CHAPTER TWO

THE ACOUSTICS OF VOWELS

2.0 Introduction

This chapter seeks to discuss the nature of acoustic specification of speech sounds especially the vowels. This objective is achieved by reviewing the literature on acoustic investigations of vowel theories.

2.1 Acoustic features of Vowels

The physical differences characterizing different speech sounds may be measured in terms of frequency, amplitude, duration, and spectral distribution. To these correspond the perceptual correlates of pitch, loudness, length, and tamber or quality.

The complex sound waves produced by human phonation may be analyzed into varying combinations of sine waves characterized by different frequencies and amplitudes. The speaker's vocal tract acts as an acoustic filter through which the glottal tone of voicing is passed. The glottal tone is composed of a fundamental frequency and its harmonics, with power falling off as frequency increases but present in substantial amounts up to about 3000 cps (cycles per second). As the shape of the vocal tract is altered, so its sound–filtering characteristics are modified. Some frequencies present in the glottal tone are stopped; others are transmitted to a greater or less degree.

While talking, a speaker never holds the shape of his vocal tract constant for more than a moment. The shape of the supraglottal cavities, and the resonances associated with them, vary almost continuously, as the frequencies stopped or passed change.

A peak of energy at a particular frequency in the output sound corresponds to a passband around that frequency in the filter system of the vocal tract, that is, a peak in the filter's frequency response; lack of energy at a particular frequency corresponds to a trough in the frequency response (Wells, 1962).

Speech sounds that are classified as vowels are found to be characterized by peaks of energy around the frequencies corresponding to the natural frequencies of the supraglottal cavities as they are during the articulation of the sounds concerned; these frequency bands are known as formants. Different vowels are characterized by different frequencies, that is, by peaks of energy at different frequencies. To the different shapes of the vocal tract and the different positions of the tongue there correspond different frequency responses and passbands, giving rise to vowel sounds of different formants.

2.2 Acoustic Characteristics of Vowel Formants

Vowels are frequently described with reference to their formant structure, which provides an indication of vocal tract resonance and therefore articulatory shape (Fant, 1960). The relationships between the first formant (F1) and the auditory quality of height, and the second formant (F2) and the auditory impression of the front/back dimension, ensures that when the first two formants of a set of vowel targets determined from a position near the midpoint of the vowel are plotted on axes, the result closely resembles the traditional auditory vowel map. Such vowel spaces, with axes F1 and F2, rely on the concept of the vowel target. The target is the vowel component least influenced by its surrounding phonetic context and is considered to be either a point in the time course of the vowel or a section of time during which the vowel position remains stable. A single point is often used to provide a representation of the target position, and for most vowels this can be assumed to be approximately mid way though the nucleus. The target theory is not without problems. Several authors have noted the problems inherent in the target theory for vowels, citing the difficulties often encountered in establishing steady state components by eye or by automatic extraction procedures (Benguerel and McFadden, 1989; Nearey and Assmann, 1986). However, Van Son and Pols (1990), are reported to have examined five different methods of identifying vowel targets and found that the use of the different methods made little difference to the results of their experiments (Cox, 1996).

The target theory is based on the traditional view that each vowel contains a relatively steady articulatory and hence acoustic component providing cue to identification. Numerous automatic classification studies based on pattern recognition techniques have demonstrated accurate separation of vowels from static spectral characteristics and therefore provide support for the theory that, at least for monophthongs, vowel target appears to be a robust cue to vowel identification (Harrington and Cassidy, 1994; Hillenbrand and Gayvert, 1993; Syrdal and Gopal, 1986; Zahorian and Jagharghi, 1993) cited by Cox (1996).

In the quest for a more reliable ways of describing speech sounds, linguists through experiments have come out with several ways of describing vowel quality. An example is the enhancement of the F1/F2 to F1/F2-F1 (Ladefoged 1982, Lindau 1975). By the use of various statistical analyses, these authors found that the F2' (F2-F1) is more closely related to the auditory concept of 'frontness' or 'backness' than the F2 dimension of the acoustic plain. The current work uses the F1/F2-F1 for the formant analysis.

The Dagbani vowel system comprises the vocalic sub-system [i, ɪ, ə, e, ɛ, a, ɔ, o, ɷ, u] referred to as monophthongs. Monophthongs can be described with reference to a single target in the time-course of the vowel.

Other vocalic sub-systems referred to as diphthongs are described with reference to two points in the time-course of the vowel to indicate extremities of gliding component. The presence of diphthongs is not confirmed in Dagbani.

Vowels may also be distinguished in terms of duration for languages that employ phonemic vowel length. The vowels of Dagbani can be classified as long and short vowels. Among the set of vowels, the pure vowels [i, e, a, o, u] have their longer counterparts [i:, e:, a:, o:, u:] doubling in words as ii, ee, aa, oo, and uu, indicating vowel length that distinguishes meaning in Dagbani. Those vowels that cannot be lengthened include [□, ə] orthographically represented as **i**; the others are [□, ɷ]. The major difference between the long and the short vowels is simply one of total vowel duration, however, the difference is relative rather than absolute as contextual and prosodic factors affect the ultimate length of the vowel. Other factors that affect segmental duration such as speaking rate, semantic emphasis, phonological/phonetic influences such as inherent segmental duration, the effect of linguistic stress, and the effect of a postvocalic consonant are important determiners of the durational characteristics of vowels.

Cox (1996) cites Chiba and Kajiyama (1941) who proposed that vocal tract length differences represented the major source of variation in vocal tract transfer between the groups (men, women and children). This hypothesis is tenable given that the standing wave characteristics of the resonant frequencies of the vocal tract tube are dictated by length to a greater degree than cross sectional dimension. Fant (1966), found that a simple scaling factor inversely proportional to vocal tract length did not adequately account for the observed differences in formant frequencies between the sexes. The magnitude of the difference depended on the vowel itself. He found that rounded back vowels required a minimal amount of correction for F1 and F2, very open unrounded vowels required maximal correction for F1, and high front vowels required minimal correction for F1. Fant provided two explanations for these findings. Firstly, that women and children have relatively shorter pharyngeal cavities and smaller laryngeal cavities than men, and that this factor would have a differential influence on formant frequencies. Secondly, that not all formants have standing wave characteristics where total length is the primary determinant of resonance characteristics. For instance, in double helmholtz resonators, such as is characteristic of back rounded vowels, overall length does not have the same magnitude of effect on F1 and F2. Fant (1966) concluded that female and male formant patterns were not related in a simple linear fashion. Fant (1975) attempted to assess the universal nature of male to female vowel formant differences in order to establish an anatomical basis for departure from a simple uniform scaling. He examined eight different languages and concluded that uniform normalization with a single scale factor could substantially reduce the sex differences in the data but non-uniform techniques were required to account for specific vowel category and formant number trends. He concludes that "the female-male formant-frequency relations are in part determined by anatomical constraints" (Fant, 1975:17), but also that "we cannot quite exclude the possibility of universal "feministic" preference in vowel qualities which might have influenced the average data" (p18).

Ladefoged (1993) is of the view that most speakers with big heads will have large resonating cavities, producing formants with comparatively low frequencies, giving rise to higher vowels; and others will have higher formant frequencies resulting in lower vowels because they have smaller vocal tracts.

It is suggestive that, assuming an optimal normalization, all vowels transcribed with the same phonetic symbol in the same language will occupy the same point in vowel space. The position of the point of a vowel corresponds to the quality of that vowel in that particular language. It is not to be assumed in consequence, however, that vowels of different languages which appear to have the same phonetic quality and which are sometimes transcribed with the same phonetic symbol are necessarily the same. Two languages might well choose different combinations of phonetic features in defining the position of, say the vowel [a] in phonetic space. Thus, the vowel transcribed as [a] in one language may have a phonetic quality that is perhaps higher or more fronted than vowel [a] in another language. For instance McClean (1969: 5); cited by Disner (1978) states that the Swedish [i] is ‘closer than the vowel in English *seen*’.