NORTHWESTERN UNIVERSITY

PHONETIC CORRELATES OF LEXICAL STRESS IN URDU

A DISSERTAION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Communication Sciences and Disorders-
Speech and Language Pathology

By

Sarmad Hussain

EVANSTON, ILLINOIS

December 1997

UMI Number: 9814230

ABSTRACT

Phonetic Correlates of Lexical Stress in Urdu

Sarmad Hussain

This work investigated how phonetic properties of Urdu segments change with stress and whether these phonetic changes support the theories that predict the effects of stress. The stimuli consisted of pairs of bi- or tri-syllabic words which contained the target segments in the initial syllables. The first syllable in each pair was stressed in one word and unstressed in the other. However, the segmental context for the target segment was held as similar as possible, within the constraints of having familiar real words. Ten block-randomized repetitions of each word, embedded in a carrier phrase, were recorded by seven native speakers of Urdu. Acoustic analysis of the recordings was done using *xwaves*. Duration, fundamental frequency (F0), relative intensity, and first two formant frequencies were measured for the six long and three short vowels of Urdu. In addition, closure duration, voicing during closure and post-release aspiration were measured for the sixteen stops in Urdu, in both onset and coda positions. The results indicated a longer duration and lower F0 (due to the alignment of a low tone) for stressed vowels. Also, high vowels got less intense and low vowels got more intense with stress. However, individual speaker data on intensity showed a lot of variation. Also, the quality of the vowels changed with stress as unstressed vowels underwent more contextual assimilation than stressed vowels. Results from stops show that the closure, voicing during closure and aspiration of aspirated (and not voiceless and voiced) onset stops increased with stress. The closure of voiceless, voiced and breathy coda stops and voicing during closure of voiced coda stops also increased with stress. The duration of closure of aspirated coda stops decreased with stress. Though many of these stress-related changes in Urdu support the Sonority Expansion or Hyperarticulation theories, there are still some data which these theories cannot explain. It is proposed that these theories should be extended to account for the variation caused by articulatory (or perceptual) constraints.

ACKNOWLEDGMENTS

Looking for an answer to a question and quite naïve to the phonological theory of sound, I walked into Janet Pierrehumbert's office for the first time about three and a half years ago. On her advice that day, I registered for my first course in phonology. Her phonology classes sparked my interest in experimental phonology and many months of her personal guidance eventually made this work possible. Therefore, I have a lot to thank her for. I also owe a lot to Bruce Smith, among other things, for his insistence to keep seeing things from a phonetic perspective, for being very accessible for discussion, and for his painstakingly careful editing of my work. I also want to thank Yi Xu for his availability for discussion and useful comments which helped me think much more clearly about some of the issues related to this work.

I would also like to thank Tom Carrell, who brought me to Northwestern, Steve Zecker for his very entertaining statistics classes and for his help with the statistical analysis for this work, John Coleman for having me at Oxford for research in Urdu stress and speech synthesis and Mike Broe for his help in certain issues in phonology.

I would also like to thank Asad Khan, Zulfikar Ali, Basir Sami, Amina Ali, Ali Rabbani, Seema Asghar, Aisha Sarmad, Beenish Khan, Hasan Shirazi and Faisal Syed, all of whom gladly volunteered to do the recordings for my experiments. I would also like to thank Stefan Frisch and Rami Nair for helpful discussions in the lab.

Finally, on a personal level, I would like to thank my wife, family and friends who have kept me sane (or insane, as they would like to think) all these years.

This research was supported by a Northwestern Graduate School fellowship. Part of this research was also supported by NSF grant no. 0830 300 TL05.

To my family

CONTENTS

FIGURES

TABLES

1. INTRODUCTION

Stress has been used by different researchers to mean different things. In this work, it refers to phonological prominence of a syllable. The difference in stress is physically manifested as syntagmatic contrasts in various phonetic properties of segments in an utterance. *Lexical stress* or *word stress* refers to this phonological prominence of syllable(s) compared to other syllable(s) within a word. This difference in prominence serves some purpose in a stress language. For example, in Greek, syllable prominence can change the meaning of a word. When the first syllable of the word *poli* is stressed, it means 'city' and when second syllable is stressed it means 'much' (Laver 1994, 522). The differences in stress are phonetically manifested by altering length, pitch, intensity and quality of different segments in a word. However, the way these phonetic properties change with stress varies across context and language.

Earlier research in English (Fry 1955, 1958, 1965, Lindblom 1963, Gay 1978) showed that stress increased the duration, pitch and/or intensity of vowels and made the quality of vowels more extreme. The perceptual work by Fry indicates that an increase in fundamental frequency (F0) is the most significant cue for stress in English. Duration also significantly increases with stress. However, both vowel quality and intensity only contribute minimally to changes in stress.

More recent research in English and other languages has shown that stress can alter these phonetic properties in other ways as well. Morton and Jassem (1965) report that either lowering or raising F0 may cue for stress in English. Nakatani and Aston (1978) find that in certain cases longer duration and not F0 changes is the primary cue for stress in English. Pierrehumbert (1980, 102-105) explains that lowering or raising F0 may cue for stress when a low or a high tone (pitch value) is articulated with a stressed syllable respectively. F0 may not cue for stress when no tone is articulated with the syllable in question. There are also differences across languages. For example, stress is normally articulated as increased vocalic duration in English (Klatt 1976, Harris 1978, Fourakis 1991, Wightman et al. 1992) and Dutch (Sluijter and van Heuven 1996). However, Lehiste (1968, 1970) reports that in Estonian unstressed vowels are longer than stressed vowels. Results from Finnish (Carlson 1980, pg. 6-7) show that short vowels are shorter in duration in stressed syllables than in unstressed syllables, and long vowels are longer in duration in stressed syllables than in unstressed syllables.

The changes caused by stress are not limited to vowels. Phonetic properties of consonants also change with stress. For example, Carlson (1980) reports that consonants in certain contexts in Finnish increase in duration with stress. Longer durations for stressed consonants have also been reported for English (Keating 1984). In addition to duration, other consonantal properties may also be effected by stress (de Jong 1995, de Jong, Beckman and Edwards 1993, Keating 1984, Pierrehumbert 1994, Pierrehumbert and Talkin 1992). These examples show that stress effects both vocalic and consonantal segments. In addition, the way these phonetic properties are changed with stress varies across languages (e.g. Keating 1984). Hence, phonetic changes caused by stress in one language cannot be determined by extrapolating results from another language. Each language should be studied individually to determine how stress changes the phonetic properties of different segments.

Little is known about how lexical stress may change the phonetic properties of segments in Urdu. A phonetic study of Urdu is, therefore, necessary to determine the phonetic effects of lexical stress. This is one motivation for the current work. Quantifying the phonetic variation introduced by stress can greatly help in understanding the phonetics of Urdu, which is not a very well studied language. As pointed out by de Jong (1995, 502), studying phonetic variation in a language with stress can help better understand the phonetics of language itself. "Prominence [or stress] manipulation can be used to ascertain what aspects of the speech signal are particularly indicative of the linguistic contrast that people are articulating. Features of the speech signal which are part of the linguistic code should be more readily apparent in stressed syllables than in unstressed syllables. Features specifically due to motor economy and organization would be more readily apparent in less stressed positions. *Thus stress can act as a diagnostic for determining the content of the linguistic code of a particular language*" (italics added for emphasis). Quantification of the phonetic effects is also necessary for development of an Urdu text-to-speech system, which is another motivation for this work.

Finally, studying the effects of lexical stress in Urdu will also help in determining how phonetic properties of segments are generally changed with stress. There are two main competing hypotheses which explain the phonetic changes caused by stress. The

Sonority Expansion theory (Edwards and Beckman 1988, Silverman and Pierrehumbert 1990, Pierrehumbert 1994) proposes that stress works to increase the sonority (see (2.8) below) of a syllable (nucleus) relative to syllable edges. The *Hyper-articulation* theory (de Jong 1995) proposes that stress increases all the distinctive phonemic gestures of the segments in a syllable. However, both these hypotheses come short of explaining all the phonetic changes caused by stress in different languages. Further work on phonetic changes caused by stress will help develop a better understanding of how these changes are motivated.

REVIEW OF LITERATURE

To investigate the changes in the phonetic properties of segments in Urdu caused by lexical stress, we need to know which syllable(s) in Urdu words have lexical stress, how lexical stress may change the phonetic properties of the segments in these syllables, and which extra-stress factors may also influence these phonetic properties (so that the stress effects can be filtered out from the collective phonetic changes). After defining what lexical stress is, these three issues are reviewed in this section

LEXICAL STRESS

Referring to an earlier example, different placement of stress on the word *poli* in Greek can change its meaning. Similarly, in Assamese if the first syllable for the word [pise] is stressed it means "he is drinking" and if the second syllable is stressed, it means "then" (Laver 1994, 522). Thus, in some languages, placement of stress can actually change the meaning of the word. In other languages, the placement of stress is more predictable. For example in Finnish, the first syllable of a word is stressed (Carlson 1980, also Hyman 1977 for other examples). In both cases, a syllable in a word may be more prominent than other syllables in the word. This "placement of phonological stress on a particular syllable within a word is a defining property of that word, and this can be referred to as *word-stress* or *lexical stress*" (Laver 1994, 511).

Lexical stress can be subjectively determined by asking the native speakers, of the language in question, which syllable of the word is most prominent. Native speakers generally agree on the same syllable for each word. Objectively, this difference can be determined by studying the phonetics and the phonology of a language. Phonetic research has shown that acoustically this difference in stress may be manifested by relative change of duration, F0, intensity and/or quality of individual segments of the word, even though the phonemic content of the words remains unaltered with the change in stress (though the phonemic content of words may also change in some cases by phonological rules triggered by phonological stress, e.g. the phonological rule which causes vowel reduction in English; Chomsky and Halle 1968, 111). These phonetic changes are discussed in detail later in this section.

Phonologically, one can also observe certain consistencies in a language which can only be explained if lexical stress is assumed. For example, lexical stress is necessary to explain how different *tunes* align with words in English. A tune is a sequence of high and low tones. Each tune is part of the English intonation system and is associated with a particular meaning (Liberman 1975, Pierrehumbert 1980). Hayes (1995) explains that a sequence of tones MLH (H, M and L represent high, mid and low tones respectively), when associated with a word, gives a percept of a question being asked (from Pierrehumbert and Hirschberg 1990). So when the speaker co-produces this tune with the word 'assimilation,' the listener hears 'assimilation?'. In this tune the final high tone always aligns with the final edge of the word. The initial tone aligns with the first syllable of the word. However, the middle tone aligns with different syllables in different words. Hayes points out that the syllable with which the second tone aligns is the one marked in dictionaries and agreed by native speakers as the main stressed syllable in the word. Three different tunes and the alignments of their tones with two different words are shown in (2.1) (adapted by Hayes 1995, 11, from Liberman 1975, Pierrehumbert 1980 and Pierrehumbert and Hirschberg 1990). As can be seen, the middle tone for each tune aligns with second syllable of the word 'preliminary' and the fourth syllable of the word 'assimilation.' These are also the syllables which carry the lexical stress for these words. Thus, alignment of a tune can be used to determine the lexically stressed syllable in English.

(2.1)

tune	pre.li.mi.na.ry	as.si.mi.la.tion
declarative	M H L	M H L
question	M L H	M L H
down-stepping	H M L	H M L

There are also some rules in English phonology which are conditioned by lexical stress. In one such rule, listed in (2.2) (Hayes 1995, 13), "word-medial voiceless stops are aspirated provided they are in the onset of a stressed syllable and are not preceded by a strident."

(2.2)

$$\begin{bmatrix}-\text{son}\\ -\text{cont}\\ -\text{voice}\end{bmatrix} \rightarrow [+\text{spread glottis}] \;/\; [-\text{strid}] \_\_\_ ([+\text{son}]) \begin{bmatrix}\text{V}\\ +\text{stress}\end{bmatrix}$$

append	[əpʰɛnd]	vs.	*campus*	[kʰæmpəs]
accost	[əkʰɔst]	vs.	*chicken*	[tʃikən]

These rules can give information about which syllable is stressed in the word. These phonological rules also indicate that lexical stress is an integral part of the phonological system of stress languages.

Recent phonetic and phonological research has shown that stress cannot be grouped together with segmental features because stress displays different properties. Hayes (1995, 30) points out that stress is different than other phonetic features because it shows a rhythmic distribution, has multiple levels of realization (vs. binary realization of phonetic features) and shows lack of assimilation i.e. stress does not spread to adjacent segments (vs. phonetic features like [voice] and [nasal], which assimilate with adjacent segments). Stress also serves a different purpose in speech than the segmental features. Segmental features define sounds (by comparison of an item with other items in the phonological inventory, i.e. paradigmatic comparison). Stress defines prominence of syllables relative to other syllables in the *metrical structure*, the structure which arranges segments into larger prosodic units like feet and words (by comparison of items in sequence, i.e. syntagmatic comparison). Thus, stress is used for aligning tunes with words and phrases (Pierrehumbert 1980, e.g. see (2.1) above) or for demarcative purposes (for example, Finnish listeners will divide the continuous speech

'XxxXxxXxxXx,' where 'x' represents an unstressed syllable and 'X' represents a stressed syllable, into a sequence of words 'Xxx Xxx Xxx Xx' because they know that the initial syllable in Finnish words is stressed).

The function of stress is then not to provide phonetic cues for segments but to help build up the prosodic structure of an utterance. According to Beckman and Edwards (1994, 2-3) "one reason why stress has been so difficult to characterize phonologically is that stress is not a paradigmatic specification like tone or vowel quality. Rather, it is a syntagmatic structural specification. It is one of the devices that a language can use to set up a hierarchical organization for its utterances." In addition, lack of "uniform and precise phonetic correlates" (Kenstowicz 1994, 550) makes stress harder to characterize. Different languages use different acoustic cues for stress. Research has shown that speakers may lengthen segments to indicate stress in English but they may shorten the segments to indicate stress in Estonian (Lehiste 1970). It is crucial to understand the variation in lexical stress and the variation in its acoustic realization to characterize lexical stress itself.

TYPOLOGY OF LEXICAL STRESS

Fixed vs. Variable Stress

Based on research by Hyman (1977), Laver (1995) divides the languages, which have lexical stress, into two broad categories: languages which have stress marked at a fixed (predictable) syllable in a word and languages which allow the stress to occur on any syllable in a word. Laver distinguishes the two types as *fixed lexical stress* languages and *variable lexical stress* languages. Of the languages analyzed by Hyman, 306 (69%) languages had fixed lexical stress and nine (2%) languages had variable stress. The stress assignment algorithm could not be determined for 113 (25%) languages and sixteen (4%) languages did not utilize stress at all. The different lexical stress types are described in more detail below.

FIXED LEXICAL STRESS

Lexical stress in a word can be fixed on a particular syllable using two different mechanisms. Most languages fix the stress with respect to the initial or the final edge of the word. Hyman's (1977) survey shows that the 306 languages with fixed lexical stress can be divided into the five groups listed in (2.3).

(2.3)

initial syllable stress, 114 languages (37%)
second syllable stress, 12 languages (4%)
ante-penultimate syllable stress, 6 languages (2%)
penultimate syllable stress, 77 languages (25%)
final syllable stress, 97 languages (32%)

Among other functions, these stress patterns serve a *demarcative* purpose, making it easier for listeners to determine where the edge of the word is.

Stress can also be fixed on a certain syllable with reference to the syllabic weight of different syllables within the word. For example, syllables in a language can be stressed depending on their relation with *heavy* syllables in the word (see (2.4) below). A syllable can be heavy due to a variety of language-specific factors. Some languages define all the closed syllables (i.e. syllables formed with onset, nucleus and coda) as heavy. Some languages consider a subset of closed syllables as heavy, i.e. only when the coda belongs to a subset of the consonants in the language. Some languages, that maintain a distinction between long and short vowels, consider only syllables with long vowels as heavy.

This 'heaviness' is sometimes conveniently represented by moraic notation. A *mora* is, according to Lehiste (1970, 44), a time unit equivalent to a single short vowel or a coda consonant. Thus, a long vowel or a vowel-consonant sequence can be represented in an abstract sense as bi-moraic, or twice as long as the short vowel. Defining this unit is useful because it simplifies the analysis of some stress languages. An example is given by Laver (1995, 304), who quotes Munro (1977) reporting the following about Uto-Aztecan languages.

> Vowel length is contrastive in most modern Uto-Aztecan languages... In some cases, vowel length is important in determining the placement of stress. The term 'second mora' ... is used to describe a situation in which the first syllable of a word is stressed if it contains a long vowel or a diphthong, but the second syllable is stressed if the first vowel is short. If long vowels count as two moras, or vowel-units, and short vowels count as one mora, it is clear that in such languages, stress always falls on the second mora of the word.

Some examples of fixed-stress languages given by Lehiste (1970, 148-149), from Hyman (1977), are listed in (2.4).

(2.4)

language		Lexical stress
Czech	→	first syllable of a word
French	→	last syllable of a word
Polish	→	penultimate syllable of a word
Macedonian	→	ante-penultimate syllable of a word
Latin	→	penultimate syllable if it has a long vowel; ante-penultimate syllable otherwise
Old Lesbian	→	penultimate syllable if the last syllable is long; ante-penultimate syllable if the last syllable is short
Classical Arabic	→	long syllable that is closest to the beginning of the word; first syllable if all syllables are short

The first four languages (Czech, French, Polish, Macedonian) use reference to a word edge to assign stress to a word. The last three languages (Latin, Old Lesbian and Classical Arabic) may use reference to a heavy syllable to assign stress to a word. Once it is determined whether there is a heavy syllable, and where it is, stress is fixed with reference to this syllable or to the edge.

Variable Lexical Stress

In languages with variable lexical stress, devising a simple algorithm to assign stress to words is not possible because these languages use free placement of stress. Laver (1994) lists Assamese, Dutch, English, Greek and Russian as some examples of variable stress languages. For example, in Greek ['po.li] means 'city' and [po.'li] means 'much.' In English there are words which have the same segmental content but stress differences can cause these words to be understood as nouns or verbs (e.g. noun/verb pair like 'permit').

This difference in stress placement can change the meaning of the word (as in Greek), change the class of word (as in English) and in some languages also change the syntactic function of the words. For example, Laver (1994, 523) gives the example of a Southwest Brazilian Language, Terena, which uses stress "to distinguish the subject from the object in a verb in an independent clause." Thus, variable lexical stress does not have the simple demarcative function as that of fixed lexical stress. The changes in stress are also associated with other functions.

Morphological vs. Rhythmic Stress

Hayes (1995) argues that not all the languages can be clearly divided into the two groups proposed by Hyman (1977) (which have been summarized from Laver (1994) in the previous sub-section). For example, Polish has penultimate stress, but also contains borrowed words in its lexicon with stress on ante-penultimate syllables. Furthermore, there are languages where the stress is *variable* but within certain *fixed* limits. For example, Hayes (1995, 31) points out that stress may fall on any of the last three syllables in Spanish. Therefore he proposes that languages may be grouped into either *morphological stress languages* or *rhythmic stress languages*.

Morphological Stress

According to Hayes (1995, 31-32), in languages with morphological stress, "stress works to elucidate the morphological structure of a word." Stress is stored as a lexical definition of the morphemes. When different morphemes are put together to form a word, the root may bear the main stress and the affixes may bear secondary stresses . For example, Hayes (1995, 32), referring to the work done by Kiparsky (1982), explains that in English "the fact that *un=bóund=ed=ness* has antepenultimate stress ... merely reflects the fact that the stem syllable is in the antepenultimate position of this word." In other cases, stress may be assigned by a complex interaction of root and affixes. Affixes can carry stress, stress removing properties or stress assigning properties. For example, a suffix might assign a stress to the last syllable of the root it attaches to. Hayes points out that, in case neither the root nor any affix has any stress or stress assigning properties, the stress is determined through a default rhythmic pattern. Indo-European, Modern Hebrew and Pashto are some of the languages which have this type of stress.

Rhythmic Stress

Hayes (1995, 31-32) explains that languages may mark stress based on phonological factors, i.e. languages may mark stress based on syllabic weight, based on distance between stresses or distance between stresses and word boundaries. These languages are classified as having rhythmic stress. Rhythmic stress languages are further classified into *bounded stress languages* and *unbounded stress langauges* by Hayes.

In bounded stress languages, a stress must fall within a fixed distance of another stress or a word boundary. In unbounded stress languages, stress may fall at an unlimited distance from another stress or a word boundary. Finnish is an example of a bounded stress language with the initial syllable always stressed (Carlson 1980), and Classical Arabic is an example of an unbounded stress language with stress on the first heavy syllable in the word (Lehiste 1970, 149).

METRICAL REPRESENTATION OF STRESS

As just exemplified, lexical stress is assigned in many different ways across languages. However, there is a single phonological theory (the metrical stress theory) which enables a coherent representation of this variation in stress (detailed accounts of this theory are presented in most of the recent phonology texts, e.g. Kenstowicz 1994, Goldsmith 1990). In this theory, the phonological prominence (or stress) of a syllable is indicated by using a *metrical grid* (Liberman 1975). Kenstowicz (1994, 553-554) explains that, "for the metrical grid, stress is neither a feature nor an inherent property of syllables. Rather, stress is defined in terms of an abstract two-dimensional array that plots metrical positions for levels of prominence. Syllable nuclei 'bear' a stress by ... associating with one of these metrical positions. In this way, stress is largely autonomous from the phonemic string." A word has three levels of prominence. These levels are represented as lines of asterisks (as in (2.5) below; notation from Halle and Vergnaud, 1987). The lines indicate the level of prominence and asterisks in each line mark the syllable which has that level of prominence. Line 0 has asterisks marked for all elements which can have stress (which includes all syllables, except certain reduced syllables in some languages). An asterisk in line 1 indicates foot level stress and in line 2 indicates word level stress. This analysis for the English words *Apalachicola*, *Tennessee*, and *hurricane* is shown in (2.5) (examples by Kenstowicz 1994, 554 following the analysis by Halle and Vergnaud, 1987).

(2.5)

```
Line 2               *                 *       *
            ____________      ___________     _________
Line 1      *   *    *        *        *      *    *
            ____________      ___________     _________
Line 0      * * *  * * *      *   *    *      *   * *
            ____________      ___________     _________
            Apalachicola      Tennessee       hurricane
```

In addition to word level stress, each syntactic constituent defines a new line in the grid (Halle and Vergnaud 1987, 263-266). This grid enhancement procedure is illustrated in (2.6) (example from Kenstowicz 1994, 554).

(2.6)

```
Line 3                          *
            _______________________
Line 2         *                *
Line 1         *          *     *
Line 0      *  *  * *     * *   *  *
            discovered Mississippi
```

Phrase level line 3 is defined when the phrase *discovered Mississippi* is formed from the words *discovered* and *Mississippi*. Only syllables which have a line *n* asterisk can be assigned a line *n-1* asterisk. Therefore, only syllables with word level stress can be assigned a phrase level asterisk. Consequently, the third syllable in Mississippi is assigned an asterisk at phrase level (assuming that the second word is stressed; the line 2 asterisk of the first word will be promoted to line 3 if this word is stressed).

Thus, metrical grid is a tool which allows representation of phonological stress on different syllables. This difference in stress between syllables has real phonetic consequences which are discussed below.

PHONETIC CORRELATES AND PHONETIC EFFECTS OF STRESS

Effects of stress have been a topic of considerable research for a variety of languages. Research has shown that stress cannot be mapped onto a single phonetic property, but is manifested directly or indirectly by changes in a suite of phonetic properties. Gay (1978, 229) noted that for English stress effects "the fundamental frequency, overall amplitude, duration, and vowel color." This section reviews how stress effects these four factors. Effects on each acoustic factor are considered individually.

However, before stress effects for each of these factors are reviewed, it is important to know the minimum changes in these quantities which the human articulatory system can produce and the human auditory system can detect. Knowing these limits is important in determining whether the statistically significant effects in production are also perceptually salient. It is also important to know which other phenomena can also effect the same phonetic properties that are effected by stress. This is necessary because unless these extra-stress effects are studied and taken into account, filtering out the influence of stress on these phonetic factors is not possible.

Duration

Perception and Production Limits

Speech segments range in duration from about 30ms to 300ms (Laver 1994, 432). Synergies of articulatory musculature can produce speech with consonant-vowel syllables at close to about 10 syllables per second (Lehiste 1970, 8). The human auditory system is well tuned to hear fine differences in duration as well. The just-noticeable-difference (JND) between two acoustic signals presented in sequence depends on the intensity, duration and frequency content of the signal, and has been found to vary between 10 to 40 ms (Klatt 1976, Fujisaki et al. 1975, Huggins 1972, Lehiste 1970, 10-13, and also Creelman 1962).

Intrinsic and Contextual Variation

However, speech normally does not occur close to these limits (except perhaps speech articulated at a high tempo). Normally a speech segment's duration ranges around its intrinsic duration which allows "relative ease of articulation and security of perceptual distinctiveness" (Laver 1994, 433). Intrinsic duration of segments is one of the factors that can influence segmental duration. Klatt (1976, 1213) reports that intrinsic durations "account for much of the variation in segmental timing in speech." The intrinsic duration of vowels is correlated with tongue height, high vowels generally being shorter than low vowels. Hillenbrand et al. (1995) report differences of as much as 30ms between low and high vowels in English. Phonological length can also change the duration of vowels (for languages that divide vowels into long and short sets). In addition, vowels preceding voiced consonants are commonly longer than vowels preceding voiceless consonants. The lengthening can increase the duration of the vowels in English and other languages by as much as 50 ms to 100 ms (Peterson and Lehiste 1960, House 1961, Klatt 1976, Laeufer 1992). Vowels are also lengthened in word and phrase final syllables. Wightman et al. (1992) report that the final lengthening is limited to the syllable rhyme (i.e. the vowel nucleus and the coda consonants). Lengthening effects have been investigated and confirmed by many researchers for many languages (e.g. Klatt 1976, Beckman and Edwards 1990, Wightman et al. 1992, and Berkovits 1993, for English; Zlatoustova 1954, for Russian; Delattre 1966, for English, Spanish, German, and French; Lindblom 1968, for Swedish).

Like vowels, consonantal duration also varies with many factors. Labial consonants are generally longer than alveolar and velar consonants. The relationship between alveolar and velar consonants is context and language dependent (Lehiste 1970, and Laver 1994). Phonological length, word final lengthening (Berkovits 1993) and adjacent vocalic and consonantal context (e.g. Klatt 1973, Smith 1978, and Weismer 1979) may all cause variations in the consonantal duration as well (also summarized in Crystal and House 1988a, Appendix B).

The research suggests that there is considerable intrinsic and contextual variation in the duration of segments. There are additional sources of variation in speech, which include variability across repetitions for the same speaker, inter-subject variability (Smith 1994), tempo (Fourakis 1991), and variability due to syntactic, semantic and pragmatic factors (e.g. Klatt 1976, Port 1981, Gerard and Dahan 1995). However, if all these factors are carefully controlled in experimental conditions, research has shown that stress also has an effect on duration. Both consonantal and vocalic segments are affected because the smallest domain of stress is a syllable.

Stress Effects

Lindblom (1960) recorded sixteen American English speakers to study correlates of stress. He found that stressed syllables in verb-noun pairs, e.g. *conflict*, are longer than the unstressed counterparts in 70% of the instances. The fact that stress increases duration of vowels in English has also been reported by Huggins (1972), Klatt (1976), Crystal and House (1988b), Anderson and Port (1994) and Fear et al. (1995). Lehiste (1970) also summarizes similar research in some other languages. Longer duration is also used as a cue for greater stress in Polish (Jassem, Morton and Steffen-Batog 1968), in French (Rigault 1962) and in Southern Swedish (Westin, Buddenhagen and Obrecht 1966). Recent work in Dutch (Sluijter and Heuven 1996) also shows that vowel duration increases with stress. Bhatia (1993) also reports increased duration with high stress for vowels in Punjabi, a New-Indo-Aryan (NIA) language.

Fry (1955, 1958) performed a series of perceptual experiments to study the effect of stress on duration (and other acoustic properties) of vowels in English. The results of the perception tests showed that longer duration gives the perception of greater stress. These perceptual results have also been confirmed by recent research (e.g. Turk and Sawusch 1996).

The effect of stress on consonant duration is less clear. Research has indicated that the duration of consonants also commonly increases with greater stress in English. Klatt (1976, 1214) reports that "lexical stress also exerts an influence on consonantal durations. Pre-stressed consonants are slightly longer than other consonants, all else being equal (Oller, 1973; Klatt, 1974; Umeda, 1975)." Similar observation are also made by Huggins (1972) and Crystal and House (1988b). Though Keating (1984) reports that closure duration of stops increases with stress, she also reports that prevoicing does not generally increase for voiced stops. As pointed out earlier, the domain of stress is at least a syllable. Therefore, both onset and coda consonants should be influenced by stress. However, the results also show an asymmetrically greater lengthening of onset than coda consonants. Reasons for this asymmetry are still not known.

Interestingly, stress also magnifies the durational effects caused by other factors. De Jong (1991) reports that vowel duration differences due to voicing of final consonants are amplified by the presence of stress. These differences in duration are further highlighted if the syllable with lexical stress also receives higher level phrasal and/or sentence stress. This is indicated by Pierrehumbert's (1994) conclusion that the degree of glottalization in V-V hiatus is influenced with the type of stress. The glottalization effect is stronger when it occurs at the onset of a syllable which has phrase level stress than at the onset of a syllable with only word level stress (also confirmed by Dilley et al. 1996).

Though research shows that duration of vowels and consonants increases with stress for a wide variety of languages, caution should be taken in generalizing this result to all languages. There are degrees, some languages depending more heavily on duration, while other languages prefer other acoustic properties, to the extent that duration does not change with stress at all. This is especially true for languages which use duration (or length) for segmental contrast, e.g. languages which use contrastive vowel length or contrastive consonantal length. As Hayes (1995, 7) points out, "languages with phonemic vowel length contrasts have been shown to avoid using duration as a correlate of stress ... a more dramatic example of this type can be found in Finnish [in which] emphatic stress can involve lengthening of unstressed, rather than stressed, syllables." It may be useful to consider the Finnish data in some detail. Carlson (1980) gives examples of some swear words in Finnish, which are listed in (2.7) below (the first syllable is always stressed in Finnish). The short vowels are reduced with increased stress, to the extent that in extreme stressed cases the short vowels are not articulated. As the short vowels shorten, the coda consonants in the stressed syllable are lengthened. Therefore, [perkele] becomes [prrrkele] under stress and not [*peerkele]. Long vowels, on the contrary, are lengthened with stress (there are no actual measurements reported by Carlson).

(2.7)

	normal articulation	stressed articulation	extremely stressed articulation
short vowels			
	vit.tu	*vitt*.tu	
	pas.ka	*pass*.ka	
	per.kele	*perr*.ke.le	*prrr*.ke.le
	mulk.ku	*mullkk*.ku	
long vowels			
	saa.tana	*saaa*.ta.na	

Carlson (1980, 6) explains that "these facts have a natural functional explanation. Given that tense [stressed] articulation is one in which articulatory targets are reached in full or even exaggerated, and that Finnish has a clear pervasive quality distinction between short and long segments, [stressed] articulation cannot lengthen short segments, but at best shorten them. On the other hand, quantity distinctions are neutralized in syllable coda, and this can be used for expressive purposes." Similarly, Lehiste (1970, 138) outlines that in Estonian, vowels in unstressed syllables are regularly longer than vowels in stressed syllables (data first reported in Lehiste 1968).

Fundamental Frequency

Perception and Production Limits

Laver (1994, 451) reports that average fundamental frequency (F0) values are 120 Hz for men, 220 Hz for women and 330 Hz for children when they are about ten years old. F0 varies typically between 50 Hz to 250 Hz for men and 120 Hz to 480 Hz for women. Within these limits, individual speakers typically exploit an octave of frequency, i.e. the highest F0 during speech of a speaker is normally double the lowest value of F0 (Fant 1956).

The human auditory range spans from about 20 Hz to 20 kHz (Gelfand 1990, 325). Flagnan and Saslow (1958) report that the human auditory system can detect changes of about 0.3 Hz to 0.5 Hz in the F0 of synthetic vowels between 80 Hz to 120 Hz (summarized from Lehiste 1970). However, generally a more conservative value of 1 Hz is quoted by researchers (Gelfand 1990, Laver 1994). This JND in F0 becomes "larger as frequency increases, and ... becomes smaller as the sensation level [or loudness] increases" (Gelfand 1990, 341). Also, an increase in F0 does not linearly correlate with its perception (i.e. pitch), especially at high frequencies. The relationship between F0 and perceived pitch is related by the *mel* scale (Gelfand 1990, 403, adapted from Stevens and Volkmann 1940). A pitch of 1000 mels is defined as 1000 Hz at 40 dB above threshold.

Intrinsic and Contextual Variation

F0 of speech is also highly variable. This variation occurs (among other factors) due to the intrinsic F0 of segments, context, intonation and pragmatic reasons (e.g. the emotional state of speakers; Laukkanen et al. 1996). Low vowels generally have lower F0 than high vowels. Hillenbrand et al. (1995, Table V) found as much as a 15 Hz F0 difference between high and low vowels in English. Whalen and Levitt (1995) list 31 different languages from eleven different language families, all of which show this behavior. One of the possible causes for the intrinsic variation of F0 for vowels is the extent of tenseness of vocal folds caused by differences in tongue height (Ohala and Eukel 1980). Therefore, this variation of F0 is more likely a universal (and not a language specific) property (Lehiste 1970, 70; Zemlin 1988, 182; Laver 1994, 455; Whalen and Levitt 1995).

F0 also changes with contextual variation. Lehiste and Peterson (1961) found that vowels had higher F0 when they followed voiceless consonants than when they followed voiced consonants. However, no variations in F0 of vowels were found when the consonants following the vowels were varied. Clark and Yallop (1990, 283) note that breathy stops and prenasal stops lower F0 of the following vowel more than plain voiced stops. Laver (1994, 477-478) reports Traill and Jackson's (1988) results for Tsonga, a Bantu tone language, that "the effect of breathy voicing on high tones can be lowering of fundamental frequency by as much as 35 Hz, and on low tones by as much as 22 Hz." Schiefer (1986) and M. Ohala (1979) report similar results for Hindi (i.e. F0 is about 20 Hz lower for vowels following breathy stops). These differences outlined for Tsonga and Hindi are substantial considering that a difference of 1 Hz is perceptible by the human ear in this F0 range and that, for tonal languages, the perception of tone can occur with F0 variation on the order of 5 Hz within the 120 Hz to150 Hz F0 range (Abramson 1961, referenced by Lehiste 1970, 80).

Stress Effects

If variation due to extra-lexical-stress effects is controlled, earlier research has indicated that stress also has an effect on the F0 of vowels. Fry (1958) reports that syllables with higher F0 are perceived as stressed in English and that the F0 cue weighs more than the durational cues for the perception of stress. Recent research has shown that these findings are true but the conclusions are misleading. An increase in F0 is not a direct cue for stress. As Pierrehumbert (1980, 102-105) explains, certain tunes (a sequence of low and high tones) are co-produced with utterances in English. The tones in these tunes influence the F0 of these utterances. These tones fall on stressed syllables, therefore, indirectly giving a low F0 value to a stressed syllable if a low tone associates with it and a high F0 value to a stressed syllable if a high tone associates with it. This is contrary to Fry's (1958) conclusions that associate only high F0 values directly with stressed syllables. The fact that stressed syllables can be associated with high or low F0 values (depending on whether high or low tone aligns with the stressed syllable) has been experimentally shown by Morton and Jassem (1965). In addition, if there are no tones associated with the stressed syllables (which is possible for the stressed syllables following the syllable with nuclear tone (Pierrehumbert 1980, 104)), the F0 does not cue for stress. As Pierrehumbert (1980, 103-104) points out, that is the reason "Nakatani and Aston (1978) report that F0 was not a cue for stress on a noun following a focused adjective." In this case, other phonetic cues are utilized for stress.

These results indicate that F0 is an indirect cue for stressed syllables only when they are associated with a tone. And depending on whether the syllable associates with a low or a high tone, either low or high F0 can cue for stress. F0 has also been reported as a cue for stress in other languages (Fear et al., 1995, for English, Fonagy, 1966, for Hungarian, Jassem, 1959, for Polish, Bhatia, 1993, for Punjabi, M. Ohala ,1986, for Hindi, Rigault, 1962, for French).

Intensity

Perception and Production Limits

Laver (1994, 502) reports that humans can hear a sound intensity of zero decibels (dB) to about 120 dB. Normal conversation is conducted at about 70 dB, quiet conversation at about 50 dB and whisper at about 30 dB sound pressure level (SPL). The quiet experienced in a deep country-side at night is about 20 dB SPL.

Loudness is the percept of intensity. Loudness of a tone depends on both the frequency of the tone and its intensity. Beckman (1986, 137) points out that different studies by different researchers, which are collectively presented in Scharf (1978), only agree on the fact that the percept of loudness increases with duration, and varies in critical duration from 10 ms to over 500 ms. The JND of loudness is also a function of frequency and intensity. According to Flagnan (1957), who conducted experiments with synthetic vowels, the JND is approximately 1 dB (summarized from Lehiste 1970).

INTRINSIC AND CONTEXTUAL VARIATION

In speech, different sounds are not articulated at similar intensities. In normal speech, high, close vowels like /i/ and /u/ are lower in intensity than the low, open vowels /a/ and /æ/. Lehiste and Peterson (1959) measured the intrinsic intensities of different vowels for American English and found that low vowels to be about 5 dB more intense than high vowels. Intensity of a vowel is also a function of the fundamental frequency, increasing when the harmonics coincide with the vocal tract resonances (or formants) (House 1959). In addition, intensity is also dependent on other articulatory factors. Sluijter and van Heuven (1996) discuss that the shape of a glottal pulse can vary the energy distribution in the spectrum, changing the spectral tilt and overall intensity of articulation. Gobl (1989) explains the intensity of articulation may also fall if the vocal folds are vibrating in a non-modal, breathy voice configuration. In addition, intensity can also vary with the emotional state of the speakers (Laukkanen et al. 1996). Intensity may also vary with the segmental context (which changes the formant transitions, duration and F0) and prosodic, semantic and pragmatic factors (Laver 1994, 504-507).

STRESS EFFECTS

Higher intensity for stressed vowels in English is reported by Lieberman (1960) but earlier research has shown that intensity is only a weak cue (at best) for perception of stress (Fry 1955, 1958, Lehiste and Peterson 1959). Recent research in English shows that stressed vowels are not always produced more loudly (Sluijter and van Heuven 1996) and that louder vowels are not perceived as more prominent (Turk and Sawusch 1996).

There are two reasons presented by Sluijter and van Heuven (1996, 2472) which explain why stress may not effect the intensity of segments. First, in production, "effort was suggested as a physical correlate of linguistic stress almost a hundred years ago ... Although these views are largely correct, they were wrong in one important respect. When more effort is expended in speech production, this results in not just greater amplitude of the (glottal) waveform, although this is certainly part of it. As we know from more recent studies, increased vocal effort generates a more strongly asymmetrical glottal pulse; the closing phase is shortened, such that the trailing flank of the glottal pulse is steep. As a result of this, the is a shift of intensity over the spectrum so that the low frequency components are hardly effected [and] that the intensity increase is concentrated in the higher harmonics only." In addition, greater effort of articulation may cause the oral vowel to be articulated more extremely (de Jong 1995), making a high vowel higher and increasing the acoustic impedance of the oral tract; thus decreasing the intensity.

Second, in perception, "of course, we need not be surprised if intensity variations should turn out to provide only a marginal stress cue. In fact, it would seem to us that intensity variation will never have communicative significance for the simple reason that intensity is too susceptible to noise. If the speaker accidentally turns his head, or passes a hand before his mouth, intensity drops of greater magnitude than those caused by the difference between stressed and unstressed syllables will easily occur" (Sluijter and van Heuven 1996, 2472).

Thus, research from different languages shows that intensity may be a possible cue for stress in certain languages, in certain contexts. However, its use is restrained perhaps due to articulatory and perceptual limitations.

Vowel Quality

Perception Limits

Using synthesized female speech, Kewley-Port and Watson (1994) and Kewley-Port (1995) found that listeners can differentiate a difference of about 14 Hz in the first formant and about 47 Hz in the second formant. The JND for F1 remains constant but the JND for F2 increases with the frequency. Hawks (1994) also varied F1, F2 and F3 in synthesized vowels, and found that perceived change is smallest when both F1 and F2 change together in the same direction. Larger changes in formants are required for perception if just F1 or F2 is varied or if F1 and F2 are varied together but in opposite directions.

Intrinsic and Contextual Variation

Vowel quality depends on the extent to which articulators deviate from their central position to articulate a particular vowel. The central position is the configuration of the articulators for schwa in English. Acoustically, this extent of articulation is translated into different formant patterns for vowels. The dispersion of the vowel formants depends (among other factors) on the number of vowels in the language. Languages with more vowels tend to have vowel articulations that are more extreme from the central position and languages with fewer vowels tend to articulate them more centrally (Lindau and Ladefoged 1986).

Other non-stress durational changes, e.g. tempo or word-final lengthening, were thought to effect vowel quality as well (e.g. Lindblom 1963). However, later research has indicated that these non-stress durational changes have minimal effect on the quality of vowels (Harris 1978, Gay 1978, Engstrand 1988, Fourakis 1991). Quality of vowels may also change with segmental context. Van Bergem (1993) shows that vowels assimilate with context, especially when unstressed.

Stress Effects

Decrease in stress may cause vowel reduction which changes vowel quality. In English, vowels may undergo phonetic or phonological reduction when unstressed. In phonological reduction, a tense vowel changes into a lax vowel when it is unstressed. Phonetic reduction, on the other hand, does not change the phonemic status of the vowel but still changes its quality. Phonological reduction is more extreme than phonetic reduction. However, other languages may not have a phonological rule for vowel reduction but may still undergo weaker phonetic reduction with change in stress (e.g. Engstrand 1988 for Swedish, Sluijter and van Heuven 1996 for Dutch).

Early research in English shows that the vowel reduction causes vowels to become more central (like schwa). However, more recent research in Dutch indicates that vowels do not simply become more central, but undergo an increasing contextual assimilation (van Bergem 1993, 1995). Conversely, stressed vowels become more extreme because they undergo less assimilation with context and are therefore hyperarticulated (de Jong 1991, 1995, de Jong, Beckman and Edwards 1993).

Earlier research by Fry (1965) shows that these quality differences are not perceptually salient. However, no recent perceptual research on quality was found and limitations of synthesis technology undermines the deductions made by Fry (which Fry himself admits while interpreting the results).

Thus, stress may also vary the quality of vowels. However, the degree of change in quality may vary with language. Also, no conclusions can be drawn about the importance of the variation in quality is to the perception of stress.

THEORIES EXPLAINING THE EFFECTS OF STRESS

Research from various languages shows that stress can directly or indirectly influence one or more of F0, duration, intensity and vowel quality of speech segments. The relative importance of these factors and the way they are influenced (whether increased or decreased with stress) is largely context and language dependent. Still there may be a common underlying process that determines how stress may change acoustic properties across all contexts and languages. Researchers have been trying to determine this process. There are currently three hypotheses which try to explain the phonetic changes caused by stress: the *Jaw Expansion* theory, the *Sonority Expansion* theory and the *Hyperarticulation* theory. This section is devoted to a review of these three theories.

The Jaw Expansion theory is quite limited in its scope, and therefore is only briefly discussed.

Jaw Expansion Theory

The *Jaw Expansion* theory, outlined by de Jong (1995), suggests that speakers tend to lower their jaw position further for stressed segments. However, as de Jong notes, this theory does not take into account the variation of other articulators with stress. In addition, this theory is primarily based on vowel production and therefore cannot explain either the consonantal changes with stress or the perceptual consequences of stress. Therefore, due to its limitations, this theory is not discussed any further.

Sonority Expansion Theory

Ladefoged (1975, 219) defines *sonority* of a segment as "its loudness relative to that of other sounds with the same length, stress, and pitch." Goldsmith (1990: 110-111) defines it as "roughly speaking ... a ranking on a scale that reflects the degree of openness of the vocal apparatus during speech production, or the relative amount of energy produced during the sound - or perhaps it is a ranking that is motivated by, but distinct from, these notions." Though Price (1980, 342) reports that "duration is a more effective cue to sonority than is amplitude... [but] amplitude may play a role when duration is ambiguous...[and] voiced segments tend to be more sonorant that hiss-excited segments, which in turn appear more sonorant than silence," Kenstowicz (1994, 254) believes that "a simple phonetic correlate to the phonological property of sonority has yet to be discovered." However, it is generally agreed by most researchers that the segments are arranged in the sonority hierarchy in (2.8). Sonority decreases from the top toward the bottom of the list. Within each category of sounds, voiced segments are more sonorous than voiceless segments.

(2.8)

Vowels
- low
- mid
- high

glides
liquids
nasals
obstruents
- fricatives
- affricates
- stops

Thus, sonority is an abstract scale which does not have any direct phonetic correlates. Sonority is increased by reducing the "acoustic impedance looking forward from the glottis" (Silverman and Pierrehumbert 1990) and/or by varying the duration of segments (Beckman, Edwards and Fletcher 1992).

Speech is a rhythmic phenomenon. The rhythm is realized with sonorant periods (vocalic regions) and less sonorant periods (consonantal regions) alternating in time. This rhythm plays a role in the perception of the prosodic hierarchy of speech (by marking boundaries and highlighting prosodic elements e.g. syllables, feet and words) (Hayes 1995). The Sonority Expansion theory is based upon the principle that stress makes this rhythm more prominent by increasing the sonority of high sonority (vocalic) regions and/or decreasing the sonority of low sonority (consonantal) regions. Making the rhythm more prominent enables a better realization of the prosodic phenomena of speech (Edwards and Beckman 1988, Silverman and Pierrehumbert 1990, Pierrehumbert 1994).

Pierrehumbert and Talkin (1992) report that they found /h/ utterances to be less sonorant, when in the onset of a syllable with nuclear stress than when /h/ was in a post-nuclear or de-accented position. Pierrehumbert (1994) and Dilley et al. (1996) also found increased intervocalic glottalization with increase in stress. Both these examples indicate that, with an increase in stress, speakers try to decrease the sonority at the onset of the syllable (either by less sonorant articulation of a consonant or by inserting consonantal material between vowels). Pierrehumbert (1994, 53) explains that in English, "the canonical sonority profile for a focus domain has a sharp rise in sonority at the beginning followed by a slow decline in sonority, spread out over more than one syllable." The slow decline in sonority in falling stress configurations, reported by Pierrehumbert, is shown by the fact that post-stressed intervocalic coda stops lenite to fricatives and fricatives to glides.

Sonority expansion also explains why low vowels become lower with stress. Lowering low vowels opens the oral tract further decreasing the acoustic impedance. The reduced acoustic impedance results in increased sonority.

However, there is some data which cannot be explained by the Sonority Expansion theory. First, the asymmetry which exists between the onset and the coda of syllables (with speakers trying to increase sonority in the onset and decrease sonority in the offset by lenition) cannot be explained by this theory. Why do not speakers also try to decrease the sonority of coda consonants at the same rate as that for the onsets, because it would also help highlight the high sonority of the nucleus? Also, if there is a possible explanation for the lenition rules in English, it may not explain why in the data from Finnish (reported from Carlson 1980 in (2.7) above) codas may become less sonorant in falling stress positions (e.g. coda /t/ in the first syllable in ['vit.tu] is articulated as ['vitt.tu] when stressed). Moreover, results for English show that articulation of high and/or back vowels becomes more extreme with stress. Thus, stress increases the [+high] and/or [+back] tongue gestures in vowels consequently increasing the acoustic impedance and decreasing the sonority (de Jong 1995, de Jong, Beckman and Edwards 1993). However, it may be that increase in duration might offset the effect of decreased sonority, because sonority is integrated over time (Beckman et al. 1992, de Jong et al. 1993).

De Jong et al. (1993, 206) also argue that if the Sonority Expansion model is correct, "features ... orthogonal to those on the sonority scale should not be directly affected by stress." However, they report that in stressed, accented articulations of the word 'put' they found increased lip activity for /ʊ/ and reduced assimilation of the final /t/ with the following segment. Similar findings are reported by De Jong (1995, 499). He indicates that "speakers enhance the articulation of nonsonority contrasts such as backness [of /ʊ/], roundness [caused by lowering and protrusion of upper lip in /ʊ/ articulation] and point of articulation [of final /t/ in the word 'put']."

In addition, the Sonority Expansion model does not explain the data from Finnish, quoted earlier in (2.7) from Carlson (1980). The stress contrast in the examples listed is increased with increasing consonantal articulation and decreasing vocalic (short vowel) articulation. Therefore, stress contrast may also decrease the sonority of syllable nucleus.

In summary, though some of the observations from the collected speech data fit the Sonority Expansion theory, there are still other observations which this theory cannot explain. These phonetic observations challenge the Sonority Expansion theory.

Hyperarticulation Theory

De Jong et al. (1993) and de Jong (1995) propose the Hyperarticulation model to explain the phonetic effects of stress. This theory is based on observations by Lindblom (1990) and Lindblom and Engstrand (1989), who suggest that speech occurs on a continuum between hypo- and hyper-articulation. The speaker aims for speech motor economy on one hand and the preservation of distinct speech output on the other (Lindblom 1983). De Jong (1995) predicts that stress shifts speech towards the hyperarticulatory end of the continuum, which increases the effort involved and "enhances perceptual clarity of the output." In addition, unlike the sonority expansion account, which restricts stress to affect only the features determining the sonority of speech, "the hyperarticulation account predicts that all phonemically distinctive contrasts will be directly affected by stress, not just sonority contrasts" (pg. 493) and that "stress should never act to decrease activity associated with production of a contrast" (pg. 502).

Therefore, the Hyperarticulation theory correctly predicts that stressed high and/or back vowels will be 'higher' and/or 'more back' than unstressed high and/or back vowels, even though this will decrease their sonority. De Jong (1995) also argues that because the articulatory gestures are more extreme and less overlapped, stress would decrease coarticulation, and increase the duration of stressed segments. Therefore, increases in segmental duration with stress also follow from this theory. In addition, because hyperarticulation predicts reduced coarticulatory overlap, this theory also explains why consonants assume more distinct targets in stressed positions. This supports a more distinct articulatory target for /t/ in 'put' and more protrusion and lowering of upper lip in the articulation of /ʊ/ in 'put', as reported de Jong et al. (1993).

In addition, de Jong (1995, 502) also explains that voicing of a coda consonant induces greater durational lengthening in stressed than in unstressed syllables because "vowel lengthening before voiced codas is a conventionalized aspect of the English language, and thus is more readily apparent with more stress." However, Pierrehumbert (1994) reports that though the 'conventionalized' glottalization word-initially in V-V hiatus is strengthened with stress, the 'conventionalized' glottalization occurring as a secondary articulation on a syllable-final voiceless stops is not strengthened. Similarly, though the word-initial /h/ articulation is strengthened with stress (e.g. in 'hogfarmer'), word-medial /h/ is less affected (e.g. in 'tomahawk') even when under nuclear stress. Moreover, Pierrehumbert also points out that the claim that stress increases vowel lengthening before voiced codas supports the hyperarticulation model is not entirely correct. According to the Hyperarticulation theory, only phonemically distinctive contrasts are enhanced. However, vowel length is not phonemically distinctive in English because "vowel length alone does not distinguish between words in English" (pg. 52). Finally, Pierrehumbert also explains that hyperarticulation cannot explain the lenition of intervocalic consonants in falling stress configurations in English.

Furthermore, Hyperarticulation theory does not explain the results from Finnish reported by Carlson (1980) and from Estonian reported by Lehiste (1970). If all the distinctive cues are hyperarticulated, then the short vowels in Finnish should also lengthen (and not shorten) with increased stress, especially when the long vowels do lengthen with stress. Also, stressed vowels should be longer than unstressed vowels in Estonian.

Again, this theory only explains some data from across languages, but cannot encompass all the available speech data. Like the Sonority Expansion theory, the Hyperarticulation theory also faces challenging issues.

THE URDU LANGUAGE

Urdu belongs to the family of New Indo-Aryan (NIA) languages, which is a sub-branch of Indo-European languages. Urdu is spoken by at least 50 million people in more than ten countries as a first or a second language (the majority of speakers are in Pakistan and India) (from Ethnologue, 12th edition, 1992 © Summer Institute of Linguistics; URL: http://www.sil.org). Urdu is similar to Hindi and both are derived from *Khari-Boli* or *Dehlvi.* Masica (1991, 27) writes concerning Hindi and Urdu that, "the ultimate anomaly in the what-is-a-language dilemma in Indo-Aryan is presented by the Hindi-Urdu situation. Counted as different *languages* in sociocultural sense…, Urdu and Modern Standard Hindi are not even different dialects or subdialects in linguistic sense. … They are different *literary styles* based on the *same* linguistically defined subdialect." Masica adds that "in terms of grammar and core vocabulary, they [Hindi and Urdu] are virtually identical; there are minor differences in usage and terminology," but he further adds, "at formal and literary levels, however, vocabulary differences begin to loom much larger (Hindi drawing its higher lexicon from Sanskrit, Urdu from Arabic and Persian), to the point where the two styles/languages become mutually unintelligible." They are also written in different scripts: Hindi in Devanagri and Urdu in modified Perso-Arabic script. Appendix A lists and explains the Urdu script.

However, some pilot work indicates that Masica's statement, that Hindi and Urdu are different only in the sociocultural sense, is not entirely true. Though having the same origins and having a very similar linguistic structure, Urdu phonetics and phonology have diverged from Hindi phonetics and phonology. The divergence is perhaps caused by the strong Perso-Arabic influence on Urdu and the strong Sanskrit influence on Hindi. For example, Hindi does not have velar fricatives while Urdu has both the voiced and voiceless velar fricatives /x, ɣ/. In Hindi these sounds are realized as a velar aspirated stop and a velar voiced, aspirated stop respectively. For example [xan] in Urdu will be [k^han] in Hindi. Some other differences in the phonemic inventories are pointed out by Kachru (1987, table 3.1). Even at the phonetic level, there are differences between the two languages. For example, Davis (1994) reports that the four stop manners in Hindi, voiceless, voiced, aspirated and breathy stops, all have different lag times (the time between the release of the stop burst and the onset of the second formant of the following vowel). She reports these lag times are significantly different and can cue to the type of stop. However, work done on Urdu (Hussain 1994) indicates that the four stops cannot be differentiated by the lag time alone. Both prevoicing and lag time are needed to separate these four stop types.

Thus, there are both similarities and differences between the two languages. These similarities and differences indicate that though the study of each can contribute to the linguistic analysis of the other, both Hindi and Urdu phonology and phonetics need to be studied in their own right.

What is already known about Urdu phonology and phonetics is summarized in this section. There has not been much research done on the phonology of Urdu. Most of the information provided in this section is through pilot work done by the author.

SYLLABIC WEIGHT

All possible syllable templates of Urdu are listed in (2.9)[1]. Open syllables with short vowels do not occur in the word final position. Also, there can be complex codas (with more than one consonant) and complex onsets in Urdu syllables, however, there are limitations on formation of these complex onsets and codas. First, the *Sonority Sequencing Principle* (SSP), which requires the onsets to rise in sonority towards the

[1] Based on pilot work by the author. Part of these results are published in Coleman, Dirksen, Hussain and Waals (1996).

nucleus and codas to fall in sonority from the nucleus (e.g. Kenstowicz 1994, 254), must be satisfied. In addition, these complex onsets and codas can contain at most two consonants. When there are two consonants in the onset, the second consonant in the onset is limited to the glides /w/ or /y/ and maybe /h/. When there are two coda consonants, the first consonant in the coda is limited to a voiceless fricative (/f/, /s/, /ʃ/ or /x/) or nasals (/n/ or /m/), and the second consonant is limited to stops. In addition, the alveolar flap cannot occur in the onset position. There may be more restrictions on the construction of these complex onsets and codas. More research needs to be done to determine the complete breadth of phonotactic constraints on syllable construction in Urdu.

(2.9)

Simple Onset	Comments
CV	short vowel, open syllable (not licensed in word-final position)
CVC	short vowel, closed syllable
CVCC	short vowel, closed syllable with consonant cluster
CVV	long vowel, open syllable
CVVC	long vowel, closed syllable
CVVCC	long vowel, closed syllable with coda consonant cluster

Complex Onset*	Comments
CCV	short vowel, open syllable (not licensed in word-final position)
CCVC	short vowel, closed syllable
CCVCC	short vowel, closed syllable with consonant cluster
CCVV	long vowel, open syllable
CCVVC	long vowel, closed syllable
CCVVCC	long vowel, closed syllable with coda consonant cluster

*Some of these templates are speculative. A detailed phonetic study needs to be done to confirm these templates.

As noted earlier, a syllable in Urdu can be formed with different number of consonants and with long or short vowels. Urdu is sensitive to the consonantal and vocalic composition of its syllables. For example, as previously stated, there are no word final open syllables in Urdu with short vowels. This sensitivity to consonantal and vocalic makeup also plays a part in determining which syllable in a word has lexical stress. Stress position is not fixed relative to a word edge in Urdu, e.g. lexical stress is not always assigned to the first, last or the penultimate syllable in a word. As shown in (2.10), the two and three syllable words can have stress on any one syllable.

(2.10)

(a)	ʔɪz.d̪ɪ.'vɑd͡ʒ CVC.CV. CVVC	'tʃəp.kə.lɪʃ CVC .CV .CVC
(b)	ba.'rat CVV.CVVC	'na.ta CVV.CVV
(c)	d̪ə.'rəxt̪ CV. CVCC	'nə.zər CV.CVC

The stress can shift from one syllable to another with a change in vowel length. In (2.10a), both words have a similar consonant-vowel structure, except that the first word has a long vowel in the final syllable and the second word has a short vowel in the final syllable. Different vowel length alters the stress; the final syllable is stressed in the first word and the initial is syllable stressed in the second word. Because stress assignment in Urdu is sensitive to the length of the vowel, long vowels have a different phonological status in Urdu than short vowels. One way to describe this difference is to represent quantity in terms of an abstract *mora*. A mora represents a time unit equivalent to a single short vowel (Lehiste 1970, 44). However, a mora is not a "species of sound but rather an elementary prosodic unit ... like the syllable ... intervening between the [syllable] and the phonemic string" (Kenstowicz 1994, 293, from McCarthy and Prince 1986, and Hayes 1989).

Using the moraic concept, a short vowel in Urdu is mono-moraic and a long vowel is bi-moraic. Moreover, the stress assignment in Urdu is sensitive to this moraic count. Stress in the words listed in (2.10b) also shifts with adding coda consonants to the syllable. Having established that stress changes with the moraic count of the syllables, it can be deduced that in Urdu coda consonants are also moraic. In addition, the examples in (2.10a,b) indicate that last syllable is stressed only when it is tri-moraic (tri-moraic syllables are supported by Hayes 1989, 1995). Otherwise a non-final syllable is stressed. In the example words in (2.10c), the first word has final syllable stressed as well. Therefore, the final syllable must have three morae. This final syllable contains a short vowel, which contributes a single mora, and therefore each of the two coda consonants must also contribute a mora each. Otherwise, the stress would fall on the first syllable as in the second word in (2.10c).

Putting it all together, in Urdu both vowels and coda consonants are moraic. Short vowels are mono-moraic and long vowels are bi-moraic. Consonant clusters in the coda of the syllables are bi-moraic. Therefore, open syllables with short vowels are mono-moraic, closed syllables with short vowels and open syllables with long vowels are bi-moraic and closed syllables with long vowels or with short vowels and a coda cluster are tri-moraic. These three quantities are phonologically distinguished in Urdu, at least when stress is being assigned.

These moraic or quantity differences are normally represented as a difference of the weight of the syllables. Mono-moraic syllables are called 'light' (L), bi-moraic syllables are called 'heavy' (H) and tri-moraic syllables are called 'super-heavy' (S) (notation of 'light,' 'medium' and 'heavy' is also utilized by some researchers). Not all languages phonologize different weights of the syllables. Also, languages which are sensitive to syllable quantity or weight may only make a two way (light vs. heavy) distinction. However, Urdu utilizes a three way weight distinction in its phonology.

LEXICAL STRESS

Equipped with information to determine the syllable weights in Urdu words, the algorithm for determining lexical stress is discussed in this section. Evidence of lexical stress in Urdu comes from *Standard Twentieth Century Dictionary: Urdu into English* (Qureshi 1992). This dictionary has stress marked in its English transliterations of Urdu words. However, there is no explanation provided about how this stress is determined. Additional indication that stress exists in Urdu comes from Masica's (1991, 121) study of NIA languages, which include Urdu. Masica states that "NIA languages are syllable or *mora*-timed rather than stressed timed, and although stress patterns differ from language to language, stress in generally predictable." These facts indicate that Urdu has lexical stress, but still they all come short of providing any clues as to how it is assigned to Urdu words.

Lack of any research on stress in Urdu and the similarity of Urdu to Hindi provides motivation to also consider the stress pattern found in Hindi. Ohala (1986, 83-84) outlines five different proposals forwarded by researchers for assigning stress to Hindi words (Grierson 1895, Dixit 1963, Mehrotra 1965, Kelkar 1968 and Sharma 1969). Additional proposals have also been forwarded by Hayes (1995) and Pierrehumbert and Nair (1996)

Going through these stress assignment proposals, a lot of variation is noticed. However, as Hayes (1995) points out, Hindi is spoken by a large mass of people, most of whom also speak another language. Thus, variation in stress is expected. However, a few generalizations can still be made from these analyses. First, all analyses assign stress starting from right and moving back to left (i.e. from the end of the word going toward the beginning). Second, most analyses (explicitly or implicitly) divide the syllables into three weight categories: light, heavy and super-heavy. Third, the last syllable (or foot in case of analyses by Hayes, and Pierrehumbert and Nair) holds a special status, avoiding stress which then falls on or before the penultimate syllable. Two more generalizations which are not very apparent also come out of these analyses. First the 'VV' and 'VC' or the 'VVC' and 'VCC' sequences behave almost similarly in stress assignment rules, meaning that a coda consonant has the same weight as a vowel (as reported for Urdu in an earlier section). And finally, there is only one stress per word in Hindi.

Some of these generalizations can perhaps also be extended to Urdu. To determine an algorithm for stress in Urdu, transliterations of Urdu words with the stress marks from *Standard Twentieth Century Dictionary: Urdu into English* (Qureshi 1992) are used. Words with all possible combinations of syllable weights are considered. These words are listed in (2.11) below. 'L' represents 'light' or mono-moraic syllables, 'H' represents heavy or bi-moraic syllables and 'S' represents super-heavy or tri-moraic syllables.

(2.11)

Syllable Structure	Urdu Words	Comments
~~L L~~		not licensed
ˈL H	'nə.zər, 'bə.ri	"sight", "acquitted"
L ˈS	nə.'d͡ʒɑt	"liberation"
~~H L~~		not licensed
ˈH H	'nɑ.tɑ	"connection"
H ˈS	nɑ.'xun	"nail"
~~S L~~		not licensed
ˈS H	'nɑtʃ.nɑ	"to dance"
S ˈS	ʔɑb.'nus	"ebony"
~~L L L~~		not licensed
L L H	---------------	no examples found
L L S	---------------	no examples found
~~L H L~~		not licensed
L ˈH H	tʃɪ.'pək.nɑ	"to stick"
L H ˈS	tʃə.rɑ.'gɑh	"pasture"
~~L S L~~		not licensed
L ˈS H	p^{h}ə.'lɑng.nɑ	"to jump"
L S S	---------------	no examples found
~~H L L~~		not licensed

'H L H	'tʃəp.kə.lɪʃ	"altercation"
H L 'S	ʔɪr.tɪ.'kab	"perpetration (of crime)"
~~H H L~~		not licensed
H 'H H	tʃɑ.'lɑ.ki	"cleverness"
H H 'S	ʔɪb.rɑ.'him	"Abraham"
~~H S L~~		not licensed
H 'S H	bər.'xɑs.təh	"who has been dismissed"
H S 'S	bər.xur.'dɑr	"son"
~~S L L~~		not licensed
'S L H	'bin.dɪ.yɑ	"extra draught animal used as support"
S L S	---------------	no examples found
~~S H L~~		not licensed
S 'H H	ʔɑb.'ka.ri	"excise duty"
S H S	---------------	no examples found
~~S S L~~		not licensed
S S H	---------------	no examples found
S S S	---------------	no examples found

The stress patterns on bisyllabic words show that if the final syllable is super-heavy, it is always stressed. If the final syllable is heavy, then the initial syllable is stressed. Final light syllables are not allowed. In tri-syllabic words, again the final syllable is only stressed when it is super-heavy. The penultimate syllable is stressed when it is heavy or super-heavy and the final syllable is not super-heavy. The ante-penultimate (or initial) syllable is stressed when it is heavy or super-heavy and the penultimate syllable is light and the final syllable is not super-heavy.

Generalizing over all words, if a final syllable belongs to {super-heavy} it is stressed, and if it belongs to {light, heavy} it is not stressed. If a final syllable is not super-heavy, the syllable belonging to {heavy, super-heavy} closest to the end of the word is stressed. And if the final syllable is not super-heavy and there are no syllables belonging to {heavy, super-heavy} preceding the final syllable, the penultimate syllable is stressed.

Comparing different sets of syllables, it is noticed that {light, heavy} pair together when considering word-final syllables (irrespective of whether the heavy syllable is closed of open) and {heavy, super-heavy} pair together when considering non-word-final syllables. This irregularity can be accounted for if it is assumed that the Urdu language only divides syllables into two groups for stress assignment purposes, mono-moraic and multi-moraic syllables. And, in addition, the final mora in the final syllable is considered *extrametrical* i.e. its weight is not counted in the syllable weight (Hayes 1995, 56-61). The final mora and not the final consonant is extrametrical because the final heavy syllable is considered as light, for stress assignment purpose, whether it is open or closed). A bi-moraic (heavy) syllable becomes mono-moraic (light) and a tri-moraic (super-heavy) syllable becomes bi-moraic (heavy) in word-final position. Though there is no motivation for this extrametricality, except that it greatly simplifies the stress assignment procedure, as shown later, there is some indirect evidence that might support its existence. Urdu does not allow word-final light syllables. One could argue that on the grounds that, if light syllables are allowed in word-final position, extrametricality would render them weightless (or mora-less), which is an oddity for a syllable. Thus, the constraint on word-final light syllables indirectly supports the extrametricality of the word-final mora.

If the extrametricality is taken into account, the super-heavy/heavy distinction becomes a heavy/light distinction in final syllable. The whole stress algorithm can then be simplified into the statement that the *last heavy syllable is stressed* (syllable weights counted after subtracting away the extrametricality effect in the final syllable). *If all syllables are light, the penultimate syllable is stressed.* The same stress assignment algorithm can be extended for words with more than three syllables.

These results show that Urdu is a fixed stress language. Yet Urdu also shows signs of being a weakly variable stress language. There are a few words in Urdu which are differentiated only by their lexical stress. For example, /'bə . ha/ (flowed, past tense) and /bə . 'ha/ (cause to flow) are phonemically identical, but differ in lexical stress. The initial syllable is stressed in the first word, and the final syllable is stressed in the second word. These word pairs are always grammatically related. Thus, lexical stress difference is utilized to indicate the change in the grammatical function. According to Hayes' (1995) classification, Urdu is an rhythmic language with unbounded, right-headed stress.

This pattern of stress assignment, explained for Urdu, is not unique. Lehiste (1970:149) describes the stress pattern of Classical Arabic as being one where stress normally falls on the "long syllable that is closest to the beginning of the word, and on the first syllable if the word consists only of short syllables". The Urdu stress algorithm is a close parallel. Masica's (1991, 121) analysis of other NIA languages also reveals that this pattern is not extra-ordinary because, "for the remaining languages (Hindi, Gujrati, Punjabi etc.) rather complicated sets of rules are necessary, involving the number of syllables, whether they are open or closed, and the nature of their vowels."

Comparing this proposed algorithm for Urdu with the various other algorithms proposed for Hindi and outlined earlier, many similarities are noticed. All the generalizations made for Hindi also hold true for Urdu. Along with its simplicity, an important addition in this algorithm is the direct reference to syllable weight (which plays a key role in stress assignment) by using morae and the use of mora extrametricality to explain the special status of final syllable in the words.

The stress algorithm proposed works for most, but not all, of the words in the lexicon. The exceptions may be caused by various reasons. First, affixations can cause stress to change irregularly when the affixes add their own stresses. Second, words borrowed from other languages can possibly have unique stress patterns. Third, some words can just be anomalous but still be an accepted form, with memorized instead of inferable stress (as indicated by the slight tendency of Urdu towards variable stress).

With this analysis, the words in Urdu have a single stress within a word. Words with syllable structure 'HLS' also show single stress on the final syllable (HL'S and not multiple stresses, e.g. *'HL'S). However, this observation is based only on analysis of Urdu words which are limited to three syllables. To determine whether there can be more than one stress per word, words with more syllables also need to be analyzed. During the first pass through an Urdu dictionary, however, no examples of words were found with syllable structures which can possibly support multiple stress (e.g. HLHH). A more detailed study of Urdu lexicon should be undertaken. However, the current work is only limited to bisyllabic (and a few tri-syllabic) words, and therefore, multiple stresses will not be investigated any further.

PHONETICS OF URDU VOWELS AND STOPS

The current study investigates the effects of stress on both vowels and consonants. The oral vowels along with the stop consonants in Urdu are studied in this work. There are various reasons for these choices. Oral vowels (and not nasal vowels) are chosen for two reasons. First, as discussed in Appendix A, the phonemic status of nasal vowels is still unclear. Second, nasal vowels are harder to analyze acoustically because the nasal passage introduces anti-resonances, which make it difficult to measure the vowel resonances. Stops are chosen over other consonants because during acoustical analysis, demarcation of stops and vowels within syllables and words is more reliable with almost discrete sections of vowel formants, closure duration and aspiration. Second, the set of stops in Urdu span all four air stream mechanisms over four different places of articulation. Stops are also very well studied in other languages, and there is plenty of literature on these segments for cross-referencing across languages. No other sub-class of consonants in Urdu has all these advantages.

Urdu oral vowels are very similar to vowels in English and other languages. In addition, there are no specific acoustic data available for Urdu vowels. Therefore, their acoustic properties are assumed to be similar to oral vowels in other languages The stops articulated at different places in Urdu have the same acoustic differences as observed in other languages (e.g. Stevens 1972, Stevens and Blumstein 1975, 1978, Kewely-Port 1983). However, research has shown that the air stream mechanisms in Urdu are produced and perceived differently from other languages (Hussain 1994, Hussain and Nair 1995). The air stream mechanisms of Urdu stops are explained below.

An overview of most of the types of stops is presented by Henton, Ladefoged and Maddieson (1992, 65). They argue that the stops can be divided into three phases: onset, closure and offset, but "phonologically only two phases- closure and release- are used; independent distinctions of features such as phonation type or articulatory manner cannot be found in the onset phase." Steriade (1993a,b) also agrees that phonologically languages can exploit both the closure and release of the stops. Hussain (1994) reports that Urdu does utilize both these parts of the stops in Urdu to distinguish between voiceless, voiced, aspirated and breathy voiced stops.

The four types of stops are contrasted phonologically by using two features, [voice] and [aspiration]. Voiceless stops are [-voice, -aspirated], voiced stops are [+voice, -aspirated], aspirated stops are [-voice, +aspirated] and breathy voiced stops are [+voice, +aspirated]. Results from (Hussain 1994 summarized in Hussain and Nair 1995) show that the [-voice] stops do not have voicing during closure, the [+voice] stops have voicing during closure, the [-aspirate] stops have an F2 lag time (time between stop release and onset of F2 of the following vowel) less than 40 ms and the[+aspirated] stops have an F2 lag time greater than 45 ms. These results are robust in production and perception of stops. Thus, using voicing during closure and F2 lag time, the four stop types can be completely distinguished in Urdu.

SUMMARY

The reviewed literature indicates that lexical stress is an abstract phonological property which is assigned to one or more syllables in each word by different mechanisms, depending on the language. Lexical stress is translated into the phonetic domain with a variety of context and language specific acoustic cues. If the stressed syllable is associated with an accent, F0 is the strongest acoustic cue for the stressed syllable for many languages. In addition, duration (of consonants and vowels), vowel quality and (to a lesser extent) intensity of vowels also change with stress. Because there are other factors which also change the same phonetic properties, only when these extra-stress factors are controlled or taken into account, can the stress effects be filtered out from the acoustic changes.

Currently, there are two competing hypotheses which attempt to explain the acoustic differences caused by stress. The Sonority Expansion theory suggests that stress works to increase the sonority features and the Hyperarticulation theory suggests that all phonemically distinctive contrasts are more distinctly articulated. Both theories explain some but not all stress related effects reported for different languages.

Urdu also has lexical stress, marked on its final heavy syllable (assuming that the final mora in Urdu is extrametrical and therefore does not contribute to the weight of the final syllable). A syllable is heavy if it has at least two morae. Morae can be contributed by long vowels (bi-moraic), short vowels (mono-moraic), or one or two coda consonants (each of which is mono-moraic).

Knowing which syllables in Urdu words are stressed and knowing how stress may effect the acoustic properties of the segments in the stressed syllable provides an appropriate background for investigating the acoustic effects of lexical stress in Urdu. The following work is devoted to this investigation.

METHODS

As the literature reviewed suggests, a lot of different vocalic and consonantal phonetic properties can be altered with stress. The present study concentrates on how a subset of these phonetic properties change with stress in Urdu. Specifically, this study investigates the change in duration, F0, intensity and vowel quality of segments in Urdu. This chapter explains how the experiments to investigate lexical stress in Urdu were set up and conducted and how the results were analyzed.

STIMULI

In choosing the stimuli, variability from extra-stress effects was controlled as much as possible. Care was taken to choose the same segmental context for all the target vowels and stops in the experiments. All words were taken from the Urdu lexicon, and where ever possible, highly familiar words were selected. There was no data available on the frequency of words in Urdu; therefore, the familiarity of the word by the author, who is a native speaker of Urdu, was used as the criterion for word frequencies. When no familiar words were found, familiarity was compromised for similar segmental context and not vice versa. These experiments were designed with a limited knowledge of the phonology and phonetics of Urdu. Though the words were carefully chosen, sometimes unexpected pronunciations or other unforeseen factors forced the original material to be altered. The progression from the original material proposed to the final material used is also explained, where relevant.

For each vowel and stop investigated, a pair of words containing the target segment was selected. The target segment, located in the first syllable in two or three syllable words, was stressed in one of the words in the pair and unstressed in the second word of the pair. For example, the word pair /ˈko.ke/ (meaning "small nails") and /ko.ˈken/ (meaning "cocaine") was chosen for the vowel /o/. The context of the target segment in the two words in each pair was kept as consistent as possible (for example, /o/ occurs between /k/'s in both words and the second vowel is /e/ in both words). Due to the limitation of using familiar, real words, exact context matches were sometimes not possible. When there were differences in context, they were minimized. This was achieved , for example, by matching the place of articulation when an exact match in both place and manner was not available.

For long vowels, the words in (3.1) were used. Target vowels are the ones in the first syllable. They are stressed in the words listed in the first column and unstressed in the words listed in the second column.

(3.1)

'd̪i.d̪ɑ	"hard work"	d̪i.'d̪ɑr	"sight"
'be.tɑ	"son"	be.'t̪ɑb	"impatient"
'bæ.tʰɑ	"sat"	bæ.'t̪ɑl	"ghost"
'pɑ.t̪ɑ	"achieve"	pɑ.'t̪ɑl	"hell"
'ko.ke	"small nails"	ko.'ken	"cocaine"
'su.bɑ	"province"	t̪u.'mɑr	"roll of papers"

The adjacent consonantal context is alike for both words in any pair, with a few exceptions. For /e/, there were no words in the lexicon with exactly the same context. Therefore, the pair listed in (3.1) was used. The stop following the stressed /e/ is the alveolar /t/ and the stop following the unstressed /e/ is the dental / t̪/. The pair was preferred over other possible pairs because the proximity of the alveolar and dental places of articulation. A similar situation exists for the vowel /æ/. In addition to the place difference, there was also a difference in the manner. The stressed /æ/ was followed by an aspirated alveolar stop. However, phonetic data on Urdu stops shows that voiceless aspirated stops and voiceless unaspirated stops differ only after the release (Hussain 1994, Hussain and Nair 1995). Therefore, the closure part of the two stops adjacent to the vowel in the stressed and unstressed case was still very similar. Though a totally balanced pair was not found for /u/[2], the consonants around the vowel did match in place and manner of articulation (to some extent)[3].

In addition to controlling the consonantal context, vowels in the second syllables were also matched for each pair of words. This controlled vowel-to-vowel coarticulation for words within a pair. Exact words with alternating stresses were not found because Urdu is a fixed stress language in which syllable weight determines the placement of stress. Therefore, syllable weight had to be altered (by adding a consonant) in order to alter the stress. However, the final consonant, added to shift the stress away from the first syllable, was quite far from the initial vowel (as far as segmental distance is concerned) and therefore was assumed to have a minimal segmental effect on the vowel of the first syllable.

The words used for short vowels are listed in (3.2). Again, the target vowels were present in the first syllable. No bi-syllabic words were found which presented the stress contrast with similar segmental context. Therefore, two syllable words and three syllable words for stressed and unstressed vowels, respectively, were chosen. For the short vowel

---

[2] The balanced pair /ˈgu.d̪ɑ/ meaning "pulp" and /gu.ˈd̪ɑm/ meaning "warehouse" was initially used for /u/. However, after analyzing the recordings of two speakers, it was found that the second word was being pronounced with vowel /o/ instead of /u/.

[3] Both /s/ and /t̪/ were voiceless apicals and both /b/ and /m/ were voiced labials. Peterson and Lehiste (1960, table II, 702) report that average vowel duration for long vowels is 313 ms before an /m/ and 307 ms before a /b/. So vowels were only slightly longer before /m/ than before /b/. Lehiste (1970, 27) also reports Elert's (1964) finding for Swedish that vowels following /t/ are about 14 ms shorter than the vowels following /s/. Thus, in /suba/ the vowel /u/ lengthens because of /s/ but shortens because of /b/ if the effects of /t̪/ and /m/ in /t̪umar/ are taken as reference. Therefore, the overall contextual effect on /u/ was assumed to be about the same in the stressed and unstressed conditions.

/ə/, a bi-syllabic word was also found. Both bi-syllabic and tri-syllabic words were used for unstressed /ə/ to compare any effects of compression caused by the additional syllable. Again, the immediate consonantal context and vowel context for target short vowels was exactly the same in stressed and unstressed case.

(3.2)

'bɪk.nɑ	"to be sold"	bɪk.'vɑ.nɑ	"to cause to be sold"
'pək.nɑ	"to be cooked"	pək.'vɑ.nɑ	"to cause to be cooked"
		pək.'vɑn	"cooked food"
'pʊg.nɑ	"to qualify"	pʊg.'vɑ.nɑ	"to cause to qualify"

Stops in both onset and coda positions of the first syllable were analyzed to find the lexical stress effects on stops in Urdu. The data listed in (3.3) was used for onset stops.

(3.3)

(a) Voiceless Unaspirated Stops

'pɑ.t̪ɑ	"achieve"	pɑ.'t̪ɑl	"hell"
't̪ɑ.ri	"take over"	t̪ɑ.'rik	"dark"
'tæ.nɪs	"tennis"	tæ.li.'fun	"telephone"
'ko.ke	"small nails"	ko.'ken	"cocaine"

(b) Voiced Unaspirated Stops

'be.tɑ	"son"	be.'t̪ɑb	"impatient"
'd̪i.dɑ	"hard work"	d̪i.'dɑr	"sight"

(c) <u>Voiceless Aspirated Stops</u>

'pʰət.ka	"came near"	pʰət.'ka.na	"to kill"
't̪ʰəp.ka	"patted"	t̪ʰəp.'ka.na	"to pat"
'kʰər.ka	"jolted"	kʰər.'ka.na	"to jolt"

(d) <u>Voiced Aspirated Stops</u>

'bʰug.ta	"suffered"	bʰug.'tan	"payment"
'd̪ʰər.ka	"palpitate"	d̪ʰə.'ɽam	"with a thud"
'dʰəl.na	"set"	dʰəl.'van	"slope"
'gʰəɽ.na	"implant"	gʰəɽ.yal	"gong"

Stops at all four places of articulation for all different types of articulation were studied (except pairs of words for voiced alveolar onsets and aspirated alveolar onsets were not found with similar segmental context; these two stops were therefore excluded from the analysis). Pairs of words with the same immediate vocalic and consonantal contexts were found for voiceless stops at all the four places of articulation. One pair used borrowed words from English ('tennis' and 'telephone'). However, both words in the pair were borrowed, so similar phonetic changes (if any) were assumed to effect the target segment in both cases. Therefore, the phonetic effects were still assumed to be caused by change in stress. No pairs for /g/ and /d/ were found in the onset position.

For the aspirated stops, pairs with bi-syllabic words were not found and therefore bi-syllabic and tri-syllabic words were used. This might have caused some compression in the closure and post-release aspiration for unstressed onset (Port 1981). However, any degree of compression could be found after analyzing the bi-syllabic/tri-syllabic pair used for the unstressed /ə/ in the list of short vowels (in 3.2). Originally, different stimuli were chosen for the bilabial and velar breathy stops[4]. However, these words had nasal consonants in the target syllables which interacted with the breathy voicing. Therefore, two other word pairs, which did not include nasals, were chosen for these breathy stops.

For the coda stops, the stimuli in (3.4) were used.

(3.4)

(a) Voiceless Unaspirated Stops

'pək.nɑ	"to be cooked"	pək.'vɑn	"cooked food"

(b) Voiced Unaspirated Stops

'səd.kəh	"alms"	bəd.'kɑr	"bad person"

(c) Voiceless Aspirated Stops

'gʊt̪ʰ.t̪ɑ	"get tangled"	t̪ʰʊt̪ʰ.'kɑr	"hate"
'gʊt̪ʰ.t̪ɑ	"get tangled"	hət̪ʰ.'yɑr	"weapon"

(d) Voiced Aspirated Stops

'tʃubʰ.t̪ɑ	"pinch"	tʃubʰ.'vɑ.nɑ	"to cause to pinch"
'bəndʰ.nɑ	"to get tied"	bəndʰ.'vɑ.nɑ	"to cause to tie"

Again, the immediate consonantal and vocalic context was chosen to be the same in each pair of words. For the aspirated stops, the first pair was used for the first five

---

[4] Originally the balanced pairs /'bʰon.kɑ/ (meaning "barked") and /bʰon.'tʃɑl/ (meaning "earth quake") for the labials and /'gʰən.tɑ/ meaning "hour" and /gʰəɽ.'yɑl/ meaning "gong" for velars were used. Analysis of recordings by the first two speakers indicated that the nasal coda consonants interacted with the breathy onsets. This has been noted elsewhere in the literature as well. Ohala (1983) and Ohala and Ohala (1993) report that the breathy voicing has similar spectral consequences as nasalization, so the two are confusable.

speakers. However, the first aspirated stop might have been interacting with the second aspirated stop in the first syllable of /t̪ʰʊt̪ʰkɑr/. Two speakers were therefore recorded with the second pair of words for aspirated codas in (3.4c). For the breathy coda, the first pair was used for the first two speakers. However, analysis showed that the labial breathy stop assimilated with the following labio-dental fricative. Therefore, the second pair in (3.4d) was used for the other five speakers. Here the closure of the breathy stop coda is nasalized because of the preceding /n/. However, both stressed and unstressed words contained the same consonantal sequence. Therefore, the effect of the nasal was assumed to be the same for both words.

The extra-segmental non-lexical stress effects were controlled by embedding each target word in the carrier phrase, listed in (3.5).

(3.5)

/ t̪um ne _____ kəha /
(you _____ said)
"you said _____"

In this phrase, the main stress fell on the target word. Within the target word, the syllable with lexical stress would attract the phrasal stress, making lexical stress more prominent. Recordings showed that the intonational pattern was also the same for all words.

Schematics of F0 contours produced by speakers and possible intonational phrases[5] (Pierrehumbert 1980) for the phrases with words /pata/ and /patal/ are presented in figure 3.1 below.

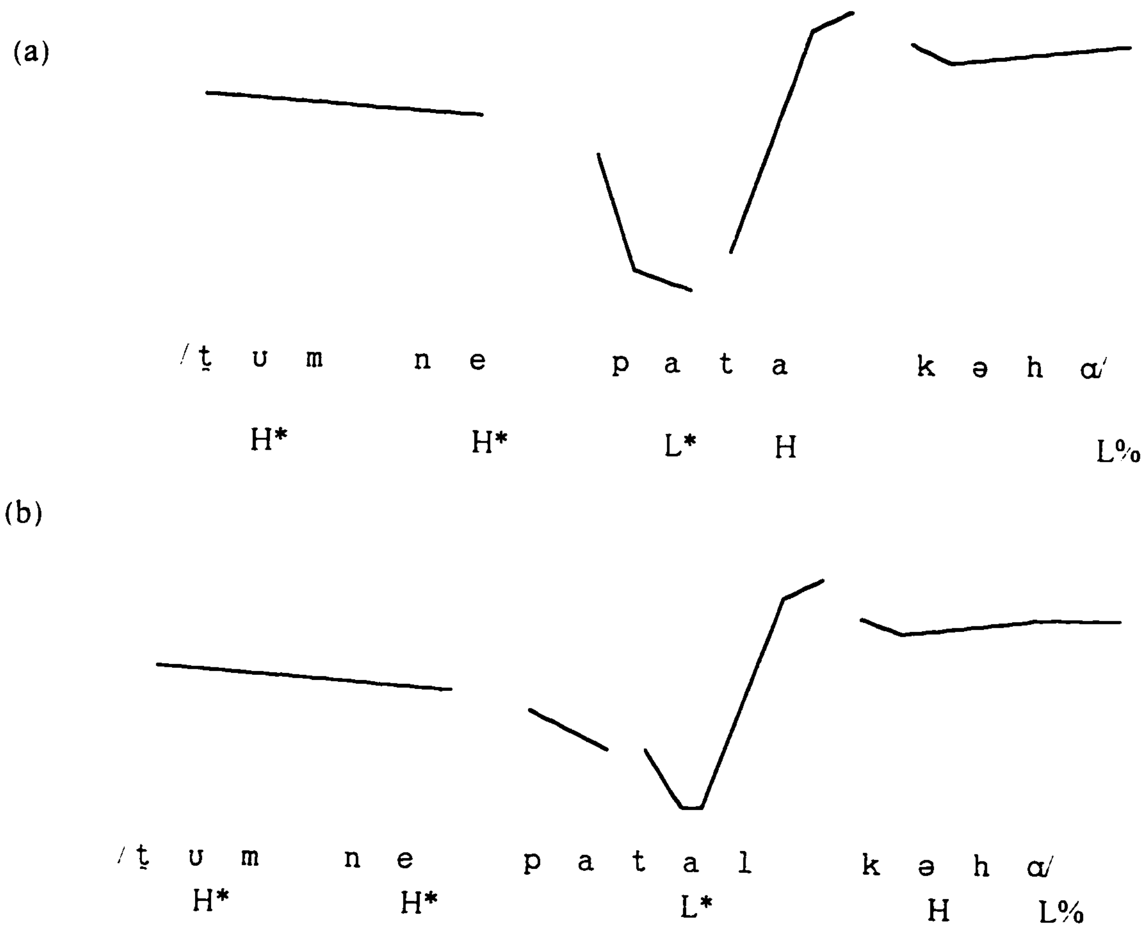


Figure 3-1: Schematic F0 contour on phrases *a.* /t̪um ne pata kəhɑ/ and *b.* /t̪um ne patal kəhɑ/.

[5] This intonational phrase is deduced from the F0 contour and need to be confirmed. Further analysis of Urdu intonation is needed to confirm the intonation pattern proposed here. The accents marked here with an asterisk are *pitch accents*, the tones which align with stressed syllables, the ones without asterisks are *phrase accents*, the tones which mark the edge of a phrase or an intermediate phrase. and the ones with a '%' are *boundary tones*, the tones which mark the end of an intonational phrase (Pierrehumbert and Hirschberg 1990).

Both the initial part (/t̪um ne/) and final part (/kəha/) of the phrase had a high F0 level. F0 fell and then rose during the target words. The minimum was temporally aligned with the stressed syllable (for more discussion, see figure 4.3 and the explanation preceding and following it). The metrical structures for the phrases with the word pairs /pata/ and /patal/ are given in (3.6) (following Pierrehumbert ,1980, as illustrated in Beckman and Edwards, 1990).

(3.6)

```
                *                          *            3, nuclear
  *     *       *           *     *        *            2, accent
  *     *       *     *     *     *        *      *     1, stress
  *     *    *  *   *  *    *     *    *   *    *  *    0, syllable
/ t̪um  ne    pata  kəha/   / t̪um  ne   patal   kəha/
```

The lowest line (line 0) represents a syllable, line 1 represents the stressed syllable in a word. Line 2 represents the syllables associated with a pitch accent and line 3 represents the syllable with nuclear stress. Embedding the words into the phrase also put the target words phrase medially, therefore avoiding strong phrase-final lengthening effects.

SPEAKERS

Recordings were obtained from seven speakers, four males (AR, BS, SH, and ZA) and three females (AA, AS, and SA). The author (SH) was one of the speakers. All subjects were native speakers of Urdu and had spent at least the first eighteen years of their lives in Pakistan, communicating in Urdu during their daily routine. Six of these speakers were from Lahore, which is located in the province of Punjab. Punjabi is also commonly spoken in Lahore and was understood by all the speakers. However, only one of the six speakers had used Punjabi actively, as a mode of communication while in Pakistan. The seventh speaker (speaker BS) was from an Urdu speaking community in Karachi. All these speakers could speak and understand English.

The speakers' ages ranged from 18 to 40 years, with an average age of 28 years. Three speakers (AA, AR and SA) still live in Lahore and were recorded while they were visiting Chicago. The other four speakers had been in the United States for at least six months, and were students at Northwestern University. Three of these speakers (AS, SH and ZA) still speak Urdu in the daily routine; the fourth speaker (BS) also speaks Urdu often in his daily routine, but not as regularly. All the speakers were judged to have normal speech and hearing.

All these speakers, except the author, were naïve about the purpose of the experiment. Two of the speakers (ZA and the author SH) had also participated in a previous experiment on Urdu stops consonants conducted by the author (Hussain 1994).

RECORDING PROCEDURES

All the test words, embedded in the carrier phrase, were written on index cards using the Urdu script (explained in Appendix A). There were five cards made for each test word. These cards were used to make five separate sets of stimuli, each set containing one card for each target word. Each set was shuffled before being recorded by the speakers. Once the five sets were read, the speakers were given a short break while the sets were reshuffled. Then they read the cards again. Therefore, ten repetitions of each word were recorded for each speaker in a block randomized design.

Before recording, the speakers were provided with the list of words to be recorded to familiarize themselves with the words. The speakers were then seated in an IAC booth. They were asked to maintain a fixed distance (of their choice) from the microphone throughout the recordings, which varied between about one and two feet. The speakers were also instructed to read the cards at a natural rate (not too slow or too fast), and to turn to the next card only after reading a card completely (to prevent noise from being generated by sliding cards while a sentence was being spoken). In addition, the speakers were also asked to speak at a comfortable loudness (not too loud and not to soft). All instructions were given in Urdu. The gain of the amplifier was adjusted for each speaker individually to get a strong signal without clipping. The microphone was placed facing them but not directly in front of them, to avoid recording any noise bursts. High quality digitized recording was done directly into a Sun workstation through and Ariel Port, at a sampling rate of 16 kHz.

ANALYSIS

SEGMENTATION

All measurements were made using *xwaves*, a speech analysis program by Entropic. A shell script was written which automatically displayed the time waveform and the spectrogram of each phrase and enabled marking the time waveform at appropriate places, using the *xmark* utility in *xwaves*. These marks were put at different places depending on the material being analyzed. For the words with long and short vowels, eight different marks were placed for each repetition of each word by each speaker. As shown in figure 3.2 below, these marks were placed at onset of the target word (mark 1), onset of the first vowel (mark 2), the middle of the first vowel (placed after half the total number of periods in the vowel, mark 3), offset of the first vowel (mark 4), onset of the second vowel (mark 5), middle of the second vowel (mark 6), offset of the second vowel (mark 7) and offset of the target word (mark 8).

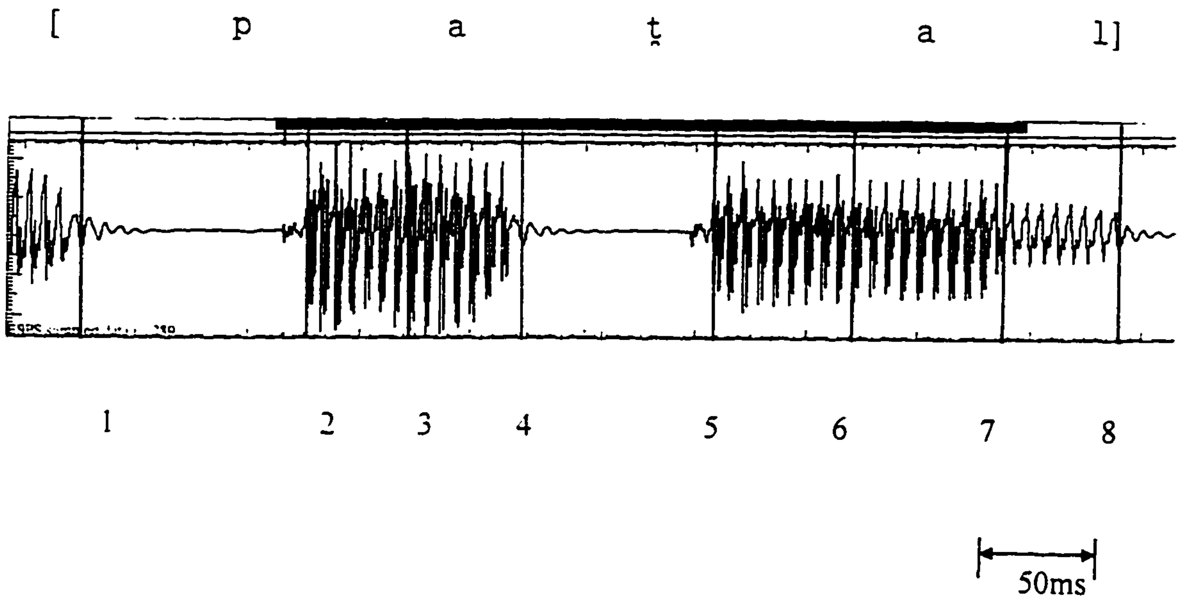


Figure 3-2: Marks placed on the waveform for long and short vowels using the *xmark* utility in *xwaves*. at *1*, word onset, *2*, first vowel onset, *3*, first vowel middle, *4*, first vowel end, *5*, second vowel onset, *6*, second vowel middle, *7*, second vowel end, and *8*, word end (recording of ZA).

Figure 3.3 shows the marks placed for the voiceless and aspirated stop onsets. These marks were placed at the onset of the word (which is the same as the onset of the stop closure, the offset of the second formant of the previous vowel, mark 1), the onset of stop release (mark 2) and the onset of the second formant of the following vowel (mark 3).

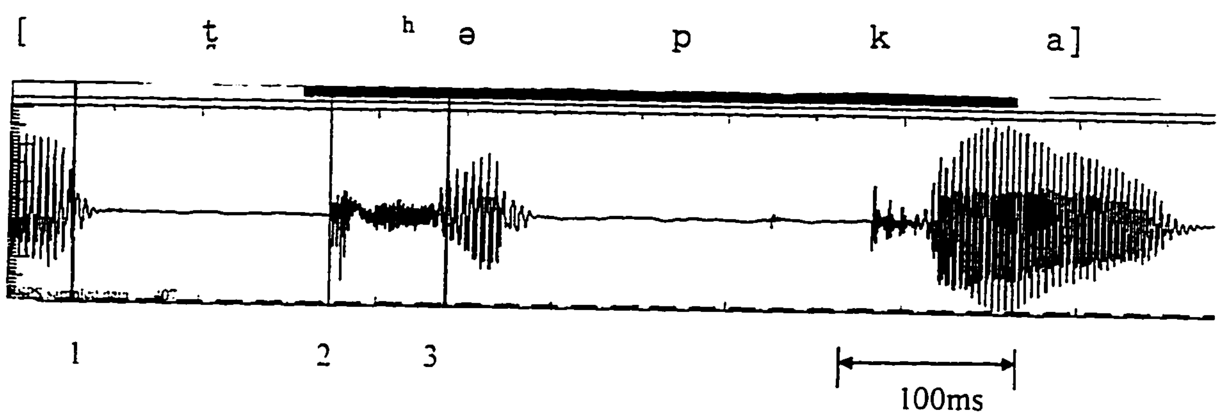


Figure 3-3: Marks placed for voiceless and aspirated stop onsets using the *xmark* utility in *xwaves* at *1*, beginning of closure, *2*. burst, and *3*, onset of F2 of the following vowel (recording of AA).

Figure 3.4 shows the marks placed for the voiced and breathy stop onsets. These marks were placed at the onset of the word (which is the same as the onset of voicing during closure, mark 1), the onset of the stop release (which is the same as the offset of voicing during closure, because the voicing continues through the closure, mark 2) and the onset of the second formant of the following vowel (mark 3).

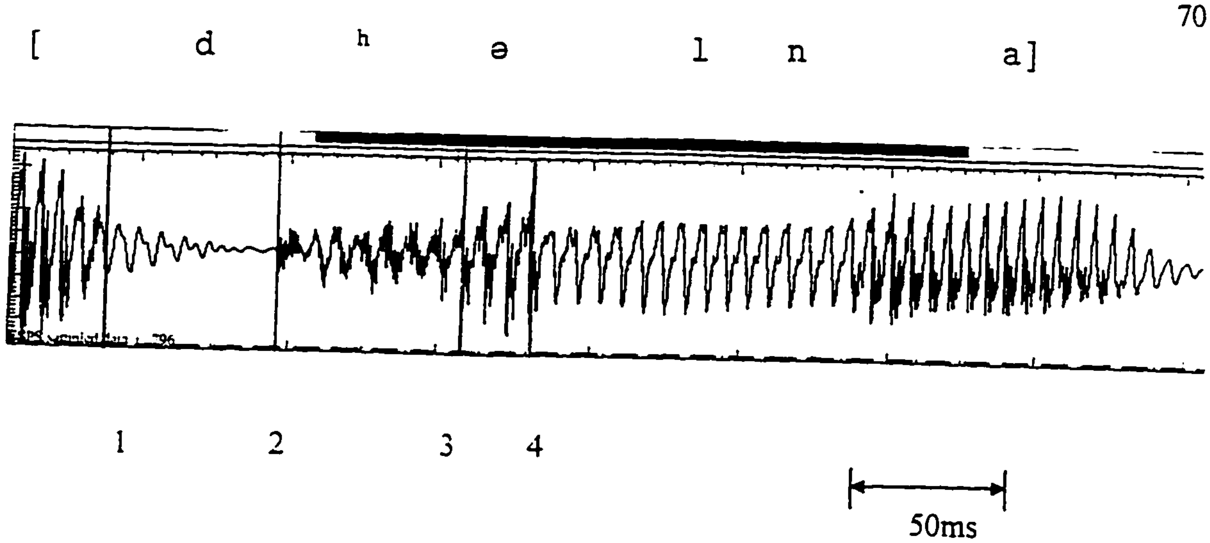


Figure 3-4: Marks placed for voiced and breathy stop onsets using *xmark* utility in *xwaves* at *1*, beginning of closure, *2*, burst, *3*, onset of F2 of the following vowel, and *4*, offset of the following vowel (recordings of ZA)

In addition to these marks, there were also five equally spaced marks placed between the release of the stop (mark 2) and the offset of the first vowel (mark 4). These marks were used to determined the extent of breathiness in the vowel after the release of the breathy stop (explained in more detail later).

For the voiceless and aspirated stop codas, the marks displayed in figure 3.5 were used. These marks were placed at the offset of the second formant of the first vowel (which is the same as the onset of the stop closure, mark 1), the onset of the stop release (mark 2) and the offset of post-release noise or aspiration (mark 3).

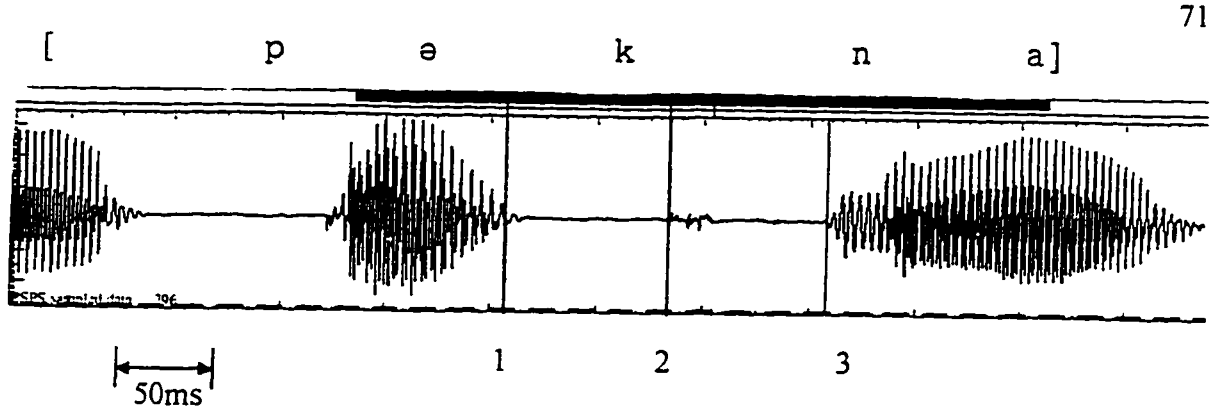


Figure 3-5: Marks placed for voiceless and aspirated stop codas using the *xmark* utility in *xwaves* at *1*, beginning of the closure duration, *2*, burst, *3*, aspiration offset (recording of AS).

For the voiced and breathy stop codas, the marks displayed in figure 3.6 were used. These marks were placed at the offset of the second formant of the first vowel (which is the same as the onset of the stop closure, mark 1), the offset of voicing during closure (mark 2), the stop release (mark 3) and the offset of post-release aspiration (mark 4).

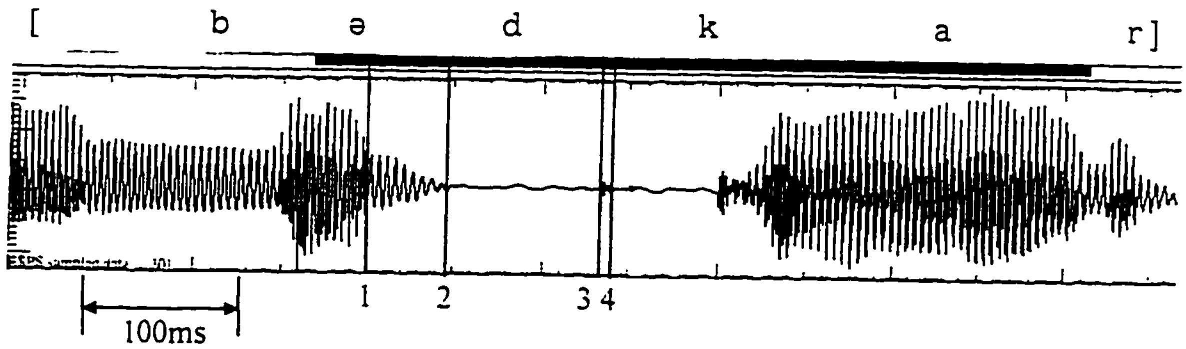


Figure 3-6: Marks placed for voiced and breathy stop codas using the *xmark* utility in *xwaves* at *1*, beginning of closure, *2*, voicing offset, *3*, burst, *4*, aspiration offset (recordings of AA).

MEASUREMENTS

For long and short vowels, durations of the first and second vowels in all words were calculated (by subtracting the time of the mark placed at the offset of each vowel from the time of the mark placed at the onset of each vowel; see figure 3.2). The *formant* utility of *xwaves* was used to calculate the F0 and the formant frequencies in the middle of these vowels (using the mark placed at the center of the vowels). The *pwr* utility of *xwaves* was used to determine the intensity in the middle of both the first and the second vowel in each word. The *pwr* utility was set to calculate the power for a 10 ms window placed at the center of the vowel (the power was calculated by squaring and adding each sample in the window and dividing the sum by the total number of samples). These procedures were automatically done after the marks had been placed, by executing a shell script. Programs in C language were written which took the output of the utilities and converted the data in a tab-delimited text format. These files were then transferred to a IBM compatible PC and analyzed using Microsoft Excel and SPSS, a statistics package.

Closure duration, voicing during closure and duration of aspiration for onset stops were also determined by calculating the difference between the times of appropriate marks (see figures 3.3 and 3.4). Closure duration is the interval between the time of the word onset mark and the stop release mark. The duration of voicing during closure is the time between the onset of voicing during closure and the onset of stop release, because the closures were completely voiced. Aspiration duration is measured by the difference between the time of the onset of following vowel (measured at the onset of its F2, as in

Davis 1994, Hussain and Nair 1995) and the time of onset of the stop burst. Thus, the post-burst frication of the stop is also included with the aspiration duration. The duration of frication at release is considerably shorter than aspiration and, therefore, its inclusion with aspiration does not significantly alter the aspiration duration, however it greatly simplifies the analysis. These values were also converted into tabbed-text format using specific C language programs and transferred to an IBM compatible PC for analysis.

The duration of breathiness of voiced aspirated stops was not easily separable from the following vowel, especially for the unstressed syllables, because instead of distinct breathy and vowel periods, the two are co-produced to give a breathy vowel. Therefore, for these stops, only the closure and voicing during closure durations were measured. However, as described earlier, an additional equally spaced five points were marked between the burst release and offset of the following vowel. Both the amplitude of the fundamental (H0) and first harmonic (H1) were measured at these five equally spaced points between the stop release and the end of the following vowel using the *sgram* utility in *xwaves*. The points were placed at percent intervals of the vowel duration (at a sixth, a third, a half, two-thirds and five-sixth of the duration between the release of the stop and end of the following vowel) rather than at fixed time intervals (e.g. at 10ms, 20ms, 30ms, 40ms and 50ms after the release of the stop). The percentage was preferred over time because vowel durations changed from one repetitions of the word to another, within speakers and among different speakers. Therefore, points placed relative to the vowel duration rather than fixed in time gave a better comparison over different repetitions across speakers. The degree of breathiness along this stop-release vowel sequence was determined by finding the ratio of the first harmonic to the fundamental (= H1/H0 or H1 dB - H0 dB). Gobl (1989, 21) reports that modal voice shows "small attenuation of frequencies below 2KHz. Higher, but still moderate degree of attenuation above 2KHz," but breathy voice shows "high attenuation at frequencies above F0 region." Therefore, in breathy voice, the second and higher harmonics contain lower energy, and hence the vowels have a weaker H1 with respect to H0, i.e. a smaller H1 to H0 ratio indicates increased breathiness. This relative breathiness was measured at each of the five points between release of a stop and the end of the following vowel. However, this analysis was limited to male speakers who have low fundamental frequency. Female speakers have higher fundamental frequency and therefore the frequency of their first harmonic falls in their F1 region. Thus for females, H1 is raised relative to H0 because of vocal tract resonance and does not truly represent the breathiness produced at the glottis.

The closure and voicing during closure durations for coda stops were measured using same techniques as for onsets (see figures 3.5 and 3.6). For post-release aspiration, duration was measured from the stop release until the offset of the aspiration noise. Different methods were tried to determine the latter point, which included setting some base threshold intensity, or using zero-crossing rate. However, none of these methods proved suitable because an appropriate threshold to demarcate the aspiration and the following silence could not be determined. Finally, the offset of noise was marked by hand, with careful inspection of the time waveform of the recordings and their spectrum.

However, this method was not found very reliable because sometimes the aspiration did not diminish monotonically with time.

RELIABILITY OF MEASUREMENTS

To determine the accuracy of measurements, a subset of data was re-measured after a lapse of about two months. Both sets of measurements were done by the author. For the data on stops, the closure durations of breathy stops in the onset position were arbitrarily chosen to be re-measured for male speakers SH and AR (a total of 320 marks were re-marked: 2 speakers x 4 places of stop articulation x 2 stress conditions x 10 repetitions x 2 marks, one at closure onset and one at stop release as in figure 3.4). The original measurements and the re-measurement results are shown in table 3.1.

The re-measured average closure durations for both speakers SH and AR for both stressed and unstressed conditions are within 1 ms of the original measurements and the standard deviations are within 2 ms. As explained earlier, the primary reason why stops were chosen to represent the consonantal inventory of Urdu was that stop and vowel sequences are easier to demarcate acoustically. The closure duration, burst, post-burst aspiration and vowel portions are quite distinct in both the time waveforms and the spectrograms. Therefore, the close agreement of first and second set of measurements is expected.

Table 3-1: Original and second measurement results for closure duration (averages and standard deviations) of breathy stops in the onset position.

Duration								
speaker	AR				SH			
stress	+stress		- stress		+stress		-stress	
measurement #	1	2	1	2	1	2	1	2
b̤^{h}	69	65	55	56	75	74	67	66
d̪̤^{h}	34	32	31	25	84	84	71	71
ḍ̤^{h}	41	39	44	44	66	67	70	69
g^{h}	46	46	40	46	75	75	59	58
average	38	37	34	35	60	60	54	53

Standard Deviation								
speaker	AR				SH			
stress	+stress		- stress		+stress		-stress	
measurement #	1	2	1	2	1	2	1	2
b̤^{h}	11	14	7	9	8	10	13	10
d̪̤^{h}	7	13	19	24	5	5	11	11
ḍ̤^{h}	12	12	15	15	6	7	10	8
g^{h}	8	9	7	10	7	7	9	9
average	8	10	10	12	5	6	9	8

In addition, marks for the vowel /a/ were measured again for the female speaker AS. A total of 160 marks were marked again (8 marks * 10 repetitions * 2 stress conditions * 1 vowel). The duration of each vowel for each repetition was deduced from these measurements. As discussed earlier, the calculation of F0, intensity and the formants were done through utilities provided in the *xwaves* package. Therefore the process did not have any subjective variation, i.e. if the same time marks were used, there would no variation in F0, intensity and formant measurements. However, some variation could occur if different time marks were used. Therefore, F0 was also re-calculated for the vowel measured for AS. The results are presented in table 3.2.

Table 3-2: Original and second measurement results for duration and F0 (averages and standard deviations) of initial and final vowels /a/ in words /pata/ and /patal/ for speaker AS.

measurement	initial V				final V			
stress	+stress		- stress		+stress		-stress	
measurement #	1	2	1	2	1	2	1	2
duration /a/	128	126	121	119	200	198	109	108
F0 /a/	211	211	232	232	214	211	257	262

std. deviation of measurement	initial V				final V			
stress	+stress		- stress		+stress		-stress	
measurement #	1	2	1	2	1	2	1	2
duration /a/	5	4	4	5	4	6	5	5
F0 /a/	4	4	9	8	8	9	15	14

Again the results of second measurements closely correspond with initial measurements. The re-measurement of duration of vowel is within 2 ms of the original measurement (less than 1% of original duration; Hillenbrand et al. 1995 found re-measurements of vowel duration within 6.9 ms reliable for English). The re-measurement of F0 was same for the first syllable vowel and within 5 Hz of the original measurement for the second syllable vowel. The standard deviations of the results are within 2 ms for duration and within 1 Hz for F0. Again, one reason for the close agreement of these measurements and re-measurements is that the data was chosen to include stop vowel sequences which can be demarcated accurately.

RESULTS

VOWELS

Effects of lexical stress on the duration, F0, intensity and quality were investigated for all the short and long vowels (except /ɔ/ as discussed before). Though results are presented together, separated statistical analyses were done for long and short vowels because they are inherently different sub-classes of vowels. The statistical results are considered significant for $p \leq .05$. Significance using this criterion and not the individual p-value of each result is quoted. For individual speaker data, see Appendix B.

DURATION

Vowels in lexically stressed syllables were longer in duration than the same vowels in lexically unstressed syllables. Because the context was carefully controlled, this difference could be attributed to the difference in stress. Figure 4.1 shows the average duration of both long and short vowels for the seven speakers. In the figure, the short vowels /ɪ, ə, ʊ/ are represented by the capital letters 'I', 'E', 'U' respectively. Long vowels are listed before short vowels and the vowels are listed going counter-clockwise in the vowel quadrilateral, starting from /i/. For the vowel /ə/, the bar titled 'unstressed 2' represents the duration in /pəkvan/, the bar titled 'unstressed' represents the duration in /pəkvana/, and the bar titled 'stressed' represents the duration in /pəkna/.

As explained earlier, for the short vowels the pair of words found were bi-syllabic for stressed short vowels and tri-syllabic for unstressed vowels. However, for /ə/ a bi-syllabic word was also found for the unstressed case. This word was included to see whether there were any durational differences due different compression effects between bi- and tri-syllabic words.

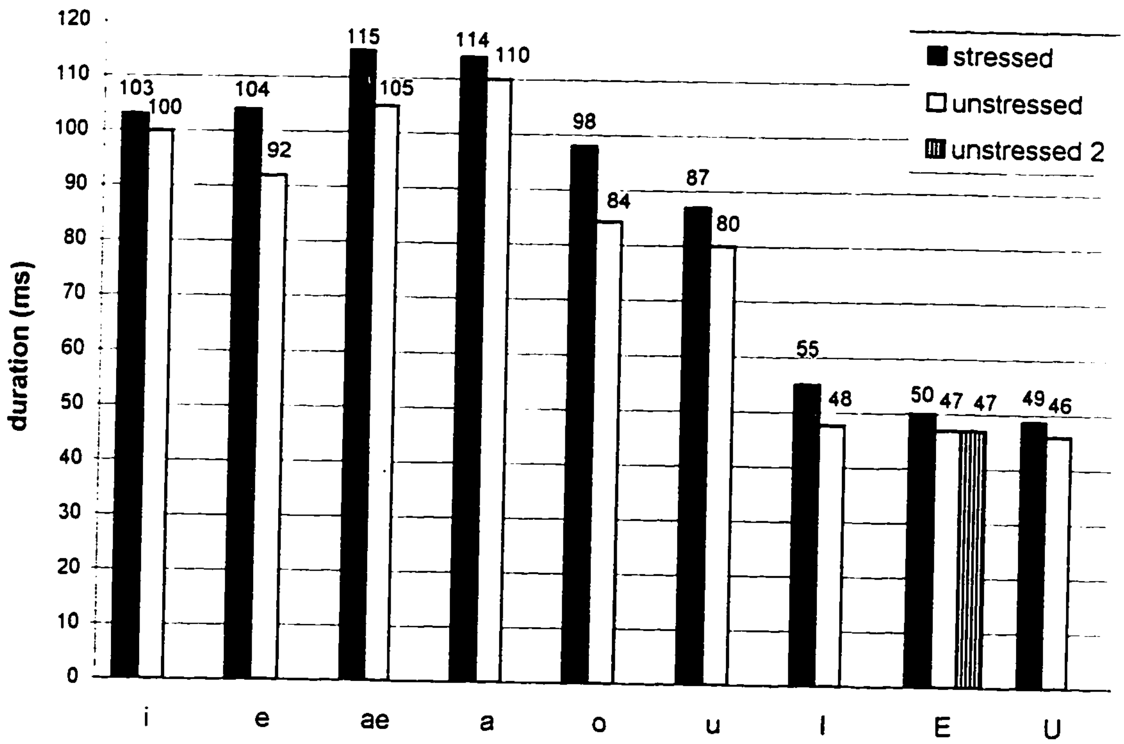


Figure 4-1: Means of duration of the first syllable long and short vowels over all speakers[6] (short vowels /ɪ, ə, ʊ/ are represented by capital 'I', 'E', 'U' respectively).

[6] These averages were for all the seven speakers, except /o/ did not include speaker AA, who mispronounced /koken/ as /kuken/, /æ/ did not include speakers AA and AR, who mispronounced /bæṯal/ as /betal/, and /u/ did not include speakers AA and AR because they were recorded for the pair /guda/ and /gudam/, which was later changed to /suba/ and /ṯumar/.

Figure 4.1 clearly indicates that the long vowels were longer than short vowels. The average durations for individual vowels over all the speakers are marked in the figure. Average duration for all stressed long vowels was 104 ms, for unstressed long vowels it was 95 ms, for stressed short vowels it was 52 ms, and for unstressed short vowels it was 47 ms (also see figure 4.2). Thus, the fact that phonologically the long vowels have twice as many morae as the short vowels surfaced quite faithfully acoustically, because the stressed long vowels and unstressed long vowels were double the duration of stressed short vowels and unstressed short vowels respectively.
The mean difference between stressed and unstressed vowels was 9ms for long vowels and 5 ms for short vowels. Though bi-syllabic words were used for both stressed and unstressed cases for long vowels, for the short vowels the stressed vowels were in bi-syllabic words, e.g. /pəkna/, and unstressed vowels were in tri-syllabic words, e.g. /pəkvana/. There was a possibility that the duration of vowels in bi-syllabic and tri-syllabic words had some differences as a result of variation in compression due to the difference in the number of syllables. A bi-syllabic word for unstressed short vowel /ə/ (/pəkvan/) was also recorded and included to determine the extent of this compression. The results showed that the durations of unstressed /ə/ in /pəkvan/ and /pəkvana/ were the same. Therefore, there was no compression due to the extra syllable in the tri-syllabic words, and the difference in duration between stressed and unstressed vowels

could be attributed to the difference in lexical stress. A possible reason for this lack of vowel compression could be that the vowel was already short and unstressed and was not further compressible. Therefore, the compression factor could be ignored in this case.

Separate paired t-tests for long and short vowels on the mean durations of all subjects showed that the stressed long vowels were significantly longer than the unstressed long vowels ($t = 4.63$, $df = 5$) but stressed short vowels were not significantly longer than the unstressed short vowels ($t = 3.25$, $df = 2$). Still the trend of increased duration with stress was present in the short vowels. The short stressed vowels were consistently longer than unstressed short vowels. The results did not achieve significance because the difference in duration was small and there were only three short vowels. Multivariate Analysis of Variance (MANOVA) tests were also conducted for each speaker for the long and short vowels. Results of the statistical analyses showed that stressed long vowels were significantly longer than unstressed long vowels for all speakers except for speaker AA (results for only three vowels /a,e,i/ were obtained from AA, because the rest of the data was discarded due to mispronunciations). Similar tests for short vowels indicated that the stressed short vowels were significantly longer than unstressed short vowels for all speakers except SA and BS. Individual speaker results and statistical analyses are presented in Appendix B.

The vowel duration and the extent of durational increase caused by stress also varied with the syllable position in the word. Figure 4.2 below shows the durational data obtained for stressed and unstressed vowels, averaged separately over vowels in different syllables. The abscissa is labeled such that the first row represents the syllable position in a word and the second row indicates whether the vowel is long or short. Thus, the bar titled "initial, long V" represents the average duration of initial syllable long vowels in all the words over all speakers. Both stressed ([+str]) and unstressed ([-str]) durations of these vowels are plotted. There were no words recorded with unstressed long vowels in word medial position, therefore, the corresponding bar is missing in figure 4.2.

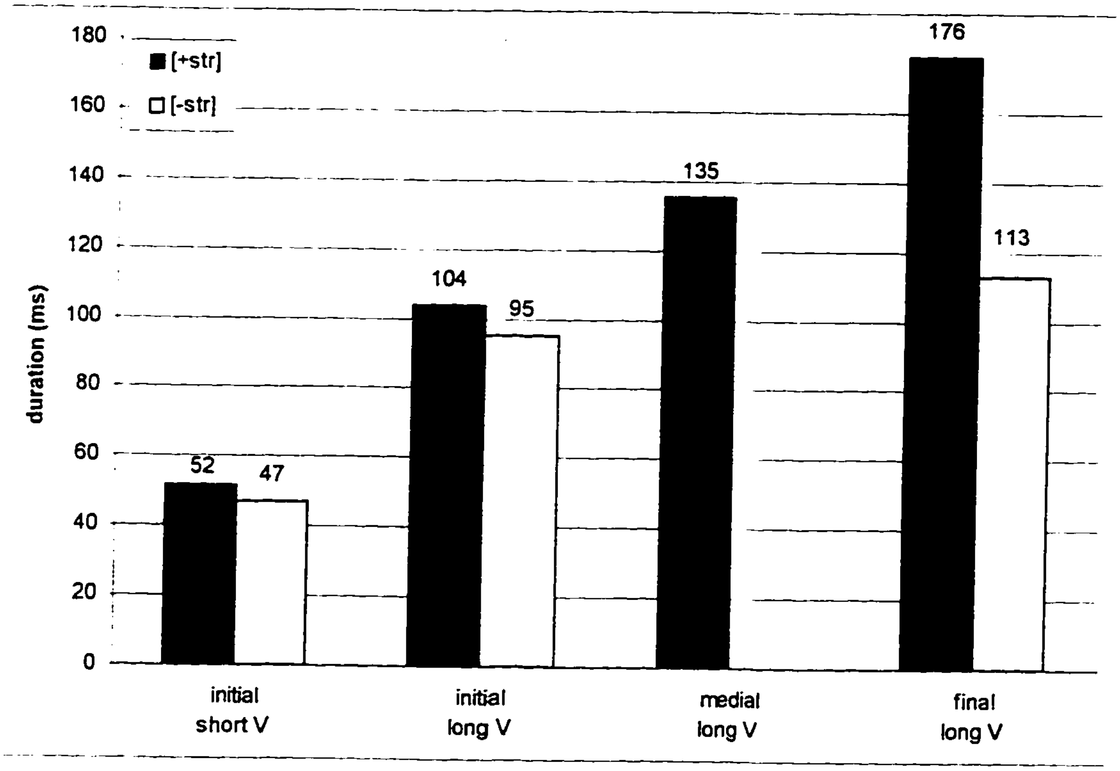


Figure 4-2: Means durations of the initial, medial and final syllable short and long vowels.

Figure 4.2 shows that initial syllable short vowels are shortest in duration. The long vowels are shortest in the initial syllable, longer in the medial syllable and longest in the word-final syllable. Vowels in word-final syllable are longest perhaps for two reasons. First, these vowels undergo lengthening because they are followed by voiced consonants (e.g. Klatt, 1976, for English, Laeufer, 1992, for French) and second, they undergo word-boundary lengthening (Oller, 1973, Klatt, 1976, and Wightman et al., 1992, for English). Though these effects have not been investigated specifically for Urdu, it is assumed that they will also effect Urdu vowels to a certain extent. The is no explanation for why the word medial stressed long vowels are longer than word initial stressed long vowels. However, the different lengthening effect by stress between initial and non-initial vowels has also been reported by other researchers, e.g. Sluijter and van Heuven (1996) for Dutch.

The data plotted in figure 4.2 also indicate that the difference in duration between the stressed and unstressed long vowels is considerably less for word-initial syllables (9 ms) than for word-final syllables (63 ms). As pointed out earlier, this increase may be partly attributable to lengthening before a voiced consonant and before a word boundary. However, these differences are perhaps also amplified by stress. For example, the lengthening before the boundary is perhaps greater for the vowel in a final stressed syllable than for the vowel in final unstressed syllable (also observed in English by Klatt, 1976, and Davis and Summers, 1989).

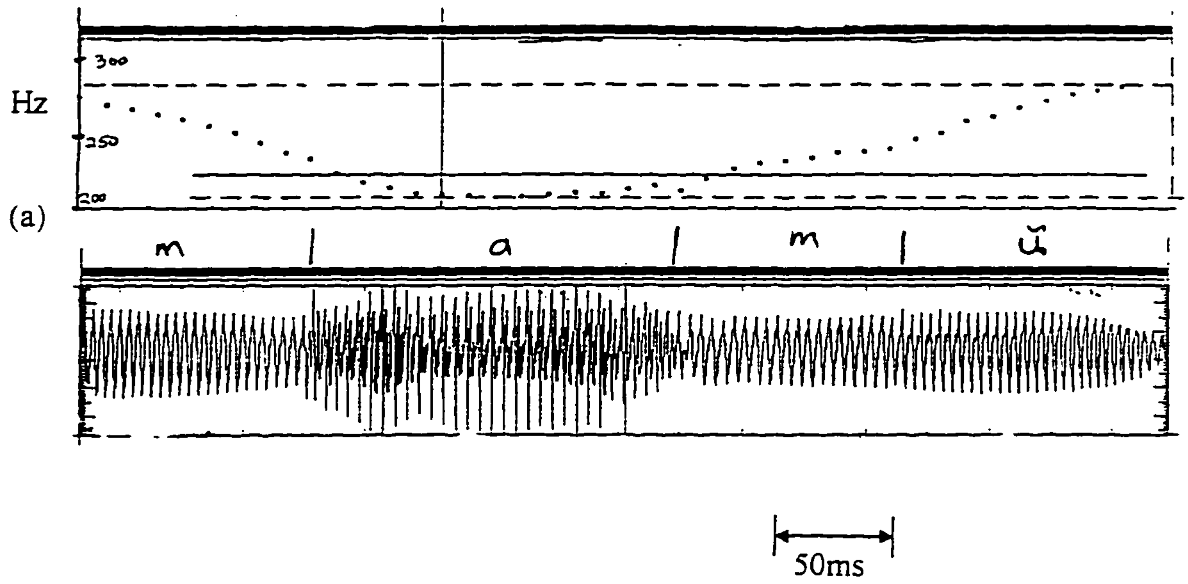
Although the sample sizes were small, it is worth pointing out that the duration and increase in duration with stress was found greater for female speakers than male speakers. The average closure duration for males (N= 4) was 105 ms over all stressed vowels and 84 ms over all unstressed vowels (a difference of 21ms). For females (N= 3) these values were 142 ms and 103 ms respectively (a difference of 39 ms). Thus, the female speakers showed a greater difference in duration caused by lexical stress, with an increase of 37% from unstressed vowel duration for stressed vowels for females and an increase of 22% for males. The average vowel durations also indicated that females had longer vowel durations than male speakers. More speakers need to be recorded to give this generalization concerning gender more validity. However, gender differences in segment durations have also been reported for English by Hillenbrand et al. (1995), who conducted a detailed survey of English which included averages for 45 male and 48 female speakers and by Diehl et al. (1996).

In summary, the duration of stressed vowels was longer than the duration of unstressed vowels by approximately 10%. The durational differences caused by stress were greater for long vowels than short vowels, greatest for word-final vowels, and smallest for word initial vowels. Also, the differences were greater for females than males.

FUNDAMENTAL FREQUENCY

As schematized in figure 3.1, the recorded words in phrases had a falling and then rising F0 contour, i.e. F0 started high at the beginning of the word, went low in the middle of the word and then went high again towards the end of the word. The low in F0 contour coincided with the first syllable for words which had stress on the first syllable and with the second syllable for words which had stress on the second syllable. This suggests that the low F0 value (which has been hypothesized to be caused by the alignment of a low tone) was always aligned with the stressed syllable, a phenomenon observed by other researchers. For example, Hayes (1995, 11) notes that "the rules linking tones to texts refer to the position of stress" (also see (2.1) above for an illustration from English). Figure 4.3 illustrates this different tone alignment found in Urdu. The figure shows the F0 contours for two words, /'ma.mu/ meaning "uncle" and /ma.'mu.li/ meaning "ordinary", as spoken by AS.

The F0 starts high at the beginning of the word. The fall starts early during the /m/ in /mamu/ and is steep to enable an early minimum because of the stress on the first syllable. The fall starts later and is more gradual if the stress is placed on the second syllable in the word, as in /mamuli/. Thus, if F0 is measured at the mid-point of the vowel, its value will be relatively lower for stressed syllable than unstressed syllable.



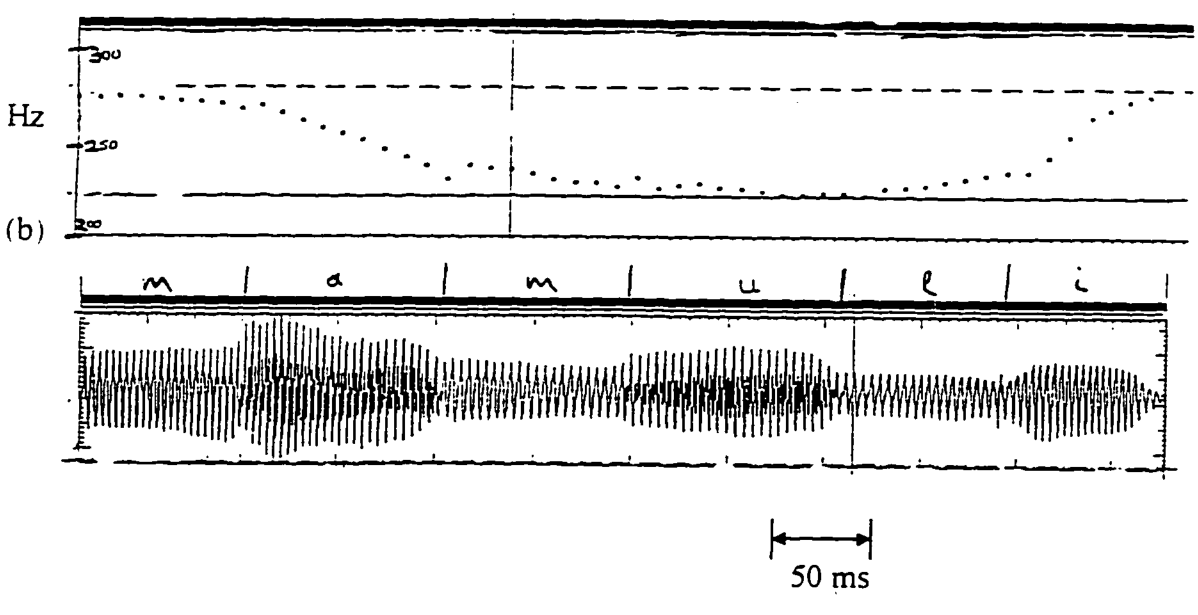


Figure 4-3: F0 contours and time waveforms of words with stress on, *a*, first syllable, and *b*, second syllable

The average F0 values obtained for six speakers for all the long and short vowels in the first syllable of the recorded words are plotted in figure 4.4. Data for male and female speakers is plotted separately because the F0 varies considerably across gender. Therefore, values averaged over both genders would not correctly represent the values of either one. As in figure 4.4, the F0 values were not much different for two or three syllable words (with short vowels in the first syllable; labeled as 'unstressed' and 'unstressed 2' in the figure). Thus, the differences in F0 in short vowels could be attributed to the difference due to tone alignment alone. The data indicated that the stressed vowels had a lower F0 than the unstressed vowels[7].

For the three female speakers, the average F0 value for stressed long vowels was 219 Hz, and for unstressed long vowels was 231 Hz. The average F0 value for stressed short vowels was 219 Hz, and for unstressed short vowels was 235 Hz. Therefore, on average, the stressed long vowels had an F0 value which was 12 Hz lower than the unstressed long vowels and the stressed short vowels had an F0 value which 16 Hz lower than the unstressed short vowels for the female speakers. For the three male speakers, the average F0 value for stressed long vowels was 128 Hz, and for unstressed long vowels was 129 Hz. The average F0 value for stressed short vowels was 137 Hz, and for

[7] BS was the only speaker for whom the stressed vowels had a higher F0 than the unstressed vowels. F0 contour of BS indicates that he may be using a LH*L pattern of tones, but more work is needed to confirm it. As pointed earlier, BS was from Karachi while all the other speakers were from Lahore. Thus, this difference could be attributed to dialectical differences between Urdu spoken in Karachi versus Lahore. The F0 values for BS were not averaged with the other male speakers in the figure 4.4. Also, data on /o/ does not include speaker AA, who mispronounced /koken/ as /kuken/, data on /æ/ does not include speakers AA and AR who mispronounced

unstressed short vowels was 139 Hz. Therefore, on average, the stressed long vowels had an F0 value which was 1 Hz lower than the unstressed long vowels and the stressed short vowels had an F0 value which 2 Hz lower than the unstressed short vowels for the male speakers.

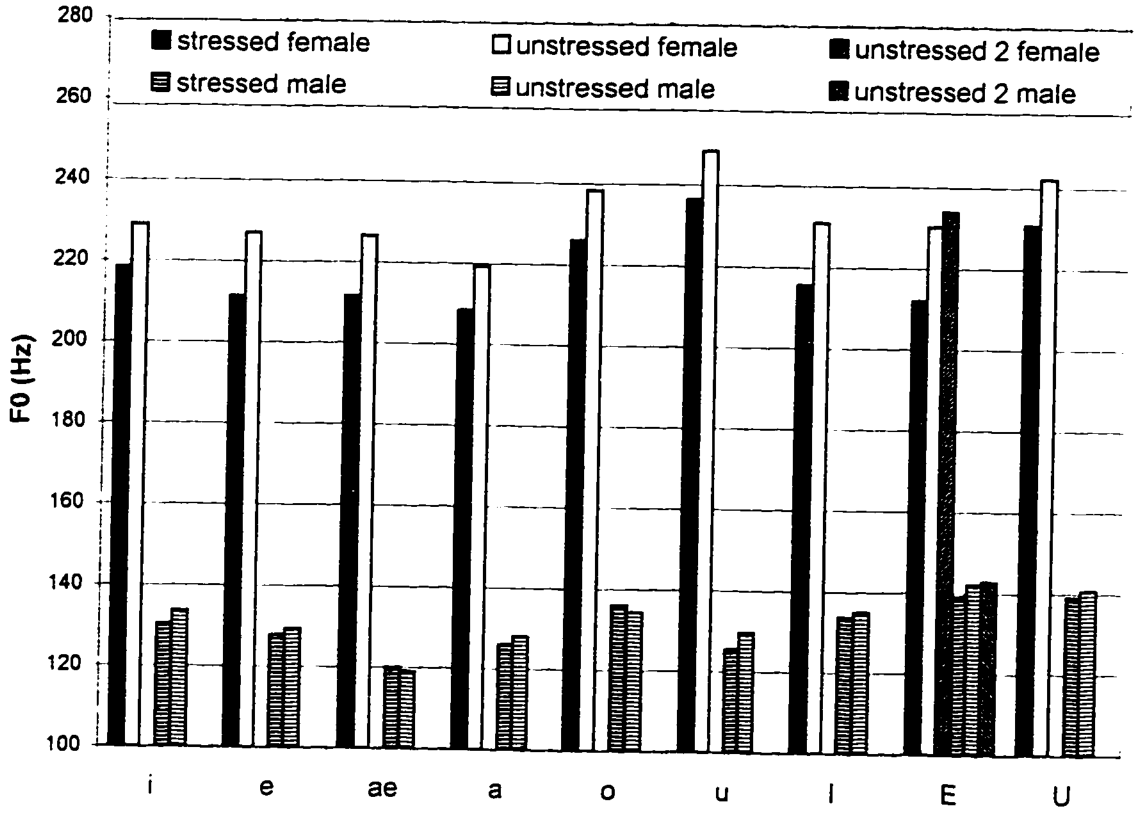


Figure 4-4: Means of F0 of the first syllable long and short vowels over all speakers (short vowels /ɪ, ə, ʊ/ are represented by capital 'I', 'E', 'U' respectively)

/bæt̪al/ as /betal/. and data on /u/ does not include speakers AA and AR because the were recorded for the pair /guda/ and /gudam/ which was later changed to /suba/ and /t̪umɑr/.

T-tests done on these averages show that for female speakers F0 was significantly lower for both the stressed long vowels (t = -15.1, df = 5) and the stressed short vowels (t = -7.23, d = 2). For males, F0 was not significantly lower for the stressed vowels than the unstressed vowels for either the long vowels (t = -1.3, df = 5) or the short vowels (t = -3.4, d = 2).

Results of MANOVAs for long vowels indicated that these differences in F0 between stressed and unstressed vowels were statistically significant for all subjects except SH, who showed little variation in F0. BS had significant differences, but stressed vowels had higher F0 compared with unstressed vowels. Results of MANOVAs for short vowels indicated that the differences in F0 between stressed and unstressed vowels were statistically significant only for the female speakers AS and AA. Results for BS were significant, but stressed vowels had a significantly higher F0 compared with unstressed vowels.

Measurements for F0 were also made for the vowels in the second syllables of each pair of words. Again F0 was lower for stressed syllables compared to unstressed syllables (because for unstressed second syllables, F0 would already be rising to attain the word final high tone, as shown in figure 4.3). The results are plotted in figure 4.5. The abscissa is labeled to show the gender and the type of vowel in the initial syllable of the word being measured. Though the initial syllable vowels were either long or short, the second syllable vowels were always long. The stressed second syllable vowels were word-medial when the word-initial syllables were short (e.g. /bɪk.na, bɪk.va.na/) and word-final when the word-initial syllables were long e.g. /pa.ta, pa.tal/.

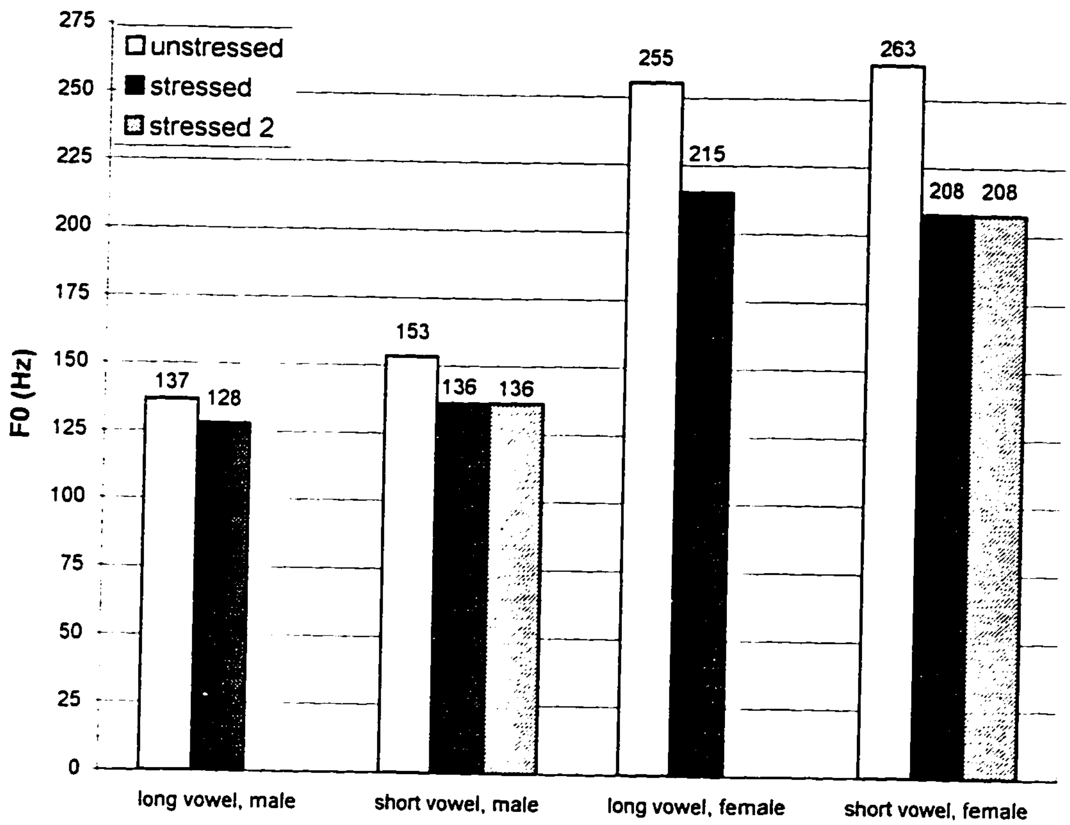


Figure 4-5: Means of F0 of /a/ in the second syllable over all speakers[6], with long vowels or short vowels in the first syllable.

Again, for the words with short vowel in the first syllable, both tri-syllabic and bisyllabic words were used. For these words, the 'unstressed' bar refers to the F0 of /a/ in words similar to /pəkna/, the 'stressed' bar refers to the F0 of first /a/ in words similar to /pəkvana/ and the 'stressed 2' bar refers to the F0 of /a/ in /pəkvan/. Figure 4.5 shows that F0 is lower for stressed vowels than unstressed vowels. Again, the two and three syllable words (labeled as 'stressed 2' and 'stressed' in the figure above) did not show any compression differences and therefore all the differences could be attributed to stress alone. The F0 for long stressed vowels was lowered by 9 Hz (7%) word-medially and 17 Hz (11%) word-finally for males and 40 Hz (15%) word-medially and 55 Hz (21%) word-finally for females. These differences are considerably greater than the differences in F0 caused by stress in the first syllable long vowels (about 1Hz for males and about 16 Hz for females).

MANOVA tests on F0 of second vowels for individual speakers show that F0 is significantly different for all speakers. All follow-up univariate tests are also statistically significant for all vowels for all speakers except for SH, who had statistically significant results only for words with first vowels /æ. ɪ. ʊ/ (details in Appendix B).

In summary, F0 was lowered to indicate stress by all speakers from Lahore. However, F0 was raised to indicate stress by the speaker from Karachi. These differences between speakers might have been caused by dialectical differences in Urdu (Urdu spoken in Karachi also 'sounds' different than the Urdu spoken in Lahore). The differences in F0 due to stress were greater for second syllable vowels than first syllable vowels. In addition, the differences in F0 due to stress were smaller for males than females.

INTENSITY

As explained earlier, though the speakers were instructed to maintain a comfortable loudness and a comfortable distance from the microphone during the recordings, no stricter measure was adopted to control the intensity variations. A ratio of the intensity of the first vowel to the intensity of the second vowel in each target word was calculated. For example, for the word /beta/, the ratio of the intensity of /e/ to the intensity of /a/ is calculated. If intensity of the stressed syllables is greater than the intensity of the unstressed syllables then this ratio will be greater for the words with a stressed initial syllable than for the words with a stressed second syllable. These ratios are plotted in figure 4.6 for all the word pairs with different initial vowels.

The data shows that the ratios of the intensities of the first vowel to the second vowel increased with stress (i.e. the filled bars were greater than white bars in figure 4.6) for only three stressed low vowels /æ, a, ə/. For the non-low vowels, the ratios were greater for unstressed vowels. This distribution of intensity indicates that perhaps low vowels become lower with stress, opening up and reducing the acoustic impedance of vocal tract, and therefore increasing in intensity. And perhaps high vowels become higher with stress, closing up and increasing the acoustic impedance of vocal tract, and therefore decreasing the output intensity. This is predicted by the hyperarticulation model proposed by Lindblom (1990), de Jong (1995) and de Jong et al. (1993). T-tests showed that these differences, for low or non-low vowels, were not statistically significant ($t = -2.39$, $df = 5$, for non-low vowels, and $t = 2.6$, $df = 2$ for low vowels).

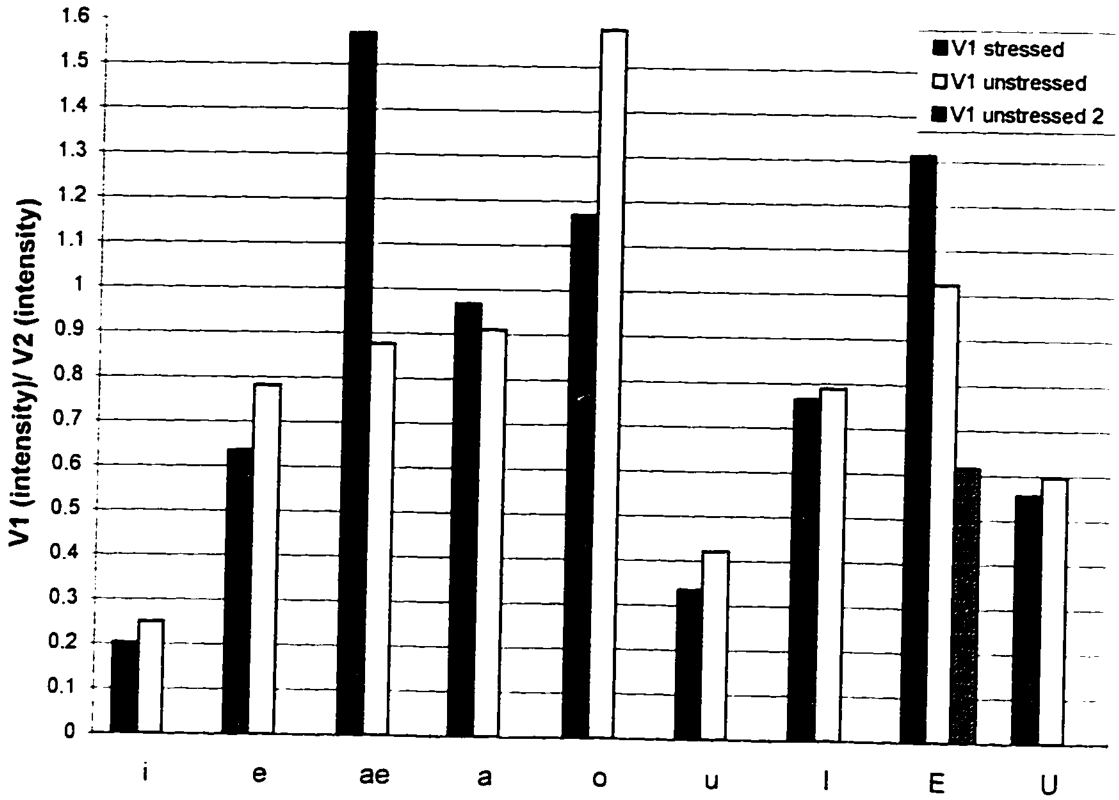


Figure 4-6: Means of ratios of the intensity of initial vowels (V1) to the intensity of second vowels (V2) over all speakers[5], with long vowels or short vowels in the first syllable (short vowels /ɪ, ə, ʊ/ are represented by capital 'I', 'E', 'U' respectively).

Mixed results were obtained on the intensity of vowels when individual speaker data is considered. While the ratio of intensity of the first vowel to the second vowel increases with stress for some vowels for some speakers, the ratio decreases for other vowels for other speakers. For example, individual speaker data shows that for /i/ intensity increases with increased stress for one speaker and decreases for six speakers; for /e/ it increases for two and decreases for four speakers; for /ae/ it increases for all the five speakers; for /a/ it increases for three speakers and decreases for four speakers; for

/o/ it increases for one speaker and decreases for five speakers; for /u/ it increases for two speakers and decreases for three speakers. These results indicate that stress does not necessarily increase the intensity of vowels.

Though female speakers show greater changes in duration and F0 with stress, the data averaged by gender shows that females have a smaller change in the intensity of vowels than males speakers. Female speakers articulate the first vowel more loudly (compared to the second vowel) than males, but the males articulate the change in intensity with stress more prominently.

QUALITY

Many researchers have proposed that stress changes vowel quality, i.e. stressed vowels have more canonical articulation and formant frequencies because they resist coarticulation with adjacent context (Gay 1978, de Jong, Beckman and Edwards 1993, van Bergem 1993). On the other hand, unstressed vowels are more susceptible to coarticulation and therefore their formant frequencies undergo more changes. Mid-vowel formant frequencies were measured to find whether this quality difference also occurred in Urdu as a function of stress. Means of first two formants are plotted in figure 4.7 for the long and short stressed and unstressed vowels for all subjects. Standard deviations for the corner vowels /i, æ, a, u/ are also marked on this figure to display the extent of variation (the standard deviations for all vowels were not marked to avoid cluttering the figure). Though the average values differ for male and female speakers (also reported by Hillenbrand 1995), they were averaged together in this plot because formant ranges for male and female speakers were found to have a considerable overlap. Detailed speaker and vowel data is listed in Appendix B.

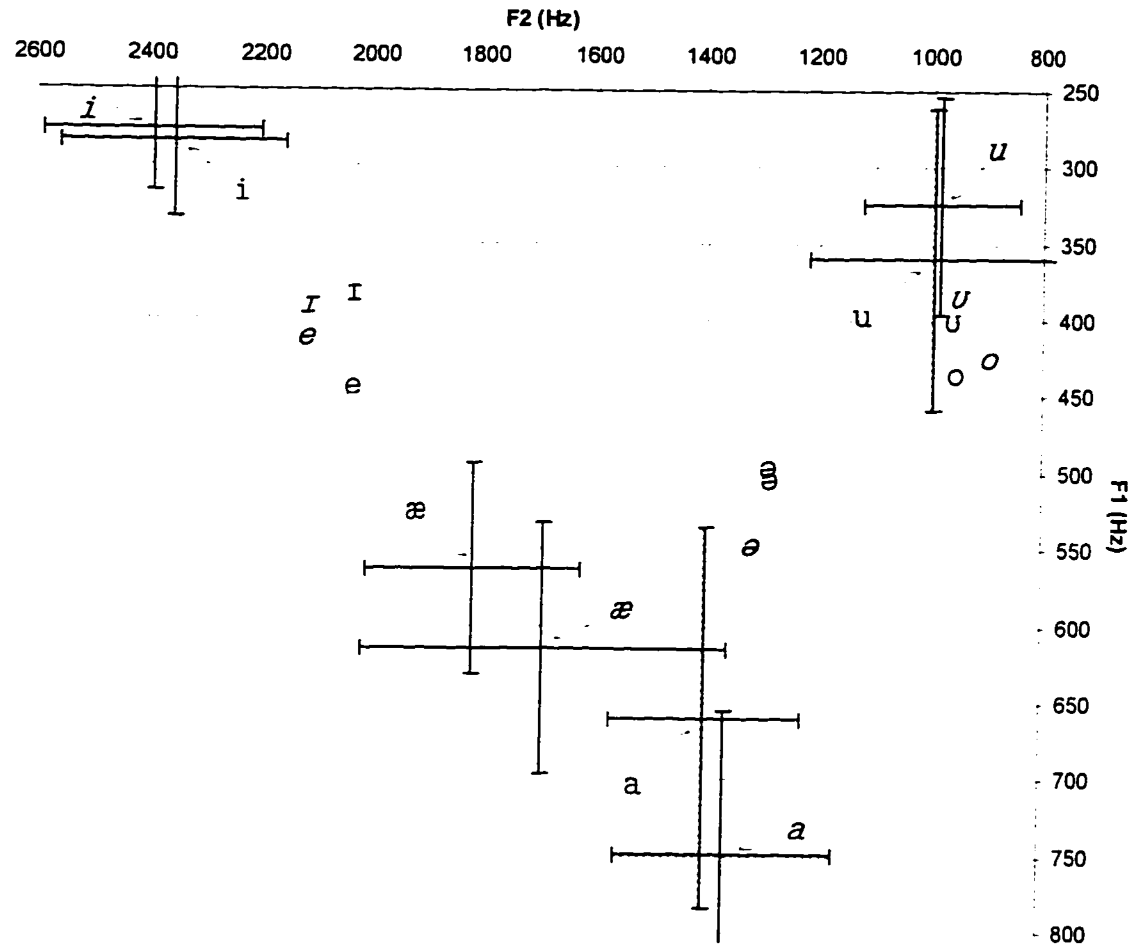


Figure 4-7: F1-F2 plot for all vowels in stressed (bold and italicized text) and unstressed (plain text) syllables, averaged over all speakers[5].

In figure 4.7, the stressed vowel averages are represented with bold and italicized symbols of the vowels, and unstressed vowel averages are represented by plain text symbols. The scales for the first formant (F1) and the second formant (F2) are arranged in such a way that the vertical scale roughly corresponds to the tongue's height and the horizontal scale roughly corresponds to the tongue's horizontal position (tongue back towards the right of the figure). Using this articulatory/acoustic scale, some observations about the vowels in Urdu can be made. The vowel /a/ is lowest, and /i/ is highest in Urdu. The vowel /o/ is most back, and /i/ is most front. There is a void space between /a/ and /o/ where perhaps the vowel /ɔ/ is articulated (due to the limited use of this vowel, it was not included in the current analyses). The short vowel /ɪ/ is between /i/ and /e/, the short vowel /ʊ/ is between /u/ and /o/ and the short vowel /ə/ is central above /a/, perhaps between /æ/ and where /ɔ/ would have been.

Also, the short vowels occupy a reduced phonetic space enclosed by the space occupied by the long vowels. Thus, phonologically short vowels are not only reduced in duration but also in quality compared to the phonologically long vowels. The phonetic space of phonologically short vowels is reduced compared to the phonetic space of long vowels perhaps because the number of vowel distinctions is reduced from five (phonologically long vowels) to three (phonologically short vowels). The fact that the phonetic space depends on the phonological contrast has been noted by Lindau and Ladefoged (1986). As pointed out in an earlier section, Lindau et al. report that the dispersion of the vowel formants depends (among other factors) on the number of vowels. Languages with more vowels tend to have vowel articulations which are more extreme from the central position and languages with fewer vowels tend to articulate them more centrally. In Urdu, two phonologically distinct vowel categories within the same language adopt the same behavior. Still, the spectral cues might not be as effective as durational cues in determining the difference between the phonologically long and short vowels (as reported for Thai by Abramson and Ren 1990).

Vowel quality changed with stress in Urdu. Similar formant values for /ə/ in the unstressed syllable for both tri-syllabic and bi-syllabic words (represented by plain text schwas in figure 4.7) also shows that the number of syllables did not affect vowel quality. Therefore, the differences for all short vowels could be attributed to stress alone. Distances (in Hz) in F1-F2 plane were calculated between stressed and unstressed vowels for a quantitative comparison of change in vowel quality with stress. The following formula was used to calculated these distances:

$$\text{distance} = \sqrt{[\,(F1_{stressed} - F1_{unstressed})^2 + (F2_{stressed} - F2_{unstressed})^2\,]}\;)$$

This formula calculates the shortest distance in Hertz between stressed and unstressed vowels in the F1-F2 acoustic space. It should be noted that the perceptual distance is perhaps different than this distance in Hertz. However, this distance still provides an objective measure to compare stress effects on different vowels. The "direction" of the change in quality is different for each vowel and can be determined from figure 4.7. These distances are tabulated in table 4.1 below.

Table 4-1: Average distance between stressed and unstressed vowels for all speakers.

vowel	distance (Hz)
i	36
e	107
æ	99
a	171
o	110
u	80
ɪ	70
ə	51
ʊ	41

On average, the distances between stressed and unstressed low vowels were greater than the distances between stressed and unstressed high vowels. Also, on average, the distances between stressed and unstressed long vowels were greater than the distances between stressed and unstressed short vowels. Therefore, the quality of the long, low vowels was most affected by stress. The quality was least affected for high and short vowels. The overall average distance between stressed and unstressed vowel was 85 Hz.

MANOVAs performed for each subject showed that the vowel-by-stress effect for long vowels was significant for all speakers except AR, and the effect for short vowels was significant for SH, ZA and AS. Results of follow-up univariate tests for each speaker for each vowel are listed in Appendix B. These results show that stress had a significant effect on vowel quality.

How the quality of vowels changed with stress was also investigated. Lindblom (1963) hypothesized that vowels tended to be articulated more centrally when unstressed (for English, the first two formants of the central vowel are at 500 Hz and 1500 Hz). Therefore, distances of stressed and unstressed vowels from this central vowel were also calculated. If the vowels were indeed centralized when unstressed, the distance between unstressed vowels and the central vowel should be smaller than the distance between the stressed vowels and the central vowel. Distances for each stressed and unstressed vowel from the central vowel are listed in table 4.2.

Table 4-2: Average distances between a 'central' vowel and stressed or unstressed vowels.

vowel	dist(stressed) (Hz)	dist(unstressed) (Hz)
i	923	885
e	625	542
æ	219	320
a	292	191
o	611	542
u	545	532
ɪ	625	541
ə	190	204
ʊ	551	539

The table shows that the unstressed vowels are not necessarily closer to the central vowel. The vowels /æ,ə/ are further from the central vowel when unstressed. Therefore, the vowels are not necessarily centralized when unstressed. One possible explanation for why the unstressed vowels did not centralize is provided by van Bergem (1993). He points out that "spectral vowel reduction could be better interpreted as the result of an increased contextual assimilation than as the tendency to centralize." Thus, the vowel formant structure change is also dictated by the adjacent stops to some extent.

Stevens and Blumstein (1978, 1359) note that "depending of the position ... of the constriction [of the adjacent consonants], the second and higher formants [of vowels]... undergo displacements." Thus, the consonantal context influences the second and higher formants of the unstressed vowel, the first formant normally shifting towards the first formant value of the central vowel (about 500 Hz). From the research on acoustics of stops (e.g. Fant 1960, 186-188, Stevens and Blumstein 1978) it can be deduced that the second formant will shift towards 1800 Hz in the context of dental and alveolar stops and towards 1200 Hz in the context of velars. The labials do not involve tongue articulation and therefore do not interfere with vowel articulation. However, the second formant is lowered by lip rounding in the context of labials. Both the onset and coda stops would influence the vowel.

Figure 4.7 shows that, for the unstressed vowels, the first formant value shifts towards the first formant value of a central vowel (about 500 Hz). Vowels which have lower F1 increase in F1 when unstressed, and vowels which have a higher F1 decrease in F1 when unstressed. The only exception is /ɪ/, for which F1 shifts away (by 9 Hz) from the central F1 position. Looking closely at the individual speaker data for /ɪ/, the F1 value decreases for AA and BS and increases for the other speakers. The results are therefore mixed, with not much change in quality overall.

The changes in F2 of different vowels with stress need more explanation because they depend on the consonantal context. In the stimuli (3.1 and 3.2), the vowel /i/ is preceded and followed by dental consonants, the vowels /e, æ, a, u/ are between a labial and a dental or alveolar consonant, the vowel /o/ is between velars, and all the short vowels are between a labial and a velar stop. As pointed out earlier, labials do not have much effect on a vowel target, except perhaps a little lowering cause by lip rounding. Dentals and alveolars would tend to pull the F2 target for the vowel towards about 1800 Hz (slightly higher for dentals than alveolars). This effect is seen in figure 4.7 for the vowels /i,e, æ, a, u /, which are in a dental or alveolar context. For /i,e/ F2 decreases towards 1800 Hz. For /æ, a, u/ an increase in F2 towards 1800 Hz is observed. This change is greater for the front vowels than the back vowel because for the front vowels, dental and alveolar consonants are all articulated with the front of the tongue resulting in more coarticulatory effects. The back vowels are articulated with back of the tongue, which is more independent of the front of the tongue and therefore there is less coarticulation.

The short vowels and the vowel /o/ are in a velar context. Therefore, F2 for these vowels will move towards 1200 Hz in unstressed syllables. This is observed for /o/ and the short vowels /ɪ,ə/. There is a very slight change for /ʊ/, perhaps because this vowel already has an F2 value close to 1200 Hz. It should be noted that the 1200 Hz value being quoted is for English, and may be slightly lower for Urdu, depending on where the velar constriction occurs for Urdu. Also, /o/ has a lower F2 from /u/ because the former is in velar context and therefore pulled back, while the latter is in labial and dental context, and therefore perhaps pulled towards the front. These observations do indicate that the vowels do not get more central when unstressed but undergo more assimilation with adjacent consonants.

The quality of vowels and the extent of change in quality of vowels with stress also differed by gender. The vowels were shifted towards higher F1 and F2 values for female speakers compared to male speakers (similar differences have been reported for English by Hillenbrand et al. 1995 and for Dutch by Sluijter and van Heuven 1996). On average, the female speakers showed greater effects of stress on quality than male speakers. Distances in Hertz were calculated between the stressed and unstressed vowels for both males and females. Distances between stressed and unstressed vowels, averaged over all vowels, were 85 Hz for females and 71 Hz for males. Detailed results for each speaker are listed in Appendix B.

SUMMARY OF VOWEL DATA

In summary, the data collected on vowels showed that there are statistically significant effects of stress on the duration of first and second syllable vowels. The duration of stressed vowels was greater than unstressed vowels especially for long, non-initial-syllable vowels. The durational differences were greatest between stressed and unstressed word-final syllables, perhaps because the lengthening-before-voiced-consonant effect and the word-final lengthening effect were also amplified by stress. The F0 of vowels was lower for stressed syllables than unstressed syllables (for six speakers), especially for second syllable vowels, due to the alignment of the low tone with the stressed syllable. F0 was higher for one speaker due to the alignment of high tone. The data obtained for vowel intensities was mixed. The averages of ratios of the first vowel intensity to the second vowel intensity were greater for stressed than unstressed vowels for low vowels and smaller for stressed than unstressed vowels for non-low/high vowels. This indicates that low vowels become more open and high vowels become more closed with stress. However, individual speaker data showed a mixed pattern. Finally, the quality of vowels also changed with stress, unstressed vowels assimilating more with adjacent context.

The data also showed that females had greater duration, F0, intensity and a different quality of vowels than males. In addition, females also exhibited a greater change in duration, F0 and quality of vowels with stress. Males exhibited a greater change in intensity with stress. However, there was a very small pool of male and female speakers (four males and three males); more male and female speakers need to be recorded to verify these results.

ONSET STOPS

CLOSURE DURATION

Means of closure durations for onset stops, obtained by averaging data over all speakers, is plotted in figure 4.8. Separate averages were calculated for all the different stops recorded. In the figure, the types of the stops are arranged in the following order: voiceless, voiced, aspirated, and breathy voiced. Within each category, the stops are listed according to their place of articulation, going from the front toward the back of the vocal tract. Also, dental stops /t̪, d̪/ are represented by 't'and 'd' respectively, alveolar stops /t,d/ are represented by 'T'and 'D' respectively and aspirated stops are represented by 'Ch' (where 'C' is a stop not specified for place and 'h' is aspiration).

The closure durations of all stressed stops in onset position were longer than the closure durations of unstressed stops in the onset position. The average closure duration was 86 ms for the stressed onset stops and 75 ms for unstressed onset stops. Therefore, with stress the onset duration increased by 15% over the unstressed duration (11 ms). Averages over different places of articulation show that closure durations were ranked as follows: labials > dentals > velars > alveolars (95ms, 86ms, 80ms and 76ms respectively for stressed syllables, and 84ms, 76ms, 67ms 65ms respectively for unstressed syllables). /k/ was not recorded in a similar vowel context as the other voiceless consonants and therefore some variation of the above hierarchy might be possible. Similar results are summarized by Laver (1994) for Danish (Fischer-Jorgensen, 1964) and Brenton (Falc'hun

1951), both reporting labials stops longer than velars and alveolars, but the relationship between alveolars and velars is not consistent and varies with context.

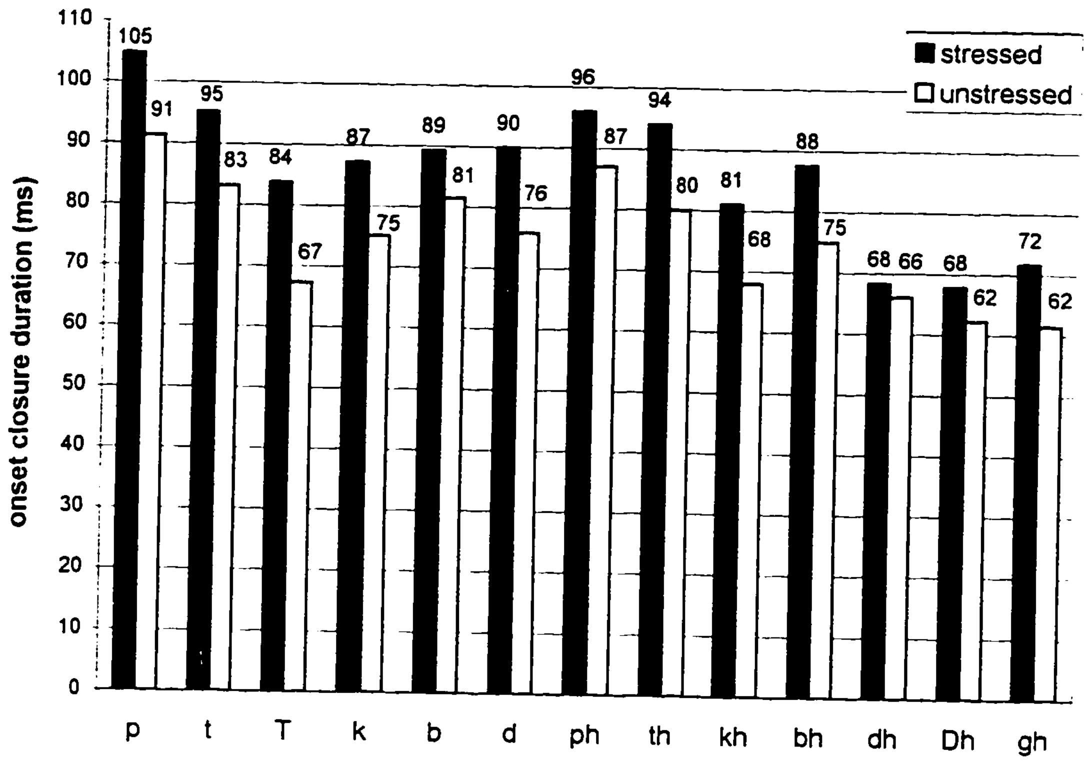


Figure 4-8: Closure durations of onset stops in stressed and unstressed syllables, averaged over all speakers.

Because stops at all the four places of articulation were not recorded for all the four stop types, average closure durations for only labial and dental stops were calculated (stops at these two places of articulation were recorded for all four stop types). The averages showed the following hierarchy of closure durations: voiceless > aspirated > voiced > breathy (100ms, 95ms, 90ms and 78ms respectively for stressed syllables, and

87ms, 83ms, 79ms and 71ms respectively for unstressed syllables). Thus, the stops which were not phonologically voiced (voiceless and aspirated stops) had longer closure durations than the stops which were phonologically voiced (voiced and breathy stops). Also, the stops which were phonologically unaspirated (voiceless and voiced stops) had longer closure durations than the stops which were phonologically aspirated (aspirated and breathy stops).

A paired t-test was done to compare the stressed and unstressed closure durations of the onset stops. Results showed that closure durations for the stops in stressed syllables were significantly longer than closure durations for the stops in unstressed syllables ($t = 10.0$, $df = 12$). MANOVA results show that the closure durations of stressed stop onsets were significantly longer than the closure durations of unstressed stop onsets for all seven subjects. Follow-up univariate tests for individual stops for all vowels and all subjects are listed in Appendix B.

Separate averages for male and female speakers were also calculated to investigate whether gender also had an effect on the extent to which the closure durations changed with stress. The average closure duration for females was longer than for males. The closure duration for males was 74 ms for stressed syllables and 68 ms for unstressed syllables. For females, these averages were 103 ms and 85 ms respectively. The relative differences were also greater for females than males. Closure duration increased by 9% with stress for males and 21% for females.

The results showed that the duration of closure of onset stops varied with both the place and the type of articulation of stops. The closure duration increased with stress and increased more for females than males.

VOICING DURATION

Both voiced and breathy onset stops had completely voiced closures in stressed and unstressed positions. Therefore, in figure 4.8 the closure durations for these stops also represent the voicing duration during closure. The voiceless and aspirated stops did not have any voicing during closure. The stop closures were completely voiced by all speakers perhaps because the stops occurred inter-vocalically and because voicing during closure is the primary cue for the feature [voice] in Urdu (Hussain 1994).

Figure 4.8 shows that voicing during closure was greater for stressed stops than unstressed stops. The average duration of voicing is 79 ms for [+voiced] stressed stops and 70 ms for [+voiced] unstressed stops. The voiced stops had longer voicing duration than breathy voiced stops (90 ms and 74 ms respectively for stressed onsets and 79 ms and 66 ms respectively for unstressed onset stops). In addition, the voicing during closure was longer for labials than other places, dental, alveolar and velars being similar in duration.

A paired t-test showed that the voicing duration of stressed voiced and breathy stops was significantly greater than the voicing duration of unstressed stops ($t = 4.82$, $df = 5$). MANOVAs done for individual speaker data showed that voicing duration was significantly longer for stressed syllables than unstressed syllables for all speakers except AR. Follow-up univariate tests for all speakers for all stops are listed in Appendix B.

Also, female speakers have a greater duration and change in voicing duration with stress than the males speakers. The voicing duration for females increases from 79 ms to 93 ms with stress (14 ms, 17% increase). And for males, it increases from 64 ms to 68 ms with stress (4 ms, 6% increase).

POST-RELEASE ASPIRATION DURATION

Values for aspiration durations after the release of the stops and before the onset of the following vowel were also calculated. The average values over all speakers were plotted in figure 4.9. Separate averages were calculated for all the different stops recorded. The breathy stops were not included because, as discussed earlier, there were no discrete boundaries between breathy and vocalic parts of the syllable for some speakers, especially in the unstressed cases. The two parts overlapped to give a breathy vowel. The dental stops /t̪ , d̪ / are represented by 't'and 'd' respectively, alveolar stops /t,d/ are represented by 'T'and 'D' respectively and aspirated stops are represented by 'Ch' (where 'C' is a stop not specified for place and 'h' is aspiration).

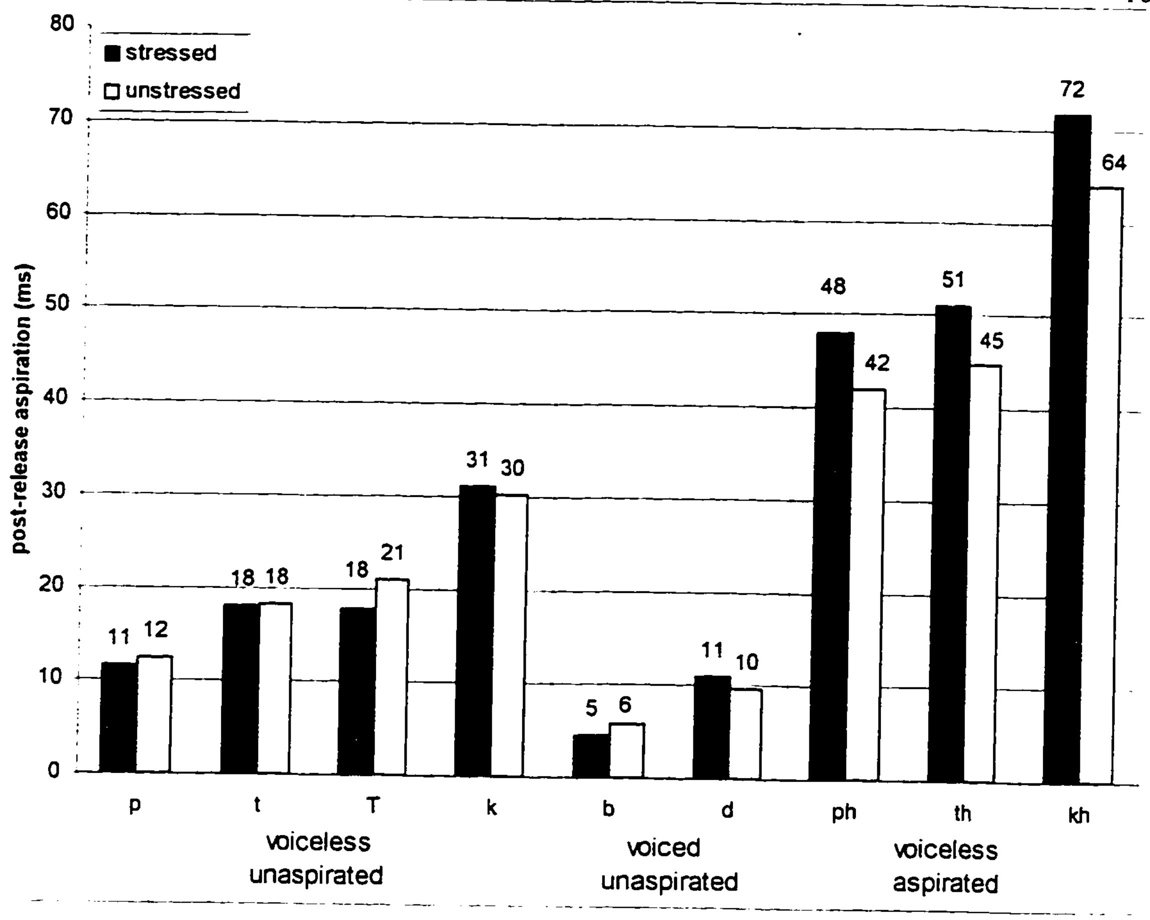


Figure 4-9: Post-release aspiration durations of onset stops in stressed and unstressed syllables, averaged over all speakers.

The average durations for post-release aspiration were 19 ms for voiceless unaspirated stops, 8 ms for voiced stops and 56 ms for voiceless aspirated stops in stressed syllables and 20 ms, 8 ms and 49 ms respectively in unstressed syllables. These averages indicate that aspiration duration was greatest for the aspirated stops and least for the voiced stops. Interestingly, stress changed the aspiration duration only for aspirated stops. Aspiration duration for the voiceless unaspirated and voiced stops remained the same for both unstressed and stressed syllables. The data also showed that aspiration duration was arranged in the following hierarchy: velars > dentals = alveolars > labials.

Statistical analyses showed that there was no significant difference between the duration of aspiration of unaspirated stressed and unstressed stops. A paired t-test for the averaged durations of aspiration (over all speakers) for aspirated onset stops showed that the duration of aspiration for stressed aspirated stops was significantly longer than the duration of aspiration for unstressed aspirated stops ($t = 5.55$, $df = 2$). MANOVAs done for individual speakers showed that aspiration duration was significantly greater for stressed than unstressed stops only for speakers AS and SA. Even though the results did not reach statistical significance, the average aspiration durations were still greater for stressed syllables than unstressed syllables for all other speakers. The detailed data and statistical results are listed in Appendix B.

Differences in duration and change in duration with stress across gender were also investigated. The post-release aspiration durations for voiceless stops and voiced stops were similar for males and females and did not change significantly with stress. However, the aspiration duration for aspirated stops was greater for females (74 ms for stressed and 59 ms for unstressed syllables) than males (56 ms and 50 ms repectively). The percent increase with change in stress was also greater for female speakers (25% or 15 ms) than males (12% or 6ms).

BREATHINESS

The breathy part of the breathy stops was not easily distinguishable from the following vowel for some speakers because the two parts overlapped and coarticulated to give a breathy vowel, especially in unstressed syllables. Therefore, the ratios of the intensities of first and the second harmonic (H0 and H1 respectively) were calculated at five equally spaced points between the release of the stop and the offset of the following vowel. As explained earlier, vowels which were more breathy would have a lower H1 intensity to H0 intensity ratio than vowels which were less breathy. For females the H1 may be amplified because it falls in the region of their first oral tract resonance (because they have a higher F0). For males, H1 is far enough from the first resonance of the oral tract that the effects of the latter on H1 can be ignored. Therefore, data from only the male speakers (AR, SH, ZA) was considered. Data from the fourth male speaker, BS, was not included because he was not available for recordings.

The five points were placed at fixed ratios of the CV duration (1/6, 2/6, 3/6, 4/6, and 5/6 of CV). However, even if the points were placed in time, similar results would have been obtained because the differences between stressed and unstressed post-release CV durations were not very large. For speaker SH the post-release CV duration was 89 ms for the stressed syllable and 77 ms for the unstressed syllable. For speaker ZA the durations were 93 ms for the stressed and 84 ms for the unstressed syllable. For speaker AR the durations were 112 ms for the stressed and 80 ms for the unstressed syllable. The results for averages over all stops for all three male speakers are shown in figure 4.10. Data from individual speakers follows the same pattern, as shown in Appendix B.

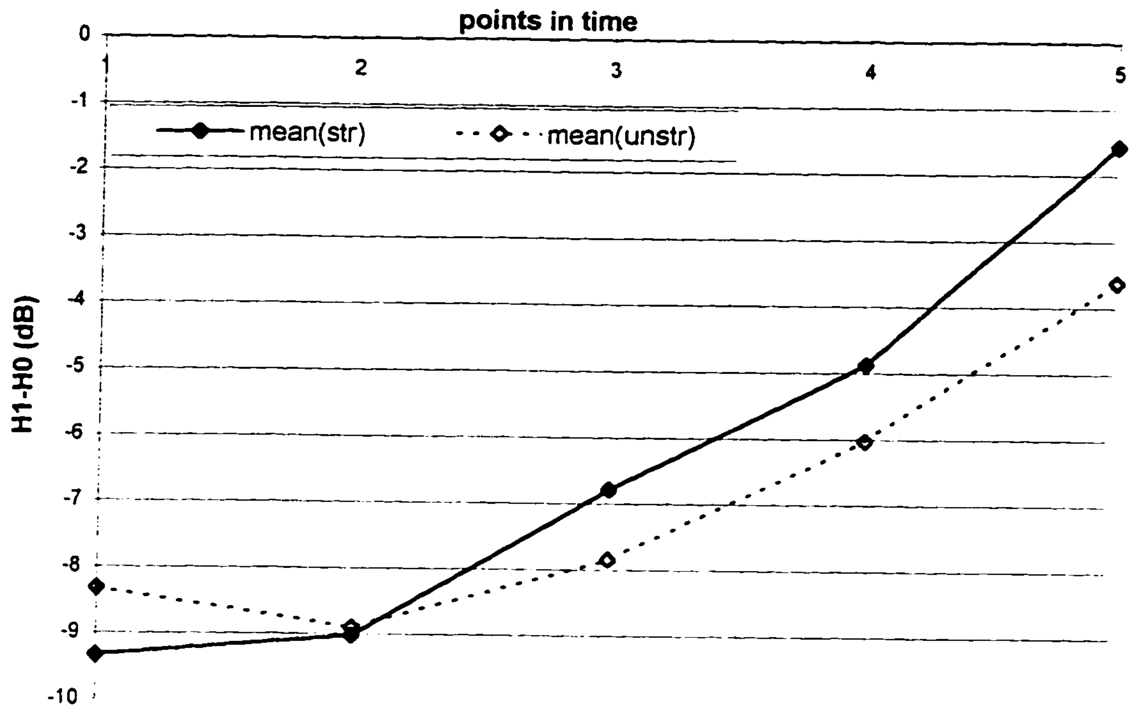


Figure 4-10: Ratio of first harmonic to the fundamental (in dB) for stressed vs. unstressed syllables at five equally spaced points, averaged for three male speakers ZA, SH and AR.

The plot shows that in both stressed and unstressed conditions, the breathiness decreased towards the end of the vowel (i.e. the ratio of H1 to H0 increased). The rise in the ratio was steeper for stressed vowels than unstressed vowels, indicating that the breathiness diminished more quickly for stressed vowels. Also, initially the vowels were more breathy when stressed and towards the end the vowels were more breathy when unstressed. Thus, the stressed vowels start more breathy than unstressed stops and become less breathy and more sonorant towards the end.

Summary of Onset Stop Data

Data from the onset stops shows that stress significantly effects the closure duration, voicing during closure, post-release aspiration and breathiness of the following vowels. The onset durations increased with stress for all stops. However, the duration of aspiration after the release of the stops only increased with stress for the aspirated stops, and not for voiced or voiceless stops. Finally the vowels following breathy voiced stops are less breathy in stressed syllables than in unstressed syllables. The durations, and the change in durations with stress, of closure of all stops and post-release aspiration of aspirated stops were also found to be greater for females than males.

Coda Consonants

Closure Duration

Means of the closure duration for different stops types in the coda position were calculated over all speakers. The means for closure durations for aspirated stops did not include the data for speakers SA and ZA. The word pair /gut̪ʰt̪a, hət̪ʰyar/ was used when these speakers were recorded. During analysis, careful listening indicated that the aspirated dental was not in the coda position and formed a complex onset in the second word (so that the word syllabified as /hə.t̪ʰyar/). The mean of the closure duration for breathy stops did not include the data from speaker AR. As discussed earlier, speaker AR assimilated the labial stop in the coda of the first syllable with the labial continuant in the onset of the following syllable, resulting in a geminated continuant, with no closure or release. This might have been caused because the speaker AR spoke at a faster than normal tempo during the recordings. In addition, the words used for breathy stops involved a complex coda (/nd^h/) for all other speakers, except AA. For these speakers, duration of nasalized closure was measured for breathy stops. The average closure durations are plotted in figure 4.11 below.

Results showed that the duration of closure of stressed voiceless, voiced and breathy stressed coda stops were longer in duration than unstressed ones. However, the average increase in duration of closure with stress for these coda stops was 5 ms, which was considerably smaller than the average increase in closure duration for onset stops, which was 11 ms. Thus, the coda stops showed a weaker effect of stress than onset stops. This asymmetrically weaker stress effect in the falling stress configuration has also been reported for English (e.g. Pierrehumbert 1994). In addition, the stressed aspirated coda stops had a shorter closure duration than unstressed aspirated coda stops. This was a consistent difference over all speakers except for speaker SH, who had longer duration of closure for the stressed aspirated stop than the unstressed ones.

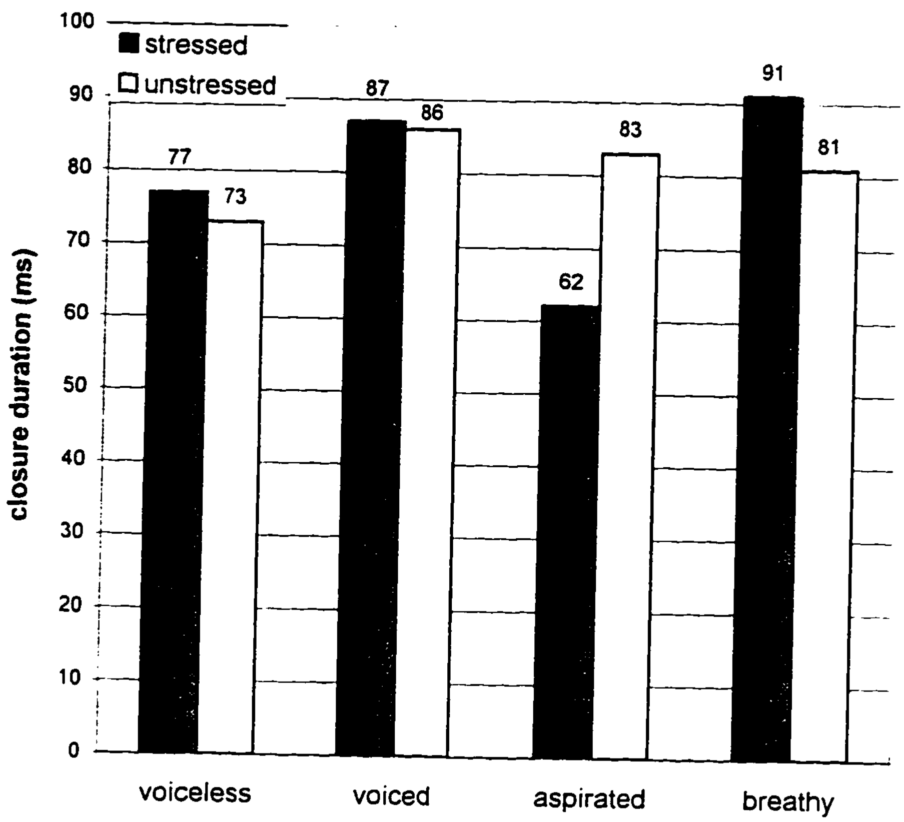


Figure 4-11: Closure durations of coda stops in stressed and unstressed syllables, averaged over all speakers.

Interestingly, the average closure durations for voiceless stops were shorter than the average closure durations for the voiced stops. In the onset position, the average closure durations for voiceless stops were greater than the average closure durations of voiced stops. This provides more evidence that there was an asymmetry between the articulation of the onsets and the codas.

A paired t-test showed that the closure duration for stressed voiceless, voiced and breathy voiced stops was significantly longer than unstressed ones (t = 2.4, df = 19; df is large because the averages over speakers were not used and individual speaker means for all the three stop types were used for comparison). The difference in aspiration due to stress did not reach significance ($t = -2.1$, $df = 3$) perhaps because SH showed an opposite change from the other speakers. Results of follow-up tests for individual speakers are listed in Appendix B.

The effect of gender on the closure duration of coda stops was also investigated. The results show that, unlike the closure durations for the onsets, the closure durations for the codas were greater for male speakers (average of 95ms for stressed and 92ms for unstressed voiceless, voiced and breathy stops, and 63ms and 96ms for stressed and unstressed aspirated stop) than female speakers (74ms, 70ms, 53ms and 67 ms respectively). The differences in closure durations between the stressed and unstressed stops are similar for both males and females, except that the differences in closure duration for aspirated stops are considerably longer for males (33 ms for males compared to 14 ms for females).

VOICING DURATION

Voicing during closure was only calculated for the voiced and the breathy voiced stops. The breathy voiced stops had a complex coda /nd^h/ (except for speakers AA and AR); therefore the closure of the breathy coda stops was nasalized. However, the duration of nasalized closure for these stops was also used in the analysis. The voicing during closure for voiced codas and the nasalized duration before breathy codas are plotted in figure 4.12.

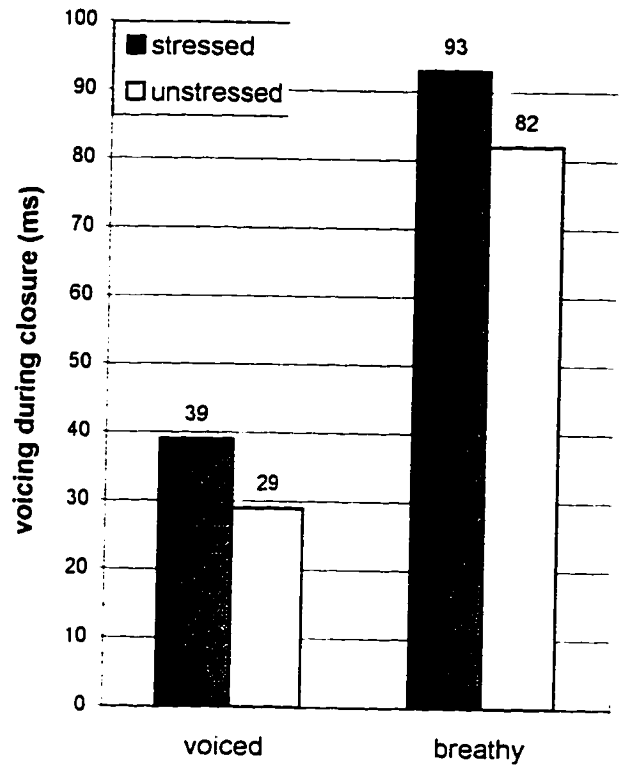


Figure 4-12: Durations of voicing during closure for voiced coda stops and nasalization before breathy coda stops in stressed and unstressed syllables, averaged over all speakers.

The results showed that stress increased the voicing during closure of voiced stops and the nasal closure for /n/ before the breathy stops. Interestingly, the closure durations of both onset and coda voiced stops were about the same duration: 90 ms for stressed voiced stops in onset position, 87 ms for stressed voiced stops in coda position, 79 ms for unstressed voiced stops in onset position, 86 ms for unstressed voiced stops in coda position. However, while the voicing during closure of onset stops was as long as the closure, the duration of voicing during closure for voiced stops was less than half of the closure duration (39 ms for stressed and 29 ms for unstressed voiced coda stops). This decrease in voicing duration in coda stops reflects the tendency of losing the contrasts in coda stops, also noted for many other languages (Lombardi 1990, Hussain and Nair 1995).

A paired t-test showed that the increase in voicing duration with stress for voiced stops was significant (df = 6.6, df = 6). Individual Analysis of Variance (ANOVA) results showed that the voicing duration significantly increased for all speakers except AA and SA. There were no statistics performed for the nasal closure before the voiced stops because it did not truly represent the voicing during closure for coda stops.

Like the closure durations of the coda stops, the voicing during closure was also greater for male speakers (53 ms for stressed and 42 ms for unstressed syllables) than females speakers (31 ms for stressed and 21 ms for unstressed syllables). However, the extent of increase with stress was greater for female speakers (about 50% for females and about 25% for males).

ASPIRATION DURATION

All the aspirated and breathy coda stops are released. Aspiration durations for only the aspirated coda stops were measured. The breathy codas were followed by /v/, which occurred in the onset of the next syllable (e.g. in /tʃubʰ. ˈvɑ.nɑ/). During acoustical analysis, it was discovered that /v/, a weak fricative with voicing, was acoustically very similar to a breathy release. Therefore, it was not possible to accurately mark where the breathy release of the coda ended and the following /v/ began. No other word pairs for breathy codas with appropriate context were found.

For the aspirated coda stops, two speakers (SA and ZA) were recorded r g the second pair of words listed in (3.4c). However, listening carefully to their recordings, it was found that these speakers were moving the coda stops to the onset of the second word to the next (stressed) syllable articulating /hə. ʼt̪ʰyɑr/ instead of /hət̪ʰ. ʼyɑr/. This assumption was also supported by the acoustic data which indicated a higher aspiration duration of the stops (about 71 ms), almost equal to the aspiration duration of stops in the stressed syllables (about 73 ms).

The first word pair in (3.4c) was used for the other speakers. The average value for post-release aspiration duration for the first pair was 68 ms for the stressed coda and 16 ms for the unstressed coda. These values indicate that coda aspirated stops were very strongly released in stressed syllables and not strongly release in unstressed syllables[8].

As pointed out in the section discussing the methods of these experiments, the duration of aspiration was hard to measure, as it sometimes did not decrease monotonically. Thus, all these factors create enough uncertainty in measurements and

---

[8] There could be another factor responsible for the greater difference in the aspiration duration of the first pair of words (68 ms for coda /t̪ʰ/ in /gut̪ʰt̪ɑ/ and 16 ms for coda /t̪ʰ/ in /t̪ʰut̪ʰkɑr/). Broe (1991) quotes Macdonell (1927, p. 55) describing shift of aspiration within a syllable from the coda to an onset for Sanskrit (which is one of the parent languages of Urdu). "If *gh*, *dh*, or *bh* or *h* are at the end of a (radical) syllable beginning with *g*, *d*, *b*, and lose their aspiration as final or otherwise, the initial consonants are aspirated by way of compensation." Though this rule does not apply directly to the first syllable of /t̪ʰut̪ʰkɑr/, there might be some similar constraint which prefers the onset and not the coda to be aspirated within the same syllable. Thus, the coda loses most of its aspiration while the onset retains it. However, there is not enough phonological research in Urdu to support or deny such a claim.

their interpretations that these results for the aspiration of codas are not considered any further.

SUMMARY OF CODA STOP DATA

The closure durations showed mixed results with a change in stress. The durations increased for voiceless, voiced and breathy stops, but decreased for aspirated stops with an increase in stress. The voicing during closure also increased with stress. These stress effects were weaker for coda stops than onset stops. Also, male speakers had longer coda durations than females speakers.

Discussion

This study was performed to investigate how lexical stress is phonetically realized in Urdu. Another motivation of this study was to investigate which of the two competing theories, that try to explain the phonetic changes caused by stress, better explains the changes caused by stress in Urdu, in particular, and cross-linguistically, in general.

The results show that lexical stress alters the phonetic properties of both vowels and consonants in Urdu. Both long and short vowels increase in duration. Interestingly, the results show that the difference in the duration of the vowels, caused by differences in lexical stress, is a function of the phonological category of the vowels. Long vowels in the initial syllable are, on average, twice as long as the short vowels in the initial syllable. The difference in duration between the stressed and unstressed long vowels is also about twice the difference between stressed and unstressed short vowels. Thus, the increase due to stress is not by a fixed duration (e.g. an increase of 10 ms) but by a fixed ratio (about 10% increase of the duration of unstressed vowels).

Durational increase due to stress was also different for vowels in different syllables. The durational increase with stress was minimum (about 10 ms) for vowels in initial syllables, longer for word-medial syllables (about 38 ms) and longest for vowels in word-final syllables (about 70 ms). The stress effect differences between final and non-final syllables can be explained by the presence of following voiced consonants and the word-final lengthening effect (Klatt 1976, Wightman et al. 1992, Berkovits 1993). Wightman et. al (1992) report that they found word-final lengthening effects only the final rhyme of a word. Such an explanation cannot be extended to the difference in increase in duration with stress between word-initial and word-medial (non-final) vowels. Therefore, this difference is caused by some other factor. At this point, no explanation can be extended to account for this difference.

Measurements also indicate that F0 is lower for stressed syllables. The non-significance of the change in F0 for short vowels can be attributed to their short duration. The durations of short vowels are on the order of 50 ms. Therefore, to realize a drop in F0 by the middle of the vowels, speakers need to change F0 within about 25 ms. Since this duration is very short, there might not have been ample time for the speakers to drop F0 of stressed short vowels significantly lower than the unstressed short vowel, before it is pulled up again to realize the word final high pitch. This assumption is supported by physiological research. Kempster, Larson and Kistler (1988, table I) tested the response time of different laryngeal muscles involved in pitch control by directly inserting electrodes into these laryngeal muscles. They report that the cricothyroid and thyroarytenoid muscles, which are involved in the control of F0, have a F0 fall time of 46.3 ms and 42.3 ms respectively. Sundberg (1979) reports even higher times for pitch changes, ranging from 60 ms to 80 ms for untrained speakers. He also reports that this duration is independent of the amount of pitch change. Similar times were found for pitch fall ranging from four semitones up to twelve semitones. Given these times, F0 in short vowels would not have enough time to fall to the target value before it would be pulled up again for the second vowel. Long vowels are twice as long as short vowels (about 100 ms) and therefore have more time to achieve the target F0. Thus, the difference is significant for more speakers for first syllable, long vowels. Long vowels in the second syllable are even longer and occur later in the word, therefore the F0 for these vowels is significantly lower for all speakers.

The data also show that for low vowels, on average, intensity increases with stress, and for high vowels intensity decreases with stress. As discussed earlier, this may indicate that low vowels become even lower with stress, decreasing the acoustic impedance of the oral tract and therefore increasing in intensity. And high vowels become even higher, increasing the acoustic impedance of the oral tract and therefore decreasing in intensity. These observations are supported by the data obtained for the change in vowel quality with stress in Urdu (plotted in figure 4.7). Also, the range of intensity ratio of phonologically short vowels is within the range of intensity ratio of phonologically long vowels perhaps because the short vowels are spread within the articulatory range of the long vowels, i.e. short vowels are more central than long vowels (see figure 4.7).

However, research has shown that articulatory control is very complex and though these generalizations show a trend, the details of intensity control are much more complicated. For example, Wood (1986) showed that an intricate mechanism involving tongue, lips and larynx is involved in the oral articulation of palatal vowels, and stress can effect all these articulatory gestures. Similarly, Sluijter et al. (1996) show a complex laryngeal adjustment with changes in stress. The output intensity is dependent on all these articulatory adjustments. That is perhaps why mixed individual speaker data is obtained on the intensity of vowels. While the ratio of intensity of the first vowel to the second vowel increased with stress for some vowels for some speakers, the ratio decreased for other vowels for other speakers. Both articulatory complexity and perceptual difficulty may make intensity an ineffective cue to stress (as also reported by Sluijter and van Heuven, 1996, for Dutch). Perceptual work is needed to establish the strength of intensity as a cue to stress in Urdu.

Finally, results show that quality of vowels is also effected by stress. The distances calculated between each vowel and a central schwa indicate that the vowels do not get more central when unstressed, as predicted by earlier research by Lindblom (1963). Instead, the analysis by vowel shows that unstressed vowels undergo more coarticulation with adjacent consonants, as reported by van Bergem (1993). Again a perceptual study needs to be performed to determine how strongly this quality difference can cue for difference in stress.

The results also show that phonetic properties of stop consonants changed with stress. The closure durations and voicing during closure increase with stress for stops in onset and coda positions, except for voiceless aspirated stops, whose closure duration increases in onset but decreases in coda position. And, for onsets the increase in aspiration only occurs for voiceless, aspirated stops. The post-release aspiration for voiceless, unaspirated and voiced, unaspirated stops does not change with stress. The post-release aspiration of coda stops could not be reliably measured. It should still be pointed out that even though these measurements were not reliable, the coda aspirated stops showed considerable aspiration after release (as much as 60ms) when in a stressed syllable. Neutralizations of coda stop-release in various languages (Lombardi 1991, Hussain and Nair 1995) indicates that perhaps the release is difficult to articulate. However, the release is well articulated in Urdu. Therefore, perhaps the decrease in closure duration for voiceless aspirated coda stops with stress is a necessary consequence of the extra articulatory effort expended for the release.

In addition, results from male speakers also showed that vowels following breathy stops in stressed syllables are less breathy than the vowels following breathy stops in unstressed syllables. From this data, it can be deduced that the breathy part of the release of the stops overlaps with the following vowel. The extent of this overlap is greater for unstressed stops than stressed stops. There is a greater extent of overlap for unstressed syllables that makes the vowels more breathy and CV sequence shorter in duration. This shows that there is increased coproduction of "four dimensional 'canonical forms'" in unstressed syllables (three dimensions of articulatory space and one dimension of time (Fowler 1980, 128, also Coleman 1992).

Unlike vowels, durational differences measured for consonants were limited to first syllable stops. Therefore, the comparison between segments in different syllables in a word, as done for vowels, could not be done for the stops. However, there are different results found for stops in different positions within the initial syllable. Stress caused a greater increase in duration of onset stops than coda stops. The average closure duration increased by 11 ms for onsets and only 5 ms for (voiceless, voiced and breathy voiced) codas with stress. Voicing duration during closure was equal to the duration of closure for the onsets but only half the duration of closure for the codas. Again, as discussed above, neutralizations of voicing and aspiration contrasts in coda stops in various languages (Lombardi 1991, Hussain and Nair 1995) indicates that perhaps these contrasts are difficult to articulate. That may explain the difference between the onset and coda results. Why this difficulty in articulation for codas may arise still needs to be investigated.

As pointed out in the Introduction, "Features of the speech signal which are part of the linguistic code should be more readily apparent in stressed syllables than in unstressed syllables" (de Jong 1995, 502). The results therefore indicates that duration, F0 and vowel quality of vowels are perhaps linguistically significant part of Urdu phonetics, and perhaps vowel intensity does not play a large role in it. For stops consonants, closure duration, voicing during closure and aspiration after release (and breathiness) contribute to the linguistic contrast.

Though Urdu and Hindi are considered to be the same language (e.g. Masica 1991), these results obtained for Urdu are different from Hindi, reported by M. Ohala (1986). Ohala found a high or rising pitch contour on the stressed syllables in Hindi, but did not find any significant durational differences caused by stress. However, she adds that "I would prefer to be cautious about this conclusion since not all possible duration measurements, using all reasonable controls, have been made" (pg. 88). Therefore, it is still unclear whether stress is phonetically realized only by changes in F0 contour of Hindi words or also by additional acoustic cues such as duration and vowel quality. A similar detailed investigation needs to be undertaken for Hindi as well to determine the acoustics correlates of stress.

These acoustic differences found for Urdu are smaller than those observed in some other languages. Sluijter and van Heuven (1996) report differences of as much as 103 ms between stressed and unstressed [ka] syllable in words *ka*non vs. ka*non* in Dutch (both words have phrasal stress, similar to the stimuli recorded for Urdu). Van Bergem (1993), studying different vowels in Dutch, reports an average difference of about 20 ms between stressed and unstressed accented vowels (in words of the type *can*dy vs. can*teen* for the target vowel /a/). In similar studies, but with nonsense words embedded in carrier phrases, Engstrand (1988) reports differences of about 40 ms in Swedish and Gay (1978) reports differences of about 20 ms in English, for initial syllable vowels. The differences in duration caused by stress in Urdu are smaller (about 10 ms for long vowels and about 5 ms for short vowels in the initial syllable) perhaps because Urdu has two phonological vowel lengths. Short vowels range from 25 ms to 85 ms in duration. These vowels cannot lengthen more with stress because they would be confused with long vowels and cannot shorten because of incompressibility (Klatt 1976). Long vowels range from 75 ms to 140 ms in duration. These vowels cannot shorten beyond a minimum length because they would be confused with short vowels and cannot lengthen because they would be confused with longer, syllable final vowels. Thus, having a phonological length contrast in Urdu reduces the effect of stress on duration of the vowels.

The results also provide data to analyze the two stress theories presented earlier, which try to explain the phonetic changes caused by stress: the *Sonority Expansion* theory (SE) and the *Hyperarticulation* theory (HA). SE proposes that stress increases the sonority related gestures in speech. HA proposes that stress increases all the distinctive phonemic gestures in speech. These theories are evaluated in light of the results from Urdu.

Results show that both phonologically long and short vowels are longer in duration when they are stressed. In addition, the vowels undergo lesser contextual influence when they are stressed. SE will predict increased duration because a longer vowel may give a percept of a more sonorant nucleus. Moreover, closure duration for the onset stops and coda stops (except voiceless aspirated stops) also increases with stress. The increased low sonority consonantal gesture also supports SE because it will increase the relative sonority of the nucleus.

However, SE predicts all vowels will be lowered by stress, as decrease in height decreases the oral acoustic impedance and makes these vowels more sonorous. However, data from Urdu (see figure 4.7) shows that high vowels /i,u/ are higher (i.e. lower F1) when stressed. In addition, the post-release aspiration for onset stops only increases for voiceless aspirated stops and not for voiceless unaspirated and voiced unaspirated stops. If post-release aspiration of onset stops has an effect on sonority, all three stops should be effected in a similar fashion. If this aspiration is unimportant to sonority, then none of these stops should be affected. Also, voicing during closure of voiced onset and coda stops also increases with stress. SE fails to explain it in terms of decrease in consonantal sonority, as voicing increases sonority (Price 1980, also de Jong 1995, 502).

Longer duration for stressed vowels also supports HA because stressed vowels are hyperarticulated, reducing coarticulatory overlap between adjacent segments and therefore increasing the duration. Lesser coarticulatory overlap also decreases the influence of the context. Thus, all the stressed vowels are more extreme in the articulatory space (figure 4.7). This also supports HA, which argues that high vowels would be higher and low vowels would be lower when stressed. Similarly, the increase in closure duration and the voicing during closure duration also support HA, which predicts increased articulatory gestures for consonants as well. Furthermore, speakers in Urdu only increase the post-release aspiration for aspirated stops and not for voiced or voiceless unaspirated stops in the onset position (this duration is not measured for breathy stops). This is supported by HA as well. According to HA, aspiration is distinctive in Urdu, therefore for the [+aspirated] (aspirated) stops, the aspiration should increase but for the [-aspirated] (voiceless and voiced) stops the aspiration should be decreased. The aspiration is already minimal for the [-aspirated] stops and therefore may be it cannot be further compressed.

However, HA does not predict why the closure duration for voiceless aspirated coda stops decreases with stress. This is because the closure duration is not phonemically distinctive in Urdu, and therefore HA theory does not predict a change in it. As earlier hypothesized, the decrease is caused to facilitate the post-release aspiration. Therefore, HA theory only indirectly predicts that, perhaps to hyperarticulate the phonemic contrast of aspiration, the closure duration is compromised.

Thus, neither theory can completely explain the complete breadth of the data. Lass (1987, 108, quoted by Laver 1994, 513-514) suggests, "the only 'universal' requirement is that [stressed syllables] be different from [weak syllables], and that this difference be perceived as a difference in prominence" by the listeners of that language. Both of these theories argue that articulating this difference normally requires *increasing* something (sonority or a distinctive gesture) perhaps because under normal circumstances speech is being articulated with *minimum* effort to produce an acceptable perceptual contrast (Lindblom 1990). However, as found out, this difference in prominence is not articulated by changing *all* contrasts and not always by *increasing* something (sonority or a gesture). The data shows that only a subset of all the phonetic contrasts increases because of possible articulatory (and perhaps perceptual) constraints.

Thus, these theories, though correctly predicting greater distinction with stress, are still too general. Both theories should be 'sharpened' to make more detailed predictions for different languages, incorporating articulatory and perceptual constraints. Though some constraints are universal (e.g. perhaps the constraint that codas have smaller stress effects than onsets), there are also some language and context specific constraints, depending on the subset of all the available contrasts a particular language employs (e.g. closure duration of coda aspirated stops might decrease for Urdu because it has distinctive post-release aspiration for these stops). A complete theory would predict all these changes with stress, given the universal and language specific constraints.

Also, the theories should explain effects at all levels of stress (i.e. line 1 stress through line 3 stress), because all these levels of stress are integrated in a single metrical representation and therefore, at some level, should have similar acoustic consequences. Earlier investigations (e.g. de Jong 1995, Pierrehumbert and Talkin 1992) have primarily studied stress differences at the phrasal level. This study has extended the perspective by attempting to explain how lexical-level stress changes may also be incorporated in these theories.

This study did not aim to investigate any gender differences in the change of phonetic properties with lexical stress. However, results do show that there are consistent differences between male and female speakers, which cannot be ignored. It has been reported earlier in the literature that female vowel durations are longer than male vowel durations (e.g. Hillenbrand et al. 1995). In addition, this study also revealed that the degree of increase in duration due to lexical stress was also greater for females. And this increase was not just in duration but also in the ratio of stressed vs. unstressed durations i.e. the increase for females is 37% of the average unstressed vowel duration compared with 22% for males. Therefore, this difference in the increase in vowel duration is unlike the difference in increase in vowel duration due to stress between long and short vowels, the latter being fixed at 10% of the length of the vowel. These durational differences can also be extended to onsets. The closures were found to be 21% longer when stressed for females and 9% longer for males. The voicing during closure increased by 17% with stress for females and 6% for males. The post-release aspiration duration increase by 25% for stressed aspirated onset stops for females and 12% for males. Effects of stress on F0 also varied with gender. Females also showed a greater change in F0 and percent F0 with stress than male speakers (44 Hz or 21% decrease in second syllable vowel F0 with stress for females vs. 9 Hz or 7% decrease in second syllable vowel F0 with stress for males; 12 Hz or 5% vs. 2 Hz or 1% respectively for first syllable long vowels; 4 Hz or 1% vs. 1 Hz or less than 1% respectively for first syllable short vowels). These differences perhaps make female speech more clear than male speech (Bond and Moore 1994). At this time, there are no explanations for these differences. However these differences may be hypothesized to be caused by physiological differences between the males and females because this phenomenon seems to occur for other languages as well (e.g. Diehl et al., 1996, and Hillenbrand et al., 1995, for English, and Sluijter and van Heuven, 1996, for Dutch). Again, it should be pointed out that the pool of speakers was small for this study (three female and four male speakers) and more speakers should be analyzed to give more validity to these generalizations.

SUMMARY

This work has investigated how different phonetic properties of Urdu vocalic and consonantal segments change with stress. The results show that the duration of stressed vowels is longer than unstressed syllables. In addition, the F0 of speakers decreases for stressed vowels (except for BS, who speaks a different dialect of Urdu). The change in intensity with stress is vowel dependent. High and/or back vowels get less intense with increased stress, perhaps due to increased high and/or back gesture which increases the acoustic impedance of the oral tract. Low and/or front vowels get more intense with increased stress, perhaps due to increased low and/or front gesture which decreases the acoustic impedance of the oral tract. Individual speaker data on intensity shows a lot of variation, which can be attributed to the variety of articulatory gestures all of which influence intensity in different ways. Also, the quality of the vowels changes with stress but vowels do not necessarily centralize when unstressed. Results show that the quality of unstressed vowels undergoes more contextual assimilation than stressed vowels. Also, the phonologically long vowels are more extreme in quality than the phonologically short vowels. Generally, initial vowels show a smaller extent of change with change in stress than non-initial vowels. Moreover, closure of onset stops, voicing during closure of onset stops, aspiration of aspirated (but not voiceless and voiced) onset stops, closure of voiceless, voiced and breathy coda stops and voicing during closure of voiced coda stops increases with stress. The duration of closure of aspirated coda stops decreases with stress. Only some of these stress-related changes in Urdu support the Sonority Expansion or Hyperarticulation theories There are also significant gender differences. Female speakers had a greater duration of vowels and stops (except coda stops), higher F0 and lower intensity of vowels and different quality of vowels than males and showed a greater change with stress in these quantities (except intensity) than males.

REFERENCES

Abel, S. 1972. Duration discrimination of noise and tone bursts. *J. Acoust. Soc. Amer.* 51: 1219-1223

Abramson, A. and Ren, N. Distinctive vowel length: duration vs. spectrum in Thai. *J. Phon.* 18.2: 79-92.

Abramson, S. 1961. Identification and discrimination of phonemic tones. *J. Acoust. Soc. Amer.* 33: 842.

Anderson, S and Port, R. 1994. Evidence of syllable structure, stress and juncture from segmental durations. *J. Phon.* 22.3: 283-315.

Beckman, M 1986. *Stress and Non-Stress Accent.* Riverton: Foris Publications.

Beckman, M. and Edwards, J. 1990. Lengthenings and shortenings and nature of prosodic constituency. In *Papers in Laboratory Phonology I: Between Grammar and Physics of Speech,* edited by J. Kingston and M. Beckman. Cambridge: Cambridge University Press: 152-178.

---. 1994. Articulatory evidence for differentiating stress categories. In *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form,* edited by P. Keating. Cambridge: Cambridge University Press: 2-33.

Beckman, M., Edwards, J. and Fletcher, J. 1992. Prosodic structure and tempo in a sonority model of articulatory dynamics. In *Papers in Laboratory Phonology II: Segment, Gesture and Tone,* edited by G. Doherty and D. Ladd. Cambridge: Cambridge University Press: 90-117.

Berkovits. R. 1993. Utterance-final lengthening and the duration of final stops closures. *J. Phon.*

Bhatia, T. 1993. *Punjabi.* London: Routledge.

Bokhari, S. 1985. *Phonology of Urdu Language.* Karachi: Royal Book Company.

Bolinger, D. 1958. A theory of pitch accent in English. *Word* 14: 109-149.

Bond, Z. and Moore, T. 1994. A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Commun.* 14: 325-337.

Broe, M. 1991. Paradox lost: a non-derivational approach to Grassmann's Law. In *Declarative Perspectives on Phonology, Edinburgh Working Papers in Cognitive Science,* edited by S. Bird. Vol. 7. Edinburgh: University of Edinburgh.

Carlson, L. 1980. Word stress in Finnish. Manuscript. Dept. of Linguistics, Massachusetts Institute of Technology.

Chomsky, N. and Halle, M. 1968. *Sounds Patterns of English.* Cambridge: Massachusetts Institute of Technology Press.

Clark, J. and Yallop, C. 1990. *An Introduction to Phonetics and Phonology.* Oxford: Basil Blackwell.

Coleman, J. 1992. The phonetic interpretation of headed phonological structures containing overlapping constituents. *Phonology* 9: 1-44.

Coleman, J., Dirksen, A., Hussain, S., Waals, J. 1996. Multilingual phonological analysis and speech synthesis. *Proceedings of Computational Phonology in Speech Technology.* Somerset: Association of Computational Linguistics.

Creelman, C. 1962. Human discrimination of auditory duration. *J. Acoust. Soc. Amer.* 34: 582-593.

Crystal, T. and House, A. 1988a. Segmental durations in connected-speech signals: current results. *J. Acoust. Soc. Amer.* 83.4: 1553-1573.

---. 1988b. Segmental durations in connected-speech signals: syllabic stress. *J. Acoust. Soc. Amer.* 83.4: 1574-1585.

Davis, K. 1994. Stop voicing in Hindi. *J. Phonetics* 22(2): 177-193.

Davis, S. and Summers, V. 1989. Vowel length and closure duration in word medial VC sequences. *J. Phon.* 17.4: 339-353.

de Jong, K. 1991. An articulatory study of consonant-induced vowel duration changes in English. *Phonetica* 48: 1-17

---. 1995. The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. *J. Acoust. Soc. Amer.* 97.1: 491-504.

de Jong, K., Beckman, M. and Edwards, J. 1993. The interplay between prosodic structure and coarticulation. *Language and Speech* 36(2,3): 197-212.

Delattre, P. 1962. Some factors of vowel duration and their cross-linguistic validity. *J. Acoust. Soc. Amer.* 34: 1141-1142.

---. 1966. A comparison of syllable length conditioning among languages. *Int. Rev. Appl. Linguistics* 4.3: 183.

Diehl, L., Lindblom, B., Hoemeke, A. and Fahey, P. 1996. On explaining certain male-female differences in the phonetic realization of vowel categories. *J. Phon.* 24.2: 187-208.

Dilley, L., Shattuck-Hufnagel, S. and Ostendorf, M. 1996. Glottalization of word-initial vowels as a function of prosodic prominence. *J. Phon.* 24.4: 423-444.

Dixit, R. 1963. The segmental phonemes of contemporary Hindi. Master's thesis, University of Texas at Austin.

Edwards, R. and Beckman, M. 1988. Articulatory timing and the prosodic interpretation of syllable duration. *Phonetica* 45: 156-174.

Elert, C. 1964. *Phonologic Studies of Quantity in Swedish*. Stockholm: Almqvist & Wiksell.

Engstrand, O. 1988. Articulatory correlates of stress and speaking rate in Swedish VCV utterances. *J. Acoust. Soc. Amer.* 83.5: 1863-1875.

Falc'hun, F. 1951. *Le Systeme Consonantique du Breton avec une Etude Comparative de Phonetique Experimentale*, Rennes: Imprimeries Reunies.

Fant, G. 1956. On the predictability of formant levels and spectrum envelopes from formant frequencies. In *For Roman Jackobson*, edited by M. Halle, H. Lunt and H. MacLean. The Hague: Mouton: 109-120.

---. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.

Fear, B., Cutler, A., and Butterfield, S. 1995. The strong/weak syllable distinction in English. *J. Acoust. Soc. Amer.* 97.3: 1893-1904.

Fintoft, K. 1961. The duration of some Norwegian speech sounds. *Phonetica* 7: 19-39.

Fischer-Jorgensen, E. 1964. Sound duration and place of articulation. *Zeitschrift fur Phonetik* 17: 175-207.

Flagnan, L. and Saslow, M. 1958. Pitch discrimination for synthetic vowels. *J. Acoust. Soc. Amer.* 30: 435-440.

Flagnan, J. 1957. Estimates of the maximum precision necessary in quantizing certain 'dimensions' of vowel sounds. *J. Acoust. Soc. Amer.* 29: 533-534.

Fonagy, I. 1966. Electro-physiological and acoustic correlates of stress and stress perception," *J. Speech and Hearing Res.* 9: 231-244.

Fourakis, M. 1991. Tempo, stress and vowel reduction in American English. *J. Acoust. Soc. Amer.* 90.4: 1816-1827.

Fowler, C. 1980. Coarticulation and theories of extrinsic timing. *J. Phonetics* 8: 113-133.

Fry, D. 1955. Duration and intensity as physical correlates of linguistic stress. *J. Acoust. Soc. Amer.* 27: 765-768.

---. 1958. Experiments in Perception of Stress. *Lang. and Speech* 1: 126-152.

---. 1965. The dependence of stress judgments on vowel formant structure. *Proceedings of the 5th International Congress of Phonetic Sciences, Munster 1964.* New York: S. Krager: 306-311.

Fujisaki, H., Nakamura, K. and Imoto, T. 1975. Auditory perception of duration of speech and non-speech stimuli. In *Auditory Analysis of Perception of Speech*, edited by G. Fant and T. Tatham. New York: Academic Press: 197-220.

Gay, T. 1978. Effect of speaking rate on vowel formant movements. *J. Acoust. Soc. Amer.* 63.1: 223-230.

Gelfand, S. 1990. *Hearing: An Introduction to Psychoacoustical and Physiological Acoustics.* New York: Marcel Dekker..

Gerard, C. and Dahan, D. 1995. Durational variations in speech and didactic accent during reading," *Speech Communication* 16: 293-311.

Gobl, C. 1989. A preliminary study of acoustic voice quality correlates. *Speech Transmission Laboratory - Quarterly Progress and Status Report* 4. Royal Institute of Technology, Stockholm.

Goldsmith, J. 1990. *Autosegmental and Metrical Phonology: An Introduction.* Oxford: Blackwell.

Grierson, G. 1895. On the stress-accent in modern Indo-Aryan vernacular. *Journal of the Royal Asiatic Society of Great Britain and Ireland* 3: 139-147.

Halle, M. and Vergnaud, J. 1987. *An Essay on Stress.* Massachusetts Institute of Technology Press, Cambridge, MA.

Harris, K. 1978. Vowel duration change and its underlying physiological mechanisms. *Language and Speech* 21, 4:354-361.

Hayes, B. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20: 253-306.

Hayes, B. 1995. *Metrical Stress Theory*. Chicago: University of Chicago Press.

Hillenbrand, J., Getty, L., Clark, M. and Wheeler, K. 1995. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Amer.* 97.5: 3009-3111.

House, S. 1959. A note on optimal vocal frequency," *J. Speech and Hearing Res.* 2: 55-60.

---. 1961. On Vowel Duration in English. *J. Acoust. Soc. Amer.* 33: 1174-1178.

House, S. and Fairbanks, 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels," *J. Acoust. Soc. Amer.* 25: 105-113.

Howes, D. 1950. The loudness of multicomponent tones. *Amer. J. of Psych.* 63: 1-30.

Huggins, A. 1972. Just-noticeable differences for segment duration in natural speech. *J. Acoust. Soc. Amer.* 51: 1270-1278.

Hussain, S. (1993). Synthesis of Urdu CV syllables using Infovox text-to-speech system. M.Phil. diss. University of Cambridge.

---. 1994. Acoustic correlates to manner of stop articulation in Urdu. Abstract in *J. Acoust. Soc. Amer.* 96.5.2: 3230.

Hussain, S. and Nair, R. 1995. Voicing and aspiration contrast in Hindi and Urdu. *Proceedings of the 31st Annual Meeting of Chicago Linguist Society*, Chicago: University of Chicago.

Hyman, L. 1977. On the nature of linguistic stress. In *Studies in Stress and Accent, Southern California Occasional Papers in Linguistics*, edited by L. Hyman, Vol. 4. University of Southern California, Los Angeles: 37-82.

Ito, J. 1986. *Syllable Theory in Prosodic Phonology*. Ph.D. diss. University of Massachusetts, Amherst.

Jassem, W. 1959. The phonology of Polish stress. *Word* 15: 252-269.

Jassem, W., Morton, J., and Steffen-Batog, M. 1968. The perception of stress in synthetic speech-like stimuli by Polish listeners. *Speech Analysis and Synthesis* 1: 289-308.

Jongman, A., Blumstein, S. and Lahiri, A. 1985. Acoustic properties for dental and alveolar stop consonants: a cross language study. *J. of Phonetics* 13.

Kachru, Y. 1987. Hindi-Urdu. In *The Major Languages of South Asia, The Middle East and Africa*, edited by B. Comrie. London: Routledge.

Kahn, D. 1976. *Syllable-based Generalizations in English Phonology*. Ph.D. diss. Massachusetts Institute of Technology.

Keating, P. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60: 286-319.

Kelkar, A 1968. *Studies in Hindi-Urdu I*, Postgraduate and Research Institute, Deccan College, Poona.

Kemplster, G., Larson, C., and Kistler, M. 1988. Effects of electrical stimulation of cricothyroid and thyroarytnoid muscles on voice fundamental frequency. *J. Voice* 2(3): 221-229.

Kenstowicz, M. 1994. *Phonology in Generative Grammar*. Cambridge: Blackwell Publishers.

Kewley-Port, D. 1983. Time-varying features as correlates of place of articulation in stops consonants. *J. Acoust. Soc. Amer.* 73: 322-335.

---. 1995. Threshold for formant-frequency discrimination of vowels in consonantal context. *J. Acoust. Soc. Amer.* 97.5: 3139.

Kewley-Port, D. and Watson, C. 1994. Formant-frequency discrimination for isolated English vowels. *J. Acoust. Soc. Amer.* 95.1: 485-496.

Kiparsky, P. 1982. Lexical phonology and Morphology. In *Linguistics in the Morning Calm*, edited by I. Yang. Seoul: Hanshin.

Klatt, D. 1973. Durational characteristics of pre-stressed word-initial consonant clusters in English. *Res. Lab. Electron., Quarterly Prog. Rep.* 108, Massachusetts Institute of Technology: 253-260.

---. 1974. The duration of [s] in English words. *J. Speech and Hearing Res.* 17: 51-63.

---. 1975. Vowel lengthening is syntactically determined in connected discourse. *J. Phonetics.* 3: 161-172.

---. 1976. Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *J. Acoust. Soc. Amer.* 59: 1208-1221.

Ladefoged, P., 1975. *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.

---. 1993. *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.

Ladefoged, P., de Clerk, P., Lindau, M. and Papçun, G. 1972. An auditory-motor theory of speech production. *UCLA Working Papers in Phonetics* 22: 48-75.

Ladefoged, P., Draper, M. and Whitteridge, D. 1958. Syllables and stress. *Miscellanea Phonetica* 3: 1-14.

Laeufer, C. 1992. Patterns of voicing-conditioned vowel duration in French and English. *J. Phon.* 20.4: 411-440.

Laukkanen, A., Vilkman, E., Alku, P. and Oksanen, H. 1996. Physical variations related to stress and emotional state: a preliminary study. J. Phon. 24.3: 313-335.

Laver, J. 1994. *Principles of Phonetics*, Cambridge: Cambridge University Press.

Lehiste, I. 1972. The timing of utterances and linguistic boundaries. *J. Acoust. Soc. Amer.* 51: 2018-2024.

---. 1970. *Suprasegmentals*. Cambridge: Massachusetts Institute of Technology Press.

---. 1968. Vowel Quantity in Word and Utterance in Estonian. *Congressus Secundus Internationalis Finno-Ugristarum, Helsinki 1965*. Helsinki: Societas Finno-Ugrica: 293-303.

Lehiste, I., and Ivic, P. 1963. Accent in Serbo-Croatian: an experimental study. *Michigan Slavic Materials* 4, University of Michigan, Ann Arbor.

Lehiste, I. and Peterson, G. 1959. Vowel amplitude and phonemic stress in American English. *J. Acoust. Soc. Amer.* 31: 428-435.

---. 1961. Some basic Considerations in the Analysis of Intonation. *J. Acoust. Soc. Amer.* 33: 419-425.

Liberman, M. 1975. *The Intonational System of English.* Ph.D. diss. Massachusetts Institute of Technology, Cambridge.

Lindau, and Ladefoged, P. 1986. Variability of feature specification. In *Invariance and Variability in Speech Processes*, edited by J. Perkell and D. Klatt, Hillsdale: Lawrence Erlbaum.

Lindblom, B. 1963. Spectrographic study of vowel reduction," *J. Acoust. Soc. Amer.* 35.12: 1773-1781.

---. 1968. Temporal organization of syllable production. *Speech Transmission Lab. Quarterly Progress Report* 2(3), Royal Institute of Technology, Stockholm.

---. 1990. Explaining phonetic variation: a sketch of the H&H theory. In *Speech Production and Speech Modeling*, edited by W. Hardcastle and A. Marchal. Dordrecht: Kluwer.

Lindblom, B. and Engstrand, O. 1989. In what sense is speech quantal. *J. Phonetics* 17: 107-121.

Lombardi, L. 1991. *Laryngeal Features and Laryngeal Neutralization*. Ph.D. diss., Massachusetts Institute of Technology.

Macdonell, A. 1927. *Sanskrit Grammar*. Oxford: Oxford University Press.

Masica, C. 1991. *The Indo-Aryan Languages*. Cambridge: Cambridge University.

McCarthy, J, and Prince, 1986. *Prosodic Morphology*. Manuscript, Brandeis University.

Mehrotra, R 1965. Stress in Hindi. *Indian Linguistics* 26: 96-105.

Morton, J. and Jassem, W. 1965. Acoustic correlates of Stress. *Language and Speech* 8, 159-181.

Munro, P. 1977. Towards a reconstruction of Uto-Aztecan. In *Studies in Stress and Accent, Southern California Occasional papers in Linguistics*, edited by L. Hyman. Vol. 4, Department of Linguistics, University of Southern California, Los Angeles: 303-326.

Nakatani, L. and Aston, C. 1978. Acoustic and linguistic factors in stress perception. *Bell Labs Technical Memorandum, Dept. of Speech and Communication Research.*

Narang, G. and Becker, D. 1971. Aspiration and nasalization in the general phonology of Hindi. *Language* 47: 646-67.

Ohala, J. 1977. The physiology of stress. In *Studies in Stress and Accent, Southern California Occasional papers in Linguistics*, edited by L. Hyman. Vol. 4, Department of Linguistics, University of Southern California, Los Angeles.

---. 1983. The phonological ends justifies any means. In *Proceedings of the 13th International Congress of Linguistics*, edited by S. Hattori and K. Inoue. New York: Academic Press.

Ohala, J. and Eukel, B. 1987. Explaining the intrinsic pitch of vowels. In *In the Honor of Ilse Lehiste*, edited by R. Channon and L. Shockey. Drodrecht: Foris.

Ohala, J. and Ohala, M. 1993. The phonetics of nasal phonology. In *Phonetics and Phonology 5: Nasals, Nasalization and the Velum*, New York: Academic Press.

Ohala, M. 1979. Phonological features of Hindi stops. *Asian Language Analysis* 1: 79-88.

---. 1983. *Aspects of Hindi Phonology*, Delhi: Motilal Banarsidass.

---. 1986. A search for the phonetic correlates of Hindi stress. In *South Asian Languages: Structure, Convergence and Diglossia*, edited by Krishnamurti. Delhi: Motilal Banarsidass.

Oller, D. 1973. The effect of position in utterance on speech segment duration in English," *J. Acoust. Soc. Amer.* 54.5: 1235-1247.

Pandey, P. 1989. Word Accentuation in Hindi. *Lingua* 77: 37-73.

Peterson, E. and Lehiste, I. 1960. Duration of syllable nuclei in English. *J. Acoust. Soc. Amer.* 32: 693-703.

Peterson, E. and Barney, L. 1952. Control methods used in a study of the vowels. *J. Acoust. Soc. Amer.* 24: 175-184.

Pierrehumbert, J. 1980. The phonology and phonetics of English intonation. Ph.D. diss. Massachusetts Institute of Technology.

---. 1994. Prosodic effects on glottal allophones. In *Proceedings of Vocal Fold Physiology: Voice Quality Control*, edited by Fujimura and Hirano. San Diego: Singular Publishing.

Pierrehumbert, J. and Hirschberg, J. 1990. The meaning of intonation contours in the interpretation of discourse. In Intentions in Communication, edited by, P. Cohen, J. Morgan and M. Pollack: Cambridge: Massachusetts Institute of Technology Press.

Pierrehumbert, J. and Nair, R. 1996. Implications of Hindi prosodic structure. In *Current Trends in Phonology: Models and Methods*, edited by J. Durand and B. Laks. Salford: University of Salford Press.

Pierrehumbert, J. and Talkin, D. 1992. Lenition of /h/ and glottal stop. In *Papers in Laboratory Phonology II: Segment, Gesture and Tone*, edited by G. Doherty and D. Ladd. Cambridge: Cambridge University Press.

Port, R. 1981. Linguistic Timing Factors in Combination. *J. Acoust. Soc. Amer.* 69.1: 262-274.

Price, P. 1980. Sonority and Syllabicity: acoustic correlates of perception. *Phonetica* 37: 327-343.

Qureshi, B. 1992. *Standard Twentieth Century Dictionary: Urdu into English.* Delhi: Photo Offset Printers.

Riesz, R. 1928. Differential Intensity of the Ear for Pure Tones. *Physical Review* 31: 867-875.

Rigault, A. 1962. Role de la frequence, de l'intensite et de la duree vocaliques dans la perception de l'accent en Francais," *Proceedings of the Fourth International Congress of Phonetic Sciences, Helsinki 1961.* The Hague: Mouton: 735-748.

Schiefer, L. 1986. F0 in the production and perception of breathy Stops: evidence from Hindi. *Phonetica* 43: 43-69.

Sharma, A. 1969. Hindi word accent. *Indian Linguistics* 30: 115-118.

Sluijter, A. and van Heuven, V. 1996. Spectral balance as an acoustic correlate of linguistic stress. *J. Acoust. Soc. Amer.* 100.4: 2471-2485.

Stevens, K. (1972). The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data, in *Human Communication: A Unified View*, (David, E. and Denes, P. eds.) McGraw-Hill, New York, NY, USA: 51-66.

Stevens, K. and Blumstein, S. (1975). Quantal Aspects of Consonant Production and Perception: A Study of Retroflex Stop Consonants, Journal of Phonetics 3: 215-233.

---. (1978). Invariant cues for Place of Articulation in Stop Consonants, JASA 64(5): 1358-1368.

Stevens, S. and Volkmann, J. 1940. The relation of pitch to frequency: a revised scale. *Amer. J. Psych.* 53:329-353.

Silverman, K. and Pierrehumbert, J. 1990. The timing of prenuclear high accents in English. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, edited by J. Kingston and M. Beckman. Cambridge: Cambridge University Press.

Small, A., Brandt, J. and Cox, P. 1962. Loudness as a Function of Signal Duration. *J. Acoust. Soc. Amer.* 34: 513-514.

Smith, B. 1978. Effects of place of articulation and vowel environment on 'voiced' stop consonant production," *Glossa* 12(2): 163-174.

---. 1994. Effects of experimental manipulations and intrinsic contrasts on relationships between duration and temporal variability in children's and adults' speech," *J. Phonetics 22*: 155-175.

Steriade, D. (1993a). Closure, release and nasal contours. *Phonetics and Phonology* 5: 401-470.

---. (1993b). *Complex Onsets as Single Segments: The Mazateco Pattern*. ms. Department of Linguistics, Ohio State University.

Stetson, R. 1951. *Motor Phonetics*. Amsterdam: North-Holland Publishing Company.

Traill, A. and Jackson, M. 1988. Speaker variation and phonation type in Tsonga. *J. Phonetics* 16: 385-400.

Turk, A. and Sawusch, J. 1996. The processing of duration and intensity cues to prominence. *J. Acoust. Soc. Amer.* 99.6: 3782-3790.

Umeda, N. 1975. Another Consistency in Phoneme Duration. *J. Acoust. Soc. Amer.* 58, S62 *(A)*.

Van Bergem, D. 1993. Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication* 12: 1-23.

---. 1995. Perceptual and acoustic aspects of lexical vowel reduction, a sound change in progress. *Speech Communication* 16: 329-358.

Weismer, G. 1979. Sensitivity of voice-onset-time (VOT) measure to certain segmental features in speech production. *J. Phonetics* 7: 197-204.

Westin, K. Buddenhagen, R. and Obrecht, D. 1966. An experimental analysis of relative importance of pitch, quantity and intensity as cues to phonemic distinctions in Southern Swedish. *Language and Speech* 9: 114-126.

Whalen, D. and Levitt, A. The universality of intrinsic F0 of vowels. J. Phon. 23.3: 349-366.

Wightman, C., Shattuck-Hufnagel, S. and Ostendorf, M. 1992. Segmental duration in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Amer.* 91.3: 1707-1717.

Wood, S. 1986. The acoustical significance of tongue, lip and larynx maneuvers in rounded palatal vowels. *J. Acoust. Soc. Amer.* 80: 391-401.

Yamuna, K. 1987. Hindi-Urdu. In *The Major Languages of South Asia, The Middle East and Africa,* edited by B. Comrie. London: Routledge: 53-72

Zemlin, W. 1988. *Speech and Hearing Science: Anatomy & Physiology*. Englewood Cliffs: Prentice-Hall.

Zlatoustova, L. 1954. Duration of vowel and consonant sounds of the Russian language. *Science Notes, Kazan State University, imeni V. I., Ul'yanova-Lenina* 114(6): 99.

APPENDIX A: URDU PHONEMIC INVENTORY AND ITS TRANSCRIPTION AND PRONUNCIATION GUIDE

PHONEMIC INVENTORY

According to Bokhari (1985, 6), there are seven long oral vowels and seven corresponding short oral vowels in Urdu (however, listed in ambiguous transcription). Kachru (1987, table 3.1) lists seven long oral vowels and three short oral. The author agrees that there are seven long oral vowels, and the inventory of short vowels is limited to only three (as proposed by Kachru) and not seven (as proposed by Bokhari). In addition, Bokhari also lists fourteen nasal vowels, which are not listed by Kachru. However, nasal vowels mostly occur as allophones of the corresponding oral vowels. The nasalization of vowels occurs due to the [nasal] feature assimilation from an adjacent consonant within the syllable. They do occur in minimal pairs with long oral vowels, but only in an extremely limited number of words and only in word final position (e.g. /hæ/, third person singular conjugation of the verb "to be" vs. /hæ̃/[9], third person plural conjugation of the verb "to be"). There is still controversy surrounding the phonemic status of these nasal vowels. (For more detailed discussions on Hindi nasal vowels, see Ohala 1983, Narang and Becker 1971 and Kelkar 1968 and Masica 1991, pg.117). The nasal vowels will not be considered in this work. The seven long oral vowels and three short oral vowels of Urdu are listed in (A.1). Among the long vowels /ɔ/ shows a very limited usage. Also, Kachru lists the front, low vowel as /ɛ/ but author believes the vowel is closer in quality to /æ/.

(A.1)

Long Vowels

i	pila	"yellow"
e	d̪ena	"to give"
æ	bætʰna	"to sit"
ɑ	gɑnɑ	"to sing"
ɔ	pɔd̪ɑ	"plan"
o	bonɑ	"to plant"
u	kud̪nɑ	"to jump"

Short Vowels

ɪ	pɪtɑ	"father"
ə	pətɑ	"address"
ʊ	pʊtli	"puppet"

[9] Even in this case, it is not certain whether the nasal vowel is distinct phonemically or is a result of nasalization of /æ/ before a coronal nasal /n/. This is being argued because in transcription the nasal vowel does not have a distinct graphemic representation, but is represented by /æ/ followed by a nasal grapheme which is also used to represent the nasal /n/ word finally.

As pointed out earlier, Bokhari (1985) lists seven short oral vowels in Urdu, corresponding to each of the long oral vowels. However, the limited data I have analyzed only gives evidence of three short oral vowels. For example, long vowels 'map' onto short vowels when grammatical categories of certain words are changed. Pilot work shows that the front vowels /i.e.æ/ map onto /ɪ/, the low vowel /ɑ/ maps onto /ə/ and the round vowels /ɔ, o, u/ map onto /ʊ/. This is illustrated in figure A.1.

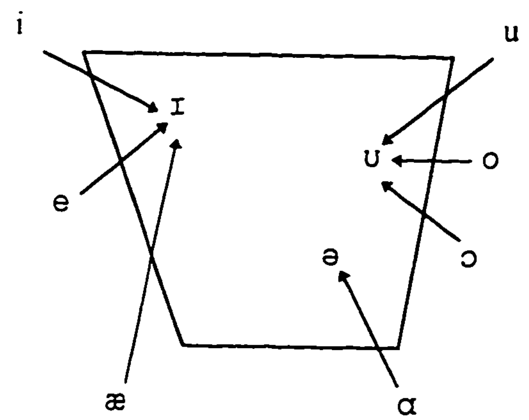


Figure A-1: Mapping of seven long vowels onto three short vowels in Urdu, illustrated in an F1-F2 vowel quadrilateral (not to scale).

Some examples of this process of vowel mapping are listed in (A.2) below.

(A.2)

Long Vowel to Short Vowel Mappings*			Base Verb
pi.na	→	pɪ.la	"to drink"
de.na	→	dɪ.la	"to give'
bæ.tʰna	→	bɪ.tʰa	"to sit"
bɑn.ta	→	bə.ta	"to distribute"
so.na	→	sʊ.la	"to sleep"
kud.na	→	kʊ.da	"to jump"

*stressed syllables are underlined

There is one exception to this three-short-vowel assertion. Abdullah mentions in the foreword of the work by Qureshi (1992) that {ʃhr}[10] meaning 'city' in Urdu is not pronounced as [ʃəhər], [ʃɪhər] or [ʃʊhər]. The vowel is similar to /æ/. The observation is correct, but there can be other possible explanations for this difference in quality. For example, the short vowel may be lengthened to a long vowel /æ/ before /h/. The hypothesis remains untested. It should also be noted here that short vowels do not occur word finally. This constraint is further discussed in the discussion of Urdu syllabification.

[10] The curly brackets '{}' are used to write Urdu grapheme transliteration. In Urdu only consonants and long vowels are written. Short vowels, represented by diacritics, have no graphemic representation. Thus, a string of consonants e.g. {ʃhr} represents a series of consonants with intervening short syllables. The vowel quality is determined by the diacritic used on the preceding consonant. For a complete guide to pronunciation from graphemes, see below.

In addition to the vowels, Urdu has thirty six consonantal phonemes. However, all these thirty six consonants, listed in (A.3) (Hussain 1993, Kachru 1987), are not those listed by Bokhari (1985). The list of consonants below and the list of consonants by Bokhari have the following differences. The labial fricative /f/, the alveolar fricative /z/, the palatal fricatives /ʃ/ and /ʒ/, the velar fricatives /x/ and /ɣ/ are not listed by Bokhari. He also lists the aspirated nasals /n/ and /m/, the aspirated lateral /l/, the alveolar flap /ɾ/ and the aspirated alveolar trill /r/. These consonants are not included in the list below because none of the examples given by Bokhari have these aspirates in the onset of words (while all other consonants, except /ɾ/, occur in word onsets). In addition, these aspirates, transcribed by Bokhari as 'Ch' (where 'C' is m,n,r,ɾ,l), always occur word medially such that there can possibly be a syllable boundary between 'C' and /h/ (e.g. /dul.ha/ "bridegroom") and therefore these aspirates can be explained as the 'C' followed by /h/. However, there is also a possibility that these aspirated consonants exist and that the words are syllabified as /du.l^ha/, with the aspirate in the onset of the second syllable. More research is needed to confirm the existence of these consonants. These aspirates are not listed by Kachru. The list also does not include the uvular stop /q/ articulated in some dialects of Urdu. This stop is increasingly being replaced by the velar /k/. This stop is listed by Kachru. In addition, the glide /w/ and fricative /v/ are allophonic in Urdu. These allophones have been listed as the fricative /v/ arbitrarily.

(A.3)

Phoneme	example	meaning
p	pɪn	"pin"
b	bɑl	"hair"
pʰ	pʰul	"flower"
b̤ʰ	b̤ʰɑlu	"bear"
t̪	t̪ɑlɑ	"lock"
d̪	d̪ɑl	"lentils"
t̪ʰ	t̪ʰɑl	"plate"
d̪ʰ	d̪ʰɑt	"metal"
t	tin	"tin"
d	dɑlnɑ	"to put"
tʰ	tʰokər	"hit"
dʰ	dʰɑl	"shield"
k	kɑl	"blackness"
kʰ	kʰɑl	"skin"
g	gɑl	"cheek"
gʰ	gʰər	"house"
t͡ʃ	tʃɑl	"walk"
d͡ʒ	dʒɑl	"trap"
t͡ʃʰ	tʃʰɑlɑng	"jump"
d͡ʒʰ	dʒulɑ	"swing"
f	fɪkər	"worry"
v	vɑr	"attack"
s	sɑl	"year"
z	zer	"defeat"
ʃ	ʃer	"lion"
ʒ	ʒɑlɑ bɑri	"hail storm"
x	xɑndɑn	"family"
ɣ	ɣussɑ	"anger"
h	hɪlna	"to move"
l	laltʃi	"greedy"
r	rɑt	"night"

ɾ	gaɾi	"car"
m	malɪk	"owner"
n	naʃtəh	"breakfast"
y	yar	"friend"
ʔ	ʔalɪm	"learned"

As outlined before, there are seven long and three short vowels and thirty six consonants in Urdu. In addition, there are more nasal vowels and aspirated consonants, which have also been proposed by Bokhari (1985), but their phonemic status still remains to be confirmed. Therefore, these proposed phonemes are not considered in this work. This appendix lists these phonemes with their transcription and pronunciation in IPA. Urdu script is written from right to left.

TRANSCRIPTION OF VOWELS

Urdu vowels are written with the help of three phonemes and three diacritics. The phonemes are 'ا' (pronounced as [əlɪf]), 'و' (pronounced as [vao]), and 'ے' (pronounced as [ye]), and the diacritics are 'َ' (pronounced as [zəbər]), 'ِ' (pronounced as [zer]), and 'ُ' (pronounced as [peʃ]). The long vowels are written using one of the three phonemes above with or without one of the diacritics on the preceding consonant. The short vowels are written with only a diacritic on the preceding consonant. However, normally these diacritics are not written in Urdu script. Therefore, the short vowels, which are represented only by these diacritics, are not explicitly written in Urdu script, but are memorized. These vowels are listed below. Because a consonant is needed to represent the vowels (as the diacritics go on the preceding consonant and not on the vowels), the vowels are written with the consonant 'ب' (pronounced as [b]).

Vowel	Letter	Example Syllable	Script
i	zer + ye	bi	بِی
e	ye	be	بے
æ	zəbər + ye	bæ	بَے
a	zəbər + əlɪf	bɑ	بَا
o	vao	bo	بو
ɔ	zəbər + vao	bɔ	بَو
u	peʃ + vao	bu	بُو
ɪ	zer	bɪ	بِ
ə	zəbər	bə	بَ
ʊ	peʃ	bʊ	بُ

TRANSCRIPTION OF CONSONANTS

Urdu consonants are derived from both Arabic and Persian and have been modified since then. Therefore, sometimes more than one letter stands for the same sound. Also, the aspirated stops do not have a unique transcription but are transcribed with the "stop + h" sequence. In addition, the three letters used for vowel transcription ([əlɪf, vao, ye]) are also used to write consonants (/ʔ/, /v, w/ and /y/ respectively). In this case, these letters are marked with the diacritics (and not the consonants before them). All these consonants are listed in isolation and with the vowel /a/ below. The consonants have a noticeably different transcription depending on whether they are connecting with an adjacent letter. The details of different transcriptions are not explained here.

Consonant	Letter		Example Syllable	Script
p	pe	پ	pɑ	پا
b	be	ب	bɑ	با
p^h	p^he	پھ	p^hɑ	پھا
b̤^{h}	b̤^{h}e	بھ	b̤^{h}ɑ	بھا
ṯ	ṯe	ت	ṯɑ	تا
	ṯoe	ط	ṯɑ	طا

d̪	d̪al	د	d̪ɑ	دا
t̪ʰ	t̪ʰe	تھ	t̪ʰɑ	تھا
d̪ʰ	----	دھ	d̪ʰɑ	دھا
t	te	ٹ	tɑ	ٹا
d	dal	ڈ	dɑ	ڈا
tʰ	tʰe	ٹھ	tʰɑ	ٹھا
dʰ	----	ڈھ	dʰɑ	ڈھا
k	kaf	ک	kɑ	کا
	qaf	ق	qɑ / kɑ	قا
kʰ	----	کھ	kʰɑ	کھا
g	gaf	گ	gɑ	گا
gʰ	----	گھ	gʰɑ	گھا
t͡ʃ	t͡ʃe	چ	t͡ʃɑ	چا
d͡ʒ	d͡ʒim	ج	d͡ʒɑ	جا
t͡ʃʰ	----	چھ	t͡ʃʰɑ	چھا
d͡ʒʰ	----	جھ	d͡ʒʰɑ	جھا
f	fe	ف	fɑ	فا
v	vao	و	vɑ / wɑ	وا

s	sin	س	sɑ	سا
	se	ث	sɑ	ثا
	saḍ	ص	sɑ	صا
z	ze	ز	zɑ	زا
	zal	ذ	zɑ	ذا
	zaḍ	ض	zɑ	ضا
	zoe	ظ	zɑ	ظا
ʃ	ʃe	ش	ʃɑ	شا
ʒ	ʒe	ژ	ʒɑ	ژا
x	xe	خ	xɑ	خا
ɣ	ɣæn	غ	ɣɑ	غا
h	he	ه	hɑ	ها
l	lam	ل	lɑ	لا / لا
r	re	ر	rɑ	را
ɾ	əɾe	ڑ	ɾɑ	ڑا
m	mim	م	mɑ	ما
n	nun	ن	nɑ	نا
y	ye	ی	yɑ	یا

ʔ	həmza	ء	ʔɑ	أَ
	ʔæn	ع	ʔɑ	عا

APPENDIX B. INDIVIDUAL SPEAKER DATA

There were seven speakers recorded, four males (AR, BS, SH, and ZA) and three females (AA, AS, and SA) which also included the author (SH). Individual data from these speakers, averaged over ten repetitions for each measurement, are presented in the tables below. The title of each table explains the measurements tabulated. Statistically significant differences ($p < .05$) are bold and italicized.

Table B-1: Means of duration of first syllable long vowels by all speakers

vowel	paṭa	paṭal	beta	beṭab	ḍiḍa	ḍiḍar	koke	koken	bætʰa	bæṭal	suba	ṭumar
stress	+	-	+	-	+	-	+	-	+	-	+	-
AA	129	132	110	108	117	122						
AR	100	95	***88***	***75***	76	83	82	78				
AS	128	121	***129***	***89***	***123***	***105***	***110***	***100***	***133***	***107***	101	93
BS	***142***	***114***	***112***	***107***	***115***	***109***	***107***	***85***	***130***	***108***	***94***	***87***
SA	112	114	100	101	120	109	***118***	***94***	111	114	93	88
SH	103	113	***102***	***90***	87	83	***87***	***71***	112	108	***73***	***56***
ZA	86	83	***87***	***72***	85	87	82	78	89	88	74	76

Table B-2: Means of duration of first syllable short vowels by all speakers

vowel	bɪkna	bɪkvana	pəkna	pəkvan	pəkvana	pugna	pugvana
Stress	+	-	+	-	-	+	-
AA	67	59	50	50	53	68	61
AR	51	42	42	39	40	40	34
AS	78	64	77	70	70	83	70
BS	49	47	58	60	61	53	58
SA	55	48	47	43	47	41	40
SH	46	34	40	34	31	26	32
ZA	40	40	38	33	29	34	27

Table B-3: Means of duration of second syllable long vowels by all speakers

vowel	paṭa	paṭal	beta	beṭab	ḍiḍa	ḍiḍar	koke	koken	bætʰa	bæṭal	suba	ṭumar
stress	-	+	-	+	-	+	-	+	-	+	-	+
AA	130	206	137	186	157	247						
AR	101	157	112	149	121	193	93	147				
AS	109	200	123	186	141	213	99	190	95	201	137	209
BS	117	161	125	176	134	164	110	127	134	149	133	172
SA	117	229	116	207	138	248	106	211	86	218	131	241
SH	106	144	108	156	120	187	89	134	83	145	120	179
ZA	99	130	93	121	103	153	93	112	84	121	110	146

Table B-4: Means of duration of second syllable vowels by all speakers

vowel	bɪkna	bɪkvana	pəkna	pəkvan	pəkvana	pugna	pugvana
stress	-	+	-	+	+	-	+
AA	138	174	132	238	178	140	183
AR	105	130	100	182	129	109	145
AS	145	165	136	230	159	141	175
BS	117	111	115	144	123	120	119
SA	117	119	105	220	124	121	149
SH	99	102	85	160	102	104	124
ZA	80	123	81	178	128	85	78

Table B-5: Means of F0 of first syllable long vowels by all speakers

vowel	paṭa	paṭal	beta	beṭab	ḍiḍa	ḍiḍar	koke	koken	bætʰa	bæṭal	suba	ṭumar
stress	+	-	+	-	+	-	+	-	+	-	+	-
AA	***196***	***205***	***202***	***215***	214	218						
AR	***133***	***138***	140	143	148	155	148	146				
AS	***211***	***232***	***213***	***236***	***222***	***240***	***221***	***243***	***204***	***225***	***234***	***250***
BS	***138***	***127***	140	137	146	139	***151***	***142***	139	138	148	145
SA	218	221	***219***	***230***	219	229	231	234	***219***	***228***	239	247
SH	102	101	104	104	101	102	113	112	101	100	104	102
ZA	143	145	139	141	142	144	147	145	139	138	***147***	***157***

Table B-6: Means of F0 of first syllable short vowels by all speakers

vowel	bɪkna	bɪkvana	pəkna	pəkvan	pəkvana	pugna	pugvana
stress	+	-	+	-	-	+	-
AA	*217*	*233*	220	239	240	237	247
AR	*154*	*160*	*162*	160	*167*	*162*	*165*
AS	*222*	*252*	*224*	260	*259*	*233*	*259*
BS	121	121	*124*	116	*116*	*127*	*120*
SA	208	208	192	192	203	222	221
SH	102	98	102	113	111	99	101
ZA	145	147	153	153	150	156	156

Table B-7: Means of F0 of second syllable long vowels by all speakers

vowel	paṭa	paṭal	beta	beṭab	ḍiḍa	ḍiḍar	koke	koken	bætʰa	bæṭal	suba	ṭumar
stress	-	+	-	+	-	+	-	+	-	+	-	+
AA	*255*	*197*	*258*	*198*	*257*	*200*						
AR	*161*	*146*	*161*	*146*	*163*	*148*	169	160				
AS	*257*	*211*	*259*	*208*	*249*	*209*	*241*	*223*	*259*	*210*	*248*	*214*
BS	150	147	142	142	148	151	151	159	*157*	*148*	*149*	*158*
SA	*255*	*219*	*251*	*218*	*256*	*220*	*267*	*233*	*259*	*222*	*259*	*226*
SH	103	101	103	102	101	101	105	105	*102*	*100*	102	101
ZA	*155*	*141*	*154*	*141*	*153*	*138*	*154*	*144*	*160*	*140*	*155*	*144*

Table B-8: Means of F0 of second syllable long vowels by all speakers

vowel	bɪkna	bɪkvana	pəkna	pəkvan	pəkvana	pugna	pugvana
stress	-	+	-	+	+	-	+
AA	*183*	*155*	*184*	*165*	*154*	*184*	*153*
AR	*267*	*199*	*261*	*201*	*199*	*268*	*199*
AS	*262*	*206*	*257*	*212*	*213*	*262*	*210*
BS	*129*	*139*	*128*	*140*	126	*132*	*137*
SA	*265*	*214*	*261*	*220*	*211*	*261*	*209*
SH	*105*	*102*	104	103	102	*107*	*102*
ZA	*168*	*148*	*172*	*149*	*153*	*174*	*148*

Table B-9: Means of intensity ratio of first syllable (long vowels)-to-second syllable vowels by all speakers

vowel	pat̪a	pat̪al	beta	bet̪ab	d̪id̪a	d̪id̪ar	koke	koken	bætʰa	bæt̪al	suba	t̪umar
stress	-	+	-	+	-	+	-	+	-	+	-	+
AA	1.86	0.93	1.31	0.75	0.49	0.46						
AR	0.57	0.83	0.30	0.49	0.10	0.14	0.86	1.07				
AS	1.06	1.74	1.79	3.40	0.31	0.57	1.43	1.64	2.16	1.69	0.34	0.88
BS	1.35	0.84	0.36	0.58	0.23	0.23	1.18	2.28	1.55	0.67	0.38	0.48
SA	0.88	0.91	0.78	1.45	0.12	0.22	1.42	2.49	1.24	0.77	0.20	0.19
SH	1.37	0.88	1.36	1.32	0.52	0.58	1.59	1.38	1.70	1.17	0.57	0.53
ZA	1.03	1.28	1.18	1.71	0.64	1.15	1.22	1.39	1.70	1.57	0.49	0.61

Table B-10: Means of intensity ratio of first syllable (short vowels)-to-second syllable vowels by all speakers

vowel	bɪkna	bɪkvana	pəkna	pəkvan	pəkvana	pugna	pugvana
stress	+	-	+	-	-	+	-
AA	1.12	1.25	1.31	1.52	1.34	1.18	1.34
AR	0.55	0.61	0.25	0.33	1.27	0.48	1.00
AS	1.75	1.67	2.36	1.30	2.08	1.66	1.39
BS	1.31	0.78	1.27	0.96	0.99	0.96	0.68
SA	0.87	0.87	0.55	0.85	0.73	0.78	0.92
SH	0.90	0.67	0.57	0.47	1.18	0.73	0.72
ZA	1.68	1.26	0.86	0.72	1.53	1.13	0.88

Table B-11: Means of formants of first syllable long vowels by all speakers

	vowel	paṭa	paṭal	beta	beṭab	ḍiḍa	ḍiḍar	koke	koken	bæt^{h}a	bæṭal	suba	ṭumar
	stress	+	-	+	-	+	-	+	-	+	-	+	-
AA	F1	***967***	***526***	***401***	***423***	224	231						
	F2	***1405***	***1485***	***2495***	***2381***	2655	2695						
AR	F1	704	685	402	402	241	241	399	400				
	F2	1173	1184	1975	1959	2295	2280	***830***	***881***				
AS	F1	697	697	***434***	***478***	351	380	***449***	***486***	611	629	***426***	***489***
	F2	***1482***	***1637***	***2261***	***2130***	***2576***	***2455***	***985***	***1184***	2014	2008	1195	1345
SA	F1	806	787	435	454	290	302	450	458	***645***	***477***	***309***	***357***
	F2	1572	1563	***2152***	***2086***	2375	2359	908	924	***1830***	***1965***	836	815
BS	F1	768	734	***433***	***530***	284	285	433	416	***728***	***622***	***333***	***367***
	F2	1346	1368	***2009***	***1801***	***2410***	***2330***	846	872	1604	1680	***1022***	***977***
SH	F1	***601***	***580***	403	402	259	266	389	389	516	506	***270***	***295***
	F2	1173	1177	1912	1868	***2211***	***2145***	860	885	1161	1617	***910***	***813***
ZA	F1	622	601	383	392	269	259	368	358	***559***	***591***	242	223
	F2	***1133***	***1188***	***1843***	***1883***	2105	2097	828	848	***1588***	***1625***	925	891

Table B-12: Means of formants of first syllable short vowels by all speakers

	vowel	bɪkna	bɪkvana	pəkna	pəkvan	pəkvana	pugna	pugvana
	stress	+	-	+	-	-	+	-
AA	F1	***404***	***355***	***572***	***505***	***489***	436	440
	F2	2349	2333	1485	1444	1463	1098	1137
AR	F1	***385***	***329***	578	526	542	372	348
	F2	1933	1870	1229	1227	1225	801	785
AS	F1	***437***	***463***	***641***	***572***	***567***	463	471
	F2	***2325***	***2194***	1565	1554	1573	1214	1172
BS	F1	432	433	***647***	651	***585***	***377***	***435***
	F2	***1950***	***1788***	1342	1379	1313	***936***	***1089***
SA	F1	***364***	***383***	406	395	415	372	389
	F2	2227	2164	1162	1175	1156	***912***	***795***
SH	F1	351	348	***508***	***438***	***442***	328	351
	F2	***1834***	***1702***	***1189***	***1110***	***1138***	859	866
ZA	F1	287	282	***491***	***406***	***407***	***288***	***337***
	F2	***1991***	***1931***	***1149***	***1119***	***1077***	***802***	***886***

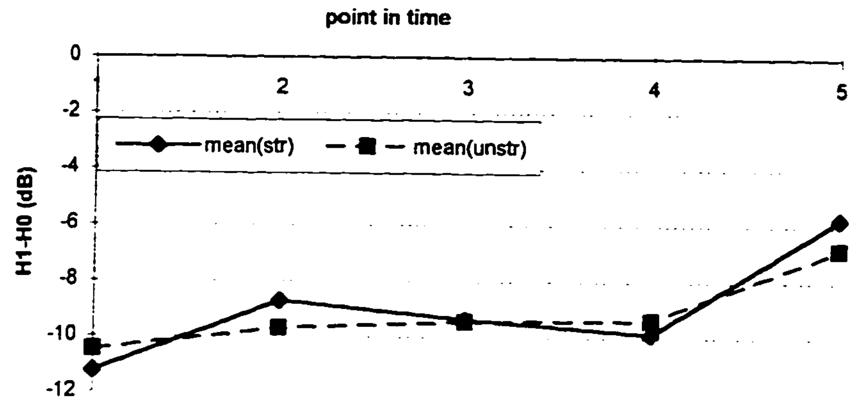
Table B-13: Means of durations of closure of stops in onset position by all speakers

stop	p		t̪		t		k		b		d		pʰ		t̪ʰ		kʰ	
stress	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
AA	***145***	***116***	***140***	***115***	***94***	***78***	***109***	***75***	***112***	***91***	***123***	***78***	***129***	***105***	***119***	***105***	***104***	***71***
AR	***83***	***72***	61	60	***59***	***38***	73	68	77	78	72	68	68	62	46	34	50	39
AS	***106***	***90***	***83***	***69***	***82***	***65***	***89***	***78***	***98***	***85***	***84***	***66***	100	91	82	77	***85***	***69***
BS	103	97	***93***	***85***	***85***	***70***	82	85	***87***	***71***	***85***	***72***	90	82	102	86	81	81
SA	***135***	***113***	***138***	***108***	***118***	***89***	***125***	***98***	106	102	124	112	106	106	***131***	***95***	106	93
SH	***79***	***70***	***63***	***53***	***72***	***64***	61	53	75	70	65	59	***93***	***80***	84	75	***75***	***62***
ZA	82	81	88	92	***76***	***67***	71	69	70	73	76	76	***87***	***80***	92	85	***66***	***60***

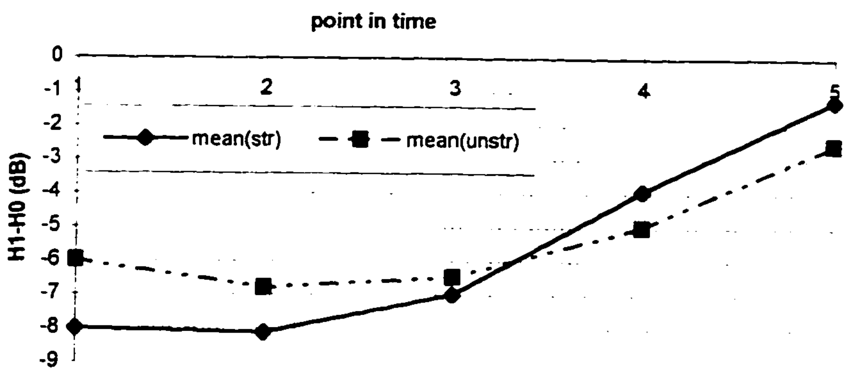
stop	bʰ		d̪ʰ		dʰ		gʰ	
Stress	+	-	+	-	+	-	+	-
AA	***126***	***90***	93	90	94	87	***103***	***80***
AR	65	57	35	30	41	42	41	45
AS	88	80	69	74	61	55	65	72
SA	95	90	***79***	***62***	***73***	***55***	***81***	***51***
SH	75	67	71	71	***66***	***70***	***75***	***59***
ZA	***76***	***66***	63	70	72	65	65	62

Table B-14: Means of durations of aspiration of stops in onset position by all speakers

stop	p		t̪		t		k		b		d		pʰ		t̪ʰ		kʰ	
Stress	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
AA	9	11	18	16	27	29	38	44	5	4	15	8	48	48	58	55	89	85
AR	8	11	14	15	12	17	25	25	5	8	13	11	37	31	46	40	47	38
AS	9	11	19	18	28	31	25	22	4	3	10	7	38	32	***51***	***36***	***75***	***61***
BS	11	11	15	16	23	30	36	31	9	9	9	11	58	51	63	50	73	67
SA	13	14	19	20	14	15	30	28	1	2	13	10	***66***	***44***	***62***	***53***	***93***	***80***
SH	17	15	24	26	13	14	39	37	8	11	13	15	39	43	35	36	60	55
ZA	13	13	14	14	12	12	25	25	6	8	6	7	52	47	46	49	***66***	***59***



ZA harm



SH harm

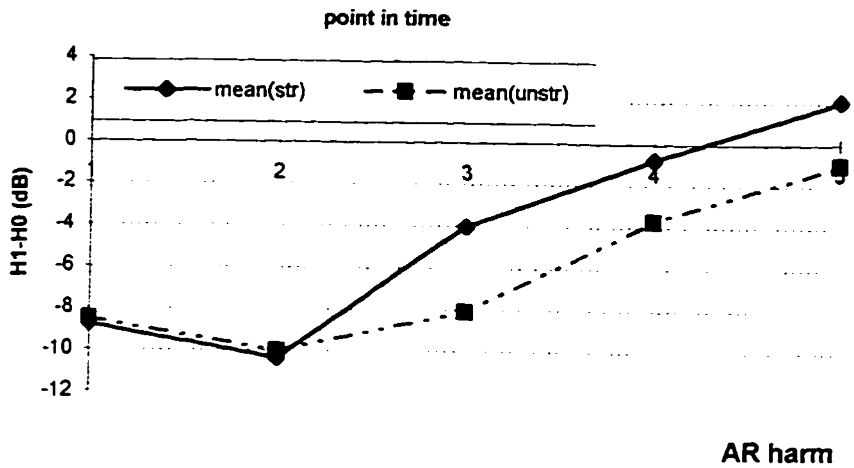


Figure B-1: Plots of ratio of first harmonic to fundamental (in dB) for stressed vs. unstressed CV syllables (where C is a breathy stop) averaged over all stops at five equally spaced points for male speakers ZA, SH and AR (see figure 4.10 and the text before it for explanation).

Table B-15: Means of durations of closure of stops in coda position by all speakers

stop	voiceless		voiced		aspirated		breathy	
Stress	+	-	+	-	+	-	+	-
AA	85	84	111	112	***64***	***91***	75	71
AR	***95***	***81***	92	99	***64***	***85***		
AS	99	89	104	103	***65***	***105***	104	103
BS	75	73	72	68			107	77
SA	73	79	108	113			***96***	***82***
SH	***67***	***57***	***71***	***57***	56	50	81	78
ZA	48	48	51	50			80	74

Table B-16: Means of durations of voicing of voiced stops in coda position by all speakers

stop	voiced	
Stress	+	-
AA	46	42
AR	*32*	*19*
AS	*50*	*38*
BS	*25*	*14*
SA	63	47
SH	*31*	*20*
ZA	*29*	*23*