	4350	4434	4082	4187	3950	4212	4016	4166	4361	4452	4217
Speaker 13F	5035	4792	4952	4391	4309	4009	4438	4722	4863	4812	4369
Speaker 14F	4126	4430	3992	3860	3786	3654	3925	4307	3904	3834	4296
Speaker 15F	5046	4999	4923	4446	4234	4253	4413	3358	4635	4665	4427
Speaker 16F	4622	4316	4435	4512	4351	4054	4531	4296	4541	4579	4260
Speaker 17F	5591	4918	4692	4621	4084	3121	4570	4890	4860	4801	5049
Speaker 18F	4791	4741	4895	4394	4756	4183	4568	4534	4926	4861	4361
Speaker 19F	4464	4260	4342	4362	4162	4012	4309	3935	3968	4864	4147
Mean	4528	4356	4238	4147	4152	3993	4194	4291	4425	4457	4227
St. Deviation	441	467	535	387	349	373	320	347	297	338	345

Table 8: Extrapolated F5 Values of Vowels

We see that the 19 participants produced a total of 209 F5 vowel tokens. Of these, only five vowel pairs exceed the acoustic distances of JND ≥ 800 Hz. This means that the RFL of F5 amounts to only 2.3%, which is statistically insignificant.

7.3 Interim Summary of Formants and Speech Intelligibility

The RFLs of F0, F1, F2, F3, F4, and F5 from the previous sections are summarized and displayed in Figure 7 as follows:

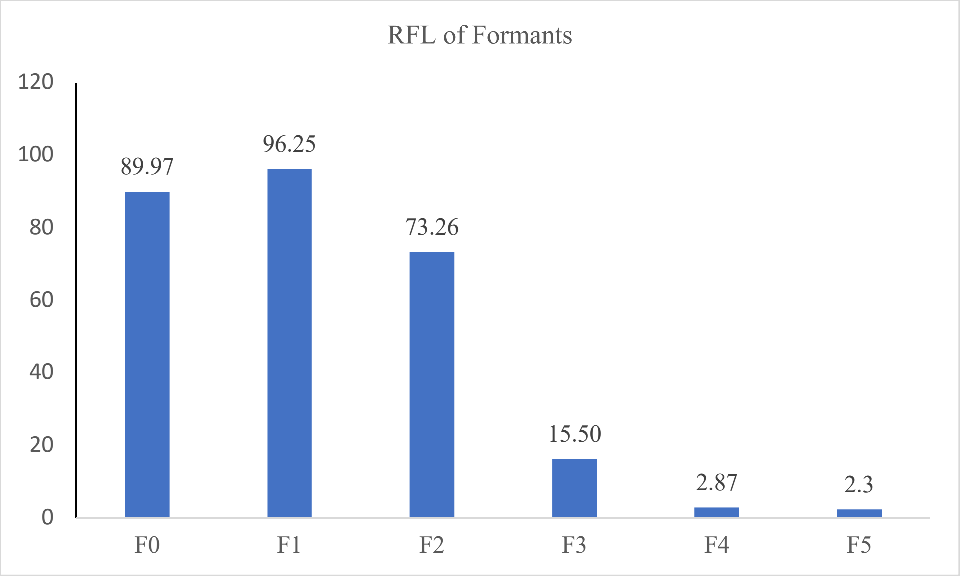


Figure 7: RFL of Formants

We see clearly that F1 and F2 have greater RFLs than F3, F4, and F5. The RFL of F1 agrees with Ladefoged and Johnson's (2015:202) observation that it alone accounts for 80% of the acoustic energy in vowels. In the dialect of American English examined in this paper, together, F1 and F2 account for 89.13% of the intelligibility of vowels. This explains why they are the two formants that are often extracted. This means that acoustic phonetic analyses of vowels in any language must obligatorily include F1 and F2 measurements.

8.0 Paralinguistic Significance of Formants

In describing languages for linguistic consumption, efforts should be placed on F1 and F2 measurements. For paralinguistic usages such as speech synthesis and voice biometrics, one will do well to extract F3, F4, and even F5 measurements. This view is in line with Reetz and Jongman (2009:184) who state that,

> Formant frequencies higher than F3 are not important cues to the identity of a vowel because they hardly vary as a function of vowel quality. Instead, for vowels, the frequencies of higher formants such as F4 and F5 seem more speaker-specific and may therefore provide information about the identity of the speaker rather than the vowel itself.

This quote indicates that higher formants are particularly well-suited for voice biometric research. Ladefoged and Johnson (2015:223) say as much in noting that higher formants "depend to a great extent on individual physiological characteristics of the speaker." For formant-based synthesis, Klatt (1980) measured F1, F2, and F3. Koffi and Petzold (2022) recommend including F4. Since information about formant-based synthesis has been around for nearly 50 years, in the remainder of the paper voice biometrics is used to exemplify an area where formants are paralinguistically relevant.

8.1 Formants and Voice Biometrics

Koffi (2023) has devoted a full paper to voice biometrics. Yet, the approach taken here is slightly different in that, here, the emphasis is on the role of formants in speaker verification. Fourcin and Abberton (2009:40) note that acoustic phonetic measurements can be used for two different purposes: either for analytical precision or for speech intelligibility. The measurements to be discussed here have already been used to determine whether or not three brothers sound alike (Koffi and Lyons 2018:85-88). It has been claimed that the three brothers are auditorily indistinguishable from each other. In fact, their own parents confuse them, and so do their friends and acquaintances. The F1, F2, and F3 measurements extracted from their speech were first used to show that the three brothers do indeed sound alike. However, in this paper, their F4 and F5 are used to show that, from the standpoint of voice biometrics, their measurements can be used to distinguish between them. For this demonstration, we focus on their vowels /a, i, u/. We also focus only on these vowels because no human language exists that lacks these three vowels (Maddieson 1984:126, 142). In the original paper, F4 and F5 were not extracted. However, they are extrapolated from F3. F4 is obtained by increasing F3 values by 8.3%. F5 is also obtained by increasing F4 by 13% (Hawk 1994:1074, 1081).

Words		**heed**	**hod**	**who'd**
Vowels		**[i]**	**[ɑ]**	**[u]**
Andrew	F4	2980	2781	2462
Donovan	F4	3440	2708	2722
James	F4	3174	2839	2556
Andrew	F5	3367	3144	2782
Donovan	F5	3887	3060	3075
James	F5	4392	3208	2888

Table 9: Formants and Speaker Identification

Cao and Dellwo (1994) have shown that F1, F2, F3, F4, and F5 can be combined in many ways for forensic speaker identification. Even though this is possible, the demonstration below relies only on using Euclidean Distance (ED) calculations to determine whether speakers sound alike or dissimilar. All things being equal, speakers who sound alike will have a smaller ED between them, whereas those who are dissimilar will have a greater distance. The ED formula is stated as below and is illustrated with the F4 and F5 of the vowels /i/ and /a/.

$$ED = \sqrt{(F4/i/-F4/a/)^2 + (F5/i/-F5/a/)^2}$$

Here is an explanation of the formula. Suppose we want to calculate the ED between the vowels [i] and [a] produced by Andrew. We subtract the F4 of his [i] (2980 Hz) from the F4 of his [a] (2781Hz) which is 199 Hz. The square root of 199^2 is 39,601. We do same for the F5 his [i] (3367 Hz) and the F4 of his /ɑ/ (3144 Hz). It yields 233 Hz. The square root of 233^2 is 49,729. We average the sum of square roots of the F4 and F5 of [i] 39601 and [a] 49729. The sum is 89330. Then we proceed with calculating the square root of 89330, which is 298 Hz. So, the ED between [i] and [a] in Andrew's pronunciation is 298 Hz. We do the same for [i] and [u], and [u] and [a]. For any two vowel pairs, an ED can be calculated and used for speaker verification. The results for the three brothers are displayed in Table 10.

Vowels Pairs		**[i]-[a]**	**[i]-[u]**	**[u]-[a]**	**ED [i-a]**	**ED [i-u]**	**ED [u]-[a]**	**ED AVG**
Andrew	F4	2980-2781	2980-2462	2462-2781	298 Hz	666 Hz	482 Hz	482 Hz
Andrew	F5	3367-3144	3367-2782	2782-3144				
Donovan	F4	3440-2708	3440-2722	2722-2708	1,104 Hz	1,083 Hz	20 Hz	735 Hz
Donovan	F5	3887-3060	3887-3075	3075-3060				
James	F4	3174-2839	3174-2556	2556-2839	1230 Hz	1,626 Hz	472 Hz	1,109 Hz
James	F5	4392-3208	4392-2888	2888-3208				

Table 10: Euclidean Distance in F4-F5 Domain

For the purposes of analysis, ED across all vowels have been averaged in the last column. Assessments of similarity and dissimilarity are based on how close or distant the averaged values are. The confusion matrix in Table 11 calculates the differences between the three brothers:

/i-u-a/ F4-F5	**Andrew**	**Donovan**	**James**
Andrew	0	253	627
Donovan	253	0	374
James	627	374	0

Table 11: Comparative Euclidean Distances

We see that ED discriminates between the three brothers in significant ways. Consequently, a voice biometric system that relies on analytical accuracy will never mistake the three brothers. This also means that if one of the brothers commits a crime while speaking, the other brothers will not be found guilty if the voice sample is analyzed even though an earwitness can confuse them. It also means that if the three brothers do banking with their voice, one cannot mistakenly access the account of the other. Even so, the ED calculations show that Andrew (253 Hz) and Donovan (374 Hz) are closer to each other than Donovan (374 Hz) and James (627 Hz). Their respective EDs are 121 Hz and 253 Hz. We also see that James is more easily distinguishable from his brothers. Artificial Intelligent (AI) systems such as those used for national security or banking rely on more sophisticated algorithms than ED. Koffi (2023:81-2) lists more than a dozen algorithms used in speaker verification.

9.0 Summary

The goals set forth in this paper have been reached in three important ways. First, the physiological bases of formants have been described in keeping with their respective resonance frequencies. The summary of the literature indicates clearly that F1 and F2 have strong anatomical bases. The anatomical correlates of F3 are less assured, but the emerging consensus is that this formant correlates with lip movements or the lowering of the velum. Unfortunately, there is no clear and definitive physiological bases for F4 and F5. Yet here also there is a firm consensus, namely that higher formants provide speaker idiosyncratic pieces of information that are particularly important in voice biometrics. The paper has underscored the linguistic relevance of formants by showing that they contribute significantly to speech intelligibility. The RFL calculations lend support to why F1 and F2 formant data have traditionally been extracted. Their combined RFLs show that they convey 89.13% of the speech intelligibility load in vowels. The RFL of F3 is only 8.53%. When F3 is used in in tandem with F1 and F2, they account for 97.29% of the intelligibility load in vowels. This explains why a linguistic analysis must extract values from all three formants. This also explains why Klatt (1980) extracted all three formants for speech synthesis. Our data confirms that F4 and F5 do not contribute much to intelligibility, only

2.71%. Yet, they are worth extracting because they carry a very important paralinguistic load that can be used in speaker verification.

ABOUT THE AUTHOR

Ettien Koffi, Ph.D. linguistics (Indiana University, Bloomington, IN) teaches at Saint Cloud State University, MN. He is the author of five books and author/co-author of several dozen articles on acoustic phonetics, phonology, language planning and policy, emergent orthographies, syntax, and translation. His acoustic phonetic research is synergetic, encompassing L2 acoustic phonetics of English (Speech Intelligibility from the perspective of the Critical Band Theory), sociophonetics of Central Minnesota English, general acoustic phonetics of Anyi (a West African language), acoustic phonetic feature extraction for application in Automatic Speech Recognition (ASR), Text-to-Speech (TTS), voice biometrics for speaker verification, and infant cry bioacoustics. Since 2012, his high impact acoustic phonetic publications have been downloaded **68,201** times (**47,355** as per Digital Commons analytics), (**20,846** as per Researchgate.net analytics), and several thousand downloads from Academia.edu, as of **February 2024**. He can be reached at enkoffi@stcloudstate.edu.

References

Alkahby, H.Y., M.A Mahrous, and B. Mamo. 1999. Mathematical Model for the Basilar Membrane as a Two-Dimensional Plate. 15th Annual Conference of Applied Mathematics. *Electronic Journal of Differential Equations*., pp. 115-124.

Baken, R. J. and Robert F. Orlikoff. 2000. *Clinical Measurement of Speech and Voice*. Second edition. San Diego, CA. Singular Publishing Group-Thomson Learning.

Bele, Irene Velsvik. 2005. The Speaker's Formant. *Journal of Voice* 20 (4): 555-578.

Boersma, Paul & Weenink, David 2016. Praat: doing phonetics by computer [Computer program]. Version 6.0.17, retrieved 21 April 2016 from http://www.praat.org/

Bradley Evan D. 2018. A Comparison of the Acoustic Vowel Spaces of Speech and Song. *Linguistic Research* 35 (2): 381-394.

Choi, Jeung-Yoon. 2012. Distinctive Features in the Representation of Speech and Acoustic Correlates. In *Advanced Topics in Biometrics*, pp.31-44. Ed by Haizhou Li, Kar-Ann Toh, Liyuan L. New Jersey: World Scientific.

Chung, Hyunju. 2020. Vowel Acoustic Characteristics of Southern American English Vowel Variation in Louisiana. *Journal of the Acoustical Society of America* 147 (1):541-553.

Crystal Thomas H and Arthur S. House. 1987. Segmental Durations in Connected Speech Signals: Current Results. *JASA* 83 (4):1553-1573.

Delattre, Pierre. 1951. The Physiological Interpretation of Sound Spectrograms *Proceedings of MLA*, Volume 66 (5):864-875.

Epps, Julien and Eliathamby Ambikairajah. 2012. Speech Characterization and Feature Extraction for Speaker Recognition. In *Advanced Topics in Biometrics*, pp.45-69. Ed by Haizhou Li, Kar-Ann Toh, Liyuan L. New Jersey: World Scientific

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

Fant, Gunnar. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton & Co.-S-Gravenhage.

Fastl, Hugo and Eberhard Zwicker. 2007. *Psycho-Acoustics: Facts and Models*. 3rd edition. Munich, Germany: Springer

Fletcher, Harvey. 1953. Speech and Hearing in Communication. New York: D. Van Nostrand Company, Inc.

Fogerty, Daniel and Larry E. Humes. 2012. The Role of Vowel and Consonant Fundamental Frequency, Envelop, and Temporal Fine Structure Cues to the Intelligibility of Words and Sentences. Journal of the *Acoustical Society of America* 131 (2): 1490-1501.

French, Peter and Louisa Stevens. 2015. Forensic Speech Science. Research Methods in Speech Acoustics. *The Bloomsbury Companion to Phonetics*, pp. 183-197. Ed by. Mark J. Jones and Rachael-Anne Knight. New York: Bloomsbury.

French N. R and J.C. Steinberg. 1947. Factors Governing the Intelligibility of Speech Sounds. Journal of the *Acoustical Society of America* 19 (1): 90-119.

Fromkin, V. R. Rodman, and N. Hyams. 2014. An Introduction to Language. 10th edition. Boston, MA: Cengage Learning.

Fourcin, Adrian and Evelyn Abberton. 2008. Hearing and Phonetic Criteria in Voice Measurement: Clinical Application. *Logopedics and Phoniatrics Vocology* 33: 35-48.

Gahl, Susanne, Yao Yao, and Keith Johnson. 2012. Why Reduce? Phonological Neighborhood Density and Phonetic Reduction in Spontaneous Speech. *Journal of Memory and Language* (Article in Print): 1-18.

Gordon, Matthew and Peter Ladefoged. 2001. Phonation Types: A Cross Linguistic Overview. *Journal of Phonetics* 29 (4):383-406.

Hawks, John W. 1994. Difference Limens for Formant Patterns of Vowels Sounds. *Journal of the Acoustical Society of America* 95 (2): 1074-1084.

Johnson, Alexander. 2022. An Integrated Approach for Teaching Speech Spectrogram Analysis to Engineering Students. *Journal of the Acoustical Society of America* 152 (3): 1962-1969/

Kent, Ray D. and Charles Read. 2002. *Acoustic Analysis of Speech*. Second Edition. Clifton Park, NY: Delmar Cengage Learning.

Kewley-Port, Diane, T. Zachary Burkle, and Jae Hee Lee. 2007. Contribution of Consonant versus Vowel Information to Sentence Intelligibility for Young Normal-Hearing and Elderly Hearing-Impaired Listeners. *Journal of the Acoustical Society of America* 122 (4):2365-2375.

Kinnunen, Tomi. 2003. *Spectral Features for Automatic Text-Independent Speaker Recognition*. Licentiate's Thesis. Joensuu, Finland.

Klangpornkun, Nittayapa, Chutamanee Onsuwan, and Charturong Tantibundhit. 2013. Predictions from Speech Banana Audiograms: Assessment of Hearing Deficits in Thai Hearing Loss Patients. *Proceeding of Meetings on Acoustics* 20:1-10

Koffi, Ettien and Jessica Krause. 2020. Speech Style Variation of Vowels in Citation Form vs. Running Speech: Intelligibility Implications for AI-Enabled Devices. *Linguistic Portfolios* 9: 60-85.

Koffi, Ettien. 2018. Differential Analysis of Lexical Pitch in Accent and Tone Languages. *Linguistic Portfolios* 7: 110-131.

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

Koffi, Ettien. 2014. The Acoustic Vowel Space of Central Minnesota English in Light of the Northern Cities Vowel Shift. *Linguistic Portfolios* 3: 2-20.

Labov, William; Sharon Ash, and Charles Boberg. 2006. The Atlas of North American English: Phonetics, Phonology, and Sound Change.

Labov, William; Ingrid Rosenfelder; and Josef Fruehwadl. 2013. One Hundred Years of Sound Change in Philadelphia. *Language* 48. 30-65.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.

Ladefoged, Peter. 2006. *A Course in Phonetics.* Fifth Edition. Boston, MA: Thomson-Wadsworth.

Ladefoged, Peter and Keith Johnson. 2015. *A Course in Phonetics.* Seventh Edition. Malden, MA: Cengage Learning.

Ladefoged, Peter, Iris Kemeny, and William Brackenridge. 1976. Acoustic Effects of Style of Speech. *Journal of the Acoustical Society of America* 59 (1):228-231.

Ladefoged, Peter and D.E. Broadbent. 1957. Information Conveyed by Vowels. *Journal of the Acoustical Society of America* 29 (1):98-104.

Lindblom, Bjorn E. F. and Johan E. F. Sundberg. 1971. Acoustical Consequences of Lip, Tongue, Jaw, and Larynx Movement. *Journal of the Acoustical Society of America* 50 (4): 1166-1179.

Lisker, Leigh and Arthur Abramson. 1964. A Cross-Language Study of Voicing in Initials Stops: Acoustical Measurements. *Word 20*: 384-422.

Mermelstein, Paul. 1978. Difference Limens for Formant Frequency of Steady-state and Consonant-bound Vowels. *Journal of the Acoustical Society of America* 63 (2):527-680.

Moon, Seung-Jae and Bjorn Lindblom. 1994. Interactions between Duration, Context, and Speaking Style in English Stressed Vowels. *Journal of the Acoustical Society of America* 95 (1):40-55.

O'Shaughnessy, Douglas. 2023. How Do Computers Understand Human Speech? *Acoustics Today*, Volume 19, Issue 4:32-40.

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. *Handbook of Acoustics*, pp. 1341-1354. Ed by. Malcom J. Crocker. New York: A Wiley-Interscience Publication, John Wiley & Sons, Inc.

Rabiner, Lawrence and Biing-Hwang Juang. 1993. *Fundamentals of Speech Recognition*. Englewood Cliffs, New Jersey: Prentice Hall.

Reetz, Henning and Allard Jongman (2009:184). *Phonetics: Transcription, Production, Acoustics, and Perception*. Malden: MA: Wiley-Blackwell.

Repp, Bruno. 1987. Integration and Segregation in Speech Perception. *Haskins Laboratories Report on Speech Research SR 89/90*, pp. 1-32.

Ryant, Neville and Mark Liberman. 2016. Automatic Analysis of Phonetic Speech Style Dimensions. *Interspeech* 2016: 77-81.

Russo, Frank A. Music Beyond Sound: Weighing the Contributions of Touch, Sight, and Balance. 2020. *Acoustics Today* 16 (1): 37-45.

Scharf, Bertram. 1961. Complex Sounds and Critical Bands. *Psychological Bulletin* 58 (3):205-215.

Stevens, Kenneth. 2000. *Acoustic Phonetics*. MIT Press.

Takemoto, Hironori, Parham Mokhtari, and Tatsuya Kitamura. 2010. Acoustic Analysis of the Vocal Tract during Vowel Production by Finite-Difference Time-Domain Method. *Journal of the Acoustic Society of America* 128: 3724-3738.

Thomas, Erik R. 2011. *Sociophonetics: An Introduction*. New York: Palgrave Macmillan.

Thomas, Erik R. and Tyler Kendall. 2007. NORM: The vowel normalization and plotting suite. [Online Resource: http://ncslaap.lib.ncsu.edu/tools/norm/]

Watt, Dominic. 2015. Research Methods in Speech Acoustics. *The Bloomsbury Companion to Phonetics*, pp. 79-97. Ed by. Mark J. Jones and Rachael-Anne Knight. New York: Bloomsbury.

Weinberger, Steven. (2019). *Speech Accent Archive*. George Mason University. Retrieved from http://accent.gmu.edu

Yaeger, Malcah. 1975. Speaking Style: Some Phonetic Realizations and their Significance. *Pennsylvania Working Papers on Linguistic Change* 1:1-61.

Zahorian, Stephen A. and Amir Jalali Jagharghi. 1993. Spectral-shape Features versus Formants as Acoustic Correlates for Vowels. *Journal of the Acoustical Society of America* 94 (4) 1966-1982.

Male Measurements of Vowels in Running Speech

The measurements for the first five speakers are the same as found in Koffi and Krause (2020:81-85). Since then additional data has been collected from male speakers.

Vowels/F0	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Running Speech	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1M	115	127	106	117	104	95	101	134	120	108	110
Speaker 2M	110	109	101	113	100	99	108	110	131	117	120
Speaker 3M	145	137	147	127	127	119	121	191	139	133	128
Speaker 4M	98	84	92	108	94	93	84	116	102	100	103
Speaker 5M	82	81	87	82	81	77	74	83	90	82	87
Mean	110	107	106	109	101	96	97	126	116	108	109
St. Deviation	23	25	23	16	16	15	18	40	20	19	15

Appendix 1: F0 of Vowels in Running Speech

Vowels/F1	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Running Speech	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1M	317	437	396	564	757	721	529	461	518	384	613
Speaker 2M	337	407	492	529	747	818	812	606	443	439	591
Speaker 3M	332	413	390	464	666	585	505	391	392	362	527
Speaker 4M	300	391	379	396	522	565	481	371	366	302	471
Speaker 5M	308	392	363	433	515	671	574	399	397	324	483
Mean	318	408	404	477	641	672	580	445	423	362	537
St. Deviation	15	18	50	68	117	103	134	95	59	53	63

Appendix 2: F1 of Vowels in Running Speech

Vowels/F2	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Running Speech	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1M	2124	1755	2140	1465	1560	1220	1171	940	1348	1355	1243
Speaker 2M	2240	1784	2122	1636	1710	1388	1776	1572	1534	1904	1409
Speaker 3M	2051	1600	2024	1480	1512	1065	871	979	1287	1230	1251
Speaker 4M	1492	1505	1984	1285	1518	1011	1135	954	1228	1851	1128
Speaker 5M	2128	1651	2080	1557	1546	1189	1350	1043	1269	1438	1303
Mean	2007	1659	2070	1484	1569	1174	1260	1097	1333	1555	1266
St. Deviation	295	114	65	130	81	147	335	268	120	303	101

Appendix 3: F2 of Vowels in Running Speech

Vowels/F3	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Running Speech	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1M	2889	2482	2838	2526	2366	2287	2428	2755	2606	2571	2401
Speaker 2M	2835	2067	2769	2699	2603	2655	2955	2940	2715	2786	2529
Speaker 3M	2603	2450	2528	2568	2207	2205	2191	2543	2483	2607	2291
Speaker 4M	2651	2779	*3068*	2839	2834	2711	2753	*2668*	1693	3128	2831
Speaker 5M	2577	2539	2587	2507	2449	2471	2670	2781	2523	2483	2529
Mean	2711	2463	2758	2627	2491	2465	2599	2737	2404	2715	2516
St. Deviation	141	256	214	139	238	221	296	146	407	255	202

Appendix 4: F3 of Vowels in Running Speech

Vowels/F4	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Running Speech	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1M	3513	3386	3366	3311	3222	3069	3392	3292	3452	3532	3276
Speaker 2M	3973	3898	3764	3801	3861	3727	4084	4186	3717	3827	3871
Speaker 3M	3542	3646	3423	3428	3853	4023	3438	3499	3310	3492	3612
Speaker 4M	3581	3889	3909	3776	3956	3796	3764	3569	3731	4041	3792
Speaker 5M	3716	3618	3742	3685	3720	3427	3808	3662	3480	3422	3838
Mean	3665	3687	3640	3600	3722	3608	3697	3641	3538	3662	3677
St. Deviation	188	213	234	218	292	369	285	333	181	261	245

Appendix 5: F4 of Vowels in Running Speech

Vowels/F5	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Running Speech	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1M	3513	3386	3366	3311	3222	3069	3392	3292	3452	3532	3276
Speaker 2M	3973	3898	3764	3801	3861	3727	4084	4186	3717	3827	3871
Speaker 3M	3542	3646	3423	3428	3853	4023	3438	3499	3310	3492	3612
Speaker 4M	3581	3889	3909	3776	3956	3796	3764	3569	3731	4041	3792
Speaker 5M	3716	3618	3742	3685	3720	3427	3808	3662	3480	3422	3838
Mean	3665	3687	3640	3600	3722	3608	3697	3641	3538	3662	3677
St. Deviation	188	213	234	218	292	369	285	333	181	261	245

Appendix 5: F5 Estimates of Vowels in Running Speech