Linguistic Portfolios

Volume 12 | Article 10

2023

THE ACOUSTIC PHONETIC PROPERTIES OF SINGLY AND DOUBLY ARTICULATED STOPS IN ANYI: A DATA-DRIVEN ANALYSIS

Ettien Koffi
St. Cloud State University

THE ACOUSTIC PHONETIC PROPERTIES OF SINGLY AND DOUBLY ARTICULATED STOPS IN ANYI: A DATA-DRIVEN ANALYSIS

ETTIEN KOFFI

ABSTRACT

This paper examines the acoustic phonetic properties of the stop consonants [p, b, t, d, k, g, k͡p, g͡b] in Anyi, an Akan language spoken in eastern Côte d'Ivoire and western Ghana. Twelve acoustic correlates (F0, F1, F2, F3, F4, VOT, intensity, duration, B1, B2, B3, and B4) are extracted from data elicited from 10 participants in order to determine which correlates are robust for speech intelligibility. The findings discussed in the paper are based on 2,880 tokens (10 speakers x 8 words x 3 repetitions x 12 correlates). The investigation serves two purposes. The first is to give an exhaustive account of stops in Anyi. The second is to prepare the language for formant-based text-to-speech synthesis. Tried and true Just Noticeable Difference (JND) thresholds are used to gauge which correlates are robust and which ones are not.

Keywords: VOT, Stops in Anyi, Short-lag VOT, Long-lag VOT, Singly Articulated Stops, Doubly Articulated Stops, Prevoicing, Anyi Language

1.0 Introduction

Anyi has six plain and two doubly articulated stops. The plain ones are [p, t, k, b, d, g] and the doubly articulated ones are [k͡p, g͡b]. Twelve correlates, F0, F1, F2, F3, F4, Voice Onset Time (VOT), intensity, duration, and bandwidths 1, 2, 3, 4 (B1, B2, B3, B4) are extracted from these consonants to satisfy two goals. The first aims at providing a comprehensive acoustic phonetic description of all the stops in the language. The second seeks to make use of the extracted measurements for building formant-based speech synthesis for Anyi. The paper is organized into two main installments. The first provides preliminary information about the acoustics of stops in general, the participants, and data collection. The second focuses on the extracted measurements and their significance for speech intelligibility.

2.0 A Succinct Review of the Acoustics of Stops

This paper follows the same methodology as Halle et al.'s (1957) paper on the acoustics of stops in American English. In that paper, they extracted F0, F1, F2, F3, and intensity data. Another paper that serves as the exemplar for this one is Lisker and Abramson's (1964) groundbreaking investigations of Voice Onset Time (VOT). Aspects of these two seminal papers will be discussed later in relevant sections. Additional insights have been gleaned from Ladefoged's (1968) and Connell's (1994) study of stops in some West African languages. The latter has been invaluable in confirming certain acoustic phonetic properties of stops in Anyi. Finally, Klatt's (1987) description of text-to-speech synthesis provides the rationale for extracting B1, B2, B3, and B4 bandwidths.

2.1 Overview of Stops in Anyi

Anyi is an Akan language spoken mostly in southeastern and central Côte d'Ivoire. Some dialects of the language enjoy a robust intelligibility with Baule with which it forms the Nzi-Tano sub-branch. There are four main sources that provide information on stops in Anyi. The first is Retord's (1980) radiocinematographic study of the production of Anyi segments. This includes brief observations about the articulatory characteristics of stops. The second is Burmeister's (1983:155-172) inventory of stop consonants in Anyi. The third is Quaireau's (1987) work that provides a listing of stops in Anyi. Finally, Koffi (1990) mentions the same stop consonants found in the three previous studies. All four sources indicate that Anyi has the eight stop consonants that can be subclassified by place of articulation (POA) and manner of articulation (MOA) as follows:

		POA						
		Bilabial	Labiodental	Alveolar	Palatal	Velar	Labiovelar	Glottal
MOA	- voice	p		t		k	k͡p	
	+ voice	b		d		g	g͡b	

Table 1: List of Stop Consonants

The distributional patterns of these segments are worth mentioning briefly before embarking on the acoustic phonetic investigation. The stops [b, t, d, k, g, k͡p] have a high frequency of occurrence in the language. However, [p, g͡b] have a very limited distribution. According to Retord (1980:291), [p] occurs only 1% of the time in the lexicon. There is no lexical minimal between [p] and [b]. In spite of nearly four decades of research on Anyi, I have come across only two minimal pairs between [k͡p] and [g͡b]. They are [k͡paá] (contract) and [g͡baá] (skin disease), [ak͡páa] (shelf) and [ag͡báa] (manioc). Yet, when verbs are negated, if the root begins with [k͡p], when the negative prefix <n-> is added, [k͡p] changes into [mg͡b] as a result of a homorganic voicing assimilation. Verb root beginning with [t] and [k] are also negated as [nd] and [ŋg] through the same process of homorganic assimilation. We also come across many nouns that have [mg͡b] in syllable onsets when the plural or the residual class marker prefix <m-> is added to the root. These morphophonemic processes cause voiced velars to occur frequently in Anyi and other Akan languages (Welmers 1973:186). Yet, these stop consonants have not undergone any significant acoustic phonetic investigation before this paper. This is the reason why we are embarking on a comprehensive study of stops in Anyi with the hope that when studies are carried out later on in other languages, there will be acoustic phonetic data available for comparisons.

2.2 Participants and Data Collection

The recordings that serve as the basis for this investigation were made in the summer of 2013. Data collection was approved by the Institutional Review Board (IRB) of Saint Cloud State University where I teach. All the participants in this study are bilingual in Anyi and French, with Anyi being their dominant language. Each participant read the words in the Table 2 three times. The recording sessions took place on the premises of the Anyi Literacy and Translation Center, otherwise known by its French acronym of CATA. The recordings were done in a quiet room with cement walls and a cement floor. The equipment used was an Olympus WS-710 Digital Recorder. The participants wore a head mounted Krome microphone with noise cancelation capabilities. The recordings were later on exported as .wav files and sampled at 44100 Hz. Praat, Version 6.1.42 was used to annotate and extract all the relevant correlates.

The database from which we are working consists of 240 recorded items (8 words x 3 repetitions x 10 speakers).

	Word	**Word**	**Word**	**Meaning**
1.	pátáà	pátáà	pátáà	backpack
2.	bàkàá	bàkàá	bàkàá	wood
3.	tàlùá	tàlùá	tàlùá	young girl
4.	dàdìɛ́	dàdìɛ́	dàdìɛ́	knife
5.	kàlɛ́	kàlɛ́	kàlɛ́	debt
6.	gàdá	gàdá	gàdá	to cover, to hide
7.	k͡páà	k͡páà	k͡páà	good
8.	g͡bàá	g͡bàá	g͡bàá	Skin disease

Table 2: Word List

An important characteristic of the data is that all the stop consonants occur before the central vowel /a/. This choice was made deliberately in order to replicate Retord's (1980:79-98, 291, 294) articulatory phonetic study of Anyi consonants. He contends that studying consonants in a uniform phonological environment helps to describe their intrinsic phonetic behavior better than if they were to occur before multiple vowels. Retord's reasoning is a good one. Yet, this methodological choice has an important limitation, as will be pointed out in 5.0. The choice of /a/ is also motivated by the fact that it is the most frequent vowel in Anyi. In the acoustic vowel space of Anyi, Koffi (2018:199) provides data showing that it is a central vowel that is equidistant between front and back vowels.

2.3 Data Extraction Procedures

Measurements were extracted only from word-initial stops in each list. Since Anyi has a canonical CV syllable structure, it goes without saying that no stops occur in syllable codas. The procedures used to extract the relevant measurements are best illustrated with the spectrograph in Figure 1 from the speech of Speaker 7M's pronunciation of [tàlùá]:[1]

[1] Speaker 7M's data is used in all the illustrations as a way of honoring him. He passed away shortly after the data was collected. He was an important member of the Anyi literacy team. Rest in peace dear friend and Anyi language expert!

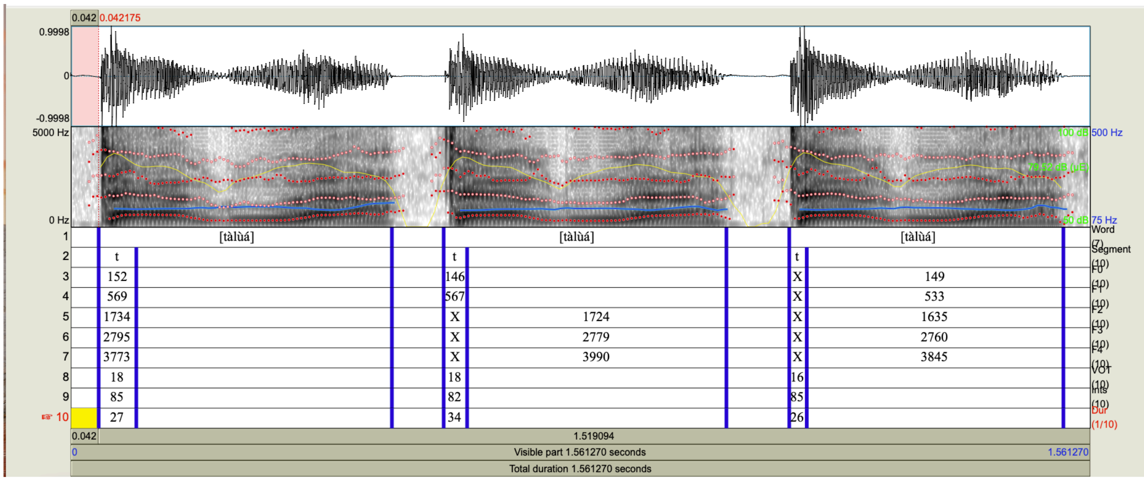

Figure 1: Illustrative Spectrograph

Speaker 7M, like the other participants, produced each word three times, as shown in the spectrograph. Boundaries were drawn around each word and around each stop consonant from which F0, F1, F2, F3, F4, VOT, intensity, duration, B1, B2, B3, and B4 were extracted. All the data was extracted manually because no bulk extraction method exists for such a wide variety of correlates. Manual extraction is onerous and time consuming. For this reason, it took several weeks to compile all the data. The "Xs" in Figure 1 indicate where measurements could not be entered because the boundaries were too small. In all such cases, the relevant measurements were entered in the row next to the "X." Subsequently, the extracted data were tabulated, as shown in Table 3.

Word	**tàlùá**	**tàlùá**	**tàlùá**	**Range**	**Mean**
Segment	[t]	[t]	[t]		
F0	152	146	149	146-152	149
F1	569	567	533	533-569	556
F2	1734	1724	1635	1635-1734	1697
F3	2795	2779	2760	2760-2795	2778
F4	3773	3990	3845	3773-3990	3869
VOT	18	18	16	16-18	17
Intensity	85	82	85	82-85	84
Duration	27	34	26	26-34	29
B1 (Bandwidth 1)	113	113	106	106-113	110
B2 (Bandwidth 2)	346	344	327	327-346	339
B3 (Bandwidth 3)	559	555	552	552-559	555
B4 (Bandwidth 4)	754	798	769	754-798	773

Table 3: Sample Tabulation

Thereafter, the relevant measurements that pertained to each stop consonants were entered accordingly. Additionally, the range of variation and the arithmetic mean for each correlate were also recorded. This procedure was repeated for each segment, for all 12 correlates, yielding a total of 2,880 stop tokens (10 x 8 words x 3 repetitions x 12 correlates).

2.4 Generalizability of the Findings

Since Atal (1972), it is now accepted that the results of any acoustic phonetic study that has a minimum of 10 participants is generalizable to the entire speech community. Himmelmann and Ladd (2008:270) contend that data from as few as eight participants is generalizable. Ladefoged (2003:67) opines that data from six participants is good enough. However, he discourages using data from three participants or less because such data only reflects speaker idiosyncrasies (Ladefoged 1968:xi). Since our study has 10 participants, it can be concluded that the findings are generalizable to the entire speech community.

Burmeister (1981) conducted a dialect survey of the Anyi-speaking areas of Côte d'Ivoire and found that the language has six main dialects that enjoy a high degree of mutual intelligibility. The participants of this study are speakers of the Morofou dialect, which is largest Anyi dialect. Nearly 75% of the Anyi speak this dialect. Moreover, it is the main dialect in which literacy materials exist.

3.0 Principles of Acoustic Phonetic Measurements

The measurements of Anyi stops are carried out and interpreted in accordance with the six guiding principles of acoustic phonetic data extraction (Baken and Orlikoff 2000:3). They are listed here as follows:

1. **Physiological Principle:** Measurements must have a known (or at least a very likely) and specific relationship to recognized aspects of speech system physiology.
2. **Relevance Principle:** A measurement must have clear relevance.
3. **Historicity Principle:** A measurement method should have a history in the literature.
4. **Ease of Understanding Principle:** Measurements must be thoroughly understood.
5. **Weariness Principle:** Never trust a computer completely.
6. **Usefulness Principle:** Measurement should be limited to situations in which it is likely to be useful.

The Physiological and Historicity Principles are the main rationale for extracting F0, F1, F2, F3, F4. These correlates were also extracted by Halle et al.'s (1957) in their study of the acoustic properties of stops in English. The Historicity Principle calls for measuring VOT because it has been extracted in a wide variety of languages for over 50 years. The Historicity and Relevance Principles are also the reasons for extracting intensity and B1, B2, B3, and B4 because Klatt (1987, 1990) measured them for his formant-based speech synthesis of English. Since, text-to-speech (TTS) synthesis is one of the goals of this paper, these correlates are worth extracting. The Weariness Principle is discussed in 4.1. The Usefulness Principle is the reason why the number of correlates is limited to 12.

3.1 The Interpretive Framework

The Relevance and Ease of Understanding Principles call on researchers to explain acoustic phonetic measurements in an accessible manner. What good does it do to anybody to pile up measurements if the reader does not know what they mean? To make sense of acoustic phonetic measurements, we turn to psychoacoustics. Fastl and Zwicker (2007:VII) define it as "the attempt to find correlations between acoustical stimuli and hearing sensations which are investigated by acquiring sets of experimental data and by models which simulate the measured facts in an understandable way." Within psychoacoustics, there is a theory called the Critical Band Theory (CBT). This theory emerged as a result of a seminal paper published in 1940 by Physicist Harvey Fletcher in which he theorized about the frequency responses of the basilar membrane to sound (speech and noise). Another physicist, Georg von Békésy, demonstrated through ingenious experiments that Fletcher's theory was based on anatomical reality. For this, Békésy received a Nobel Prize in Medicine/Physiology in 1961.

3.2 The Appeal of JND Thresholds

The advantage that CBT has over other theories is that nearly 100 years of experimentation have yielded important auditory thresholds called Just Noticeable Differences (JNDs). These thresholds are authoritative and serve as the basis for interpreting acoustic phonetic measurements. They are used ostensibly in audio engineering products and are endorsed by reputable national and international regulatory bodies such as the American National Standards Institute (ANSI), the International Standardization Organization (ISO), and the International Electrotechnical Commission (IEC) for the manufacturing and testing of audio products.[2] JNDs are worth relying on to interpret acoustic phonetic measurements because "they are widely used to establish a baseline for measurements" (Everest and Pohlmann 2015:23).

The groundbreaking acoustic phonetic research carried out at the Bell Research Laboratories from the 1920s to the 1980s led to the discoveries of many of these important JNDs. Unfortunately, the specialized jargons and the migraine-inducing formulas justifying them are beyond the grasp of average linguists. Consequently, these findings have not yet made their way into the general phonetics literature. Koffi (2021:37-41) has provided a good summary of the JNDs of F0, F1, F2, F3, F4, intensity, duration, and many others in an attempt to popularize them for use in linguistic analyses.

When JNDs are used, they obviate the need of a statistical interpretation of the data because for a JND to qualify as such, it must meet and exceed 75% of correct responses (Stevens 2000:225, Houtsma 1995:271). According to Fastl and Zwicker (2007:VII) JNDs have a distinct advantage over other forms of statistical analyses because they help to gauge acoustic phonetic facts "in an understandable way." Labov et al. (2006) have relied on the JNDs of F1 and F2 to interpret sociophonetic data. Since they do not explain these JNDs, many of their readers do know where the 60 Hz and the 200 Hz threshold that they relied on to motivate their analyses originated from. Simply put, JNDs are the results of CBT-led research carried out at the Bell Research Laboratories.

[2] JNDs are often mentioned in the documentation of audio products, but average users do not pay attention to them because they appear in fine prints and because average consumers do not understand the jargon. For instance, the packaging of the head mounted Krome microphone used in the recording lists the sensitivity specification of 110 dB with ± 3dB. Other specifications such as the driver diameter: 40 mm, the impedance rate of 32 Ohms, and the frequency response: 20 Hz-20KHz do not mean much to people. Yet, these are important JNDs that manufacturers have to adhere to.

4.0 Data Analysis and Processing

Acoustic phonetic investigations generate huge amounts of data. It is therefore challenging to organize the data in a logical manner, because as the saying goes, "There's more than one way to skin a cat." Eight words, three repetitions of each word by 10 participants generate 240 recorded items. When the 12 correlates are added, this leads to a corpus of 2,880 tokens. It is unwieldy to discuss such a large volume of data in a single paper. Yet, I want this paper to be comprehensive so that readers have all the relevant information about Anyi stops in a single paper. The most logical way to present the information is to highlight one acoustic phonetic correlate at a time. So, in the remainder of the paper, in each subsection, a JND is discussed, underscoring its robustness or lack thereof to the intelligibility of the stops in Anyi.

4.1 Robustness Analysis of F0

F0 is also called pitch. Pitch data is worth extracting in the study of stops because it provides prima facie evidence for the differentiation between voiced and voiceless stops. The JND of F0 for distinguishing these two categories of segments in general, and stops in particular, can be stated as follows:

JND for Pitch Detection

The minimum pitch detection algorithm in Praat is set at 75 Hz. Any segment whose F0 is < 75 Hz is deemed voiceless.[3]

The Weariness Principle (see 3.0) applies here because one cannot blindly trust the pitch detection algorithm in Praat or any other acoustic phonetic software package to make an accurate determination of pitch. Yet, the JND of < 75 Hz is a good starting point. If one wants to be completely sure that a segment is voiceless, one must defer to the Voice Report function in Praat. As imperfect as pitch tracks are, they provide some interesting insights about the acoustics of stops. The spectrographs in Figures 2 and 3 are used to illustrate this point.

When we examine the spectrograph of the two [d]s in <dàdìɛ́>, we notice that there is a blue line (pitch track in Praat) even before the initial [d] is produced. This is an indication that the vocal folds began vibrating well before. In fact, 77 msec before the tip of the tongue of the speaker touched the alveolar area to produce [d], the vocal folds began vibrating. Now, when we look at the spectrographic behavior of the second [d], we see that the pitch tract (blue line) runs throughout and into the adjacent vowel. This means that the vocal folds vibrated all the way through.

[3] When Praat renders a *"pitch undefined"* result, I enter 74 Hz for pitch by convention. This means that Praat could not detect any pitch. Entering 0 Hz for *"pitch undefined"* is not correct because, according to Fry (1979:68), the smallest detectable pitch that the human vocal folds can produce is 60 Hz. If one wishes, one could enter 60 Hz for all instances of *"pitch undefined."* It does not matter what number one enters so long as it is between 60 Hz and 74 Hz. For the purposes of calculations, a digit must be entered, otherwise the arithmetic means and standard deviations cannot be calculated.

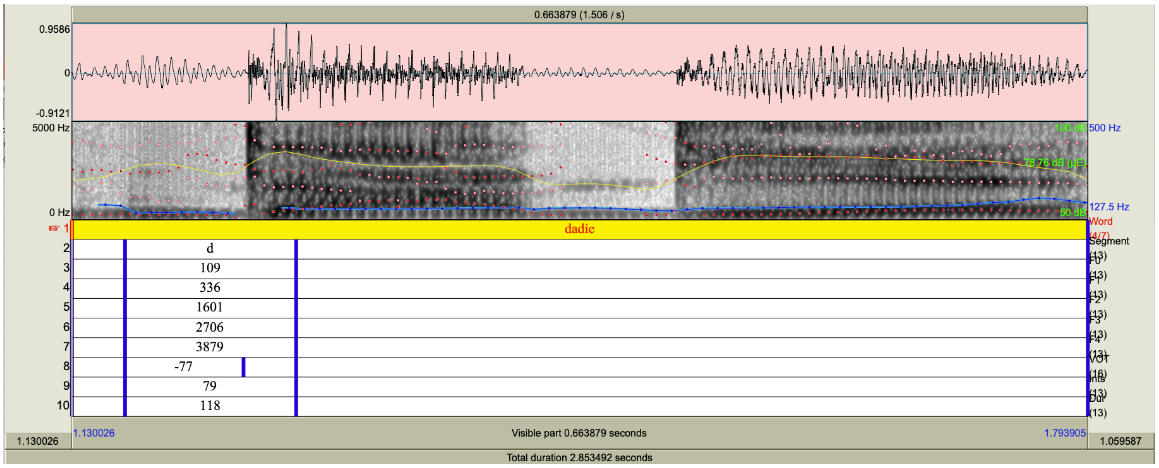

Figure 2: Illustration of Voicing

Now, let's contrast the spectrographic behavior of the initial [d] in <dàdìɛ́> with that of the initial [t] in <tàlùá>, as shown in Figure 3.

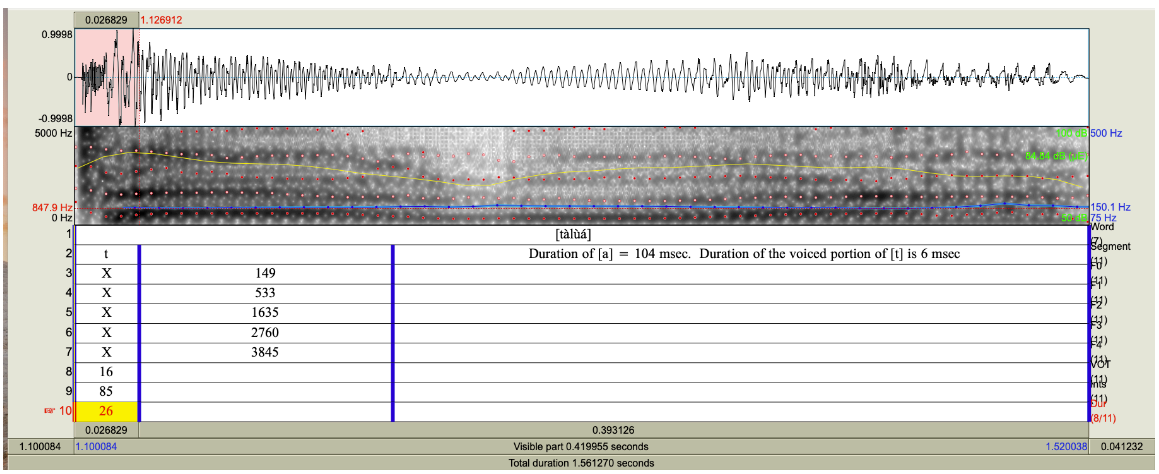

Figure 3: Illustration of Prevoicing [t]

We see that the entire duration of [t] is 26 msec. We also see that 6 msec of it is voiced (blue pitch tracker). The presence of pitch means that a tiny portion of [t] is voiced. If this is the case, why is [t] classified as a voiceless stop in Anyi (see Table 1)?

Phoneticians refer to this tiny portion of voicing as **prevoicing**. It has been found that in some languages, when voiceless consonants occur immediately before vowels, a tiny fraction of the voiceless consonant becomes voiced. The amount of voicing is usually so small that the naked ear cannot perceive it. Impressionistically, the segment is perceived as voiceless even though a tiny fraction of it is voiced. Since many languages have not been investigated from the perspectives of acoustic phonetics, we do not know how many languages have prevoicing of voiceless segments. Yet, we know that Anyi is in a good company. Katz (2013:252) notes that French has prevoicing. Connell (1994:459-60) lists five Nigerian languages in which prevoicing has been observed. On average, prevoicing lasts 10 msec, but in some instances, it can last up to 26 msec. It is because of prevoicing that Halle et al. (1957:107) concluded more than half a century ago that voicing is not a reliable feature for differentiation between stops in American English. Their reasoning goes as follows:

> During the period of the closure *the vocal cords may or may not vibrate*; if they do, we have a voiced stop; if they do not, we have a voiceless stop. Although in many instances the presence or absence of voicing serves to distinguish /b/, /d/, /g/ from /p/, /t/, /k/, *in English voicing is not crucial to this distinction*. The essential difference between these two classes of stops lies in the fact that in the production of the latter more pressure is built up behind the closure than in the former. This difference in pressure results in higher intensity and accounts for the well-known fact /p/, /t/, /k/ bursts are often followed by aspiration, which is not present in the case of /b/, /d/, /g/. *Since the traditional role of vocal cord vibrations is thus relatively less important, the traditional terms "voiced" and "voiceless" seem somewhat inappropriate and will not be used here*. [Italics not in original. They are added for emphasis].

The measurements in Table 4 also show that vocal fold vibration does not discriminate between voiced and voiceless stops in Anyi[4]:

	F0 Measurements							
Words	**pátáà**	**bàkàá**	**tàlùá**	**dàdìɛ́**	**kàlɛ́**	**gàdá**	**k͡páà**	**g͡bàá**
Segments	[p]	[b]	[t]	[d]	[k]	[g]	[k͡p]	[g͡b]
Speaker 1M[5]	74	111	74	102	120	74	138	106
Speaker 2M	74	95	74	88	89	84	94	82
Speaker 3M	108	117	143	138	164	126	168	120
Speaker 4M	108	111	96	136	115	83	136	126
Speaker 5M	98		101	124	98	103	167	109
Speaker 6M	74	110	141	106	74	106	123	107
Speaker 7M	144	116	149	113	153	112	158	117
Speaker 8M	172	164	225	165	237	161	233	166
Speaker 9M	123	97	129	92	93	105	145	94
Speaker 10F	120	147		246	183	207	244	205
Average	**109**	**118**	**125**	**131**	**132**	**116**	**160**	**123**
St. dev.	32	27	3	46	51	40	46	36

Table 4: F0 Measurements

The arithmetic means show that both voiced and "voiceless" stops are voiced in Anyi. In other words, F0 is not a robust correlate for differentiating between stops in regard to voicing. The so-called voiceless stops are voiced because of prevoicing. The only reason why hearers perceive some stops as voiceless is because the amount of voicing in these "voiceless" stops is below the **40/60 threshold**. The following percentages allow a three-

[4] Stevens (2000:339) notes that "Glottal vibration may continue for 10 to 20 msec following closure but with a greatly reduced amplitude. At the release of closure, vocal fold vibration resumes 5 to 15 msec following the release, the amount of delay depending on the rate of opening of the constriction." We see clearly that for almost all the speakers, with the exception of Speakers 1M, 2M, and 6M, the vocal folds vibrate for a tiny amount even during closure. This is so even for "voiceless" stops. Rampant prevoicing is the cause.
[5] The suffix "M" and "F" after each name stands for "male," and "female."

way differentiation in the amount of voicing that can be perceived by the naked ear (Smith 1997:478 and Gradoville 2011:68-69)[6]:

1. If 40% or more of the duration of a segment is voiced, that segment is perceived auditorily as voiced.
2. If 11-39% of the duration of a segment is voiced, that segment is perceived auditorily as devoiced.
3. If 0-10% of the duration of a segment is voiced, that segment is perceived auditorily as voiceless.

Again, the amount of prevoicing in [p, t, k, k͡p] is so small that the segments are perceived as voiceless by the naked ear. For instance, in the case of [t] in <talua>, prevoicing of [t] amounts to only 23.07%. Since it is below the 40% threshold, the naked ear perceives [t] as voiceless. If one wants to be hyper technical, one can say that Anyi (and probably other Akan languages) does not have voiceless stops, as is the case for the five Nigerian languages discussed by Connell (1994:459-60). However, such an extreme position is unwarranted. So, we will continue to use the impressionistic classification of [p, t, k, k͡p] as voiceless and [b, d, g, g͡b] as voiced consonants, with the proviso that these labels are misnomers.

4.2 Robustness Analysis of F1

F1 correlates with mouth opening along three degrees of aperture: high, mid, and low. F1 measurements that are ≤ 400 Hz indicate that the mouth is barely open, while measurements that are ≥ 700 Hz correlate with a widely opened mouth. According to Koffi (2018:199), the mean F1 and F2 of /a/ are respectively 925 Hz and 1486 Hz, which means that in producing this vowel, the mouth is opened wide, but the tongue is in a neutral position.[7] With this in mind, let's gauge the intelligibility of stops in light of the measurements in Table 5 and of the JND of F1:

---

[6] Abramson and Whalen (2017:81) refer to the 50% criterion for stops. Even though the 40/60 threshold was initially proposed for fricatives, there is now ample evidence that it also applies to stops.

[7] At the 54th Annual Conference on African Linguistics in Storrs, CT, from June 11-14th 2023, a couple of participants suggested that taking formant measurements only from the onset to the end of closure of voiceless stops may cause the formant measurements to be incorrect. It is true that Klatt (1980:978) extracted formant measurements 45 msec into the vowel. However, this is not necessary for Anyi, since voiceless stops are not aspirated and also because of prevoicing. I redid a few measurements from various speakers to see if measuring from the onset of stops until 20 msec into the following vowel /a/ is indeed a better method. It turned out not to be because doing so skewed the formant measurements and caused F1 measurements to be greater, i.e., wider mouth opening because a small portion of the F1 values of /a/ leaked into the formant measurements. Because of widespread prevoicing, glottal vibration takes place even before the release of burst. So, there is no reason to include glottal periods of the following vowel into the extraction of /p, t, k/. One participant raised concerns about the F1 measurements of /p, t, k/ because they appear to be higher than those reported elsewhere. However, the measurements are correct and are in keeping with those reported in Stevens (2000:339, Figure 7.13) when the vowel /a/ follows /p, t, k/ in English. The F1 values reported in Klatt (1980:987) are considerably lower probably because he averaged them over 15 vowels (see page 986). Even so, the principle that voiceless stops involve a wider mouth opening than their voiced counterparts is the same in Anyi as in English.

Auditory Discrimination on the F1 Frequency Bandwidth

Of two speech signals **A** and **B**, **A** is perceived as auditorily distinct from **B** if there is a difference of 60 Hz or more between them. [8]

In assessing the relevance of F1, it is best to compare stops that have the same place of articulation. In other words, [p, b] should be compared with each other, [t, d], [k, g], and [k͡p, g͡b]. Cross-category comparisons are fruitless because doing so would be tantamount to comparing apples and oranges.

	F1 Measurements							
Words	**pátáà**	**bàkàá**[9]	**tàlùá**	**dàdìɛ́**	**kàlɛ́**	**gàdá**	**k͡páà**	**g͡bàá**
Segments	[p]	[b]	[t]	[d]	[k]	[g]	[k͡p]	[g͡b]
Speaker 1M	646	410	545	453	696	766	461	307
Speaker 2M	800	916	771	817	883	723	511	425
Speaker 3M	791	373	656	417	735	390	452	361
Speaker 4M	663	429	893	306	665	510	540	303
Speaker 5M	694		527	422	623	373	502	304
Speaker 6M	859	635	724	386	1023	419	412	241
Speaker 7M	690	357	556	387	941	475	464	311
Speaker 8M	790	374	693	707	855	458	460	353
Speaker 9M	734	471	548	383	811	495	526	284
Speaker 10F	912	544		372	918	494	728	416
Average	**757**	**501**	**657**	**465**	**815**	**510**	**505**	**330**
St. dev.	87	116	20	163	131	131	87	57

Table 5: F1 Measurements

Care should be taken in interpreting F1 measurements. Smaller Hz values correlate with smaller mouth opening, while bigger Hz values indicate that the mouth is wide open. Judging by the measurements, we see that F1 is a very robust cue for differentiating between pairs of stops that share a common place of articulation because the JND is greater than the 60 Hz required for audibility. The arithmetic means show that voiceless stops involve a somewhat wider mouth opening than their voiced counterparts. The differences are respectively 256 Hz between [p] (757 Hz) and [b] (501 Hz), 192 Hz between [t] (657 Hz) vs. [d] (465 Hz), 305 Hz between [k] (815 Hz) vs. [g] (510 Hz), and 175 Hz between [k͡p] (505 Hz) and [g͡b] (330 Hz).

The interspeaker variability analysis confirms this overall pattern. Out of 98 possible cases, there are only three instances where the F1 value of the voiceless segment is smaller than that of its voiced counterpart. Those three cases are [t] vs. [d] produced by Speakers 2M and 8M, and [k] and [g] by Speaker 1M. In other words, the **relevant functional load**

[8] Rabiner and Juang (1993:152) list slightly different set of JNDs. The JND of F1 is 62 Hz, that of F2 is 158 Hz, for F3, the JND is 355 Hz, while the JND of F4 is 480 Hz. It is important to keep in mind that the differences between these JNDs and those used in the paper do not amount to much on the 1/3 frequency bandwidth. It is universally accepted that the 1/3 frequency bandwidth replicates as faithfully as mathematically possible how the naked ear processes frequency data.

[9] I did not realize that Speaker 5M skipped [bàkàá] and that Speaker 10F skipped [tàlùá] until I returned from fieldwork.

(RFL) of F1 is 96.93%. This means that the degree of mouth opening contributes to the intelligibility of stops. In all instances, voiceless stops were produced with a somewhat wider mouth opening than voiced ones. It is also worth noting that doubly articulated stops [k͡p] and [g͡b] call for the least amount of mouth aperture. Even so, even among them, the mouth is slightly more open for [k͡p] than for [g͡b]. The measurements indicate conclusively that F1 is a **very** robust correlate in the intelligibility of stop consonants in Anyi.

4.3 Robustness Analysis of F2

F2 correlates with tongue positions in the mouth (Ladefoged and Maddieson 1996:196). When the tongue is thrusted forward, F2 value increase. When the tongue is retracted, the F2 values decrease. The rule of thumb is that F2 ≥ 2000 Hz is indicative of fronting, while F2 values of between 1600 Hz and 1400 Hz are indicative of a neutral tongue movement, which means that the segment in question neither front nor back, while an F2 ≤ 1400 Hz correlates with tongue retraction. The JND for audibility on the F2 frequency bandwidth is stated as follows:

Auditory Discrimination on the F2 Frequency Bandwidth

Of two speech signals **A** and **B**, **A** is perceived as auditorily distinct from **B** if there is a difference of 200 Hz or more between them.

When the information in Table 6 is viewed in light of this JND, it can be concluded that F2 does not discriminate between pairs of plain stops because the acoustic distances between [p] (1561 Hz) and [b] (1537 Hz), [t] (1769 Hz) and [d] (1756 Hz), or [k] (1729 Hz) and [g] (1748 Hz) are all below the JND of 200 Hz.

The interspeaker variability analysis shows that the only exceptions are the pairs [t] and [d] produced by Speakers 2M and 4M, and the pair [k] and [g] produced by Speaker 5M. In other words, in 54 out 60 instances (90% of the time), F2 does not discriminate between plain stops. So, the RFL of F2 is 10%, which is not much.

	F2 Measurements							
Words	**pátáà**	**bàkàá**	**tàlùá**	**dàdìɛ́**	**kàlɛ́**	**gàdá**	**k͡páà**	**g͡bàá**
Segments	[p]	[b]	[t]	[d]	[k]	[g]	[k͡p]	[g͡b]
Speaker 1M	1546	1480	1637	1694	1829	1719	1738	1951
Speaker 2M	1992	1928	1881	1573	1832	1595	1613	2124
Speaker 3M	1279	1395	1906	1778	1659	1841	1357	1697
Speaker 4M	1410	1376	1839	1492	1679	1702	1356	1207
Speaker 5M	1387		1841	1987	1638	1949	1613	2247
Speaker 6M	1798	1826	1837	1924	1865	1975	1982	1844
Speaker 7M	1368	1375	1697	1702	1626	1677	1064	1625
Speaker 8M	1396	1424	1738	2106	1747	1812	1203	1653
Speaker 9M	1722	1441	1550	1504	1683	1675	1555	1926
Speaker 10F	1719	1595		1818	1733	1540	1688	1631
Average	**1561**	**1537**	**1769**	**1757**	**1729**	**1748**	**1516**	**1790**
St. dev.	233	183	35	205	86	143	272	295

Table 5: F2 Measurements

The arithmetic mean shows that F2 discriminates between the doubly articulated labiovelar stops [k͡p] (1516 Hz) and [g͡b] (1790 Hz) because the acoustic distance of 276 Hz. The interspeaker variability analysis shows that 7 of the 10 (70%) participants produced them differently. Only Speakers 4M, 6M, and 10F produced them similarly. In articulatory terms, [g͡b] (1790 Hz) is more fronted than [k͡p] (1516 Hz) because the acoustic distance of 274 Hz between them is greater than the JND of 200 Hz required for audibility.

The F2 measurements confirm a claim repeated by Connell (1994:443) that "labial-velars tend to resemble more labials than they do velars." We verify this to be true because [p] (1561 Hz) and [k͡p] (1516 Hz) have the same F2 value but [g͡b] (1790 Hz) and [b] (1537 Hz) do not. The difference between them is 253 Hz, which exceeds the JND of 200 Hz. In fact, [g͡b] (1790 Hz) resemble [g] (1729 Hz) more than it does [b].

What could possibility explain this exception? In Anyi, as in many African languages, the voiceless labiovelar [k͡p] is more frequent than its voiced counterpart [g͡b] (Connell 1994:441). In Fante (Ladefoged 1968:52) and Twi (Ladefoged 1968:52), two very close cousins of Anyi, [k͡p] and [g͡b] do not exist. Nzema (Ladefoged 1968:50), which is somewhat intelligible with the Anyi dialect of Sanvi, has [k͡p] but lacks [g͡b]. The Indenie dialect of Anyi lacks both [k͡p] and [g͡b]. Could it be that Anyi Morofu borrowed [g͡b] from Baule? If so, then [g͡b] is a newer sound than [k͡p]. Could this also be the reason why [p] is like [k͡p], but [g͡b] is not like [b]? The brief discussions about the distribution patterns of [k͡p] and [g͡b] in 2.1 can help explain this irregular behavior of [k͡p] and [g͡b]. Even so, we conclude that F2 is robust for the intelligibility of the doubly articulated stops [k͡p] and [g͡b] but not for the singly articulated stops [p] and [b], [t] and [d], and [k] and [g].

4.4 Robustness Analysis of F3

The articulatory correlate of F3 is lip rounding or unrounding. When the lips are rounded, F3 values fall below 2500 Hz. The mean F3 value of [a] reported by Koffi (2016:127) is 2506 Hz. The JND of F3 is stated as follows:

Auditory Discrimination on the F3 Frequency Bandwidth

Of two speech signals **A** and **B**, **A** is perceived as auditorily distinct from **B** if there is a difference of 400 Hz or more between them.

The acoustic measurements displayed in Table 6 show that F3 does not discriminate between pairs of stops that have the same place of articulation. For example, the F3 distance between [p] (2658 Hz) and [b] (2680 Hz) is only 22 Hz, which is a far cry from the 400 Hz required for auditory discrimination. The same goes for the acoustic distance of 25 Hz that separates [t] (2808 Hz) and [d] (2783 Hz), or the 12 Hz difference between [k] (2664 Hz) and [g] (2676 Hz).

	F3 Measurements							
Words	**pátáà**	**bàkàá**	**tàlùá**	**dàdìɛ́**	**kàlɛ́**	**gàdá**	**k͡páà**	**g͡bàá**
Segments	[p]	[b]	[t]	[d]	[k]	[g]	[k͡p]	[g͡b]
Speaker 1M	2573	2507	2744	2697	2745	2395	2838	3145
Speaker 2M	2991	2939	2985	2554	2695	2246	2881	3288
Speaker 3M	2634	2463	2719	2783	2636	2618	2353	2767
Speaker 4M	2490	2650	2980	2583	2364	2588	2509	2378
Speaker 5M	2493		2875	3057	2440	2966	2816	3250
Speaker 6M	2714	2956	2838	2924	2747	2900	2940	2865
Speaker 7M	2513	2551	2778	2846	3105	2914	2420	2846
Speaker 8M	2751	2664	2769	2957	2506	2770	2670	3017
Speaker 9M	2687	2584	2584	2541	2764	2491	2638	2859
Speaker 10F	2734	2814		2888	2646	2874	2912	2865
Average	**2658**	**2680**	**2808**	**2783**	**2664**	**2676**	**2697**	**2928**
St. dev.	154	169	10	182	206	246	212	265

Table 6: F3 Measurements

The interspeaker variability analysis shows that only Speaker 2M exceeded the F3 JND between [t] and [d], [k] and [g], and [k͡p] and [g͡b]. Speaker 5M did also differentiate between [k] and [g], and [k͡p] and [g͡b]. However, for 8 of 10 participants, F3 is not a robust correlate. Furthermore, in 10 of 98 (89.79%) instances, lip position does not differentiate among stops in Anyi. The RFL of 10.21% shows that lip rounding does not contribute much, if anything at all, to intelligibility.

4.5 Robustness Analysis of F4

The Physiological Principle would normally exclude F4 from consideration because it is not entirely clear which parts of the speech organs correlate with F4. Ladefoged (2006: 187, 205) correlates it with the size of the speaker's head. Here is how he said it:

> My head is larger than that of American English speaker, so all my formants are slightly lower, p. 187. 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 vowels being compared.

Cao and Dellow (1999) correlate F4 with the supralaryngeal cavity and offer the following explanation:

> The mechanisms for F4 or F5 are more complicated and relatively little investigated. One possible interpretation is that F4 and F5 are sensitive to the laryngeal cavity (LC) shape (when LC is shortened, F5 and F4 increase). More recently, Takemoto et al. found that F4 was mainly determined by the LC geometry.

Indeed, Takemoto et al. (2006) associate F4 more firmly with the laryngeal cavity than the size of one's head. Regardless, F4 is included among the correlates worth extracting because Klatt (1987) and Klatt (1990) deem it indispensable for speech synthesis. The JND of F4 is stated as follows:

Auditory Discrimination on the F4 Frequency Bandwidth

Of two speech signals **A** and **B**, **A** is perceived as auditorily distinct from **B** if there is a difference of 600 Hz or more between them.

The data displayed in Table 7 shows clearly that F4 is not a robust cue for differentiating between stop consonants in Anyi. In other words, the size of the speaker's head and/or laryngeal geometry do not matter for the auditory perception of stops.

	F4 Measurements							
Words	**pátáà**	**Bàkàá**	**tàlùá**	**dàdìɛ́**	**kàlɛ́**	**gàdá**	**k͡páà**	**g͡bàá**
Segments	[p]	[b]	[t]	[d]	[k]	[g]	[k͡p]	[g͡b]
Speaker 1M	3785	3813	4004	3909	4163	4021	3834	4269
Speaker 2M	4239	4250	3973	3820	3973	3838	3928	4169
Speaker 3M	3936	3030	4188	3951	4400	4391	3915	4116
Speaker 4M	3739	3814	4040	3634	4019	4045	3635	3843
Speaker 5M	3769		3878	4285	4273	4320	3859	4442
Speaker 6M	3811	4068	3688	3998	4189	4036	4041	3965
Speaker 7M	3607	3701	3869	3960	4205	4107	3661	3948
Speaker 8M	3952	3860	4256	4053	4170	4030	3851	4186
Speaker 9M	3957	3935	3783	4238	3973	4071	3843	4067
Speaker 10F	3855	3934		4217	4012	3939	3947	4012
Average	**3865**	**3822**	**3964**	**4006**	**4237**	**4079**	**3851**	**4101**
St. dev.	170	133	25	201	156	164	124	174

Table 7: F4 Measurements

The interspeaker variability analysis shows that only Speaker 3M differentiated between [p] and [b] on the F4 frequency bandwidth out of 98 tokens with the same place of articulation. Consequently, the RFL of F4 for intelligibility is 2.04%, which is negligible.

4.6 Robustness Analysis of Duration

By duration, we mean the length of the entire stop consonant, from beginning to end. The JND for auditory perception of duration is stated as follows:

Auditory Discrimination in Duration

Of two speech signals **A** and **B** lasting less than 200 msec, **A** is perceived as auditorily distinct from **B** if there is a difference of 10 msec or more between them.[10]

[10] The JND for duration changes for segments that last longer than 200 msec. However, since none of the stops exceeds 200 msec, we stay with this JND.

Duration is perceived by the naked ear partly on a logarithmic scale and partly on a linear scale. We are using the JND of 200 msec as the standard segmental duration because none of the segments produced by the participants lasted more than 200 msec. A cursory look at the data shows that the voiced segment in each pair is longer than its voiceless counterpart. This is so for 9 out 10 speakers. Again, Speaker 2M is the exception. Even for him, [g͡b] is longer than [k͡p].

	Duration Measurements							
Words	**pátáà**	**bàkàá**	**tàlùá**	**dàdìɛ́**	**kàlɛ́**	**gàdá**	**k͡páà**	**g͡bàá**
Segments	[p]	[b]	[t]	[d]	[k]	[g]	[k͡p]	[g͡b]
Speaker 1M	27	55	20	72	44	53	69	116
Speaker 2M	29	38	45	27	33	32	49	109
Speaker 3M	21	104	16	62	21	69	42	123
Speaker 4M	17	67	115	61	35	78	32	82
Speaker 5M	25		24	106	30	149	65	192
Speaker 6M	40	56	42	89	58	92	56	125
Speaker 7M	30	82	29	100	30	119	38	137
Speaker 8M	24	66	33	62	32	90	38	104
Speaker 9M	49	81	41	104	49	141	63	185
Speaker 10F	32	65		91	31	102	40	179
Average	**29**	**68**	**40**	**77**	**36**	**92**	**49**	**135**
St. dev.	9	11	1.3	25	10	37	13	37

Table 8: Duration Measurements

Pairwise duration analysis shows that voiced segments are considerably longer that their voiceless counterparts for segments that have the same place of articulation. A graph display shows the comparison more strikingly, as shown in Figure 6:

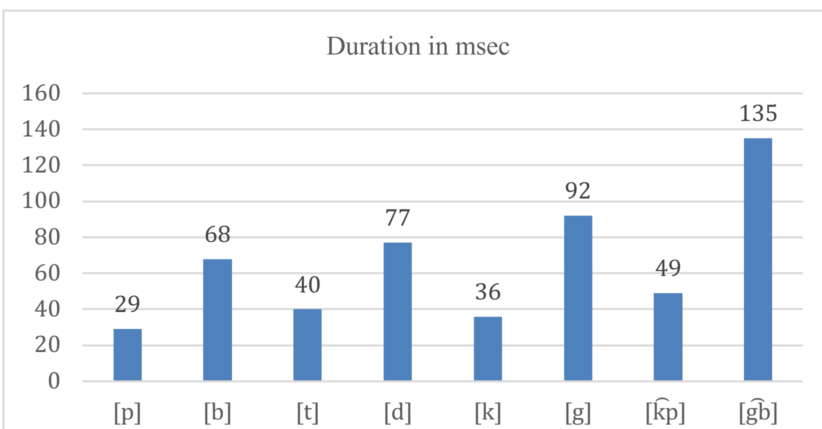

Figure 6: Duration Measurements

In fact, based on the JND of duration, we see that voiced stops are at least twice as long as voiceless ones. The composite arithmetic means underscore this fact. Collectively the four voiced segments [b, d, g, g͡b] last 372 msec while the voiceless stops [p, t, k, k͡p] last 154 msec. Therefore, on average, voiced segments last 2.41 times longer than voiceless one. The ratio is even greater for [g͡b] and [k͡p]. The voiced [g͡b] lasts 2.75 times longer than the voiceless [k͡p].

The interspeaker variability analysis shows clearly that this is the case for all speakers, except for the two instances when Speakers 2M and 4M produced [t] longer than [d]. In other words, in 2 out of 98 cases (97.95%), voiced stops are longer than voiceless ones. Duration is, therefore, a **very** robust correlate that helps to discriminate between pairs of stops that have the same place of articulation.

Connell (1996:457-8, Tables IV and V) reports that in Igbo and Obolo duration is also a robust cue, but in the opposite direction. Their voiceless segments last longer than their voiced ones. Do doubly articulated stops [k͡p] and [g͡b] last longer than the singly articulated ones? Connell (1994:444, Table 1) reports mixed findings. Maddieson and Ladefoged (1989:119) provide measurements for [g͡b] (132 msec) and [b] (128 msec) in Yoruba, showing that [g͡b] is not longer than [b] because the 4 msec that separates them is below the threshold of audibility. Anyi is different from these languages. In Anyi, voiced stops are unmistakably longer than voiceless ones. In this regard, Anyi behaves like many other languages in which voiced segments are longer than voiceless ones. Furthermore, the doubly articulated [k͡p] (49 msec) is longer than the singly articulated [p] (29 msec), and [g͡b] (135 msec) is twice as long as [b] (68 msec).

4.7 Robustness Analysis of VOT

Cho and Ladefoged (1999:225) define VOT as "the interval between the release of an articulatory gesture, usually a stop, and the beginning of vocal fold vibration." VOT is so important in acoustic phonetics that Ladefoged (2003:94) opines that "Any description of the phonetic structures of a language should include an account of the VOT." Kent and Read (2003:150) explain that the "basic appeal of VOT is that it is a single acoustic measure that may correlate with voicing contrasts in *all* relevant natural languages (emphasis added)." Indeed, this has been the case since 1964 when Lisker and Abramson published their influential paper on the VOT of 11 languages (Deutch, Tamil, Puerto Rican Spanish, Hungarian, Cantonese, English, Eastern Armenian, Thai, Korean, Hindi, and Marathi). The 50th anniversary of the "discovery" of the VOT was celebrated a while back. Yet, to date there is no VOT data on Anyi or any Ivorian language that I know of.

Let's begin our investigation of VOT in Anyi by examining Figures 4 and 5. Figure 4 is the same as Figure 3 used to illustrate prevoicing in 3.0. In Figure 4, we see that the entire duration of [t] is 26 msec, but the VOT is only 16 msec. This means that the interval of time that elapsed when the tip of the tongue touched the alveolar ridge and the onset of vocal fold vibration is 10 msec.

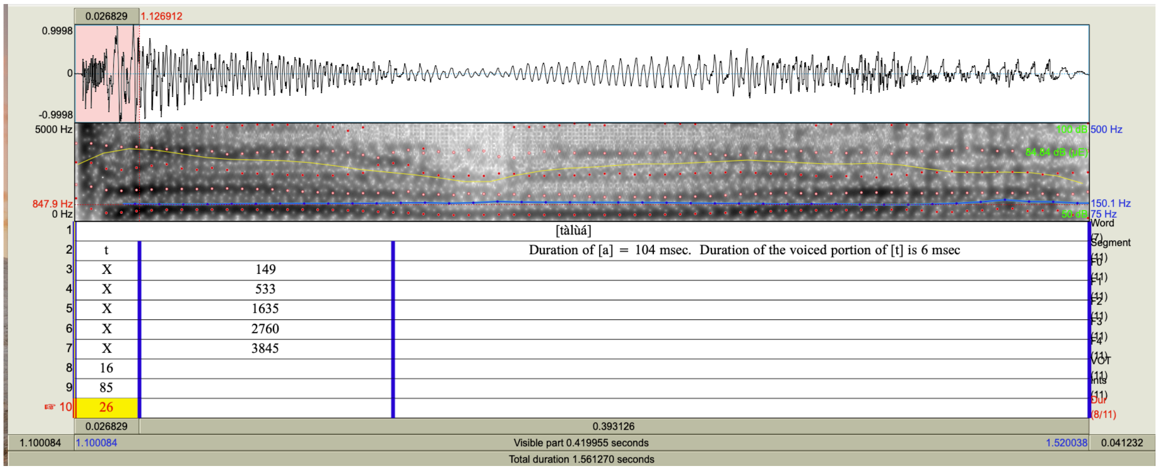

Figure 4: VOT Measurement

Now, let's compare the VOT pattern in Figure 4 with the one in Figure 5 in which the focus is on the initial [d] in [dàdìɛ́]. Figure 5 is also a repeat of Figure 2 seen earlier in 3.0.

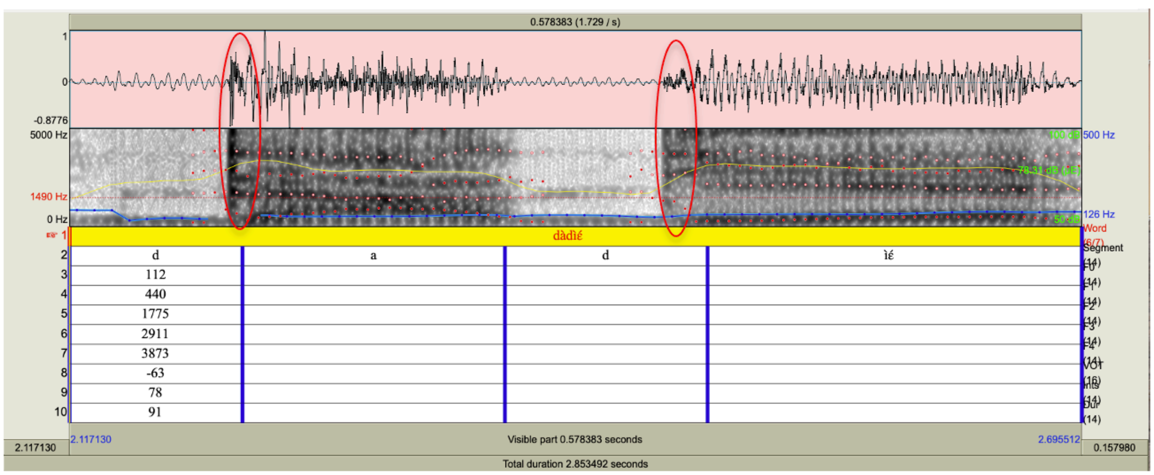

Figure 5: Negative VOT

There are two ovals that correspond to the [d] sounds. The dark vertical line inside the oval corresponds to the closure when the tip of the tongue came into contact with the alveolar ridge. We see that even before the tip of the tongue touched this area, the vocal folds had started vibrating for the initial [d]. In other words, 63 msec before the tip of the tongue touched the alveolar area, vibration had already started. When the vocal folds start vibrating before the articulators come into positive contact with each other, we have a **negative VOT**. So, the VOT of the first [d] in [dàdìɛ́] is -63 msec. If we examine the second [d], we see that negative VOT is so prominent that the vocal folds do not stop vibrating from the end of [dà] all through to the beginning of [d] in the syllable [dìɛ́]. Katz (2013:252) refers to languages that have negative VOT for voiced segments as "true voicing languages." Spanish, French, and Russian are among the languages in which voiced stops have negative VOT. The measurements in Table 9 show that, except for Speaker 2M, all the participants produced [b, d, g, g͡b] with a negative VOT.

	VOT Measurements							
Words	**pátáà**	**bàkàá**	**tàlùá**	**dàdìɛ́**	**kàlɛ́**	**gàdá**	**k͡páà**	**g͡bàá**
Segments	[p]	[b]	[t]	[d]	[k]	[g]	[k͡p]	[g͡b]
Speaker 1M	11	-39	14	-42	34	36	52	-87
Speaker 2M	13	20	36	17	20	25	31	-85
Speaker 3M	9	-85	12	-46	13	-51	32	-64
Speaker 4M	16	-29	25	12	25	25	4	-58
Speaker 5M	16		12	-84	19	-103	42	-143
Speaker 6M	21	-39	23	-72	52	-74	38	-109
Speaker 7M	22	-62	17	-68	18	91	25	104
Speaker 8M	17	-63	26	-38	23	55	27	-86
Speaker 9M	41	-53	27	-66	38	-109	41	-138
Speaker 10F	25	-50		-74	19	-86	26	-126
Average	**19**	**-44**	**21**	**-46**	**26**	**-19**	**31**	**-79**
St. dev.	9	9	0.46	35	11	73	12	70

Table 9: VOT Measurements

We see that 9 of the 10 participants (90%) produced negative VOT. When the speakers of a language produce negative VOT as is the case here, this has a direct consequence on how they produce voiceless stops. In such cases, the VOT of voiceless stops is very short. Katz (2013:252) alludes to this as a universal tendency that governs the VOT of stops. Here is his explanation:

> If a language sets voiced sounds to be so negative in VOT, then the voiceless counterparts doesn't have to be strongly voiceless. For instance, French has a voiced/voiceless, two-way opposition, like English. Similar to Spanish and Russian, French uses very pronounced, prevoiced VOTs for its voiced sounds. On the other hand, its voiceless utterances are actually produced with short-lag VOTs.

Short-lag VOT is a term used to describe languages such as French, Anyi, and others in which the VOT of voiceless stops is very short.

The take-away is that in Anyi, VOT is a **very** robust correlate that differentiates between pairs of voiced and voiceless stops that have a common place of articulation. The VOT distance between pairs of voiceless and voiced stops in every case is greater than the JND of 10 msec. This confirms the view expressed by Kent and Read (2003:150) and others that "[the] basic appeal of *VOT* is that it is a *single* acoustic measure that may correlate with voicing contrasts *in all relevant natural languages* (emphasis added)." The averaged VOT of [b, d, g, g͡b] is -188 msec, while that of [p, t, k, k͡p] is 97 msec. In other words, the VOT of voiced stops is 1.93 times longer than that of voiceless stops. The VOT of the doubly articulated [g͡b] is 2.54 times longer than that of [k͡p]. It can be said Anyi talkers and hearers rely on the robustness of VOT cue to encode and decode messages involving stops consonants.

4.8 Robustness Analysis of Intensity

Intensity, more precisely sonority, has to do with the relative loudness of one speech sound in relation to another. Intensity research on English segments has shown that individual speech sounds have their intrinsic loudness (Koffi 2020:2-27). If we just focus on the arithmetic means for each individual segment in Table 10, we come to a similar conclusion for stops in Anyi. However, care should be exercised when interpreting intensity measurements because the naked ear does not perceive intensity on a linear scale but on a logarithmic one. Therefore, we must make use of the following JND in gauging the robustness of intensity in the intelligibility of stops:

Auditory Discrimination in Intensity

Of two speech signals **A** and **B**, **A** is perceived auditorily as distinct from **B** if there is a difference of 3 dB or more between them.

Given this JND, intensity is robust only for [t] (81 dB) and [d] (77 dB) on the one hand, and [k] (79 dB) and [g] (76 dB) on the other. In other words, when one listens to Anyi speakers, one will find that their [t]s are louder than their [d]s, that their [k]s are also louder than their [g]s. However, their [p]s (80 dB) and [b]s (78 dB) and their [k͡p]s and [g͡b]s are auditorily the same.

	Intensity Measurements							
Words	**pátáà**	**bàkàá**	**tàlùá**	**dàdìɛ́**	**kàlɛ́**	**gàdá**	**k͡páà**	**g͡bàá**
Segments	[p]	[b]	[t]	[d]	[k]	[g]	[k͡p]	[g͡b]
Speaker 1M	85	80	87	84	84	79	84	81
Speaker 2M	80	78	81	80	80	78	78	78
Speaker 3M	82	79	79	79	82	77	82	79
Speaker 4M	84	82	85	82	81	81	85	82
Speaker 5M	74		74	69	73	75	77	77
Speaker 6M	72	69	73	67	70	68	71	73
Speaker 7M	85	82	84	78	81	76	83	80
Speaker 8M	79	77	79	76	79	76	82	81
Speaker 9M	88	84	89	84	86	81	88	83
Speaker 10F	79	79		76	78	75	81	77
Average	**80**	**78**	**81**	**77**	**79**	**76**	**81**	**79**
St. dev.	5	5	0.53	5	4	3	4	2

Table 10: Intensity Measurements

The interspeaker variability analysis reveals that for stops with the same place of articulation, the voiceless segment is louder than the voiced one on a linear scale. For each such pair, in 31 out of 37 instances (83.78%), the participants produced the voiceless stop higher in sonority than the voiced counterpart. In five instances (13.51%), voiceless and voiced stops have equal sonority. In the case of [k] (73 dB) and [g] (75 dB), Speaker 5M produced the former with a greater sonority than the latter. Speaker 6M did the same for [k͡p] (71 dB) and [g͡b](73 dB). Except for these two instances (5.40%), the overall pattern is that in Anyi, voiceless stops are linearly louder than voiced ones. This general pattern can be displayed visually in Figure 7:

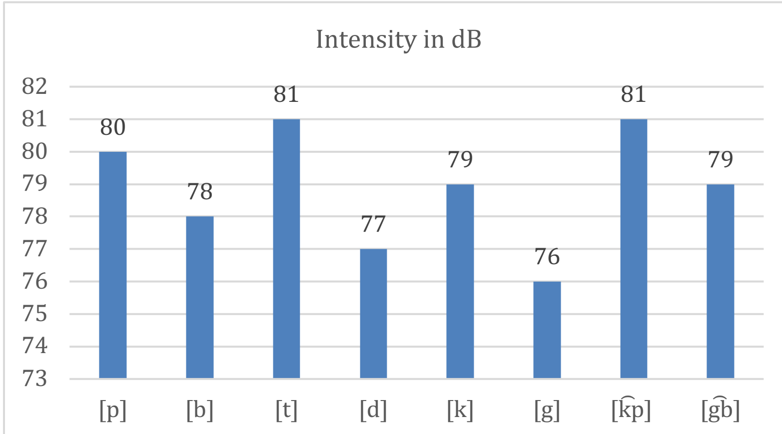

Figure 7: Sonority Profile of Stops

The linear scale in Figure 7 does not mirror the logarithmic (auditory scale) for [p] (80 dB) and [b] (78 dB) and [k͡p] (81 dB) and [g͡b] (79 dB) because the sonority differences are not auditorily perceptible, since the naked ear cannot detect intensity difference below 3 dB. However, the sonority differences between the [t] (81 dB) and [d] (77 dB) and between [k] (79 dB) and [g] (76 dB) are auditorily robust because they are respectively 4 dB and 3 dB.

Finally, Ladefoged and Maddieson (1996:259) describe doubly articulated stops as being very loud sounds. By their description, one is under the impression that they are louder than singly articulated stops. However, our findings do not support this view. The intensity of [k͡p] (81 dB) is not auditorily greater than that of [p] (80 dB), nor is it louder than that of [k] (79 dB). As for [g͡b] (79 dB), it is not auditorily different from [b] (78 dB). Yet, the difference between [g͡b] (79 dB) and [g] (76 dB) is auditorily salient because it reaches the JND of 3 dB. Even so, as noted by Connell (1994:443), "Labial-velars tend to resemble more labials than they do velars." This is confirmed by the intensity measurements. This may explain why doubly articulated stops in Anyi are not louder than their singly articulated counterparts.

4.10 Bandwidth Measurements for Speech Synthesis

Bandwidth measurements are rarely provided in acoustic phonetic studies because they are not important in human-to-human verbal interactions. Yet, they are extremely important for speech synthesis. For discussions of bandwidths for speech synthesis, see Koffi and Petzold (2022). Suffice it to quote Rabiner and Juang (1993:153) to illustrate this point, "Unlike the frequency JNDs, the bandwidth JNDs do not show clear dependence on either the bandwidth itself or the formant frequency." For the measurements in Table 11, I did not extract bandwidth data directly from Praat. Instead, I relied on the JND of 20 to 40% discussed by Rabiner and Juang (1993:152) to help estimate the bandwidths of each stop consonant. The JND of bandwidths can be stated as follows:

Auditory Discrimination in Bandwidth

The bandwidth of any segment is based on 20% of the arithmetic mean of its corresponding formant frequency.

The JND of 20% is chosen because smaller bandwidths are better for intelligibility than bigger ones. Bandwidths are measured in Hz. Table 11 lists the bandwidths of Anyi stops.

	B1	B2	B3	B4
[p]	151	312	531	773
[b]	100	307	536	764
[t]	131	353	561	792
[d]	93	351	556	800
[k]	163	351	532	884
[g]	102	349	535	815
[k͡p]	101	303	539	771
[g͡b]	66	358	585	820

Table 11: Bandwidth Estimations

These estimates are considerably higher than those of English in Klatt (1980:975, 986-987). I have not found bandwidth measurements for any African language. Therefore, I do not know if the data in Table 11 is typical or not. For English, Klatt reported no sound with a bandwidth value of less than 30 Hz and nor a bandwidth value greater than 800 Hz.[11] However, we see that the B4 measurements of the voiced stops [d, k, g, g͡b] are higher than 800 Hz in three instances.

5.0 Limitations

The methodology adopted for this study, which extracted all acoustic correlates of stops that occur before the vowel /a/, has advantages and limitations. On the plus side, we understand the acoustic properties of stops when they precede the same vowel /a/. On the downside, Anyi has nine oral vowels /i, ɪ, e, ɛ, a, o, ɔ, ʊ, u/ and six the nasal vowels /ĩ, ɛ̃, ã, ɔ̃, ʊ̃, ũ/. A more comprehensive study would have examined the behavior of stops in all these different phonological environments because it is well known that some acoustic properties of stops change depending on the vowels they precede. This limitation notwithstanding, the study has the merit of being the first to have examined the behavior of stops in such a comprehensive manner in a West African language. Others are invited to supply what is missing.

6.0 Summary

Twelve acoustic correlates are extracted to help gauge their robustness in the intelligibility of stops in Anyi. These findings lead to two separate conclusions. First, 1) duration, 2) VOT, and 3) F1 are indisputably robust correlates. The first two go hand in hand. Because Anyi is a true voicing language, its voiced stops are longer than its voiceless ones. As a result, VOT is a very strong correlate. Voiceless stops are short-lagged whereas voiced stops are long-lagged with considerable negative VOTs. Nine out of 10 participants differentiate between voiced and voiceless stops on the basis of VOT and duration. Additionally, but unexpectedly, F1 is also a robust correlate. The adverb "unexpectedly" is used because I have not come across any study that has made this point before. In Anyi, both the arithmetic means and the interspeaker variability analyses show that speakers open

[11] These are presumably based on men's bandwidths. However, Klatt (1990) claims that they work for both male and female voices.

their mouths slightly more widely for voiceless stops than voiced stops ones. This pattern applies to pairs of stops that have the same place of articulation. The difference in F1 contributes to auditory differential perception because in every case, the acoustic distance between voiced and voiceless stops is greater than the JND of 60 Hz.

The analyses have also revealed that 1) F0, 2) F3, and 3) F4 are not robust for audibility. In every case, the acoustic distances between voiced and voiceless stops with the same place of articulation are less than the JNDs required for audibility. Klatt (1980:987) shows that F2 and F3 are not robust cues by reporting the exact same measurements for voiceless and voiced stops. He did not provide data on F4. However, it is worth underscoring that F2 is robust for encoding and decoding differences between [k͡p] (1516 Hz) and [g͡b] (1790 Hz).

Intensity and F2 are a mixed bag. The voiceless stops [t, k, k͡p] tend to be louder than their voiced counterparts [d, g, g͡b]. However, [p] is not louder than [b]. This is not very surprising since [p] is not a high frequency segment in Anyi. Speakers do not need to differentiate it from [b] because, as noted in 2.1, the former occurs only 1% of the time. Ladefoged and Maddieson (1996:259) report that doubly articulated [k͡p] and [g͡b] are louder than singly articulated stops. However, our intensity measurements do not support this finding for Anyi. Instead, what we see is that [k͡p] (81 dB) [t] (81 dB) are the loudest sounds in the language. F2 is not robust for singly articulated stops but it is for doubly articulated ones.

The bandwidth measurements in B1, B2, B3, and B4 show that Anyi differs markedly from English. This is to be expected since they belong to two different language families. What we need are measurements from West African languages to see how Anyi compares with genetically related languages.

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 **54,717** times (**37,140** as per Digital Commons analytics), **17,577** (as per Researchgate.net analytics), and several thousand downloads from Academia.edu, as of **February 2023**. He can be reached at enkoffi@stcloudstate.edu.

References

Atal, B. S. 1972. Automatic Speaker Recognition Based on Pitch Contours. *Journal of the Acoustical Society of America*, 52 (6):1687-1697.

Baken, R. J. and Robert F. Orlikoff. 2000. *Clinical Measurement of Speech and Voice*. 2nd Edition. San Diego, CA: Singular Publishing Group.

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/

Burmeister, Jonathan. 1983. L'Agni. *Atlas des Langues Kwa de Côte d'Ivoire*, ed by. G. Herault, pp. 155-172. Abidjan, Côte d'Ivoire: Université d'Abidjan, Institut de Linguistique Appliquée et Agence de Coopération Culturelle et Technique.

Burmeister, Jonathan. 1981. *Anyi-Baule Dialect Survey Notes*. Unpublished paper.

Cao, Honglin and Volker Dellow 1999. The Role of the First Five Formants in Three Vowels Mandarin for Forensic Voice Analysis. International Conference on Phonetic Sciences. https://icphs2019.org/icphs2019-fullpapers/pdf/full-paper_796.pdf.

Cho, Taehong andPeter Ladefoged. 1999. Variation and Universals in VOT: Evidence from 18 Languages. *Journal of Phonetics* 27:207-229.

Connell, Bruce. 1994. The Structure of Labial-velar Stops. *Journal of Phonetics* 22:441-476.

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

Fastl, Hugo and Eberhard Zwicker. 2007. Pschoacoustics: Facts and Models. 3rd Edition. New York: Springer.

Fry, Dennis. B. 1979. *The Physics of Speech.* New York: Cambridge University Press.

Gradoville, M. (2011). Validity in measurements of fricative voicing: Evidence from Argentine Spanish. Online Version, retrieved from http://www.lingref.com/cpp/larp/5/paper2635.pdf.

Halle, Maurice, G.W. Hughes, and J.P. A. Radley. 1957. Acoustic Properties of Stop Consonants. *Journal of the Acoustic Society of America* 29 (1):107-116.

Himmelmann, Nikolaus P. and Robert D. Ladd. 2008. Prosodic Description: An Introduction for Fieldworkers. *Language Documentation and Conservation* 2: 244-274.

Houtsma, Adrianus J. M. 1995. Pitch Perception. Hearing: Handbook of Perception and Cognition. Second edition, ed. by Brian C. J. Moore, pp. 267-295. New York: Academic Press.

Katz, William F. 2013. *Phonetics for Dummies*. Hoboken, NJ: John Wiley & Sons, Inc.

Klatt, Dennis H. 1987. Review of Text-to-Speech Conversion for English. *Journal of the Acoustical Society of America 83 (3):737-793.*

Klatt, Dennis H. 1980. Software for a Cascade/Parallel Formant Synthesizer. *Journal of the Acoustical Society of America 68 (3):971-995.*

Klatt, Dennis and Laura Klatt. 1990. Analysis, Synthesis, and Perception of Voice Quality Variations among Females and Male Talkers. *Journal of the Acoustical Society of America 87 (2):820-857.*

Koffi, Ettien. 2021. *Relevant Acoustic Phonetics of L2 English: Focus on Intelligibility*. Boca Raton, FL: CRC Press.

Koffi, Ettien. 2020. A Comprehensive Review of Intensity and its Linguistic Applications. *Linguistic Portfolios* 11:2-28.

Koffi, Ettien. 2018. The Acoustic Vowel Space of Anyi in Light of the Cardinal Vowel System and the Dispersion Focalization Theory. In Jason Kandybowicz, Tavis Major, Harold Torrence & Philip T. Duncan (eds.). *African Linguistics on the Prairie*: Selected Papers from the 46th Annual Conference on African Linguistics 191-204. Berlin: African Language Press. DOI:10.5281/zenodo.1251730

Koffi, Ettien. 1990. *The Interface between Phonology, Morpho(phono)logy and the Standardization of Anyi Orthography*. Ph.D. Dissertation: Indiana University, Bloomington, Indiana.

Koffi, Ettien and Mark Petzold. 2021. A Tutorial on Formant-Based Speech Synthesis for the Documentation of Critically Endangered Languages. *Linguistic Portfolios* 11:26-55.

Labov, William., Sharon Ash, and Charles Boberg. 2006. *Atlas of North American English:Phonetics, Phonology, and Sound Change.* New York: Mounton de Gruyer.

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

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

Ladefoged, Peter. 1968. *A Phonetic Study of West African Languages: An Auditory-Instrumental Survey.* Cambridge, UK: Cambridge University Press.

Ladefoged, Peter and Ian Maddieson. 1996. *The Sounds of the World's Languages.* Malden, MA: Blackwell Publishers.

Lisker, Leigh and Abramson, Arthur. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20* (3), 384-422.

Maddieson, Ian and Peter Ladefoged. 1989. Multiply Articulated Segments and the Feature Hierarchy. *UCLA Working Papers in Phonetics* 72:116-138.

Quaireau, André. 1987. *Description de l'Agni: Des Parlers Moronou, Ndénié et Bona*. Thèse de Doctorat d'Etat. Université de Grenoble III.

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

Retord, Georges L.A. 1980. *Étude Radiocinématographique des Articulation de l'Agni-Sanvi*. Thèse de Doctorat d'Etat. Paris, France: Librairie Honoré Champion.

Smith, Caroline L. 1997. The Devoicing of /z/ in American English: Effects of Local and Prosodic Context. J*ournal of Phonetics* (25): 471-500.

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

Takemoto, B.G, S. Adachi, and T. Kitamura. 2006. Acoustic Roles of the Laryngeal Cavity in Vocal Tract Resonance. *Journal of the Acoustical Society of America* 120 (4):2228-2232.

Welmers, Wm. E. 1973. *African Language Structure*. Los Angeles: University of California Press.