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
Università di Torino
10100 Torino, Italy

ABSTRACT

A professional speaker read four replications of 34 meaningful 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: sinalephe, 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 postverbal 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 phonoetical 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 y would have made thorough comparisons impossible in any case. As typical
examples of no-phase-boundary contexts, we chose cells of nominal predicate structures, like *dovecito andare* ("you should go") and *il senz' soldi* ("She felt abandoned"). On the contrary, we took as a typical phase-boundary the separation between subject and predicate. SA, SP, SV, in sentences like *Anna inizia il concerto* (*Anna began the work*). The list of word boundary contexts is shown in Table III; in the left column vowel sequences are given in graphic 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 speech rate of about six syllables per second. The recordings were digitally converted using a 12 bit A/D converter with 12 kHz sampling rate. Broad-band digital spectrometers were generated by the FFT computation, with a 2 sec frame rate, and the signal was analyzed by means of interactive software and videographic facilities, which helped the operator to locate the exact boundary location. Temporal values were measured on both spectromers and waveforms by applying widely accepted rules whenever possible ([7], [6]). The following acoustic parameters were measured:

1. Onset target, glide (transition) and offset target durations of the second formant, in all the cases
2. Onset and offset frequency locations of first and second formants, in [a] and [e] contexts of section 4.

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 maximum maximum transient region from the region of coarticulation with the consonant, thus enabling us to determine, with great accuracy throughout the corpus, a boundary point for subsequent temporal measurements (see Figure 1).

### Table I - Diphthong versus hiatus within word.

Subsegment durations expressed as percentage of whole duration, and mean value (sec. x 1000). Mean value, SD: Standard Deviation. *x* in most dictionaries, it is transcribed as [ux], but in currently spoken Italian and also by our speaker it is pronounced [wux] (see also Malnati [3, p.86] about *lutus*).

Consequently, we think that purely acoustic data of our speaker show that it is not necessary to maintain the traditional distinction between "vowel" ([i], [u], and [w]) and "diphthongs" ([iu], [uw], and [wux]).

In order to verify the correctness of our conclusions, we performed a t-test analysis (see e., [2, p. 362]) considering the mean values of the [high] vowels ([i], [u], and [w]) and the two diphthongs ([iu], [uw], and [wux]), and the statistical test of Pearson's correlation coefficient, being the calculated values lower than the critical value for [i] = 0.01, [u] = 0.02 and for [w] = 0.10, showing that there is statistically the same difference between the mean durations only for the first two couples. The analysis of the results will be divided into two main sections (4.2.1 and 4.2.2) corresponding to the two parts of the corpus.

### 4.1. Lexical items

The whole duration of the bi-vocalic sequences and the percentage of the onset target, transition and offset target in each context. Data clearly suggest that target duration is quite short for [i], since it reaches 17% of the whole duration. Moreover, in the case of unstressed [i], [u], whole duration is also very short. In other words, even 100 msec, in every other context it is greater than 145 msec. This result proves evidence for the traditional transcription of rising diphthongs and consequently for the phonological representation of [i], [u], and [w] as sesquionicophones. 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 simple vowels. The percentage of the target duration and in falling stressed sequences ([ait], [ait], and [wux] is different from that of any other unstressed vowel, ranging between 20% and 30%. Finally, in all falling unstressed sequences ([arait], [arait], and [wux] targets are even considerably longer than the [a] target.

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

It is interesting to note that bi-vocalic sequences stressed on the second formant have longer durations than those stressed on the first (see again Table I). Together with the shortness of sesquionicophones 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 correspondences of duration and diphthong/hiatus contrast (see Bond [10]) for this hypothesis, as regards English) is not born out by evidence.

The rate of change of F2 transition (Hz/sec) is a further acoustic parameter taken into account in the literature (see, among others, Lohiste and Resterum [8], Guy [11], Borzoei Marloqui [12]). We measured this parameter in [a] and [i] contexts of lexical words. Consistent differences between stressed diphthongs and the remaining sequences are not apparent, but only some cases of publish.

Similar results were also found in Spanish ([12]). As it is known, F2 Frequency locations of vowels show great variability depend on the context of consonant [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. As a result, we have observed that the F2 of [a] in [e] allophone and in [e] the F2 of [a] was 1722 Hz, and in 1796 9, compared to the mean value of 1655 ± 20 typical of our speaker in prototypical monosyllabic contexts. Even more drastic cases are present in the word boundary data. For instance, the expected [a] of *bentà* indicated in isolation was irregularly perceived as [w] (which is not surprising due to its formant pattern: F1 563 ± 34, F2 2631 ± 34).

As Lindblom and Studdert Kennedy [14] have demonstrated, the identity of the final 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 rate of adja.

### Table III - Phonetic events at word boundary.

Subsegmental duration and whole duration, in msec. Mean Value, SD: Standard Deviation. *x* in most dictionaries, it is transcribed as [ux], but in currently spoken Italian and also by our speaker it is pronounced [wux] (see also Malnati [3, p.86] about *lutus*).

In the absence of phrase boundary, the acoustic realization is often very far from the graphic transcription. When both segments are unstressed, the first undergoes a strong length reduction. Consequently, the negative effect associated with the nasality of the vocalic nucleus is not included in the table. For example, in the noun phrase data in [43] the [a] context lasted about 70 msec in all repetitions and its measured and per-
ceived formant pattern was that of a single vowel lying between [ᴇ] and [ɪ]. 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 sì 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 semivocals [ɛ, ɪ], but not semivowels [ɻ, ʊ], 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].

ACKNOWLEDGMENTS

This work was partially supported by an EEC (European Economic Community) contract in ESPRIT "SPIN" Project N. 64.

REFERENCES