DURATION AND FORMANT FREQUENCIES OF ITALIAN BIVOCALIC SEQUENCES

PIER LUIGI SALZA

Dept. of Speech Synthesis
CSELT S.p.a.
10148 Torino, Italy

GIOVANNA MAROTTA

Laboratory of Linguistics
Scuola Normale Superiore
56100 Pisa, Italy

DAVIDE RICCA

Institute of Glottology
Universita' di Torino
10100 Torino, Italy

ABSTRACT

A professional speaker read four replications of 34 meanigful sentences containing ia, ai and ua, au bivocalic sequences, under different stress conditions, both in lexical words and at word boundaries. The following measurements were made on each test sequence: onset and offset frequency locations of first and second formants, durations of onset target, glide (transition), and offset target.

Results show that within word clear duration differences exist between diphthongs and vowel sequences in hiatus only in the onglide (rising) case. At word boundary the discriminant issue is the presence or absence of a phrase boundary, according to which word final vowel is preserved or reduced, respectively. In the latter case, the stress carried by final vowel is also lost.

1. INTRODUCTION

Bivocalic combinations are very common in Italian, both within words and at word boundaries. Traditional grammars usually classify the realizations permissible within word as diphthongs or as bisyllabic vowel sequences (hiatus). For word boundaries, they basically account for two phenomena: sinaloephe, i. e. the fusion of the two elements, and elision, where the final vowel of the word is dropped.

Moreover, as for diphthongs within words, traditional Italian phonology makes a distinction between onglide (rising) and offglide (falling) diphthongs, see e. g. Battaglia and Pernicone [1], Romeo [2], Muljačić [3], Tagliavini and Mioni [4], Canepari [5]. The presence of one segment marked [+high] and the presence of stress on the [-high] vowel are invoked as the main conditions for diphthong realization. Nonetheless, also in unstressed contiguous vowels contrasting pronunciations are meant to occur depending on morphological and phonological rules, see e. g. Romeo [2] and Muljačić [3]. In conclusion, the studies quoted above point to the phonetic transcription scheme shown in the left column of Table I (where [a] is selected as [-high] full vowel).

Experimental work undertaken in the past year aimed at measuring the segmental durations which a speaker performs to produce contrasting diphthongs and vowel sequences in a wide corpus of nonsense words [6]. The results indicated that stressed structures could barely be differentiated in diphthongs and hiatuses, as long as reduction phenomena were not present; in unstressed cases only onglide diphthongs were clearly characterized by very short durations of the first element, namely [j] and [w], while no strong evidence was shown for postvocalic reduction.

In the present work a further assessment of the pertinence of the classification adopted by traditional phonology in lexical words, and of its possible extension at word boundary, is carried out on the basis of durational and spectral measurements in a carefully selected corpus of meaningful sentences.

2. SPEECH MATERIAL

The corpus consists of two sections: a) words from the lexicon; b) pairs of lexical words including vowel combinations at word boundaries. In both sections, [a] was chosen as [-high] full vowel. In the following, the stress mark will be reported for the lexical stress and will precede the stressed syllable.

Section a). All the items to be tested are placed in the same syntactical and intonational context, corresponding to a simple meaningful sentence having the following structure: V SN2 Adv, with the subject omitted (as frequently happens in Italian). The target word always holds the post-verbal SN2 position and the syntactical role of direct object. The other two main stresses in the sentence are kept in the same position throughout, as we employed the third singular form of the simple past stressed on the last syllable for V, and modal adverbs ending in -mente for Adv. The test words represent typical cases of hiatuses and diphthongs, as witnessed by phonetical transcriptions or morpho-phonemic rules found in previous standard works on Italian phonology, quoted above. The list of test words is shown in Table I.

Section b). As for word boundary, no clear-cut rules about diphthong or hiatus-like pronunciation can be inferred from the literature. Thus, we attempted to assess the possibility of different pronunciation strategies depending on the existence versus absence of a phrase-boundary. This was done exhaustively only for ia and ai combinations, not only for obvious reasons of economy, but also because the lack in the Italian lexicon of words ending with unstressed u would have made thorough comparisons impossible in any case. As typical

examples of no-phrase-boundary contexts, we chose Aux-V or nominal predicate structures, like dovresti andare ("you should go") and si senti' adatta ("she felt beloved"). On the contrary, we took as a typical phrase-boundary the separation between subject and predicate, SN | SV, in sentences like Anna inizio' il compito ("Anne began the work"). The list of word boundary contexts is shown in Table III: in the left column vowel sequences are given in graphemic form.

3. EXPERIMENTAL METHOD

The list was read four times by a professional speaker and the speech was recorded on high quality video cassettes. The speaker was required to use great care in order to maintain a constant mean speaking rate of about six syllables per second. The recordings were digitally converted using a 12 bit A/D converter with 12 kHz sampling rate. Broadband digital spectrograms were generated by the FFT computation, with a 2 msec frame rate, and the signal was analysed by means of interactive software and videographic facilities, which helped the operator to determine segment boundary location.

Temporal values were measured from both spectrograms and waveforms by applying widely accepted rules whenever possible [7], [8]. The following acoustic parameters were measured:
1. Onset target, glide (transition) and offset target durations of the second formant, in all the items;
2. Onset and offset frequency locations of first and second formants, in ia and ai contexts of section a).

At times, the starting point of F2 transition was partially masked by the raising towards the locus characterizing the neighbouring consonant, especially for stop consonants. In these cases, we took the behaviour of F1 into account, which showed a clear maximum separating the transition region from the region of coarticulation with the consonant, thus enabling to determine, with good consistency throughout the corpus, a boundary point for subsegmental duration measurements (see Figure I).
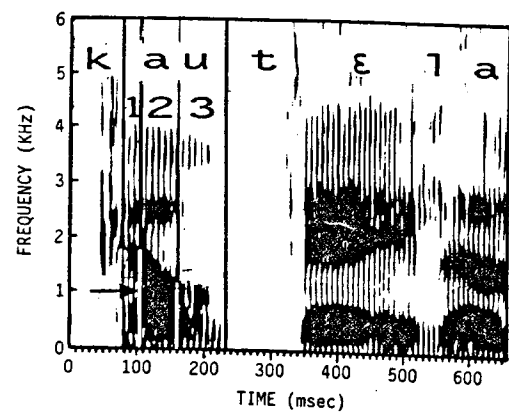

Figure I - Example of target identification strategy in the contiguous vowels [au] of the word cautela, by using the F1 maximum criteria described in the text. ——➤ : F1 maximum; 1: onset target; 2: transition; 3: offset target.

4. RESULTS

The analysis of the results will be divided into two main sections (4.1. and 4.2.) corresponding to the two parts of the corpus.

4.1. Lexical items

Table I shows the whole duration of the bivocalic sequences and the percentage duration of the onset target, transition and offset target in each context. Data clearly suggest that target duration is quite short for [j,w], since it reaches 17% of the whole duration at most. Moreover, in the case of unstressed [ja, wa], the whole duration is also very short, i. e. lower than 100 msec, while in every other context it is greater than 145 msec. This result provides evidence for the traditional transcription of rising diphthongs and consequently for the phonological representation of [j,w] as semiconsonantal phonemes. On the other hand, the traditional transcription of falling diphthongs is not confirmed by the data: their whole duration is substantially similar to that of sequences classified as hiatuses; the target duration of i and u in falling stressed sequences (daino, fauci) is not different from that of any other unstressed vowel, ranging between 20% and 30%; finally, in all falling unstressed sequences (arcaicita', aizzatore, cautela, baulino) the i and u targets are even considerably longer than the [a] target.

PHONETIC TRANSCRIPTION	ITEM	% ONSET TARGET MV	SD	% TRANSITION MV	SD	% OFFSET TARGET MV	SD	WHOLE DURATION MV	SD
ja	pianale	16	3	53	5	31	5	96	6
'ja	biada	.15	2	25	1	60	2	243	4
ia } ai	sciatore	32	3	34	2	34	1	147	5
ai }	arcaicita'	27	3	34	3	39	5	171	9
ai }	aizzatore	28	3	31	3	41	1	179	13
i'a	sciata	25	2	25	2	50	1	269	6
'ia	diade	47	1	29	3	24	2	222	6
'ai } 'ai	daino	45	7	30	4	25	3	218	17
a'i	Caino	21	2	24	3	55	1	266	11
wa	qualita'	17	2	46	8	37	8	84	6
'wa	inattuato*	15	2	36	5	49	5	216	15
ua	dualismo	43	7	34	10	23	4	165	13
ay } au	cautela	18	2	37	4	45	5	145	15
	baulino	27	1	38	3	35	3	171	15
u'a	zuavo	20	2	25	3	55	2	250	8
'ua	tua	54	3	29	2	17	2	253	2
'ay } 'au	fauci	48	2	24	3	28	2	242	24
a'u	baule	20	2	33	4	47	4	257	12

Table I - Diphthong versus Hiatus within word. Subsegmental durations expressed as percentage of whole duration, and whole duration in msec. MV: Mean Value. SD: Standard Deviation. *: in most dictionaries it is transcribed as [ua], but in currently spoken Italian and also by our speaker it is pronounced [wa] (see also Muljačić [3, p.86] about luttuoso).

Consequently, we think that purely acoustic data of our speaker show that it is not necessary to maintain the traditional transcription with "semi-vowels" [j,w] (see the second column of Table I).

In order to check the statistical validity of our conclusions, we performed a T-test analysis (see e. g. Welkowitz et al. [9, p.159]) on the mean durations of the [+high] vowels. The results, as detailed in Table II, are quite satisfactory, since the calculated values are lower than the critical one for [i↔i], [u↔u] and higher for [j↔i], [w↔u], showing that there is statistically significant difference between the mean durations only for the last two couples.

ITEM	t value	ITEM	t value
ja ↔ ia	10.98	wa ↔ ua	11.50
'ja ↔ i'a	8.19	'wa ↔ u'a	4.61
ai ↔ ai	1.29	au ↔ au	2.81

Table II - T significance test computed for the mean durations of the [+high] vowel of each couple. At the confidence level P<0.01, the critical value is 3.71.

It is interesting to note that bivocalic sequences stressed on the second segment have longer durations than those stressed on the first (see again Table I). Together with the shortness of semiconsonants indicated above, this datum would suggest a trend towards weakening of the first element of the sequence. This tendency becomes even stronger at the word boundary, as will be shown later.

The transition length does not show any clear trend, apart from a longer percentage duration in the case of unstressed sequences. The hypothesis concerning correlation between transition duration and diphthong/hiatus contrast (see Bond [10] for this hypothesis as regards English) is not borne out by evidence.

The rate of change of F2 transition (Hz/msec) is a further acoustic parameter taken into account in the literature (see, among others, Lehiste and Peterson [8], Gay [11], Borzone de Manrique [12]). We measured this parameter in ia and ai contexts of lexical words. Consistent differences between assessed diphthongs and the remaining sequences are not apparent, and the data are not worth publishing. Similar results were also found in Spanish [12]. As it is known, F2 frequency locations of vowels show great variability depending on consonantal context [13]; but even when bordering consonants were kept in the same articulatory class, many of our measurements showed steady targets well above the mean values of our speaker. For example, in un aizzatore and in diade the F2 of [a] was 1722 ± 20 and 1786 ± 9, respectively, compared to the mean value of 1656 ± 20 typical of our speaker in apicodental nonsense contexts. Even more dramatic cases are present in word-boundary data. For instance, the expected [a] of Renata indica heard in isolation was invariably perceived as [e], which is not surprising due to its formant pattern: F1 567 ± 34, F2 1851 ± 51.

As Lindblom and Studdert Kennedy [14] have demonstrated, the identity of a vowel sound is likely to be determined not only by the formant patterns at the point of closest approach to target values but also by the direction and rate of adjacent formant transitions which could compensate for possible undershoot effect. Thus, both the formant pattern and the F2 rate of change per se are likely to have little or no significance. A good proof of these phenomena is provided by our data about i in lexical items. Only stressed [i]'s reach the F2 target value typical of our speaker, about 2500 Hz, while for shorter i's the F2 steady states are located at much lower frequencies. A clear positive correlation between F2 steady-state duration and its frequency location is indeed detectable, as shown in Figure II: the linear correlation coefficient r is .915, which is higher than the critical value 0.798 (at the confidence level P<0.001), so that r is statistically significant [9, p.182].
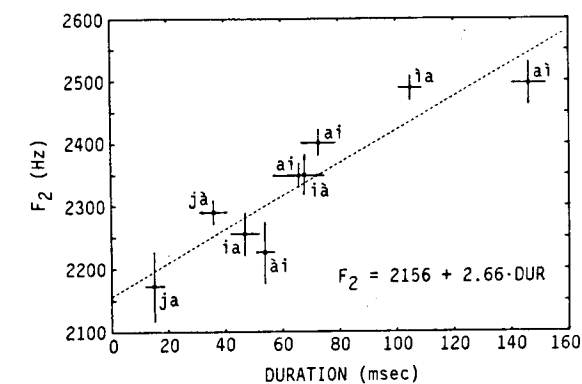

Figure II - F2 frequency location of i in lexical words as a function of its F2 steady state duration. ——— : Standard Deviation.

$$F_2 = 2156 + 2.66 \cdot DUR$$

4.2. Word boundary items

Turning now to the bivocalic sequences at word boundary, it must be noted first of all that our data concern ia and ai sequences only, as explained in the Introduction. Here a new factor is important, namely the presence versus absence of syntactic boundary. Moreover, a new type of vowel sequence is possible: one where both contiguous segments bear graphical stress. The mean absolute values of subsegmental durations of sequences at word boundary are shown in Table III. In this case, in fact, absolute values appear more suitable than percentage values for pointing out the relevant feature of our data, that is to say the weakening effect on the first target (word final vowel).

GRAPHEMIC FORM	ITEM	ONSET TARGET MV	SD	TRANSITION MV	SD	OFFSET TARGET MV	SD	WHOLE DURATION MV	SD
ia	dovresti andare	16	3	49	9	30	9	87	11
i|a	Giovanni ando'	34	4	52	4	41	5	127	6
ai	sembrava iniziato	10	1	52	7	31	7	94	10
a|i	Anna inizio'	59	21	56	14	65	2	182	7
'ia	si senti' adatta	40	10	53	2	55	8	151	17
'i|a	il salmi' ando'	72	9	59	9	56	15	188	16
'ai	sara' idiota	34	13	74	8	58	13	167	20
'a|i	l' attivita' inizia	96	18	63	9	49	5	205	15
i'a	saresti acida	59	5	52	4	135	2	246	1
i|'a	Giovanni agita	87	9	66	4	106	11	260	23
a'i	restava intimo	53	5	58	3	79	12	191	14
a|'i	Renata indica	43	6	66	15	91	7	198	17
'i'a	si senti' acida	66	9	59	5	133	6	261	6
'i|'a	Mimi' agita	123	6	99	7	123	2	346	6
'a'i	sara' ibrida	30	10	85	6	90	13	206	9
'a|'i	la verita' indica	118	11	77	5	103	5	299	12

Table III - Phonetic events at word boundary. Subsegmental duration and whole duration, in msec. MV: Mean Value. SD: Standard Deviation. | : phrase boundary

In the absence of phrase boundary, the acoustical realization is often very far from the graphemic transcription. When both segments are unstressed, the first undergoes a strong length reduction. Complete elision occurred in many cases not included in the table. For example, in the noun phrase data iniziale the ai context lasted about 70 msec in all replications and its measured and per-

ceived formant pattern was that of a single vowel lying between [e] and [i]. The presence of a syntactic frontier acts as a strong protection against the quoted weakening phenomenon, keeping both targets within comparable durations.

It is interesting to observe that when the boundary is not present, even the graphical stress does not protect the first element from the reduction process: in si senti' adatta and sara' idiota the first target is shortened to 40 msec on the average, while in the corresponding phrase boundary contexts the same target lasts about twice as long. These effects are also quite clear in the case of two-stress combinations: in the absence of syntactic frontier the word final vowel loses its stress completely, whereas in the opposite case it remains well preserved.

5. CONCLUSIONS

From the experimental analysis carried out on bivocalic sequences within lexical word in connected speech, we can claim that the traditional distinction between diphthongs and hiatuses in Italian is not confirmed by purely acoustic data: while for onglide (rising) diphthongs it is possible to establish specific durations different from those typical of hiatuses, offglide (falling) diphthongs do not differ from vocalic sequences classified as hiatuses. Thus, a phonetic transcription based on acoustic data should recognize semi-consonants [j,w], but not semivowels [i̯,u̯], which appear to be the same as the corresponding unstressed vowels.

The hypothesis concerning a possible extension of the distinction between diphthong and hiatus to the context of word boundary has not been proved, since in sentences new factors occur, in particular presence or absence of syntactic phrase boundary between the two test words. Depending on whether this boundary exists or not, word final vowel is quite preserved or reduced, respectively. In the latter case, the stress carried by the final vowel is also lost.

On the other hand, data suggest that the Italian speaker, tending to avoid production of two full contiguous vowels, chooses to reduce the first segment of the vocalic sequence, even when it is marked [-high].

REFERENCES

[1] S. Battaglia and V. Pernicone, GRAMMATICA ITALIANA, Loescher, Torino, 1951.
[2] L. Romeo, "A phonemic inventory of the Italian bivocalic sequences", FORUM ITALICUM, 1968, 2, 117-143.
[3] Ž. Muljačić, FONOLOGIA DELLA LINGUA ITALIANA, Il Mulino, Bologna, 1972.
[4] C. Tagliavini and A. M. Mioni, CENNI DI TRASCRIZIONE FONETICA DELL' ITALIANO, Patron, Bologna, 1972.
[5] L. Canepari, INTRODUZIONE ALLA FONETICA, Einaudi, Torino, 1979.
[6] P. L. Salza, "La durata dei suoni nelle sequenze vocaliche dell' Italiano", ELETTRONICA E TELECOMUNICAZIONI, 1986, XXXV, 27-34.
[7] G. E. Peterson and I. Lehiste, "Duration of syllable nuclei in English", J. ACOUST. SOC. AM., 1960, 32, 693-703.
[8] I. Lehiste and G. E. Peterson, "Transitions, glides and diphthongs", J. ACOUST. SOC AM., 1961, 33, 268-277.
[9] J. Welkowitz , R. B. Ewen and J. Cohen, INTRODUCTORY STATISTICS FOR THE BEHAVIORAL SCIENCES, Academic Press, London-New York, 1982.
[10] Z. S. Bond, "The effect of varying glide durations on diphthong identification", LANGUAGE AND SPEECH, 1978, 21, part 3, 253-263.
[11] T. Gay, "A perceptual study of American English diphthongs", LANGUAGE AND SPEECH, 1970, 13, 65-88.
[12] A. M. Borzone De Manrique, "Acoustic analysis of the Spanish diphthongs", PHONETICA, 1979, 36, 194-206.
[13] F. E. Ferrero, "Caratteristiche acustiche dei fonemi vocalici italiani", PAROLE E METODI, 1972, 3, 9-34.
[14] B. E. F. Lindblom and M. Studdert Kennedy, "On the role of formant transitions in vowel recognition", J. ACOUST. SOC. AM., 1967, 42, 830-843.