
The acoustic analysis of speech: A precursor to better speech performance and perception

George E. Onwudiwe

Abstract

Speech production and speech analysis require careful study in order to help communication and language study. To do a thorough job and a near error proof analysis requires that authentic approach would be employed. Mere ear perception has been observed to be insufficient, inconsistent and unreliable that a considerably more reliable approach needed to be sought for, hence the introduction of acoustics in speech analysis. This approach has been applauded by many modern linguists, including Ladefoged (1962, 2003); Pickett (1980); Ashby and Maidment (2005); Clark, Yallop and Fletcher (2007), among others. This paper is therefore a contribution to the call for the adoption of the acoustic analysis in the study of speech, an approach that accounts for the titbits in speech using modern computerised machines that identifies and clears all confusions in any individual's speech. Data for this discussion were collected from relevant scholarly literatures on acoustic analysis of speech. This paper adopts the descriptive approach in the discussion. The acoustic investigation is considered better in this era of digitisation. The paper therefore recommends that acoustic analysis should necessarily be used to authenticate auditory perception and then enhance performance.

Introduction

Until very recently, speech analysis had been conducted mainly through ear perception. Although this method enjoyed very wide and unperturbed acceptance and practice, it was fraught with several guesses and inconsistencies and as such produced many results that are either erroneous or inconsistent. In fact, it was far from being authentic. Human speech act is full of fluctuating varieties. Besides that, some speakers are incompetent in the observance of correct acoustic principles of speech production. Other speakers exhibit individual idiosyncrasies and show of style while speaking. Languages themselves share varying nuances of pitch pattern.

Due to these variations and idiosyncratic habits of speakers, meaning of speech is at times unclear and confusing. Hence, at times, what is perceived by the ear is not correct and this often leads to misconception. This is however not to discredit ear perception; after all, for any meaningful machine interpretation of speech to take place, ear perception must take the front stage. No wonder Westerman and Ward (1990), as well as Laver (1994) argued that ear perception must precede every speech analysis.

Nonetheless, despite the support and campaign for ear perception by some scholars, notably Westerman and Ward (1990), Laver (1994), Harley (1995) modern researches have stiffly challenged and possibly overthrown sole dependence on ear perception for speech analysis. Clark, Yallop & Fletcher (2007:230) aptly state, "It is characteristic of human perception that the sensations we experience in response to stimuli rarely correspond directly with the values we derive from measurement of those stimuli". For instance, they observe that the human auditory system is capable of responding to an enormous range of sound intensities with the upper end ranging more than a million

times greater than the lowest perceivable intensity. Therefore they opine,

Not only does this lead to some very inconvenient numerical values, but, given the nature of perception, the figures do not relate very well to the perceptual effects of differences in intensity...The relation between perceived loudness and acoustic intensity is more nearly logarithmic...Hence the most convenient way to express intensity so that it relates to perceived loudness is as a logarithmic ratio,...

(p. 230).

Similarly, pitch, as the perceived period or frequency of a sound wave is determined largely by the fundamental frequency of the sound. But, the relationship between pitch and fundamental frequency is nonlinear and varies with the frequency involved. Besides, Clark, Yallop & Fletcher further contend that our sensitivity to changes in the frequency of a sinusoidal tone, that is, our pitch discrimination varies as we move up the audible frequency scale. To check this, they proclaim that a perceptual unit called the MEL was devised to represent equal increments of pitch and relate them to frequency. This is one great advantage of acoustic approach to speech analysis.

This study therefore is significant in many respects. Hence, to avert, or at least manage pitch variation in languages, as well as understand different nuances in languages, and also account for individual speech habits and linguistic incompetence, it is very necessary and safer to analyse or decompose sound into its acoustic parameters to identify the fluctuating frequency of speech. Consequently, this paper strongly argues in favour of the acoustic approach to speech analysis as this will curb to a very large scale

problems of misconception which mere ear perception may create for the speech analysts.

Bussmann (1996:4) affirms: “Many recent phonological investigations make extensive use of the concepts and terminology of acoustic phonetics”. Therefore, in this modern age of computerisation, acoustic approach to speech analysis will be a welcome and result-oriented option to speech analysis. Its adoption will stimulate greater interest in experimental phonetics, particularly in developing languages including the African languages and, in so doing improve their standards and match them with other more developed languages of the world. Acoustic analysis will also aid better performance in speech as well as promote better perception and communication.

This paper is expected to immensely benefit language analysts. It will also be of great benefit to journalists, newscasters and language teachers. The paper equally hopes to expose the gains of the acoustic study of speech as well as demystify the phobia that often surround the Acoustic phonetic studies and thereby stimulate students’ interest in the area of phonetic studies. The paper adopts the descriptive approach in presenting its argument. This is because, the approach is considered more relevant and convenient to the discussion; and is viewed to be more error-proof.

Relevant literatures were read from where relevant information presented in this discussion were collected, while illustrations were based on the data collected also mainly from these literatures. It is therefore hoped that the arguments marshalled in this paper will be convincing enough for its purpose as to draw many readers to the use of the acoustic approach in the analysis of speech and speech sound as well as improve performance, perception and communication.

Conceptual framework

Here, major concepts or theories that are relevant to this discussion are illustratively explained in order to provide better apprehension for the reader.

Acoustics

Acoustics is a branch of physics which is devoted to the study of sound. It adopts instrumental techniques of investigation, especially electronics. When adopted into language study and analysis, it helps in the investigation of the physical features of speech, and is hereby referred to as the ***acoustics of speech***. Hence, Crystal (2003:7) says: “Its importance to the phonetician is that acoustic analysis can provide clear, objective datum for investigation of speech – the physical ‘facts’ of utterance”. This implies “the analysis of the acoustic characteristics (such as amplitude, **quantity**, and frequency) by means of electronic instruments” (p. 4).

Crystal (2003) then asserts that acoustic evidence is always sought for to support articulatory or auditory phonetic analysis. This evidence therefore projects acoustics as an important instrument for major aspects of speech synthesis.

Acoustic phonetics

When acoustic evidence is applied in linguistic study of speech, it becomes phonetics, specifically acoustic phonetics. Bussmann (1996) defines acoustic phonetics as: “Branch of general **phonetics** that investigates the physical properties of the acoustic structure of **speech sounds** according to frequency (**pitch**), **quantity** (duration), and intensity (spectrum)”. Bussmann’s definition of acoustic phonetics highlights the major exponents of speech which when handled well, production and perception of speech become

easy and better. Of course, reverse becomes the case when not considered or properly handled.

Finch (2000) corroborates Bussmann (1996), adding that as we speak, the air molecules around are disturbed in such a manner that they oscillate (move back and forth). Each complete back and forward movement is technically called a **cycle**. These cycles are then plotted with the aid of instruments to arrive at the particular **frequency** of a sound; and the more cycles that occur in one second, the higher the frequency. Finch (2000:34) therefore enumerates some gains in the acoustic investigation of speech sounds to include:

- determination of the voice frequency of different genders and ages.

This claim is justified by Laver (1994) and Clark, Yallop and Fletcher (2007) in Onwudiwe (2020) with the categorisation of the voice frequency of different genders, including children as in example (1) below:

(1) <i>Laver (1994) classification</i>	<i>Clark, Yallop and Fletcher (2007) Classification</i>
Adult males: Av. 120 Hz.	80 – 200 Hz.
Adult females: Av. 220 Hz.	150 – 300 Hz.
Children: Av. 330 Hz.	200 – 500 Hz.

These classifications indicate variations in the frequency of woman's voice *vis-à-vis* that of the man, on the one hand; and that of children against those of the adults. For the women, it is clear here that the voice frequency is about twice higher than that of the man. No wonder Finch asserts that through acoustic phonetics, it can be discovered that "on the average the frequency of a woman's voice is twice that of the man's". Although we can conjecture this fact through mere auditory perception as most women are known

to operate on higher pitch when they speak than the men, this cannot be scientifically accounted for. Hence, it is shrouded in guessing.

In like manner, children's voice frequency is noticeably higher than those of the adults when they speak. Thus, looking at the examples in (1), it is evidently clear that an average voice f_0 of 330 Hz and 200 – 500 Hz respectively are approximately between twice and three times higher than those of their adult counterparts. This scenario is demonstrated in children's high-pitched speech.

Further evidence of variation in voice frequency of men and women in discourse is equally accounted for by Laver (1994) in Onwudiwe (2020) as shown in example (2) below:

- (2) Adult men: 50 – 250 Hz.
Adult women: 120 – 480 Hz.

By this classification, a clear distinction is made between the voice frequencies in an ordinary speech and those in discourse. This certainly makes it understandable why good speech and normal discourse go the way they do.

- Finch (2000) also reveals that through acoustic phonetics, “it is also possible to measure the frequency of individual speech sounds, both vowels and consonants, and show that they have their own distinctive resonance”.

In vowel harmony in Igbo for instance, many scholars describe the patterning of the vowels without recourse to acoustic analysis. Hence, the eight vowels in the standard Igbo are generally classified as ‘a’ group of vowels and ‘e’ group of vowels; or light and heavy vowels respectively. However, Pike (1947) propounded a phonetic feature for analysing vocoid performance which was described as “tongue root position in vocoid articulations”. In 1963, Stewart an African phonologist, investigated Pike's

hypothesis in his work on Twi where he observed first that vowel phonemes could be classified into two sets occurring in mutually exclusive words in the process called vowel harmony, Stewart (1967). As Laver (1994:289) reports: “The pronunciation of one such set, Stewart suggested might involve adjustments of the root of the tongue, and this could be tested by x-ray investigations”. Following Stewart’s conversation with Ladefoged on the hypothesis, during which he urged Ladefoged to investigate the physiological basis of distinctions that then were ascribed to differences of lax and tense muscle tension, Ladefoged came up with this conclusion which Laver (1994) again reports thus:

Ladefoged then independently reached the conclusion, from x-ray investigations of vowel harmony in West African language Igbo, that a tongue-root advancing feature was indeed instrumental, at least in Igbo, for creating audible and distinctive differences of vocoid-quality(p. 289).

The tongue-root hypothesis as used for the Igbo vowel harmony is represented in the chart below (Adapted from Emenanjo, 1978):

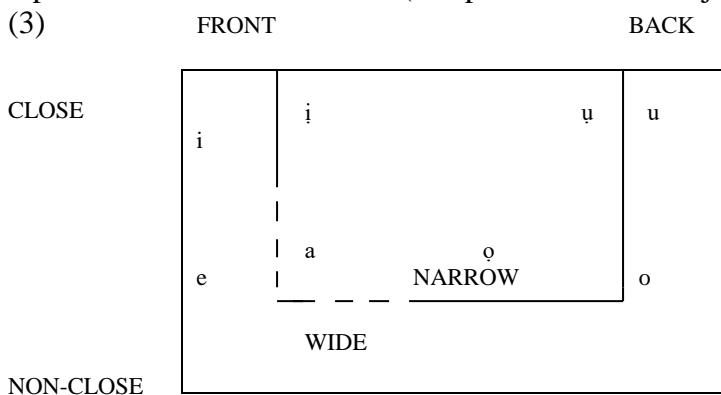


Chart showing x-ray display of vowel harmony in Igbo

According to Laver, adjustments of the root of the tongue in this way constitutes a topographical aspect of articulation, which “has come to be called **advanced tongue root** (often abbreviated to **ATR**, mostly by phonologists using it to characterize vowel-harmony effects)”. However, “Ladefoged uses the terms ‘wide’ and ‘narrow’ for advanced versus non-advanced tongue-root positions, reflecting the differences in the front-to-back dimension of the pharynx”. Emenanjo (1978:6) then comments:

For a very long time now scholars of West African languages have been studying the various mechanisms responsible for the patterning of vowels into the two harmony sets. Earlier investigations by Ladefoged (1964) and Stewart (1967), showed that the position of root of the tongue conditions the system.

Thus, Ladefoged (2006:181) notes “...we can say that a vowel sound contains a number of different pitches simultaneously. There is the pitch at which it is actually spoken, and there are the various overtone pitches that give it its distinctive quality”. As he goes on to observe, each vowel is distinguished from another by differences in these overtones; and each of the vowels has “three overtone pitches”. The lowest pitch, according to him is formant one, represented as F1 “could be heard most easily when the vowels were produced with a creaky voice, while the second formant, F2 “goes down in pitch, as can be heard more easily when these vowels are whispered”. Formant three, F3, he says “adds to quality distinctions...”. He however concludes his observation thus: “If you check a complete set of vowel positions...with this technique, you should hear the pitch of the first formant going up for the first four vowels and down for the second four vowels”.

In our own case here, the complete set of vowels are [i, ɪ, e, a, ɔ, o, u, ʊ]. These vowel sounds represent the vowels in the chart in (3) above.

These explanations and illustrations show the importance of acoustic phonetics in the analysis of speech sounds and the efforts earlier made to authenticate any analysis made outside acoustic procedure. We should note that similar acoustic analysis can be made for consonants as presented in the IPA (International Phonetic Alphabets). Without these experimental approach to the analysis of speech sounds, our analysis would have been mainly conjectural. Hence, Finch (2000) asserts:

Phoneticians use a **spectrograph**, a machine designed to analyse/decompose sound into its acoustic parameters, to capture the fluctuating frequency of speech. This produces a **spectrogram**, a kind of chart which shows frequency in terms of relative degrees of light and dark. On the basis of these kinds of experiments phoneticians can establish the acoustic structure of speech and demonstrate the distinctiveness of particular segments (p.34).

Consequently, Clark, Yallop & Fletcher (2007) record that prior to the establishment of spectrograph in the 1940s, it was a very tedious task to embark on an acoustic analysis of speech. To dare this was also restricted by dearth of equipment. Hence, formant structure and its auditory qualities of speech sounds were very little explored in natural speech. So, they state: “Given the problems of providing a reliable auditory description of vowel quality ..., the availability of an ostensibly objective technique of acoustic analysis, free from the bias of the human observer was an important step in phonetic and phonological description” (p. 264).

This is the main stand and focus of this paper as information can be better managed and appreciated via this approach.

Presently, sound waves are mainly handled in digitised form. Hence, Ashby and Maidment (2005:29) state:

The tracks on a CD, or **wav files** in a computer, are simply long strings of numbers representing waveforms sampled at regular intervals. The **sampling rate** controls what frequencies will be preserved when the wave is reconstructed. Basically you have to sample at a rate that is at least twice the highest frequency you need to show.

Explaining this exercise further, they say that a CD works at a sample rate of more than 40 kHz, enabling it to provide a ‘hi-fi’ sound to 20 kHz or so. Two major advantages of digital analysis of speech signals have been identified by Clark, Yallop & Fletcher (2007:258). In their words,

The first is that once the signal has been digitally encoded and stored, it can be edited, processed, measured, manipulated and filed with far efficiency than is possible with analogue instruments and an ordinary tape recorder. The second is that the analysis itself can be more easily varied to give optimum time- and frequency-resolution properties.

The contention by these scholars has given great credence to the option of acoustic analysis of speech perception over mere ear perception. The advantages enumerated above equally necessitate the option of this approach for speech study and analysis.

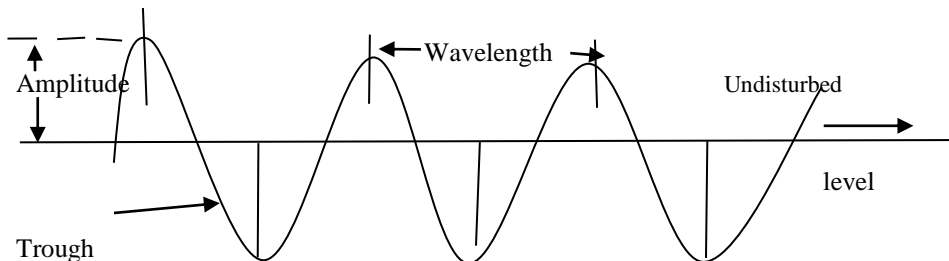
It should then be noted that in deciding what to record requires basically acquiring the knowledge of phonetics and understanding the phonology of the language of study. Hence, Ladefoged (2003) suggests that when looking for sets of words “...another major source is the speakers of the language themselves”. This, according

to him, is to confirm that the words exist. He also recommends the Digital Audio Tape (DAT) for recording of data, and further lists four properties to be looked for in a good recorder to include a good frequency response (the range of pitches the system can record); a good signal/noise ratio (the range of loudness); reliability and user-friendliness; and the possibility of using the recorder for a long time. All these properties, he claims can be found in the DAT.

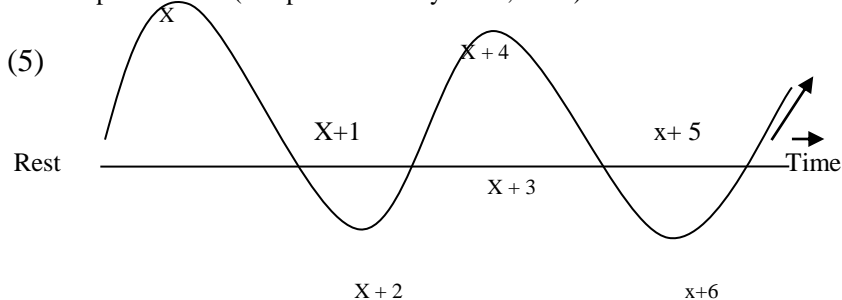
Acoustics of speech production

Speech as sound results from vibration of vocal chords. Vibration itself is a factor of some source of energy that generates it in a form of waves. But for any sound to be audible, three criteria need be satisfied, viz. “propagating medium”, “frequency” and “amplitude”. (Clark, Yallop and Fletcher 2007). Anyakoha (2010:248) defines wave as “a disturbance which travels through a medium transferring energy from one point to another without causing any permanent displacement of the medium”. The source of every wave is vibration. Waves such as sound are transferred through a process known as wave motion. Sound wave is an example of mechanical wave; and a mechanical wave is an example of waves that require material mediums for their propagation. Below are examples of wave.

(4)



Wave representation (Adapted from Anyakoha, 2010)



A waveform indicating the pressure wave built up when particles move. The horizontal line represents the passage of time. (Adapted from Crystal, 1997).

The propagating medium, such as air, metal, liquid is what the sound will travel through. The medium provides a link between the source and destination of sound. If there is no lead, no sound can be heard as sound does not travel in a vacuum. Besides, the medium is the property of sound that is relative to the sensitivity of sound to the ear. This property is frequency of vibration which ranges from very rapid to very slow. As explained by Clark, Yallop and Fletcher (2007:205), “The ear detects only a certain range of these frequencies, commonly down to about 20 vibrations per second and up to about 20,000 vibrations, per second...”

Ashby and Maidment (2005) also describe frequency as the number of repetitions (or cycles) per second, formerly expressed in cycles per second, (c.p.s or c/s). But presently, the unit used is Hertz (Hz), hence, 1Hz = 1 c.p.s. The range of vibration differs from individual to individual, and is affected by age. However, inconsistency in phonation from cycle to cycle may have some

effect on vocal quality. Hence, they say, “all speakers seem to exhibit some inconsistency in duration from cycle to cycle of phonation” (Clark, Yallop and Fletcher, 2007:235). This inconsistency gives rise to “pitch jitter” which is at its highest at the start of phonation before a voiceless consonant, and then decreases in the syllable peak. Fundamental frequency, which is the frequency of vibration of the larynx in phonation is measured from the speech waveform.

The waveforms of voiced and voiceless sounds are different. Ashby and Maidment (2005) explain that when the pattern of a wave repeats regularly in time, the wave is regarded as a “periodic wave”. A period runs from one clearly identifiable point on the wave to the next place where the point occurs. Therefore, “one period of a simple (sine) wave contains one upwards-and-over excursion, and one downwards-and-up-again excursion, returning to the zero line. The length of one period is the periodic time, T ” (Ashby and Maidment, 2005:28).

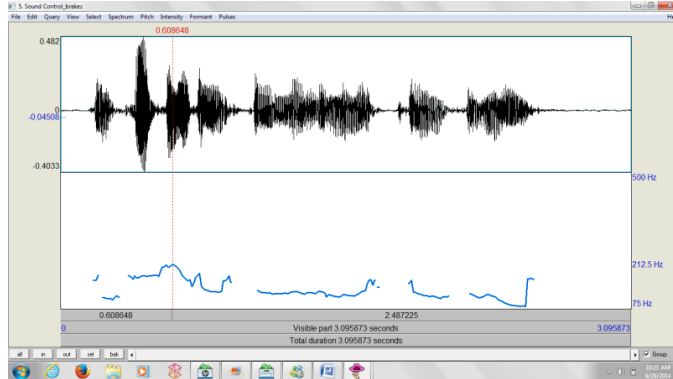
Apart from frequency, another phonetically important property of sound waves is duration. Durations being considered may be as small as a fraction of one cycle of a periodic waveform, or may be one complete period of vibration, or may even be far longer. All this depends on the nature of utterance being considered. Hence, Clark, Yallop and Fletcher (2007:223) state, “in some instances, we want to determine the duration of a whole word or utterance, or even the duration of a silence such as may occur in the closure of voiceless stop”. This is one important characteristic of acoustic study of speech.

However, Ashby and Maidment (2005:127) contend, “...the duration of a particular sort of segment is not fixed. It will vary depending on the context in which the segment appears”. They further report that the last sound or two at the end of an

utterance are generally prolonged giving rise to what they refer to as “pre-pausal lengthening”. Also, a syllable-initial consonant is known to be generally longer when alone, but considerably shorter when preceding another consonant in a cluster. Still illustrating the effect of context on duration of a sound, Ashby and Maidment further posit that the duration of a vowel may be influenced by whether the following consonant is voiced or voiceless. They observe this phenomenon in most English accents and refer to it as “pre-fortis clipping”. Other factors that influence segment duration that they observe include “overall speech rate, and degree of stress placed on syllable”.

The duration of the utterance is measured using reference markers on the waveform that have meaningful relationship to the phonetic structure that is being measured. This often involves displaying the waveform on a computer screen using speech editing and analysis package. Clark, Yallop and Fletcher (2007) argue that this process does not really yield a reliable and consistent means of measuring larger durations and determining the appropriate threshold of intensity that mark the start and the end of the speech to be measured. Hence they say, that a combined process of “time and frequency domain information” was advocated. See example (6) below:

(6)



A waveform indicating time and frequency for an utterance

Another criterion that needs to be satisfied for a sound to be audible is amplitude. The amplitude of a wave is a measure of the size of the pressure variations (or eardrum movements) (Ashby & Maiment 2005). Variation in air pressure results to loudness of a sound, and loudness is the auditory property that is correlated with amplitude. Hence, Ladefoged (2003) says that loudness of a sound can fairly well be determined by reference to intensity (a measure of acoustic energy) which he says, is its acoustic counterpart. In other words, Ladefoged is imputing that intensity depends on the amplitude of the sound wave.

Therefore, to measure the intensity of a sound requires taking the amplitude of the waveform at each moment in time during a window, squaring same and finding the mean of all the points in the window, and then taking the square root of the mean. So, Ladefoged concludes, “The power of a sound is the square of this mean”(p.90). The implication of the aforesaid is that large movement of the source of sound produces a loud sound which invariably necessitates exertion of more energy during production of the sound.

The amplitude of the sound refers to the strength of each peak of pressure; the rate at which the peak (of pressure) occurs refers to pitch. The difference between high and low-pitched sounds is that the higher-pitched sound is making a greater number of variations per second than the lower-pitched sound. Consequently, Ladefoged (1962:18) states, “the variation in air pressure in any sound that has a definite pitch must form a pattern which is repeated at regular intervals”. See Example (4) above. For Pickett (1980), the air molecule motions of speech sounds can be described aptly by considering them to be made up of simple oscillation. He illustrates the motions of speech sounds with the motions of a pendulum and says,

An ideal pendulum, one that has no friction, moves back and forth in simple harmonic motion. The pendulum motion is very regular in time, but still it involves many different speeds and positions. ... this motion can be represented very simply by relating it to uniform motion on a circle. (p. 14)

Pickett’s illustration above presents two-value parameters for the description of pendulum motion thus: one giving the ‘rate of motion’, and another giving the ‘size of the motion’. These two values can be translated for our purpose here to ‘the time, or period’, for one complete revolution, and ‘the size of the motion’. A plot in time of the motion of ideal pendulum, he continues, gives a wave called “sine wave”; and its sinuous form is the form in time for all simple harmonic motions. Continued Pickett: “The period of a sine wave is the time for one complete cycle ... related to the rate of oscillation, is often given as the *frequency of repetition* of the cycle in a unit of time”. The frequency which he simply describes as “the reciprocal of the period” is given in cycles per second or Hertz (Hz). Also, “the distance or amount of sine wave

motion is called the *amplitude*”, which corresponds to the extent of the oscillation from the resting position.

For any sound wave, therefore, the extent or amplitude of the motion and its form describe the sound completely, in physical terms. Therefore, when a sound is received by the ear, the extent of air molecule motion “determines the loudness of the sound heard”, while the form of the motion “determines the timbre or quality of the sound heard”. Pickett then declares: “The heard quality and loudness of sound are very important to us and these are very neatly represented by means of the frequency and amplitude of simple harmonic motions” (p. 18).

Pickett (1980) as Clark, Yallop & Fletcher (2007), Anyakoha (2010) had earlier declared that the “propagating medium” for sound transmission to the ear is air, and describes sinusoidal sounds (the form in time for all simple harmonic motions) as pure tones. He, however, notes that vibrations (of sounds) can be affected by “damping” forces which make the vibrations to die out gradually in time. These forces of “damping” include air friction. All these acoustic variables of sound, viz. medium, frequency and amplitude with their correlates contribute to the relevance of any sound, both in production and perception. These variables further authenticate the genuineness of analysis of speech production using the acoustic approach.

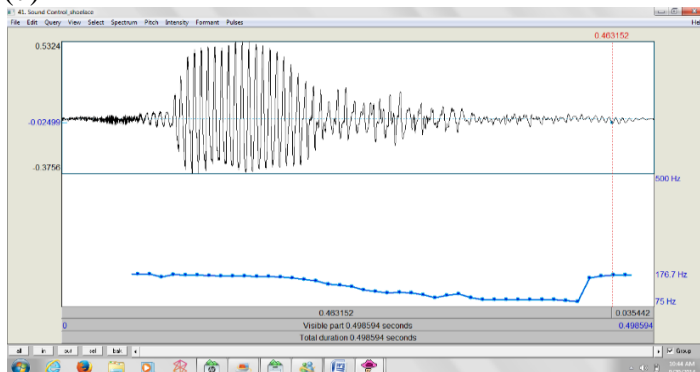
Discussion of acoustic analysis of speech

In this section, acoustic analyses of speech are demonstrated. This is to justify the aim of the paper.

Pitch variation

In line with the argument in favour of acoustic study of speech, Onwudiwe (2015) in his study of interface of tone and intonation using Igbo and English native speakers chose the acoustic (experimental) approach because he believes that that will help him better to identify the suprasegmental features they both share. Onwudiwe also states one of the major objectives of his study thus: “The study equally sets out to ascertain whether the native Igbo speakers encounter problems speaking English (intonation language), as well as identify the nature of such problems” (p.139). Let’s illustrate this with these experimental diagnosis of different Igbo native speakers of the compound word *shoelace* adapted from Onwudiwe (2015) in Example (7) below:

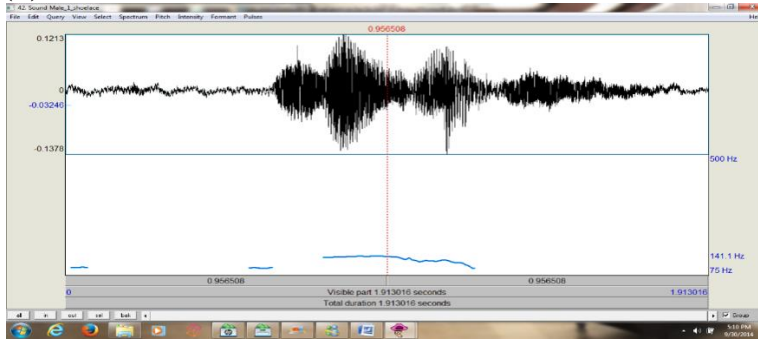
(7)



7(i) Sound wave, F_0 and time duration of Control utterance of *shoelace* [ʃu:lɜ:s]

The compound word, *shoelace* is produced in periodic cycles of 0.463152 secs and 0.035440 secs, with a total time duration of 0.498594 secs. The highest pitch value is 176.7 Hz. Clearly, the first syllable carries a greater prominence than the second with the contour indicating a rise on the second element.

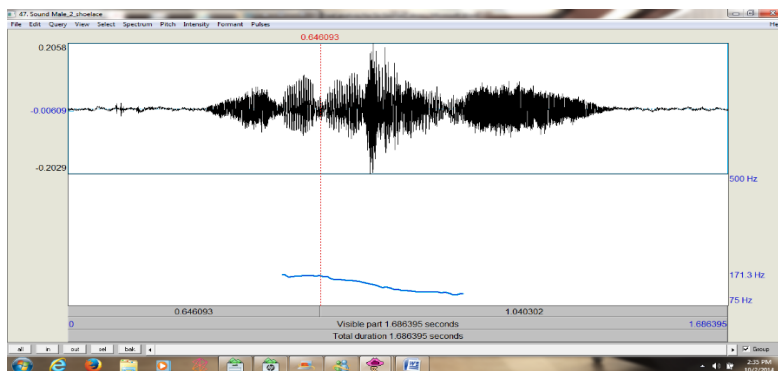
(ii)



7(ii) Sound wave, F_0 and time duration of Male 1 utterance of *shoelace* [ʃu:lɛɪs]

The Male 1 utterance of the compound was done in periodic cycles of 0.956508 secs and 0.956508 secs. The total time duration is 1.913016 secs, while the pitch value is 141.1 Hz. The pitch contour, however, presents more prominence on the first element and a downward movement on the second element indicating non observance of the fall-rise signalled by the diphthong at the end of the word.

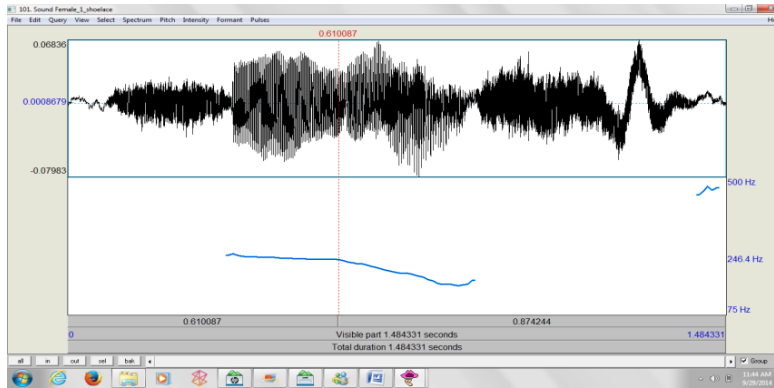
(iii)



7(iii) Sound wave, F_0 and time duration of Male 2 utterance of *shoelace* [ʃu:lɜ:s]

The periodic cycles used in uttering *shoelace* are 0.646093 secs and 1.040302 secs. The total time 1.686395 secs; and the pitch value is 171.3 Hz. The pitch contour rather indicates a glossed articulation of the syllables, although there seems to be a portrayal of stress at the inception of the utterance and a downward movement of the pitch at the end.

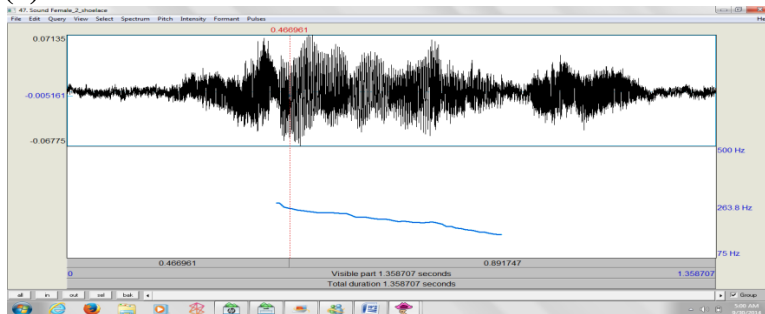
(iv)



7(iv) Sound wave, F_0 and time duration of Female 1 utterance of *shoelace* [ʃu:lɛs]

Female 1 uttered *shoelace* in periodic cycles of 0.610087 secs and 0.874244 secs. The total time duration is 1.0484331 secs, while the pitch value is 246.4 Hz. As in other Consultants, the pitch contour shows one stretch articulation, and it also begins on a high pitch which drops and rises perhaps in observance of the diphthong.

(v)



7(v) Sound wave, F_0 and time duration of Female 2 utterance of *shoelace* [ʃu:lɛs]

Female 2 presented *shoelace* in periodic cycles of 0.466961 secs and 0.891747 secs. The total time duration is 1.358707 secs. The highest pitch value measures 263.8 Hz. The pitch contour presents a straight string of utterance, which begins on a high pitch and gradually drops which actually indicates MT interference as Igbo does not have diphthongs.

The sound waves in (7 i) gives the example of the Control pronunciation, while (7 ii - iv) are those of none native speakers of English, who are native speakers of the Igbo language. The Igbo speakers are in two sets of two males and two females. This is to give enough variables for good analysis.

The result of the acoustic correlates of utterances for the word *shoelace* by the various speakers is as follows:

Control (English native speaker)-	Pitch value	=176.7	Duration	= 0.498594
	(Hz)		(ms) is	
Male 1 (Igbo native speaker)		= 141.1		= 1.913016
Male 2 (Igbo native speaker)		= 171.3		= 1.686395
Female 1 (Igbo native speaker)		= 246.4		= 1.484331
Female 2 (Igbo native speaker)		= 263.8		= 1.358707

The above result presents the different variables that bring about variation in performance by speakers as made manifest in the different sound waves (7i – v). Specifically, they show that Male 2 and Female 1 renditions resemble that of the English (Control), especially in observance of the fall-rise tone at the end of the utterance which is indicative of the diphthong /ei/.

However, the other Consultants' utterances of the compound, *shoelace* notably Male 1 and Female 2 end on fall instead. This is perhaps an evidence of interference of their native language, Igbo which is a tone language, and does not have

diphthongs in its sound inventory. Also clearly manifest is the difference in pitch variation displayed by the different speakers, particularly Male 1 and Female 2, who also are Igbo native speakers. This must be as a result of their different linguistic backgrounds. The variation in voice frequency of males and females is also evident as earlier discussed above. Furthermore, the varying speech duration of the different speakers acoustically account for their individual differences and speech habits.

Speech melody

Speech melody as the ‘train of relative pitch values that the listener perceives in the succession of syllables that make up the utterance’ is found in all languages, but their patterns differ. As Abercrombie (1967:107) observes:

In phonological analysis and description of the patterns of speech melody of both tone and intonation languages, it is not *absolute* that is of importance... Not only are both based on patterns arising from pitch fluctuation, but in both it is the position of the points in the pattern *relative to each other* that counts, not their frequency in terms of number of vibrations per second.

Corroborating Abercrombie, Laver (1994) notes that rather, it is dependent on one hand on a relative perceptual judgement the listener makes based on the general range of pitch the speaker’s voice is believed to move. Also, the pitch-value of a given syllable in a train of syllables in connected speech judged relative to the pitch-values of its immediate neighbours contributes to the yardstick for the determination. These pitch-values could be high, low, mid, etc. Therefore Laver concludes thus:

The **melody** of a speaker’s voice on any given occasion is thus a matter of the train of relative pitch values that the

listener perceives in the succession of syllables that make up the utterance, within the framework of the speaker's assumed **pitch-range**. (p. 457)

The speaker's range of pitch can be determined on either of these parlances: the "organic range" of the speaker's voice, the speaker's current "paralinguistic range" and the "linguistic range".

Despite a speaker's **speech range**, people use different melodic patterns which are distinctive and peculiar to languages. Hence, it can be found in all languages, and it helps to classify languages.

The melodic pattern of speech (speech melody) is therefore a consequence of fluctuation in the pitch of the voice, otherwise called "voice gesture" (Abercrombie 1967). The speaker's estimated range of pitch (voice gesture, speech melody) as earlier illustrated in Examples (1), (2) and (7), with detail in (7) also suffice here. To analyse the melody of a speech, nay language and speaker's pitch range; and to better interpret the consequences such as the ones listed above is not a matter for auditory perceptual analysis, rather it would require acoustic investigation.

Tonal downdrift

Pitch value of tones in utterances is not easily discernible via auditory perception. The pitch value of such tonal phenomenon as **downdrift** is one of such cases. Hyman (1975:154) defines downdrift as

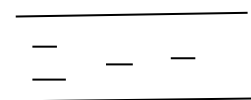
the process which results in high tones after low tones being phonetically less high than any preceding high tone in the utterance, which can be regarded as a type of assimilation with the low tone influencing the height of the succeeding high tone.

Emenanjo (1978:15) corroborates Hyman (1975) but with a disparity. In discussing downdrift as a feature of the Igbo sound

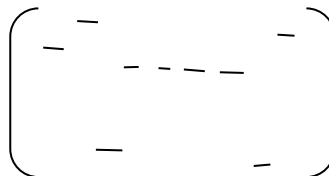
system, Emenanjo defines the phenomenon as the automatic and predictable lowering of tones as a result of their position in the utterance; and then explains the procedure thus: “As a result of downdrift the pitch of a high tone is lowered by one step every time it follows a non-initial low and that of a low tone is lowered every time it follows a high”.

Downdrift has also been interpreted as an intonational use of pitch as it can affect a whole utterance. Its effect on a tone language, according to Welmers (1973) gives the language a **terraced-level** effect. Hence, (Emenanjo 1978:16) posits: “When, therefore, in an utterance there is a series of alternate high and low tones, because of the automatic lowerings of these tones, the utterance progresses down like the levels of a terrace or the steps of a staircase...”

Secondly, downdrift has been observed to have the capability to produce several types of intonational patterns. For instance, Kelly (1969:159) reports that in Urhobo of Southern Nigeria, “successive high tones remain at the same pitch-height, and successive low tones maintain a common level”. Then Laver (1994:473) observes: “High tones interspersed with low tones drift downwards, but the low tones interspersed with the high tones are represented by Kelly as maintaining their standard baseline”. This procedure does not, however, agree in its entirety with the system in Igbo, in South Eastern Nigeria as reported by Emenanjo (1978) above. Therefore, to unravel this type of procedural disparity would require acoustic analysis as represented in Example (8) below:

- | | | | | |
|-----|-----|---|------|--|
| (8) | (i) |  | (ii) | Anyị agawala nzuko
L H L H H H H L H
3 1 4 2 2 2 2 5 3 |
| | | [óðibó] 'banana' | | |

Tonal downdrift in Urhobo
(Adapted from Laver 1994)



Tonal downdrift in Igbo
(Adapted from Emenanjo 1978)

Example 8 (i) and (ii) both present the downward drift of high tones when interspersed with low tones. But 8 (ii) further demonstrates the disparity of one-step downward lowering of high and low when following low and high tones respectively. Therefore, downdrift phenomenon can better be determined from such acoustic investigation than through mere auditory perceptual analysis.

Analysis of loudness

The loudness of the speech of individuals varies. In the same vein, the reasons for the variation are also different. Laver (1994) enumerates the factors germane for variability of the loudness of individuals to include sociolinguistic accent-community which the individual is a member of; social convention which prescribes different ranges of loudness for males and females. Other factors are the organic nature of the speech apparatus in every speaker and phonation types.

Whichever factor that is responsible for the loudness of a speech, loudness has been described as a perceptual physical correlate of frequency which relates to the physical concept of intensity which in turn is proportional to the amplitude of oscillations of air molecules in sound waves passing through the atmosphere (Laver, 1994; Clark, Yallop and Fletcher, 2007).

The loudness in a speaker's speech can be perceived auditorily but the degree of the difference from one speaker to another, or from one social milieu to another cannot be determined through ear perception alone. Hence, Laver (1994:501) states "Intensity (or power) of a sound is usually measured in terms of a scale whose units are called **decibels**". Decibel is abbreviated as **db**, and according to Laver, 0 db corresponds to "the sound-pressure level (SPL) of a reference sound, namely one close to the absolute limit of detectability by the average listener of a sound whose frequency is 1 kHz" (p. 502).

Above scientific investigation of loudness of speech sound places the most intense sound one can hear without physical damage to the auditory apparatus at 120 db SPL. Then he warns: "Exposure to 120 db SPL can be tolerated for a very short time before grave risk of permanent damage". In all, acoustic or scientific investigation of the intensity of sound will provide the logarithmic report of differences in the loudness of individuals' speech sounds. It will also help to provide a limit to the intensity of sound that is tolerable to the ear.

Articulations

In order to better account for articulation of speech sounds, especially the complex articulations and avoid guess work, instrumental techniques are adopted. Hence, Ashby and Maidment (2005) say: "Some more objective techniques were available... [and] it has been possible to record the movements and positions of a speaker's articulators with a fair degree of accuracy". For instance, instrumental techniques are necessary to clearly account for coarticulations, double articulations and secondary articulations. The Igbo labial-velar implosives [ɓ] and [ɓ̥] which involve a minimum of four articulators in the oral cavity with all of

them acting simultaneously requires instrumental analysis for better interpretation.

Accounting for the acoustic and physiological analysis of coarticulation, Ashby and Maidment (2005:129) have this to say: “There are a number of different ways of investigating coarticulation and showing how the vocal tract changes shape over time. Spectrograms show the changing acoustic patterns, or direct measurements can be made of articulator positions”. By so doing, the spectrogram reflects “the coarticulatory variations in precise point of contact between the articulators”. Example (7) above also apply here as illustration.

In like manner, “electropalatograms” show the physiology of speech production; such as presenting the hold, closure and points of articulation. “X-ray studies” are another way to account for the physiology of speech production. It also shows the various hold phases of different sounds, as well as the position of the tongue and locations in the course of articulation of speech sounds. However, the X-ray study, though considered very effective is rarely used presently due to its radiation side effects.

Conclusion

This paper has strongly argued that acoustic analysis is a far better approach to the analysis of speech, both because of its proven authenticity and for modernity. In present times, human societies are getting more sophisticated, and so are the mediums of communication and explanation of the activities of the modern society. The analysis and interpretation of complex pitch patters of languages, as well as complex idiosyncratic speech styles of the modern world are better accounted for using the instrumental method of speech analysis which is the acoustic analysis. For instance, pitch variation is not so simple and clear as to be well

analysed using mere ear perception. Same goes for speech melody, downdrift loudness and articulations, especially complex articulations and coarticulation. This will authenticate the auditory perception. It will also instigate more interest in the linguistic study of our languages, and enhance better communication and comprehension.

Granted though that the acoustic approach has its challenges and shortcomings, its advantages and prospects far outweigh the performance of the ear perception or any other known approach for speech study and analysis, hence its strong advocacy by this paper.

References

- Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- Anyakoha, M. N. (2010). *New School Physics for Secondary Schools*. Onitsha: Africana First Publishers.
- Ashby, M. & Maidment J. (2005). *Introducing Phonetic Science*. Cambridge: Cambridge University Press.
- Bussmann, H. (1996). *Routledge Dictionary of Language and Linguistics*. London: Routledge.
- Clark, J.; Yallop, C. & Fletcher, J. (2007). *An Introduction to Phonetics and Phonology*. Maldene: Blackwell Publishing.
- Crystal, D. (2003). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Emenanjo, E. N. (1978). *Elements of Modern Igbo Grammar*. Ibadan: University Press.
- Finch, G. (2000). *Linguistic Terms and Concepts*: New York: Palgrave Macmillan.
- Harley, T. A. (1995). *The Psychology of Language: From Data to Theory*. Hove: Psychology Press.

- Ladefoged, P. (1962). *Elements of Acoustic Phonetics*. Chicago: University of Chicago Press.
- _____ (1964). *A Phonetic Study of West African Languages*. Cambridge: Cambridge University Press.
- _____ (2003). *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Maldene: Blackwell.
- _____ (2006). *A Course in Phonetics (Fifth Edition)*. Boston: Wadsworth Cengage Learning.
- Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Onwudiwe, G. E. (2015). "Tone Intonation Interface: An Acoustic Analysis of Igbo Speakers of English". A PhD Dissertation of Nnamdi Azikiwe University, Awka.
- _____ (2019). "Mma Nsokwasi nke: Nkamma nà ụkpuru nghoṭa okwu" in *Odezuruigbo: An International Journal of Igbo, African and Asian Studies, Vol. 3, No. 1*.
- Pickett, J. M. (1980). *The sounds of speech communication: A Primer of Acoustic Phonetics and Speech Perception*. Baltimore: University Park Press
- Stewart, J. M. (1967). "Tongue-root in Akan Vowel Harmony"; *Phonetica* 16:185 - 204.
- Westerman, D. & Ward, I. C. (1990). *Practical Phonetics for Students of African Languages*. London: Kegan Paul International.

George E. Onwudiwe
Department of Igbo, African and Asian Studies
Nnamdi Azikiwe University, Awka
ge.onwudiwe@unizik.edu.ng