CONTRIBUTIONS TO NEPALESE STUDIES
CNAS, TRIBHUVAN UNIVERSITY

Vol. 12, No. 2
April 1985

Acoustic Analysis of the Maithili Diphthongs

Sunil Kumar Jha
S.M.B.M.Campus, Rajbiraj
Tribhuvan University

Introduction

There have been various studies on diphthongs (e.g. Hibbitt, 1948; Trager and Smith, 1951; Gleason, 1955; Lehiste and Peterson, 1961; Malmberg, 1963; Gay, 1968; Gerber, 1972) during the previous four decades or so. But the interpretations and definitions of diphthongs given by different researchers do not always agree. For example, some researchers (e.g. Hibbitt, 1948; Gleason, 1955; Catford, 1977) hold that a diphthong is a blend of two perceptually different vowels within one and the same syllable; others (e.g. Trager and Smith, 1951) hold that it is a sequence of vowel and semivowel; while still others (e.g. Lehiste and Peterson, 1961; Malmberg, 1963; Gay, 1968; Gerber, 1972) hold that it is a vocalic syllable nucleus containing two 'target' positions.

It is primarily due to the different criteria followed in the investigation of diphthongs — whether articulatory, whether articulatory, perceptual, acoustic, or some combination of these — that different researchers have resorted to different definitions of what a diphthong is. On the whole, most researchers seem to agree that in point of fact a diphthong may, indeed, consist of two discrete 'elements', with a relatively rapid transition between them. According to most of them, a diphthong may be described and identified in terms of its beginning and ending points, using the categories for monophthongs with the assumption that the articulators, in their movement, take the shortest path between these points.

But the question is: Is a diphthong really a pair of monophthongs, end-to-end? It seems that in the articulatory domain a diphthong is a single event characterized by the change of articulator position. In the perceptual domain, however, there is the question whether diphthongs are identified as single events — i.e. as individual vocalic syllable nuclei — or as dual events. Put slightly differently, the question still is: Is the perceived diphthong different from the perception of a pair of vowels? It is with the help of some attributes gathered from the results of a spectrographic study of the Maithili diphthongs that we shall make an attempt to find out some possible answers to this question.

Procedures

For the purpose of the present study, a list of test utterances containing the two Maithili diphthongs, /ə i/ and /ə u/, was prepared,

keeping the diphthongs both in isolation[2] and in word contexts. Table I lists these test utterances:

Table I: Test utterances containing the Maithili diphthongs spoken both in isolation and in word contexts.

Diph-thongs	Initial	Medial	Final
/ə i/	/'əik^{h}ən/'at this time'	/kəil/'of a white colour'	/del'kəi/'(a third person) gave (something to another third person)'
/ə u/	/'əuk^{h}ən/'even	/kəul$^{(h)}$/'small mill'	/del'kəu/'(a third person) gave (something to a second person)'

All these test utterances were recorded under three different speaking rates: (1) at a slow speed (i.e. recorded in isolation); (2) at a moderate speed (i.e. recorded in words); and (3) at a fast speed (i.e. recorded in words put in a normal conversational sentence context, the frame of the sentence being /'p^{h}ero . . . kəhi'yəu/ 'Please say . . . . again').

Nine tokens of each test utterance were recorded in a soundproof studio at Essex University. All these recordings were make on a Revox B 77 tape-recorder. Of the nine tokens of each test utterance, the first two as well as the last two tokens were ignored, and all the remaining five tokens from the middle were used to obtain visual records in the form of spectrograms. Wideband spectrograms and wideband 'sections' (i.e. short-term spectrum) of each of the five tokens were made on a Sona-Graph 6061-B Sound Spectrograph, Kay Electronic Co., Pine Brook, N.J.

The formant frequency of each token was measured from wideband spectrograms, since these were found to offer the most convenient way of locating the centre frequency of each formant. Duration measurements were made from all the five tokens of each test utterance as recorded in various contexts. Formant frequency and duration measurements were made to the nearest 20 Hz and 5 ms, respectively.

Results and Discussions

The two main objectives of the present study were to determine:

(1) the average frequency of F_1 and F_2 for each of the two elements (i.e. of the 'onset' and 'offset' steady states) as well as the rate of change of F_2 for each diphthong as recorded under the three speaking rates; and

(2) the duration of the onset and offset steady states, the duration of the F_2 transition from the initial element to the final element as well as the overall duration of the two diphthongs in each one of the three speaking rates.

The 'steady state' was defined as the time interval during which formants remained parallel to the time axis; the 'transition state' was considered in terms of the rate of change of the formant frequency in a given time interval (i.e. Hz per ms).

Formant Frequency Measurement Results and Discussions

The average frequency values for the onset and offset steady states of the two diphthongs as well as the rate of change of F_2 for each diphthong are all given in Table II[3]. The table shows the effects of duration conditions on F_1 and F_2 target[4] frequencies. For both diphthongs, there is little consistent change in onset frequency positions for the three duration conditions. But both F_1 and F_2 offsets show consistent changes in frequency levels across changes in duration. For /əi/, F_1 offset frequency values are higher and F_2 offset frequency values are lower as the diphthong duration

Table II: Average formant frequencies (in Hz) for the onset and offset steady states of the Maithili diphthongs and their F_2 rate of change for the three conditions of speaking rate.

Diphthong	Rate of speaking	F_1		F_2		F_2 rate of change (in Hz/ms)
		Onset	Offset	Onset	Offset	
/əi/	slow	485	255	1090	2450	12.95
	moderate	480	310	1120	2170	12.35
	fast	475	360	1140	1880	12.33
/əu/	slow	475	310	1080	650	4.78
	moderate	470	350	1100	775	4.64
	fast	465	395	1110	840	4.50

decreases under the three conditions of speaking rate; for /əu/, both F_1 and F_2 offset frequency values are higher for conditions of shorter duration. Another important point to note is that the rate of change of F_2 for each diphthong is relatively constant across changes in duration.[5] This rate of change of F_2 for the diphthong /əi/ has a range between 12.33 and 12.95 in Hz per ms; for /əu/ the rate of change has a range between 4.50 and 4.78 in Hz per ms.

To show the effects of these changes on the identity of the 'targets', we first plot in Figure 1 the average F1 F2 frequency values for the three steady state vowels /ə i u/ spoken in isolation.

The average F_1 and F_2 frequency values for these vowels are given in Table III[6]. In Figure 1 the three steady state vowels are shown by three bold dots. The average F_1 and F_2 frequency values for the steady state onset and offset positions of the diphthongs /əi/ and /əu/ as spoken under the three conditions of speaking rate are plotted on an F_1-F_2 grid in Figure 2 against the three boldly dotted positions of the steady state vowels: /ə i u/:

Table III

Vowel	F_1	F_2
/ə /	480	1088
/i /	258	2456
/u /	308	626

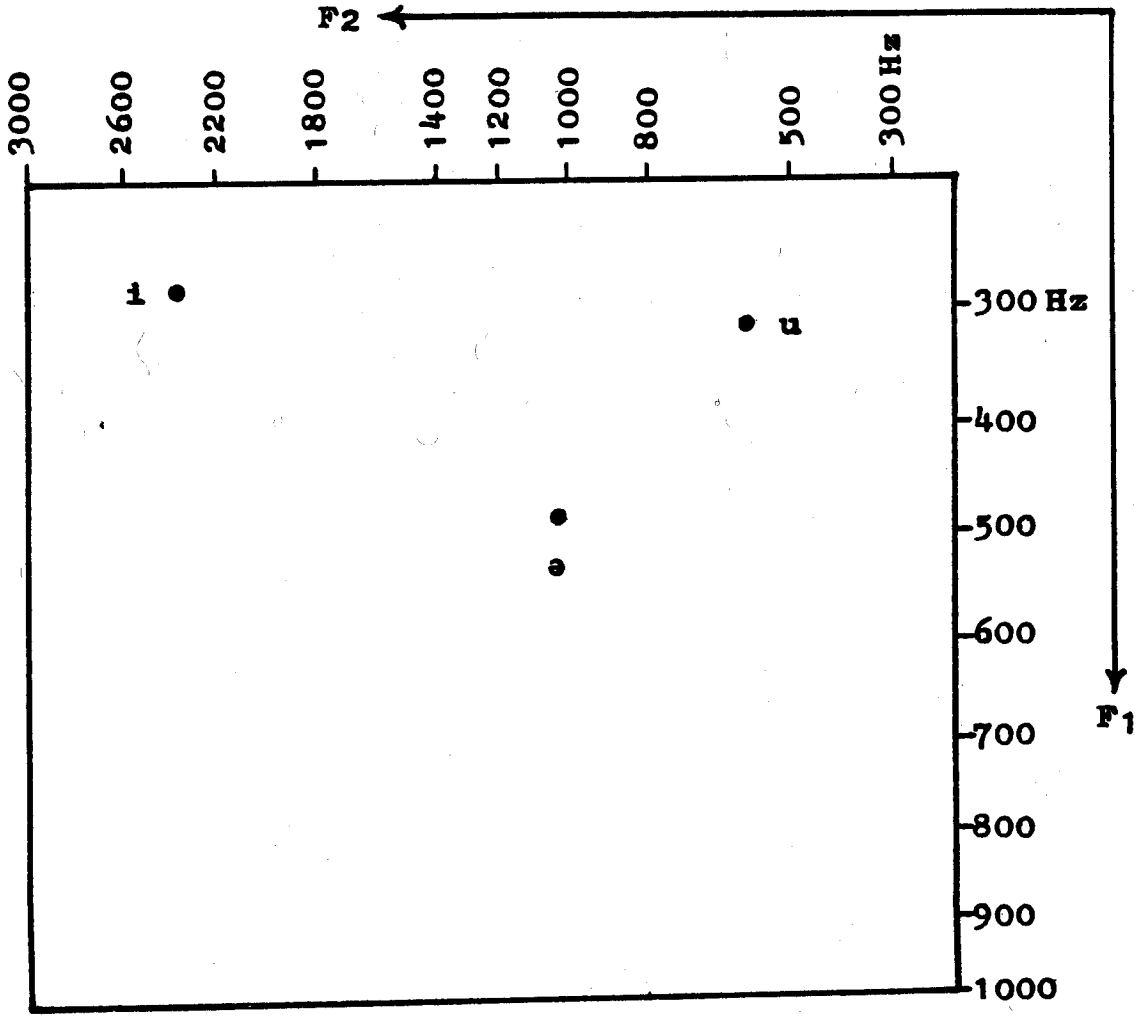

Figure 1: Average frequency values of F_1 and F_2 for the three Maithili steady state vowels: / ə i u/.

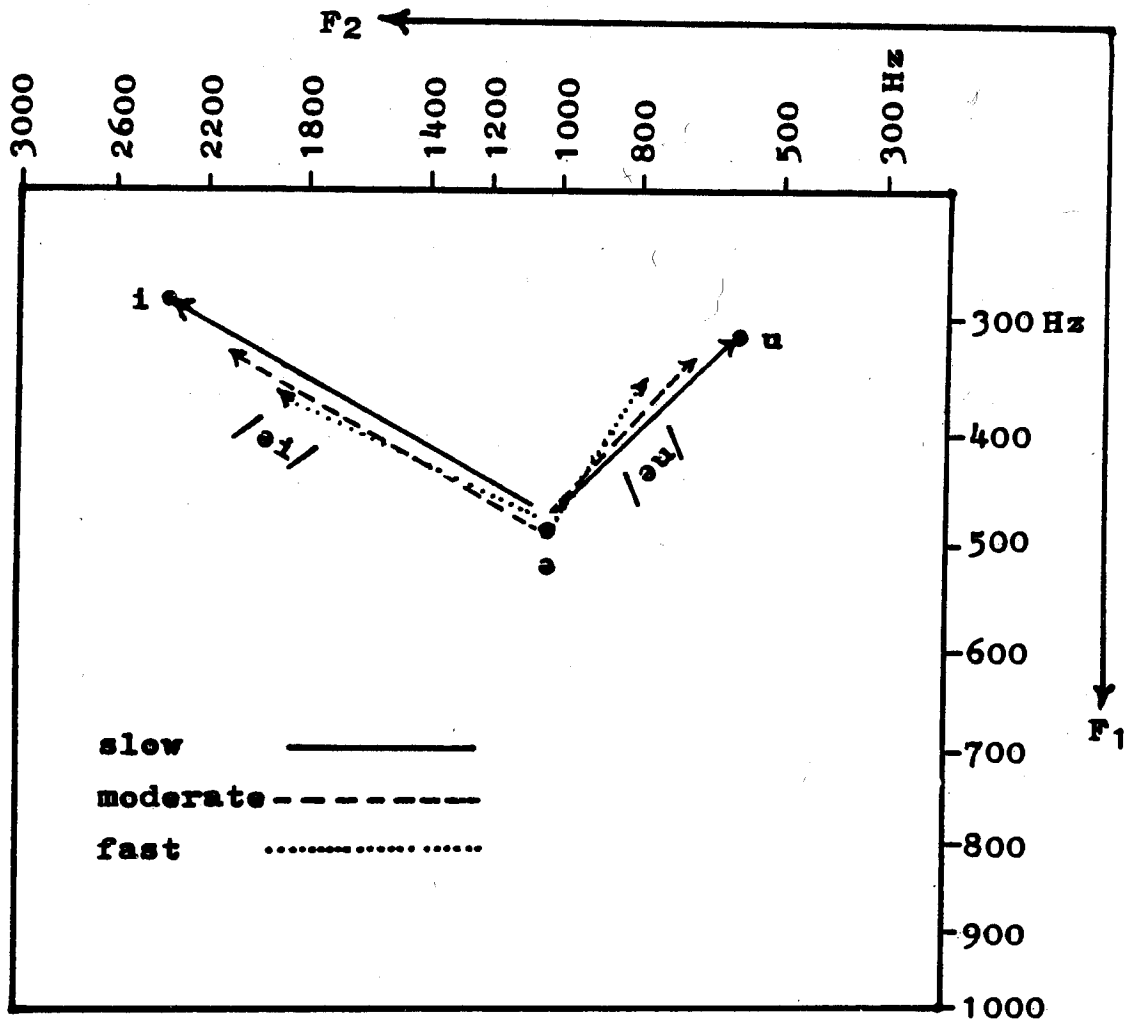

Figure 2: Average F_1 and F_2 values for the steady state onsets and offsets of the Maithili diphthongs / əi/ and / əu/ as recorded under the three conditions of speaking rate plotted on an F_1-F_2 grid against the three boldly dotted positions of the steady state Maithili vowels / ə i u/.

Figure 2 embodies:

(i) 'corresponding' start 'target': ;

(ii) 'corresponding' end 'target': i or u;

(iii) direction of movement: ———— ;

(iv) amount of movement: i.e. length of ———— ; and

(v) variation of rate of movement: i.e.

connected ———— vs broken – – – – – vs
dotted ..... lines.

As Figure 2 shows, each of the two diphthongs displays a clear change of offset identity; onset positions, on the other hand, are more or less preserved across changes in duration. The figure further illustrates that each diphthong follows a similar movement-direction across changes in duration with the shorter utterances terminating before reaching their targets. In other words, the plots in Figure 2 clearly display that it is only in conditions of slow speaking rate that the offset targets of the diphthongs /əi/ and /əu/ are reached, while in conditions of both moderate and fast speaking rates the offset targets are never quite reached. In general, the faster the conditions of speaking rate, the quicker the termination of the offset movement-direction of each diphthong.

Thus, the results of our spectrographic study show that the Maithili diphthongs /əi/ and /əu/ are each characterized by two attributes:

(1) a comparatively fixed onset steady state; and

(2) continuously as well as systematically changing glide elements, the main property of this systematic continuous change being a very similar rate of change of F_2 across changes in the three duration conditions.

Offset steady states are found least prominent in these diphthongs - in fact, each diphthong displays a clear change of offset identity across changes in duration conditions. Hence, if we assume - as the symbols traditionally used to write the diphthongs /əi/ and /əu/ would suggest - that the offset targets of /əi/ and /əu/ are the steady state vowels /i/ and /u/, respectively, then it is only in conditions of slow speaking rate that these offset targets are reached, while in faster conditions of speaking rate the offset movement-direction of each diphthong always terminates before reaching its target.

Duration Measurement Results and Discussions

In addition to the frequency of vowel formants, other parameters involved in the perception of the Maithili diphthongs - like the diphthongs of other languages, for that matter - concern their duration. Therefore, studies on the duration of these diphthongs were also made. The results of these duration measurements are given in Table IV. This table shows average durations (in ms) of the onset steady state, the offset steady state, the F_2 transition as well as the total duration of the two diphthongs as recorded under the following three conditions of speaking rate: slow, moderate and fast.

To show exactly what these measurements refer to, Figure 3 displays spectrograms of two words - i.e. /['əik^{h}ən]/ 'at this time' and /[kəhi'yəu]/ '(please) say' - containing the two diphthongs. Both words were spoken at a moderate speed.

Table IV: Average durations (in ms) of the onset steady state, the offset steady state, the F_2 transitions and the total duration of the two Maithili diphthongs as recorded under the three conditions of speaking rate.

Diph-thong	Rate of Speaking	a	b	c	Total
		Onset	F_2 transition	Offset	
/əi/	slow	50	105	80	235
	moderate	25	85	40	150
	fast	10	60	20	90
/əu/	slow	40	90	60	190
	moderate	20	70	30	120
	fast	10	60	15	85

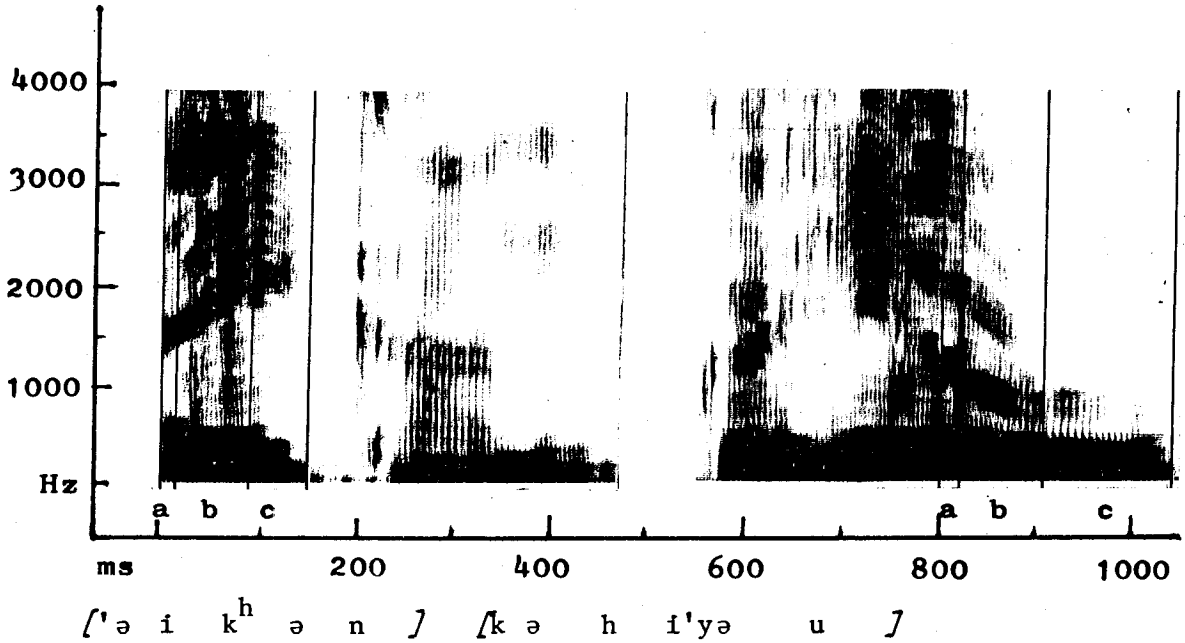

Figure 3: Spectrograms of two Maithili words, /['əik^{h}ən]/ and /[kəhi'yəu]/, spoken at a moderate speed.

A time scale (with an arbitrary starting point) is shown in ms along the bottom of the horizontal axis; the vertical axis shows the frequencies in Hz. in Figure 3:

(i) 'a' embodies onset steady state;

(ii) 'b' embodies F_2 transition; and

(iii) 'c' embodies offset steady state.

The average F_2 frequency values for the two diphthongs given earlier in Table III and the average durations for the same diphthongs given in Table IV are graphically illustrated in Figure 4. In Figure 4, too, 'a', 'b' and 'c' embody onset steady state, F_2 transition, and offset steady state, respectively. Connected lines of the figure show the two diphthongs spoken at a slow speed; broken lines show them as spoken at a moderate speed; and dotted lines show them as spoken at a fast speed.

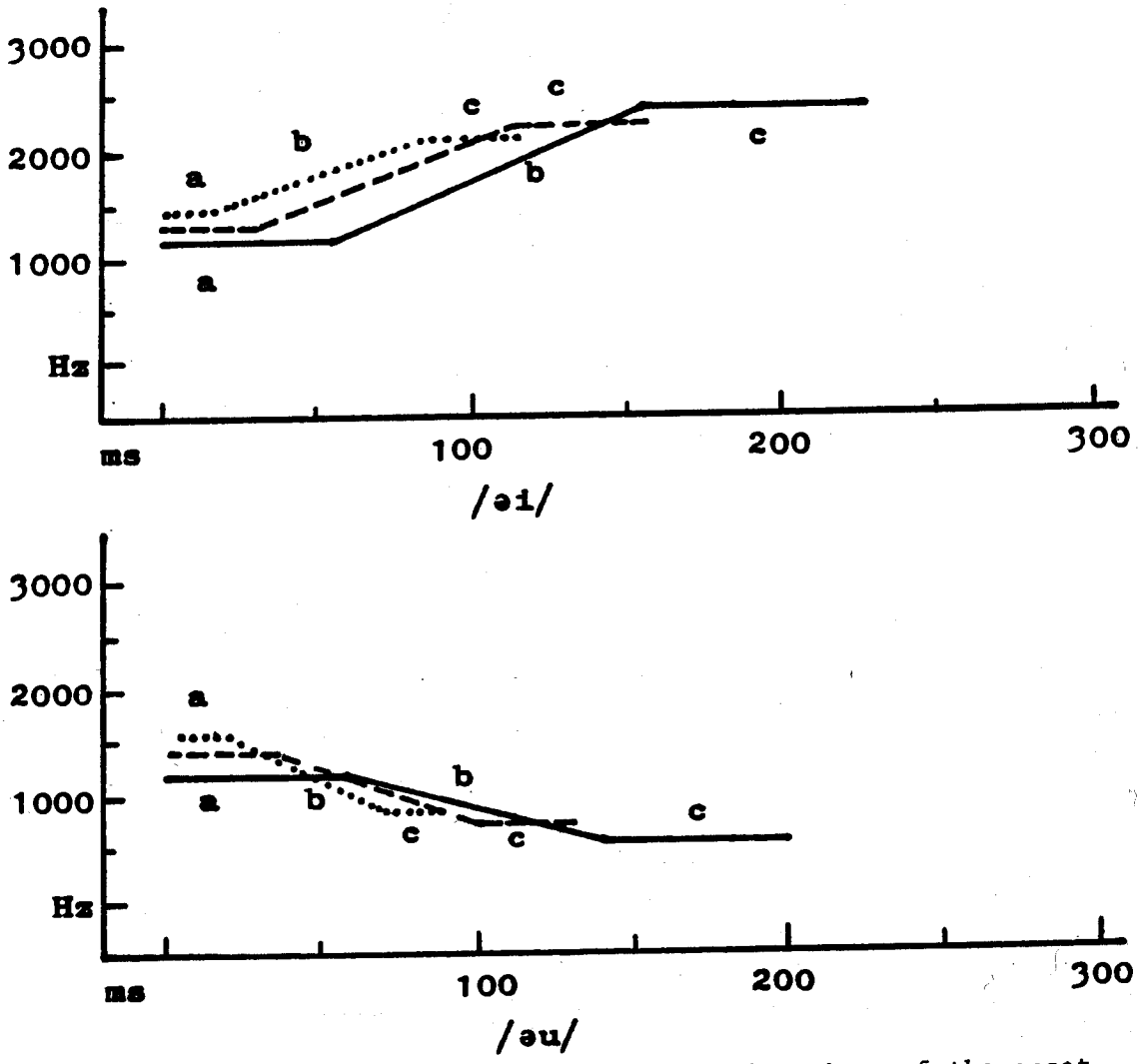

Figure 4: Average F_2 frequencies and average durations of the onset steady state, the F_2 transition and the offset steady state of the two Maithili diphthongs recorded under the three conditions of speaking rate: slow (connected lines), moderate (broken lines) and fast (dotted lines).

All the above tables and figures concerning the two diphthongs show that a change in the speaking rate from slow to fast brings about a reduction in the total duration ranging between 38 and 64%. The data also shows that the onset durations for both diphthongs are less than their offset durations. Although both onset steady state and F_2 transition durations decrease as the speaker's rate of production increases, the significant point to note about our data is that the decrease in F_2 transition duration is always accompanied by an increase in F_1 offset frequency and a decrease in F_2 offset frequency for /əi/ but an increase in F_2 offset frequency for /əu/. And there is no distinct modification of the rate of change of F_2 either in /əi/ or in /əu/, as spoken in all three conditions of speaking rate -- i.e. the rate of change of F_2 remains very stable in both diphthongs. This stability is a fact of the production. It seems that the speaker maintains the stability here because he believes this section of the diphthong to be that most relevant for diphthong-perception.

When recorded under moderate and fast speaking rates, both diphthongs were placed in all word positions: initial, medial and final. The formant frequency and duration measurements of these diphthongs do not indicate any consistent initial or terminal consonant effects on target frequency levels, or initial consonant effects on duration. However, there was some slight terminal consonant voicing-effect on duration. The increased duration of /əi/ and /əu/ preceding a voiced consonant - i.e. in the Maithili words /kəil/ 'of a white colour' and /kəul$^{(h)}$/'small mill' - is accomplished primarily by a lengthening of the steady state onset. F_2 transition and steady state offset durations are also slightly longer.[8]

Our analysis of the acoustic characteristics of the two Maithili diphthongs reveals a certain movement along the particular path between the zones corresponding to two different vowels - i.e. between /ə/ and /i/ or /u/ - in the vowel space. It has been observed that in a consonant diphthong-consonant context this movement is performed in five steps:

(i) transition from the consonant to the initial element of the diphthong;

(ii) steady state (or 'target') of the initial element - i.e. onset steady state;

(iii) transition from the initial target-element to the final target-element of the diphthong;

(iv) steady state (or 'target') of the final element - i.e. offset steady state; and

(v) transition from the final target-element of the diphthong to the following consonant.

These steps are more evident in a slow speaking rate; when the speaking rate is increased, they tend to shorten.

Thus, from the point of view of the production, it may be inferred that the articulatory gesture for the Maithili diphthongs begins at a rather fixed position in the vowel space, which is very similar to the position held by the Maithili steady state vowel /ə/. This position is held for a particular length of time (relative time, that is), the absolute duration of which is determined mainly by the rate of speaking and/or the presence or absence of a following voiced consonant. The articulators then begin to move toward a final target position at a given speed, the final target position being similar to the one that the steady state vowel /i/ or /u/ has. If the rate of speaking is slow, the final target is reached and the gesture is completed; if, on the other hand, the rate of speaking is fast, the movement, while still on course, is cut off before reaching the final target.

Hence, if we recall our earlier question: Is a diphthong really a pair of monophthongs, end-to-end?, our spectrographic evidence from Maithili suggests an answer which could be put as yes and no: yes, if the speaking rate is slow; no, not really, if the speaking rate is fast. The spectrographic measurements of our study, especially those of formant positions and overall movement, do suggest that each of the two Maithili diphthongs, /əi/ and /əu/, does have a first element which is the starting point and a second element which is the point in the direction of which the glide or movement is made. Our results clearly suggest that, in general, the faster the speaking rate, the quicker the termination of this movement of the articulators in the vowel space.

Conclusion

Form our findings and from what the literature has to say on this topic, it may be concluded that the phonetic features emerging from the consideration of diphthongs in a particular language are not necessarily sufficient or optimal to characterize the phenomenon in general. Hence, for example, in Spanish (e.g. Manrique, 1979) the prevalence of one steady state over the other depends on the position of the open vowels; in English (e.g. Lehiste and Peterson, 1961; Gay, 1968; Gerber, 1972) it is the onset steady state which is always most relevant; while our data suggest that the most significant attributes of the Maithili diphthongs are two: (1) their onset frequency position and (2) their second-formant rate of change, which is always very stable in these diphthongs across changes in duration conditions. This stability is a fact of the production. It seems that the speaker maintains the stability here because he believes this section of the diphthong to be that most relevant for diphthong-perception.

NOTES

1. This paper forms part of the fourth chapter of Jha's (1984) unpublished doctoral dissertation. The author is grateful to both Professor Marcel A. A. Tatham and Dr. Jacques Durand of the Department of Language and Linguistics at Essex University (England) for their detailed criticisms, comments and suggestions on various versions of this study which have led to its improvements both in style and in content.

2. It should be pointed out that both Maithili diphthongs, /ə i/ and /ə u/, are used as morphemes in this language.

3. For the detailed frequency values for each of the total recorded tokens, together with the test utterances and with the average and range of all five tokens of each test utterance, see Jha (1984: 330-332).

4. In this study 'target' positions are defined as the ones where both 'onset' and 'offset' maintain their own distinct steady states - i.e. the states in which the formants remain parallel to the time axis for a particular time interval.

5. Measurements of F_1 rates of change for these diphthongs were not done because of possible measurement error effects.

6. See Jha (1984: 311-312) for the detailed frequency values of every token of the steady state vowels /ə i ə u/ as spoken in isolation.

7. For the detailed duration measurement values for each token of all these test utterances, together with their average and range, see Jha (1984: 333-335).

8. See Jha (1984: 334-335) for more information with regards to the duration of /əi/ and /əu/ preceding the voiced consonants /l/ and /l^h/ in the words /kəil/ and /kəul$^{(h)}$/, as spoken under both moderate and fast speaking rates.

REFERENCES

Catford, J.C., 1977. Fundamental Problems in Phonetics. Edinburgh: Edinburgh University Press.

Gay, T., 1968. "Effect of speaking rate on diphthong formant movements". Journal of the Acoustical Society of America, 44, 1570-1573.

Gerber, S.E., 1972. "Perception of segmented diphthongs". In Rigault, A. and Charbonneau, R. (eds.), Proceedings of the Seventh International Congress of Phonetic Sciences. The Hague: Mouton and Co., 1972, 480-492.

Gleason, H.A., 1955. An Introduction to Descriptive Linguistics. New York: Henry Holt and Co.

Hibbitt, G.W., 1948. Diphthongs in American Speech. New York: (privately published).

Jha, S.K. 1984. A Study of Some Phonetic and Phonological Aspects of Maithili. Unpublished Ph.D Thesis, University of Essex.

Lehiste, I. and Peterson, G.E. 1961. "Transitions, glides and diphthongs". Journal of the Acoustical Society of America, 33, 268-277.

Malmberg, B. 1963. Structural Linguistics and Human Communication. Berlin: Springer-Verlag.

Manrique, A.M.B., 1979. "Acoustic analysis of the Spanish diphthongs". Phonetica, 36, 194-206.

Trager, G.L. and Smith, H.L., 1951. An Outline of English Structure. Norman, Oklahoma: Battenburg Press.