CHAPTER TWO

BACKGROUND AND REVIEW OF RELATED ACOUSTIC STUDIES

This chapter discusses the background of the study with reference to published literature. It also discusses some acoustic studies in other languages, which may be relevant to the present study.

2.1 Theoretical Background of the Study

Phonetics as a branch of linguistics deals with how human speech sounds are produced, transmitted and perceived. It "is …the indispensable foundation of all study of language, whether that study is purely theoretical or practical as well…" (Sweet 1877 cited by Catford 1988). The truth of this statement is apparent because linguists have tried, over the years, to provide phonetic evidence for all the claims they made about speech sounds. In recent years, linguists and students of language have made a tremendous headway in the scientific study of speech sounds; among these is the acoustic analysis of speech sounds and especially vowels through spectrographic analysis. By the use of speech analysis equipments, which are modeled to correspond to the human articulatory system, sounds produced are changed into visual representations of the sounds, which are analyzed, and authentic conclusions drawn about them. These visual representations are normally in the form of horizontal bars with the relative intensity of each component frequency shown by a darkened mark which is conventionally referred to as formants. Formants are the characteristic overtones of the vowels (Ladefoged 1982). They are narrow bands in the acoustic spectrum in which energy is concentrated during the production of speech sounds.

The vocal cords, during speech production, release consistent amount of air that generate sound waves. These sound waves are transmitted to our ears through the air in the form of vibrations. The sounds we hear are combinations of the simplest kind of vibration known as the sinusoidal or sine waves. This is to say the greater parts of the sounds that come to our ear are the result of complex sound waves. The lowest frequency sine wave component is the fundamental frequency (F0) and the next two are referred to as the first and second formants.

The frequency of each formant is related to the volumes of the cavities in front (oral cavity) and behind the constrictions (pharyngeal cavity) of the vocal tract. They are dark bands on a wideband spectrogram, which corresponds to a vocal tract resonance. When we talk, we make different speech sounds by configuring the vocal tract differently. So we can infer the shape of the vocal tract (articulation) from the pattern of the formants and resonance in the spectrogram. Frequency is the number of vibrations or cycles the sound makes per second and it is measured in Hertz (Hz) (Akpanglo-Nartey 2002:57). The first formant (F1) is the lowest formant but has the most intense sound pressure. Generally, it corresponds to vowel height. It is related to the volume of the pharyngeal cavity and to how tightly the vocal tract is constricted. The second formant (F2) is related to the length of the oral cavity and corresponds to vowel backness. The farther back the vowel, the lower the F2 (Kent & Read 1992:92) cited by Shank & Wilson (2000). In the vowels, F1 can vary from 200 Hz to 1000 Hz and F2 from 850 Hz to 2500 Hz (Ladefoged 1982). Phoneticians usually recognize at least two formant regions as uniquely characterizing the different vowels. A larger number are however present but have less strength.

Linguists, since the introduction of the sound spectrograph, have carried out experiments to see what changes can be made to the system so as to bring good and authentic results. For example, Lindau (1978) and Ladefoged (1982, p.177-80) by the use of statistical analyses replaced the F2 dimension by $F2^1$, that is F2-F1. The reason, as they claim was that the $F2^1$ is more directly related to the auditory concept of ‘frontness’ or ‘backness’ than the F2 alone (Clark and Yallop 1990:241). Ladefoged and Maddieson (1990) have also reviewed articulatory features that may differentiate vowels phonetically in the world’s languages. Their view agrees with that of Lindau. They propose three contrastive degrees of “tongue advancement (front, central and back) and five degrees of height”.

The attempt to provide a scientific evidence for classifying speech sounds have led to the development of powerful acoustic models of the vocal tract. There have been attempts to compare the predictions of the models with the articulatory acoustic relationships familiar from modeling studies of the vocal tract to see if they are actually observable in speakers’ utterances, an example being the “human articulatory synthesizer” approach followed by Ladeofged & Bladon (1982). Experiments with these synthesizers show that articulatory – acoustic relations qualitatively conform to expectations from modeling studies, and suggest that speakers may indeed exploit articulatory strategies that have been claimed to be acoustically advantageous. Coker, Umeda and Browman (1973) cited by Ladefoged (1978) also discussed the use of computer programs that synthesize speech. They showed that it is possible to use articulation specifications to produce intelligible English. They made input of a string of phonetic segments into the speech synthesizer and were able to program them into articulatory parameters, which were able to produce English sounds.

It has been noted that the auditory differences in individual vowel qualities that help us to give them the appropriate phonological descriptions are “determined by the frequency distribution of the first three formants” (Clark and Yallop 1990:241and Barry Truax 1999). Apart from vowel identity, the acoustic features of vowels also depend on the physiological make up of the speaker’s vocal tracts.”…Some speakers with big heads will have large resonating cavities, producing formants with comparatively low frequencies; and others will have higher formant frequencies because they have smaller vocal tracts (Ladefoged, 2001:39). This is very true considering the vocal tracts of men and women. For a male voice the frequency of the vocal cord vibration may be between 80 and

200Hz whiles that of a woman may go up to about 400Hz (Ladefoged 1982:169). Also relating to this is the fact that speakers' habitual articulatory settings are different. For instance a group of speakers of a particular dialect of a language may produce a speech sound in palatal setting while another dialect produce it in a velaric setting. They may sometimes use a breathy voice or protruded tongue or always use a relatively slow rate of articulation. F1, F2 and F3 are all, for example, also noted to undergo changes in pharyngealized environments.

The influence of consonant context on the formant frequencies of the vowels is also evident; the tongue body in the position of the consonants can affect the position of the vowel (Papçun, 1976:8 and Stevens, 1998: 260-261). To this effect, Shank and Wilson (2000) investigated with the aim of documenting and precisely describing the acoustic effects of the glottal stop /ʔ/ and the pharyngealized glottal stop /ʕ/ on adjacent vowels in Nuu-chah-nulth, with the assumption that such effects are a primary cue to their differentiation. Recorded data was digitized on an iMac Computer Sampling at 44kHz and analyzed using Praat 3.8.64 software program. Formant averages for each of the vowels /a/, /i/ and /u/ were measured in nine different environments. They concluded from the results that the formant value of the vowels adjacent to the plain glottal stop /ʔ/ is different as compared to the "pharyngealized glottal stop" /ʕ/ in Nuu-Chah-nulth. The latter causes a greater rise in F1 and a drop in F3 than the glottal stop. The linguistic indication of this is that consonantal context may somehow affect the quality of a vowel.

Other factors such as dialectal, social, or generational differences within a language community may also substantially alter the positions of particular vowels in the F2/F1 plane. Such factors may even act like differences between languages, in that certain vowels may appear in one dialect or social group but absent from another. Peterson and Berney (1952) cited by Rosner and Pickering (1994: 49), conducted a study which had, men, women and children produce American English monophthong, /i, I, ɛ, æ, ɑ, ɔ,ʌ, u, o,∪/ twice in an /hvd/ environment. Formant frequencies and bandwidths were measured on spectrograms and plotted on the F2/F1 acoustic plane. Some of the points were scattered all over the plane indicating that, the vibration or frequency of each speaker is different. Also relating to that is a study conducted by Brückl and Sendlmeier (2003). They carried out an acoustic analysis and perception test on the changes that occur in female voices due to ageing. The analyzed data was recordings of 56 female speakers of different ages. The recorded speech samples included sustained vowels, read speech and spontaneous speech. They noted, among many other findings that increasing amplitude perturbation is an indicator of increasing age even on the basis of spontaneous speech and that there is significant change in articulation rate of spontaneous speech. Based on sustained vowels, the frequency tremor intensity index indicates age more accurately than F0 and amplitude perturbation.

However these differences as always observed by linguists are not significant in altering vowels quality. As Yallop and Clark (1991, pp.117) stated, “such differences usually do not affect the articulation of individual speech sounds in a particular or selective way, but are global properties that contribute to a total impression of voice quality”. Phonetic quality is that set of properties that help to differentiate one sound from another adequately. It is those properties that distinguish one speech sound from those properties that result from personal differences of individual speakers. As Papçun (1980:1) put it, “phonetic quality may be considered as the residue which remains of speech sounds when individual speaker differences are eliminated”.

2.2 Review of Related Acoustic Studies

Experimental phonetic work on the vowels of Ewe is not common. However, there have been many studies conducted on acoustic properties of vowels, which may be relevant to this study. Eric Zee (1978) conducted an investigation into the effects of tone on vowel quality in Taiwanese Chinese. He explored how formant frequencies change when a vowel is produced with different tones in natural speech. Five Taiwanese vowels, [i, e, a, ɔ, u] as produced by five native speakers were investigated. Formant frequency values of F1, F2, and F3 associated with the low and high tone of each vowel collected every 10msec from LPC data with a window size of 25.6msec were measured. Various statistical methods were used for normalization and the results plotted in an acoustic space. The formant frequencies associated with the high or low tones for all the vowels investigated occupied distinct acoustic vowel spaces for all the speakers in different directions. The implication of this is that vowels are affected by tonal difference, for different vowels and for different speakers.

Disner (1978) presented an acoustics study on six Germanic languages. Her aim was to consider improved ways of describing vowel quality precisely and systematically and attempt an illustration as to how the findings can be use in a ‘natural’ explanation of phonological processes within those languages and also “to provide a descriptively adequate model of the sound patterns” in those languages to aid the linguists in drawing conclusions about them. It was clear in Disner’s study that if we were to find that vowels in all languages had similar phonetic quality, any language would be equally able to provide useful insight into the nature of phonetic features and how they participate in the determination of vowel quality. However it would not be clear whether languages belonging to a single family could be considered one in this respect. A particular sound may differ between languages along the lines of any of the various parameters that determine vowel quality.

In a related study, Godinez (1978) investigated using spectrographic analysis the underlying phonetic parameters that characterize the vowels of the Romance languages to see the features that identify the vowels of each of the individual languages and to make objective statements concerning their similarities and differences. Vowels of 25 male speakers of some Romance languages consisting of Mexican Spanish (6 speakers), Argentine Spanish (4 speakers), Peninsular dialects of Spanish (6 speakers) and 9 speakers of Brazilian Portuguese were measured. They were put into groups and some made to read words, containing the vowels in one frame, and others in isolation. The recorded data was submitted to a spectrographic analysis by means of a computerized LPC (Linear Predictor Coefficient) to determine formant frequency values. F1, F2 and F3 were measured and presented in tables. The values were plotted on formant charts with F1 on abscissa and F2 – F1 on the ordinate. The target areas of the vowels [o], [u] and [a] of Peninsular Spanish were seen to be closer than the others. Their back vowels have larger acoustic space than those of Peninsular Spanish. The sound [e] in Argentine is higher than that of Mexican Spanish and [a] was noted to be a back vowel in Spanish rather than central. Vowel [e] and [ɛ] were discovered to have a considerable overlap in Brazilian Portuguese. There were other differences, which show considerable variants in the phonetic features of the vowels of the Romance languages.

In a related study, Moosmüller and Granser (2003) investigated three dialects of Albanian Language to see their relationship with the standard Albanian. 9 male subjects (3 each from North Albanian, Middle Albanian and South Albanian) were asked to read a list of words.The stressed vowels in the list of words were analyzed. The recorded material was digitized at 22,05kHz, 16 Bit using the acoustic workstation STX. All prosodically strong vowels (1170 in total) have been labeled and frame-by-frame, formant frequency contours were produced. The first three formants were measured using the LPC with 26 coefficient, a pre-emphasis of 0.9, a frame width of 46ms and a 2ms frame shift. For speaker normalization, a readjustment of the labanov normalization procedure as proposed by Disner [(1980), was chosen. One-way ANOVAs were calculated for each vowel. The results indicated that all vowels show statistically significant regional differences. The two central vowels of Albanian, [a] and [ə] additionally revealed a high range of variation from front to back articulation for /ə/ and from central to back articulation for /a/.

Another study by Huffman (1997) aims at extending the knowledge of the acoustic difference between vowels in American English by providing an acoustic analysis of the Vowels of Long Island. It was to examine the relative importance of time-varying formant change in characterizing vowel differences in Long Island English. Monolingual English speakers from Long Island read words with the same consonants but different vowels such as “heed”, “hid”, “head” and “had” and by using computerized speech analysis software, formant values and formant measurements of 20%, 50% and 80% time point of the vowel duration was taken for each vowel. A statistical technique called discriminant analysis was used to evaluate how the vowel formant values of different words compared at the different time points. Formant values measured at the time point of 50% or

midpoint did not appear to show any difference in the vowels. But those measured at different time points, such as 20% or 80% showed clearly the particular vowel or word being used. She concluded that, two vowels produced at the same part of the mouth might be different in that one is a little higher or lower or one is a little front or back than the other.
A paper of the result of an acoustic study of vowel sounds in three regional variants of Danish in spontaneous speech has also been presented by Ejstrup and Hansen (2004). The monophthongs in spontaneous speech of 18 subjects (9 men and 9 women) of three Danish regional variants were measured to see whether there are any systematic acoustic differences in the realization of the same allophones in these three dialects. By the use of Praat 4.0.13 Software, formant frequencies were measured with Praat's Formant (Burg) facility. The results indicated that there was the tendency for the speakers of Herning, one of the investigated dialects to diphtongize the /o:/ as either [ou] or [ɔu] depending on the environment. /r/ colouring was noted to have been one of the clearest regionally determined phonetic differences with these regional dialects. It was generally concluded that there is the tendency of the vowels of Danish to cramp together in the upper part of the acoustic vowel space.

Qin Yan, & Vaseghi (2002) also make a comparable study of British and American accents, with the aim of quantifying their effects on speech recognition. American English was found to achieve 31% less error than British English in matched accent conditions from the results. Durations at the start and the end of sentences in British Accent were found to be shorter than that of America. This may be due to the fact that British speakers always tend to pronounce last syllable (especially for consonants) than Americans. It was also found that British speakers possess much steeper pitch rise and fall pattern and lower average pitch in most vowels. The implication of these findings is that speakers of different dialects will, by all means, exhibit different articulatory properties.

Al-Masri and Jongman (2004) conducted an acoustic study to find out what the acoustic correlates of emphasis are in Jodanian Arabic. The study attempted to gain a better understanding of the acoustic correlates of emphasis in that language. A list of four sets of minimal word pairs differing only in the terms of emphatic vs. plain distinction was prepared and the words embedded in the carrier sentence 'say…Once again'. 5 male and 3 female native speakers of the northern dialect of Jordanian Arabic were recorded. The recordings were digitized using Praat speech analysis software. F2 was measured for every vowel in every stimulus word. Duration of plain emphatic consonants was also measured and the data averaged across all repetitions. It was concluded in this study that acoustic correlate of emphasis is a lowering of F2 for the vowels in the same syllable as the emphatic consonant. A gender effect was observed, showing that emphasis is more pronounced in females than in males in this dialect. They however concluded that there is the need to investigate more dialects in order to really understand the phenomenon of acoustic correlates of emphasis in that language.

Taehong and Ladefoged et al (2000) compared the vowels of Cheju with those of standard Korean. The study investigated the acoustic qualities of monophthong vowels of Cheju as spoken by rural and urban speakers of Cheju mostly in their 60-70s. All the speakers understood Korean well and were all literates. The Cheju vowels /i, e, æ, o, u/ were placed in three sets of words and recorded. The first set has the vowels after a velar stop; the second after a dental-alveolar stop and the third set contains additional words for comparison of e/æ. F1 and F2 were measured using the Kay Elemetrics Multispeech system. A steady state portion was found for each token near the mid-point of the vowel, and superimposed LPC and FFT spectra were calculated with 30ms and 2.5.6ms frames, respectively. The formant values were determined from LPC Spectra, using the FFT Spectra as a supplementary check. One obvious thing was that quite a few Cheju speakers make a distinction between /e/ and /æ/, and o and ʌ.There was a strong evidence that the rural speakers of Cheju have an e/æ distinction in their clear speech, whereas the urban speakers, even in their 60s do not have. Also the merging of [e/æ] and [o/ ʌ] are under way in Cheju among both rural and urban speakers.

In another related study, Watt and Tillotson (2001) conducted a study to look for acoustic evidence with a view to matching acoustic cues to the auditory impression of /o/ fronting in Bradford English. The study also investigated contrasts in /o/ fronting to see whether they are features of phonological systems of older British English speakers and if they could be used for a phonemic distinction. Seven Bradford English speakers (5 female, 2 male), all from the City of Bradford, ranging in age from 17 to 75 years were recorded. Each speaker read aloud in as natural a way as possible a list of some 100 isolated words, plus 8 short phrases containing target phonological variables. Some of the speakers were made to read a second time in rows from top to bottom rather than down each column from left to right. The recordings were sampled at a rate of 11025Hz into sentimentrics speech station 2, a spectrographic analysis software package, running within a Windows NT environment. The target words were then isolated and labelled. Formant values were extracted from LPC (Linear Prediction Coding) spectral envelopes generated from the approximate midpoints of target vowels on broadband spectrograms. The formant data was then presented in the form of cross-plots of F1 values against the difference between F2 and F1 in an acoustic space. The results indicated that the acoustic signal of the fronting of the target of /o/ exist in Bradford English. The fronting which involves a shift away from the periphery of the vowel space toward a more central region is most advanced among the younger speakers. There was also some evidence of sporadic fronting among the older female speakers and it is young women who tend to introduce innovations of /o/ fronting in Bradford English.

Morrison (1994) with the aim of investigating the production of /e/ and /o/ in contexts in which Navarro (1918 and 1965) claimed that these mid-vowels would have open and close allophones also recorded two highly educated professionals, one male and one female from Madrid. Pairs of words that exemplified the contexts in which Navarro (1965) claimed open and close allophones would occur were chosen. The words were presented to the participants in written form in the carrier sentence “well she/he said _______ to him/herself”. There were a total of 24 sentences presented 10 times in random order. Acoustic analysis was performed by the use of Praat 3.9. The beginning and the end of each vowel of interest was manually marked and labeled. F1 and F2 of the vowels were then measured at the center using the Burg LPC auto-correlated-tracking algorithm with a step size of 10ms, and window size of 35ms. Maximum number of formant were four for the female speaker’s /e/, five for her /o/ and the /e/ of the male speaker and six formants for his /o/. Two statistical procedures were applied to the first two-formant values of the data. Cluster analysis was used to determine whether the vowels clustered in groups that correspond to their contexts and discriminant analysis was used to determine the degree of differentiation between groups of contexts identified by the cluster analysis. The results of the F1 and F2 values were standardized to z scores (mean 0 and SD of 1). The values obtained were plotted in F1 – F2 acoustic space. The statistical analysis of the first two formants of the mid vowels indicated that, Spanish does not have open and close allophones of the mid-vowels. If the Spanish mid-vowels should be divided into allophones at all, front versus retracted /o/ allophones and close-fronted versus central versus open-retracted allophones are better labels.

Ladefoged (1982) presented a spectrographic analysis of eight American vowels. In that study he noted a considerable amount of differences which one can use as a basis for conducting an acoustic study of other languages. Michael A. Stokes (1998) published a paper that discusses vowel perception that stem from interaction between individual formants. He explained that those relationships categorize and distinguish each vowel and associate a formant frequency with a specific articulatory gesture. He further noted that spectrograms are a useful tool for acoustic-phonetic analysis of formants because they separate individual formants making detailed analysis of vowel components possible.