On Sensitivity to Durational Modifications in Italian and English

Pier Marco Bertinetto & Carol A. Fowler

In this paper we examine the sensitivity of Italian and English speakers to artificial modifications in the duration of unstressed vowel. The hypothesis on which our research is based is that speakers of stress-timed languages (like English) are relatively insensitive to the durational compression of unstressed vowels, whereas speakers of syllable-timed languages (like Italian) are more sensitive to such alterations. This depends on an important feature of stress-timed languages, as distinct from syllable-timed ones, namely the tendency to reduce unstressed vowels. Supposedly, vowel reduction consists in (among other things) the implementation of an articulatory process involving a high degree of coarticulation between adjacent segments. Thus, even when reduction is carried to an extreme, unstressed vowels do not disappear altogether, but remain discernible throught the adjacent segments. Conversely, speakers of syllable-timed languages tend to preserve to a higher degree the autonomous identity of all vocalic segments (i.e., to resist reduction), regardless of their accentual status.

The experiment is based on six pairs of Latinate words, of the type *superfluous/superfluo*, which are relatively similar, from the segmental and prosodic point of view, in the two languages considered here. As predicted, Italian speakers show significantly greater degrees of reaction to artificially produced alterations in the duration of unstressed vowels. This effect is stronger (again as predicted) respecting words pronounced in isolation than respecting words produced in sentence context.

A pilot experiment on the effects of the artificial lengthening of unstressed syllables was also performed. In this case, the speakers of either language seem to react with almost equal sensitivity.

1. *Introduction.*[1]

The alleged distinction between stress-timed and syllable-timed (and mora timed) languages has lately received very much attention by phoneticians and phonologists, despite several attempts to dismiss the

[1] We wish to aknowledge the assistance of Gianna Faccioli Riconda and Maddalena Agonigi in performing part of the statistical computations, and of Francesca Brogelli and Valentina Rossi in tabulating the data.

question as either irrelevant or untestable [e.g. Beckman 1982; Roach 1982; Scott *et al.* 1985]. The obvious importance of the rhythmic factors for general theories of language production, as well as for the development of speech synthesizers, abundantly justifies the persistent interest in this topic.

Needless to say, the existence of a dichotomy such as the one referred to above, does not contradict the existence of a common basis for all human languages. It can be reasonably assumed, for instance, that all human beings, regardless of the language they speak, share the same inclinations in terms of central control of some typical rhythmical phenomena, such as: (i) the predilection for some well-formed (i.e., more natural, Gestaltic) rhythm schemata [Martin 1972; Essens 1986; Benguerel & D'Arcy 1986], or (ii) the presence of a system of rhythmical expectancies [Huggins 1972]. In addition, it must be presumed that some basic properties of the language production mechanism envisaged by action theorists [Fowler 1977], such as the alternation of specialized coordinative structures for the separate (but partly overlapping) production of vowels and consonants, or the cyclic regulation of vowel production (with special regard to stressed vowels), may be common to all natural languages.

This, and in particular the last point, has obvious consequences for a general theory of linguistic rhythm: it suggests that a naive conception of stress-timing (in terms of very regular alternations of stress-beats) is implausible. In fact, if that were the case, we would be put in the very uneasy situation of having very rhythmical languages on the one side (the so-called stress-timed languages), and on the other side languages which would appear quite a-rhythmical, if viewed in the light of the cyclic regulation of stressed vowels. This should be carefully considered by those who have repeatedly hastened to celebrate the isochrony funerals, whenever faced with experimental data which point to the imperfect temporal recurrence of stresses in English, or in other alleged stress-timed languages. On the contrary, it would be highly unnatural if stresses were very evenly spaced in some languages, and quite unevenly spaced in others, for this would destroy the common rhythmical basis of all human languages. It follows, then, that the real issue concerns the different degree with which the various languages conform to the common tendency to evenly regulate, within limits, the distance between stresses (or tone-accents, where this is the relevant feature, as in Japanese). It also follows, for that matter, that even a naive conception of syllable-timing (according to which, in some languages syllables should receive the same, or nearly the same, duration) is equally implausible: indeed, if such language were to exist, it would strongly tend to reject any kind of regular recurrence of stresses/accents, unless very special circumstances occurred (i.e., unless the language in question presented, besides fixed stress, only words of the same syllabic extension: a very unlikely situation). But once we get free of the unnatural

expectation that some languages are extremely regular, we are in a better position to try to understand the discrepancies that exist between them, and ultimately to grasp the real nature of linguistic rhythm.

As to these discrepancies, we believe that abundant evidence has been collected that the various languages do differ in their rhythmical behaviour, although sometimes not quite in the expected way. Since this is not the place for a full inventory of the differences that have been observed, a short summary will suffice (notice however that we shall not attempt here to provide a thorough account of the various lines of research which have been pursued in connection with the isochrony problem; on this topic, cf. Bertinetto [1988]).

First of all, recall the existence of the various systems of versification, most of which are based on the (more or less regular) alternation of prosodic elements, which have an obvious bearing on the problem of (the different kinds of) isochrony, such as stress, syllable and mora.

Second, although any attempt to find support for the very naive conception of stress- and syllable-timing failed since the very beginning, several authors have collected clear evidence on the divergent durational behaviour of different languages [Bertinetto 1977, 1981; Hill *et al.* 1978; Pointon 1980; Dauer 1981; Hoequist 1983*a*, 1983*b*; Duez & Nishinuma 1984/85; Fletcher 1985; Bond & Fokes 1985; Kohler 1986]. There is also some evidence that speakers of various languages have some systematic intuition about the rhythmic tendencies of their own and other languages [Miller 1984]. Of special interest to us is research which points to the different compression effect produced on the stressed vowel by the following unstressed ones, for they show that the amount of compression, although present, is much smaller in Italian than in English [Farnetani & Kori 1982; Vayra, Avesani & Fowler 1984; Marotta 1985]. This substantiates the idea that the core of the isochrony problem, in terms of inter-language discrepancies, lies in the definition of different tendencies and different degrees inside the same dimensions, rather than in defining radically opposite orientations. The number of phenomena which make the various languages similar is admittedly far larger than the number of phenomena which make them different. On the other hand, the sum total of several quantitative differences is likely to produce a qualitative discrepancy; and the evaluation of such qualitative discrepancies (if any) is the ultimate goal of this domain of the phonetic sciences.

Third, languages differ in many aspects related to their phonological structure. Among the most relevant ones, Bertinetto [1977] indicated:

(a) the presence *vs* absence of vowel reduction;
(b) the complex *vs* simple syllabic structure;
(c) the different effect of tempo accelerations (predominant compression of unstressed syllables *vs* tendentially proportional compression of both unstressed and stressed syllables);

(d) the tendentially elastic *vs* tendentially rigid position of lexical stress (i.e. large *vs* restricted use of the so-called 'rhythm rule', which shifts the position of the stress under certain conditions).

Essentially the same factors were pointed out by Dauer [1981]. Notice that factors (a-b) jointly conspire towards an opposition in terms of difficult *vs* easy syllabification, whereas factor (d) is a powerful tool for equalization of interstress intervals. Thus (indeed, unsurprisingly) the two basic prosodic components which are relevant to the isochrony issue (syllable and stress) are directly involved in that part of the phonological structure which is responsible for the rhythmical behaviour of the languages. For further observations concerning the relevance of phonological factors to isochrony, see for instance Bolinger [1981], Major [1981, 1985], Brakel [1985].

The appeal to phonological factors might be interpreted in two opposite ways, one unsympathetic and one sympathetic. The unsympathetic view would consist in saying that recourse to phonology was made only after failing to find strong phonetic evidence for isochrony in the durational patterns of the languages. Thus, this line of investigation would appear, to the critical observer, as a kind of exorcism, pursued under tenacious theoretical biases (i.e., that speech *must* be rhythmical, although it often does not look so). Under the sympathetic view, one would suggest instead that after all a link between phonological and phonetic structure does exist, although its consequences are often difficult to grasp, so that some ingenuity must be used in order to make them apparent. The present paper will make the assumption that such a link does exist.

In particular, we shall examine the sensitivity of Italian and English native speakers to artificial modifications of the duration of unstressed vowels. This proceeds from point (a) above: in fact, the effects of vowel reduction are apparent both in spectral structure and in duration. It can be hypothesized that speakers of stress-timed languages are relatively insensitive to durational compressions of unstressed vowels, whereas speakers of syllable-timed languages are relatively more sensitive to such alterations. Conversely, speakers of syllable-timed languages might prove to be relatively tolerant of a lengthening of unstressed vowels, whereas speakers of stress-timed languages are expected to react more promptly. This would be consistent with whatever we know about the phonetic and phonological structure of Italian and English. The experiment reported below tests this hypothesis.

Of course, it would be meaningless to collect the kind of judgements that are of interest to us, using a non-native language as target. Therefore, it was necessary to build two separate sets of stimuli, one for the Italian and one for the English subjects. This is in itself undesirable, for it makes the comparison between the two languages more difficult. The only solution consisted in trying to find very similar materials in the two languages,

exploiting the possibility offered by pairs of latinate words, such as *political* vs *politico*. In a case like this, we would like to know whether the shortening of the first (unstressed) vowel is equally tolerated by native speakers of the respective language.

Obviously, the application of an artificial shortening procedure, leading eventually to 'total' deletion of an unstressed vowel, is never completely harmless: it would be naive to believe that all and only the intended vowel is actually deleted from the original word. Each segment conveys not only information about itself, but also come cues pertaining to adjacent segments. Indeed, this characteristic is so pervasive, that theories of speech perception have been devised in which this is explicitly assumed as a fundamental feature of the phonetic processing [e.g. Fowler & Smith 1986]. This explains why the stimuli corresponding in our experiment to the final step of shortening are still heard as containing some residue of the target vowel (although the vowel itself does not appear in the speechwave: cf. fig. 6 and 10), as though this vowel were somehow hidden behind the adjacent segments which were originally coarticulated with it. Notice however that, rather than embarassing, this situation is highly desirable for testing the hypothesis under investigation in this paper. In fact, what is at issue here is precisely the alleged existence of different coarticulatory strategies in two different languages (stress-timed English *vs* syllable-timed Italian): it can be assumed that one important aspect of vowel reduction consists in the implementation of an articulatory process which involves a high degree of coarticulation among adjacent segments. In other words, the speaker of vowel reduction languages supposedly makes use of a strong coproductive strategy, so that unstressed vowels tend to remain hidden behind adjacent segments, rather than being deleted altogether (although this can be the final result of a diachronic process).

That this is so is shown, in particular, by Price [1980], where stimulus continua such as *prayed-parade* and *plight-polite* were submitted for judgement to native speakers: it was found that the most effective cue to the discrimination between the two items of each continuum was the duration of the liquid consonant. This outcome was possible precisely because, admittedly, English speakers do not necessarily attribute an autonomous articulatory role to the unstressed vowel of the first syllable of *parade* and *polite*: the presence of such vowels can be easily suggested by the longer duration of the liquid itself, which is strongly coarticulated (i.e. coproduced, in the sense of 'action theory') with the unstressed vowels. Conversely, the speaker of a (syllable-timed?) non-vowel-reduction language, is likely to adopt a coarticulatory strategy which does not emphasize the coproduction of unstressed vowels with adjacent segments, although some amount of coproduction must of course exist for obvious physiological reasons. These speakers tend to preserve (to some extent) the autonomous identity of any vocalic segment, regardless of its accentual status: the systematic ambiguity

of a pair like *prati* - *pirati* (= "lawns" - "pirates") would be absolutely unconceivable in Italian, although occasionally this kind of misunderstanding might occur in casual speech. The speakers of non-vowel-reduction languages are therefore likely to be relatively more sensitive to modifications involving the duration of unstressed vowels.

2. *Description of the material.*

For the purpose of the present experiment, the following six pairs of words were chosen (the underlined vowels are those which underwent modification during editing):

1. superfluous	superfluo
2. political	politico
3. stochastic	stocastico
4. esoteric	esoterico
5. mesopotamic	mesopotamico
6. canonical	canonico.

The target vowels belong to different syllables: to the first syllable in pairs 1-3, to the second syllable in pairs 4-5, to the third syllable in pair 6. Notice further that the target vowels in the first syllable immediately precede the main stress; the target vowels in the second syllable immediately precede the main stress in one case (pair 4), and another unstressed syllable in the remaining case (pair 5); finally, the target vowel in the third syllable follows the main stress. It should also be noticed that in pairs 4-5 the English words present on the first syllable a secondary stress, which is considered to be more prominent than that of the Italian words [Bertinetto 1981]: however, this should not cause any consequence in the present experiment.

More important for us is the syllabic structure that emerges through the eventual deletion of the target vowel. In almost every case, this deletion leads to acceptable syllabifications: this is not true, however, for pair 3, for the sequence /stk/ is normally not acceptable as a tautosyllabic sequence in English, and is even virtually unpronounceable for an Italian speaker.[2] Pair 3 therefore provides a severe test for the hypothesis under investigation. As for pair 6, the deletion of the target vowel produces a sequence which, although apparently acceptable, does not correspond to the actual

[2] It should be noticed that in casual speech English admits of this kind of deletions: Osika *et al.* [1975], for instance, quote examples such as *p'tassium*, *d'stroy*, adding that the first syllable is one of the favourite sites for vowel deletions, if it precedes the stressed syllable. It is quite possible, however, that pronounciations such as *p'tassium* contain some cues for the correct decoding: for instance, the first stop might be fully exploded, contrary to the normal behaviour of English speakers in dealing with such clusters.

pronounciation (especially for Italian speakers): the sequence /nk/ is in fact normally produced with a velar nasal, which is not the kind of nasal present in the given items.[3]

A further point which must be made with respect to the same pair 6 is that, for an unfortunate circumstance, the indication of the target vowel was mistaken in the form presented to the English subjects: i.e., the vowel underlined in the form was that of the first syllable, instead of the third one. Because of this, the results concerning pair 6 must be taken with caution: English and Italian subjects did not perform a strictly identical task.

The six pairs were pronounced by two native speakers: the Italian speaker was one of the two authors, the English speaker a North-Eastern American male. Each word was read both in isolation and in context. Here follows a list of the sentences:

— Yes, that was a VERY superfluous explanation
— Yes, this is a TOUGH political confrontation
— Yes, that was a REALLY stochastic process
— Yes, that was a VERY esoteric matter
— Yes, this is a TRUE mesopotamic legend
— Yes, that was a VERY canonical verdict.
— No, non ERA superfluo spiegartelo
("no, it wasn't superfluous to explain it to you")
— No, non È politico il confronto
("no, this confrontation is not political")
— No, non ERA stocastico il processo
("no, that process wasn't stochastic")
— No, non ERA esoterico il problema
("no, that problem wasn't esoteric")
— No, non È mesopotamica la leggenda
("no, this legend is not mesopotamic")
— No, non ERA canonico il verdetto)
("no, that verdict wasn't canonical").

The words in capital letters were read with emphasis, in order to obtain a nearly flat contour on the final part of each sentence. This was done with the purpose of reducing, as far as possible, the intonational discrepancies among the target words (notice that target words follow the emphasized words immediately). In spite of this, some kind of language-bound intonational difference was observed, in as much as the English speaker found it slightly uncomfortable to achieve a flat contour on the last part of each sentence. Nevertheless, after practice, this speaker managed to find a convenient way of reading these sentences with a fully natural intonation, and an almost flat final contour.

Each isolated word and sentence was read, by the appropriate speaker, at least three times. A major problem of the experiment was the different duration of target vowels from word to word (both within the same language and between languages).

[3] Notice, though, that in English the sequence /nk/ is not altogether ruled out, provided there is a juncture between the two consonants (cf. *concord*).

In an attempt to handle this problem, the most suitable tokens were chosen, so that the words of each pair should undergo the minimum manipulation in order to achieve the same (or nearly the same) duration on the target vowels. This kind of preliminary manipulation, made necessary by the need of applying the same shortening steps on the target vowels of each pair, was performed on the words listed in table I.

English	*Italian*	*Isolation*	*context*
superfluous		+ 8.0 ms	
stochastic		+ 18.0 ms	
esoteric		+ 9.4 ms	+ 9.6 ms
	mesopotamico	+ 13.8 ms	+ 10.9 ms
	canonico	+ 11.8 ms	
canonical			+ 19.0 ms

Table I. Duration increments applied to the target vowels of the specified words in order to obtain the 'derived originals' (cf. text for explanation).

Altogether, preliminary manipulation was performed on four English *vs* two Italian words, and on five isolated tokens *vs* three tokens in context, in order to create the base forms of the words onto which further shortening and lengthening modifications were performed. Technically, the manipulation consisted of the reduplication of one or two glottal pulses, chosen in the middle part of the vowel. In all cases but one (see below), this modification did not cause any perceivable noise. Although this manipulation is in itself undesirable, it was necessary to introduce it in order to obtain comparable judgements on the shortenings of the target vowels of each pair.

The shortening applied to the base words was obtained by signal editing, using standard procedures available at Haskins Laboratories. There were four steps of shortening in isolated words (with the exception of the pair *canonical - canonico*, whose stressed vowel was too short to allow four steps), and three steps of shortening in the words pronounced in context. The actual extent of the shortening steps varied from case to case, according to the duration of the target vowel; thus, by the expression 'shortening step' we refer to a proportional value rather than to an absolute value, each step corresponding to one fourth (vz. one third) of the original duration of the vowel. In addition to the normal shortening procedure, the pair *superfluous - superfluo* (in the version pronounced in isolation) also underwent a procedure of lengthening in four steps. The duration of the target vowels is reported below, in the detailed description of the materials.

As a way of illustration of the strategy adopted by the experimenters, fig. 2-6 show the result of the shortening procedure as applied to the target vowel of the word *superfluous*, pronounced in isolation. Here, as in all instances to be discussed below, each figure shows only the relevant portion of the token, i.e. the last portion of /s/, the target vowel, and the initial (and gradually increasing) portion of /p/. Notice that, as said above, the stimulus corresponding to the fourth and final shortening step (cf. fig. 6), although apparently bereft on any sign of the target

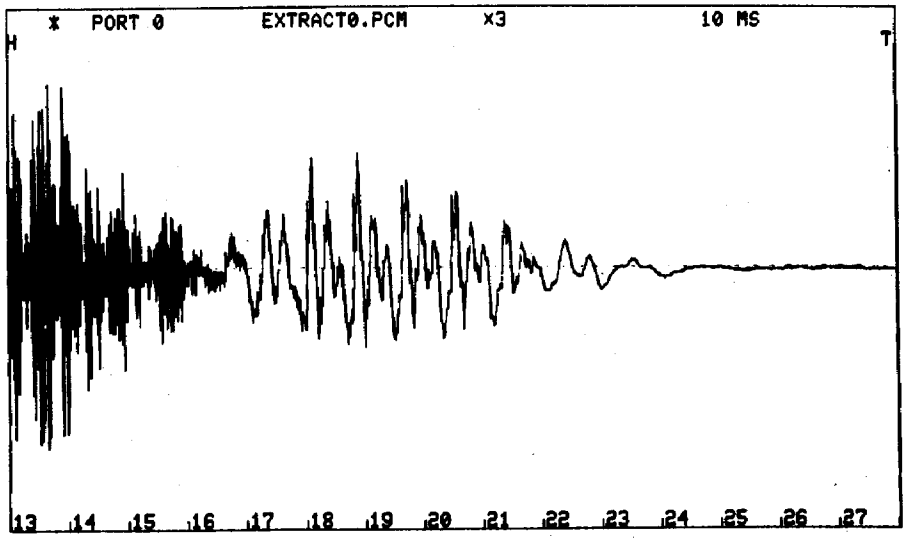

Fig. 1. Display of the original target vowel of the word *superfluous*.

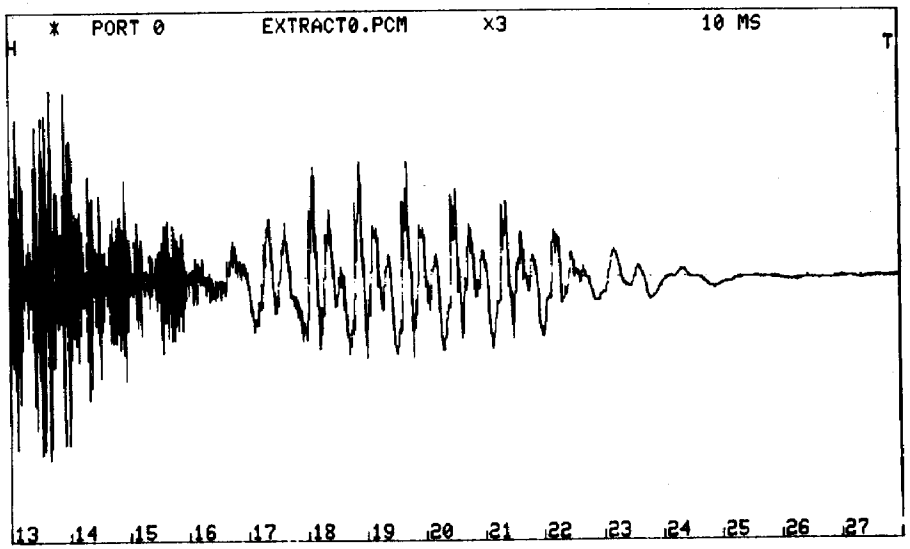

Fig. 2. Display of the 'derived original' target vowel of *superfluous*.

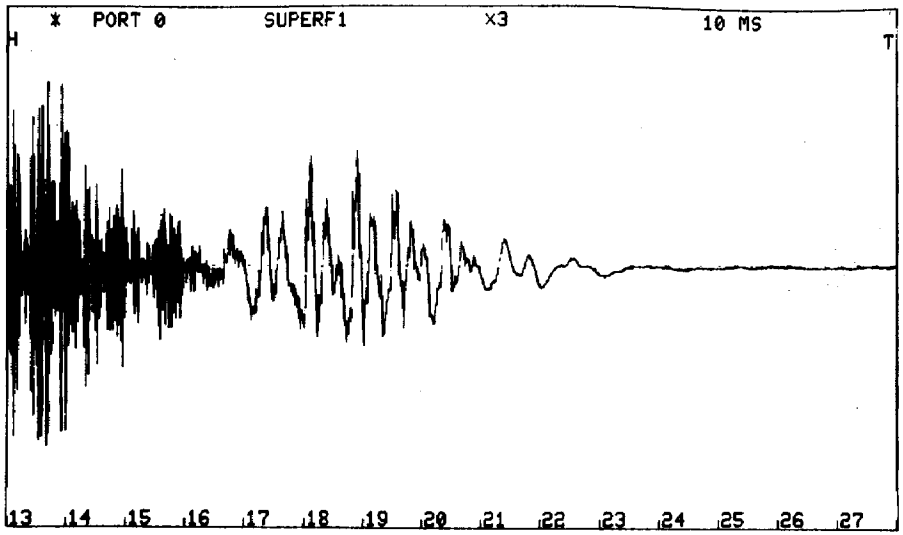

Fig. 3. First shortening degree of the target vowel of *superfluous*.

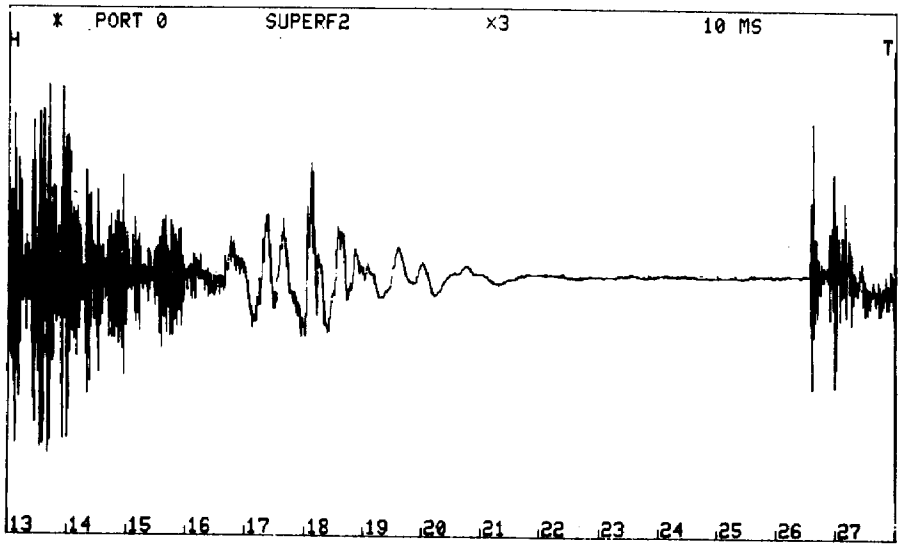

Fig. 4. Second shortening degree of the target vowel of *superfluous*.

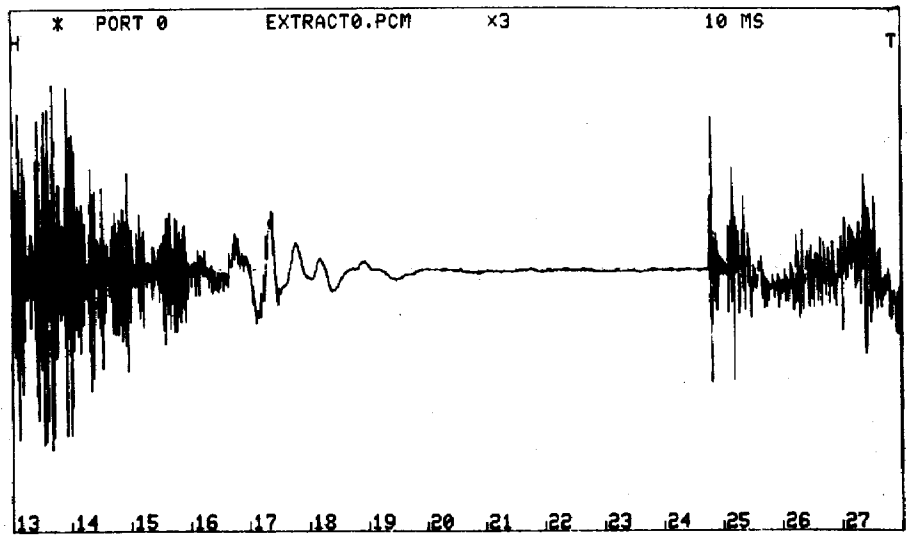

Fig. 5. Third shortening degree of the target vowel of *superfluous*.

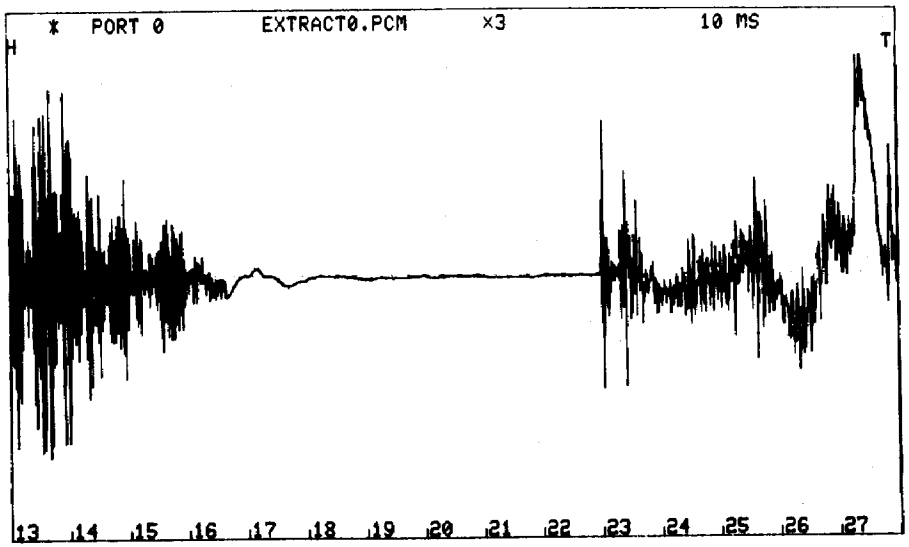

Fig. 6. Forth shortening degree of the target vowel of *superfluous*.

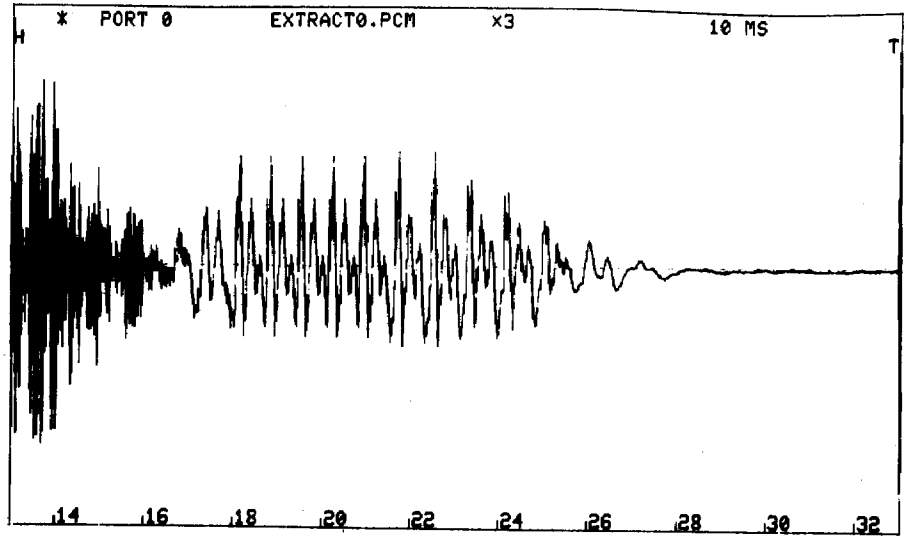

Fig. 7. First lengthening degree of the target vowel of *superfluous*.

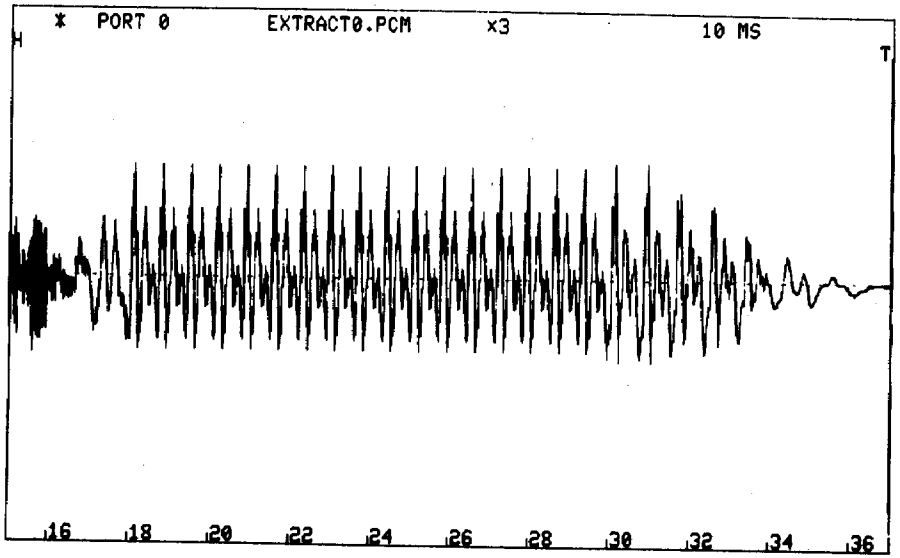

Fig. 8. Forth lengthening degree of the target vowel of *superfluous*.

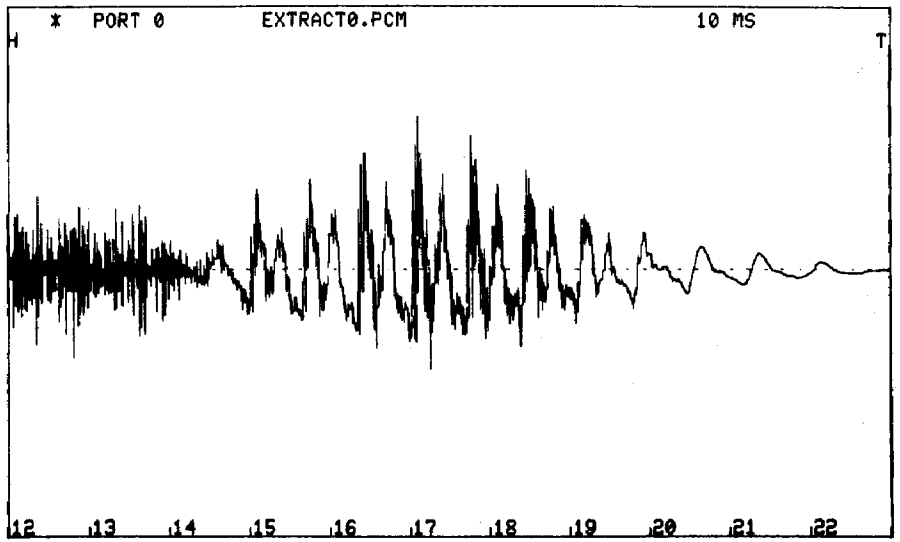

Fig. 9. Display of the original target vowel of the word *superfluo*.

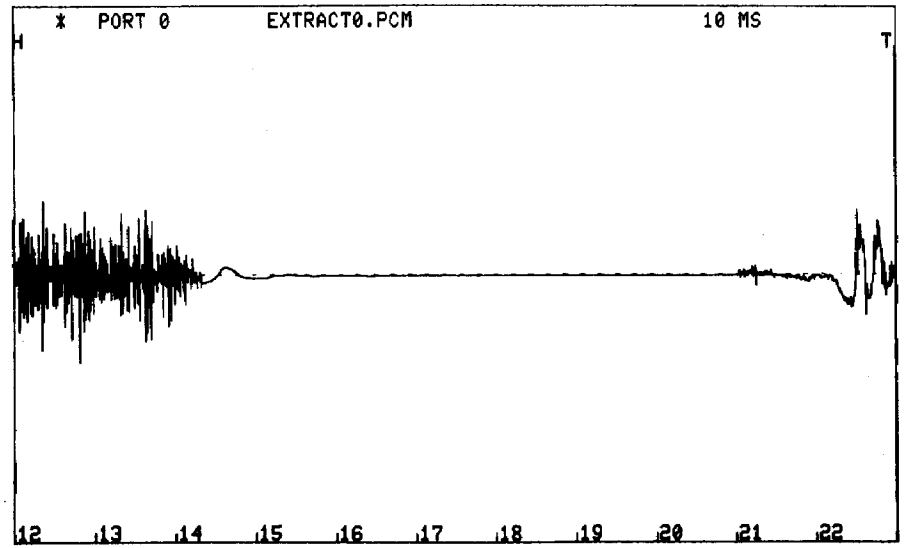

Fig. 10. Forth shortening degree of the target vowel of *superfluo*.

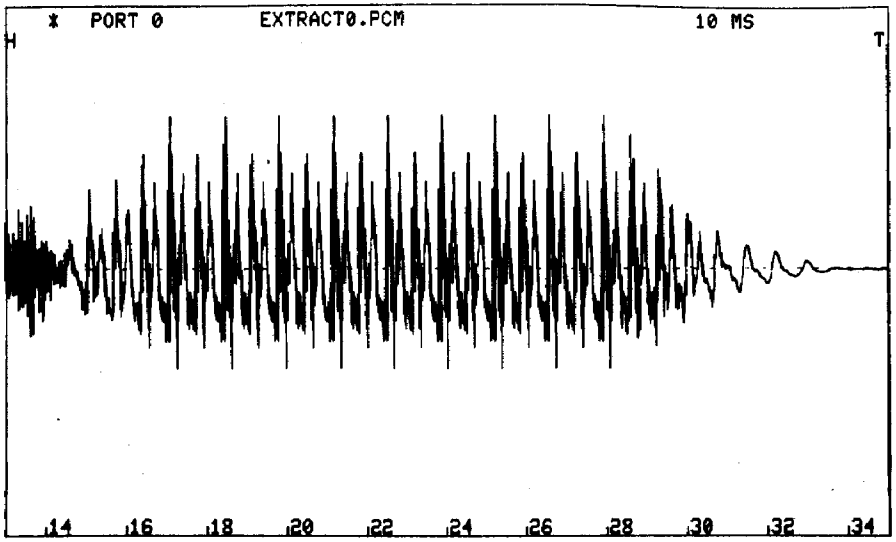

Fig. 11. Forth lengthening degree of the target vowel of *superfluo*.

vowel, still conveys somehow its presence through the adjacent consonantal segments which are coarticulated with it. Fig. 1, on the other hand, shows the original version of the item *superfluous*, from which the stimulus represented in fig. 2 was derived, for reasons explained above. Stimuli like the one in fig. 2 will be called from now on 'derived originals', in order to remind that although they are used as the starting point of the shortening and lengthening procedures, they are in fact derived for the sake of the experiment from a true original item.

Fig. 7-8 exhibit, respectively, the first and the final lengthening step obtained from the 'derived original' of fig. 2. Notice that the lengthening procedure consisted in the iteration of whole glottal pulses, taken from the middle portion of the original vowel, in order to avoid the intrusion of click noises which might prove troublesome for the listening test. The shortening procedure, on the other hand, was obtained by deleting the relevant portion of the target segment from the middle of the vocalic section of the speechwave, as displayed in each figure. The software used for this purpose (Haskins's "Wendy") takes care by itself of the necessary interpolations between points on either side of the cutting spot.

Finally, fig. 9-11 show a selection of the corresponding stimuli built for the Italian cognate *superfluo*. Fig. 9 shows the original item (in this case, a true original); fig. 10-11 show the final outcome of the shortening and, respectively, the lengthening procedures as applied to this item. As can be seen, the durations of the target vowel are in any instance very similar among each pair of corresponding stimuli.

	English items	*Italian items*
	superfluous	*superfluo*
(in isolation)	OD = 63.6; DD = 71.6 step = 17.9	OD = 75 step = 18.7
(in context)	OD = 54.2 step = 18	OD = 55.1 step = 18.4
(lengthening, in isol.)	DD = 71.6 step = 28; D max = 183.6	OD = 75 step = 27; D max = 183
	political	*politico*
(in isolation)	OD = 100.6 step = 25.1	OD = 98.3 step = 24.5
(in context)	OD = 84.2 step = 28	OD = 84.8 step = 28.2
	stochastic	*stocastico*
(in isolation)	OD = 59; DD = 77.6 step = 19.4	OD = 74 step = 18.5
(in context)	OD = 55.7 step = 18.6	OD = 59.5 step = 19.8
	esoteric	*esoterico*
(in isolation)	OD = 67; DD = 76.6 step = 19.1	OD = 75.2 step = 18.8
(in context)	OD = 59.7; DD = 69.3 step = 23.1	OD = 70.9 step = 23.6
	mesopotamic	*mesopotamico*
(in isolation)	OD = 66.5 step = 16.6	OD = 51.5; DD = 65.3 step = 16.3
(in context)	OD = 78.7 step = 26.2	OD = 67.4; DD = 78.3 step = 26.1
	canonical	*canonico*
(in isolation)	OD = 57 step = 19	OD = 46.8; DD = 58.6 step = 19.5
(in context)	OD = 57.4; DD = 76.4 step = 25.5	OD = 76.3 step = 25.4

Table II. Duration values, in msec., of the target vowels of the test words, pronounced in isolation and in context. OD = original duration; DD = derived duration; step = magnitude of the shortening step (in the case of *superfluous*/*superfluo*, there is also a lengthening step; Dmax = maximum duration of the target vowel after the forth lengthening step). Cf. text for additional explanations.

Table II provides the basic details about the construction of each series of stimuli. The figures reported below refer to the duration, expressed in ms, of the target vowels: 'OD' stands for original duration, 'DD' for the modified duration obtained for the so-called 'derived originals'; 'step' stands for the extent (in ms) of the shortening (vz. lengthening) steps; 'D max' stands for the maximum duration obtained through the lengthening procedure applied to the single pair *superfluous - superfluo*.

The modification of the originals produced, in some cases, some effects, which (although reportedly non-pervasive) were perceived by the speakers. In particular, it must be noted that in one series of stimuli, the one generated from the word *stochastic*, the shortening procedure resulted in the introduction of a click noise in the speech signal. Attempts to eliminate the noise did not succeed: however, since the click was not perceived on the target vowel, it was decided to preserve the stimuli. In addition, the /s/ of *mesopotamico* and *esoterico* was judged too sibilant by most of the Italian listeners: this was evidently due to the idiosyncratic pronounciation of the speaker. One subject reported being slightly disturbed by the possible ambiguity between the final stage of shortening of *esoterico* (the actual word) and *isterico* (= "hysteric").

The listening tapes (one for each language) contained two iterations of each stimulus. The stimuli were randomized over the complete series of test words and iterations, for a total of 114 tokens. The test consisted of two parts, as explained in the instructions given to the subjects (identical for both groups of listeners, with the only exception of the language):

INSTRUCTIONS

Part I

This is a perceptual experiment. You will be listening to single words and words in sentences, which have been modified by computer. The modification takes place on a single vowel in a key word in each sentence.

The vowel is underlined on the answer sheet as follow:

superfluous

The underlining means that the modification is localized on the first vowel.

We want to learn through your judgments how good the speech is. Use the following categories:

VG = very good;
G = good;
T = tolerable (by this we mean something that is a possible way to say the word, but not the one that is heard very often);
B = bad;
VB = very bad.

To express your judgements, circle the relevant character(s) on the answer sheet immediately beside each item.

Notice that we do not know how these words and sentences sound. It is possible that all of them will sound VG (very good) or at least T (tolerable); therefore you should judge every item in itself, and not worry about using all the categories in the scale.

There are 114 items in all. The tape takes about 20 minutes.

Each item will be repeated four times successively to help you make your judgments. You must provide an answer to each item; if you are not sure, take a guess.

The sentences you will hear are the following:

Yes, that was a VERY superfluous explanation.
Yes, that is a TOUGH political confrontation.
Yes, that was a REALLY stochastic process.
Yes, that was a VERY esoteric matter.
Yes, that was a VERY canonical verdict.
Yes, this is a TRUE mesopotamic legend.

Part II

You are now going to listen again to the same tape.

Although in the previous instructions it was stated that each token had been modified, that was in fact not true of all of them. Therefore we would like to know whether you can tell which tokens have actually been modified.

Notice that most of the times the modification (when indeed it occurs) shortens the underlined vowel, but in a few cases the vowel is lengthened.

The choice will be between the following categories:

O = original;
M = modified.

The relevant character must be circled on the answer sheet beside each item.

After you have made that decision, you have to specify in addition the degree of confidence of your judgment by selecting one of the following numbers, whose meaning is here indicated:

1 = pure guess;
2 = uncertainty;
3 = certainty.

Circle the number on the answer sheet, next to the capital letter that you will have just circled.

Remember that you must judge every item in itself, and not worry about using each category an equal number of times.

Thus, overall there were six independent variables and three dependent ones. The independent variables are:

'language' = English and Italian (corresponding to the two sets of stimuli described in table II);
'subjects' = 14 psychology undergraduate students at Dartmouth College, N.H.; and 14 undergraduate students at Pisa University, with different regional backgrounds, although with a prevalence of Tuscan subjects;
'word' = 6 words for both languages (cf. table II);
'context' = each word occurred both in a sentence frame and in isolation (cf. decription above);

'shortening' = each 'original' word was submitted to the editing procedures described above, which produced additional tokens; the vowels undergoing shortening (in one case, also lengthening) was always unstressed (notice that the number of shortening steps, and their dimensions, differ according to the individual words and contexts: cf. table II);

'repetition' = 2 repetitions for each stimulus (the randomization involved of course both repetitions within each language series).

The dependent variables are the following:

'quality' = a 5-point rating scale, ranging from 1 ('very good') to 5 ('very bad'); this task was the object of part I of the listening test;

'nature' = a forced choice, in which the subjects had to tell whether the given stimulus was the 'original' one, or had been 'modified'; this task, and the following, were performed in part II of the listening test;

'confidence' = a 3-point rating scale, ranging from 1 ('pure guess') to 3 ('certainty').

3. *Results.*

The data were statistically processed by means of the ANOVA procedure made available by the SAS software package. Preliminary investigations suggested dropping two of the variables ('repetition' and 'words'). This was done for the sake of simplicity, considering the complexity of the design: however, in the next section we shall have something to say on the variable 'word'.

3.1. *'Quality' judgements.*

3.1.1. *Shortenings.*

Let us first examine the data concerning the 'quality' judgements, starting with shortenings (cf. table III). Recall that the hypothesis on which this test is based was that Italian subjects would be more sensitive to durational changes than English subjects. In fact, this turns out to be true: the overall judgements of the Italian speakers are altogether more severe, i.e. more displaced towards the 'bad' side of the 5-point scale. In addition, it was hypothesized that words in isolation would be more prone to be judged

Factors and interactions			*Quality*	*Nature*
L	E		2.19	−0.11
	I		2.79	−0.36
C	Ct		2.18	0.52
	Is		2.75	−0.87
S	0		1.60	1.70
	1		1.78	1.12
	2		2.57	−0.76
	3		3.33	−1.95
	4		4.12	−2.77
L x C	E	Ct	1.96	0.53
	E	Is	2.38	−0.65
	I	Ct	2.41	0.51
	I	Is	3.11	−1.08
L x S	E	0	1.66	1.58
	E	1	1.67	1.16
	E	2	2.09	−0.36
	E	3	2.79	−1.73
	E	4	3.48	−2.76
	I	0	1.54	1.81
	I	1	1.90	1.08
	I	2	3.04	−1.16
	I	3	3.88	−2.18
	I	4	4.76	−2.79
C x S	Ct	0	1.48	2.37
	Ct	1	1.69	1.56
	Ct	2	2.36	−0.19
	Ct	3	3.20	−1.74
	Is	0	1.71	1.02
	Is	1	1.87	0.68
	Is	2	2.77	−1.33
	Is	3	3.47	−2.27
	Is	4	4.12	−2.77
L x C x S	E	Ct	0	1.51	2.19
	E	Ct	1	1.60	1.28
	E	Ct	2	1.96	0.19
	E	Ct	3	2.76	−1.51
	E	Is	0	1.81	0.97
	E	Is	1	1.74	1.03
	E	Is	2	2.23	−0.92
	E	Is	3	2.82	−1.95
	E	Is	4	3.48	−2.76
	I	Ct	0	1.45	2.54
	I	Ct	1	1.79	1.85
	I	Ct	2	2.76	−0.57
	I	Ct	3	3.64	−1.77
	I	Is	0	1.62	1.08
	I	Is	1	2.01	0.32
	I	Is	2	3.32	−1.74
	I	Is	3	4.13	−2.58
	I	Is	4	4.76	−2.79

Table III: shortenings. Mean values of the 'quality' and 'nature' ratings, exhibited by single factors and interaction of factors. The symbols must be read as follows: E = English, I = Italian, Ct = in context, Is = in isolation, 0 to 4 = degrees of shortening; for the abbreviations of the factors, cf. the text.

unnatural by both language groups when undergoing shortening modifications: this too turns out to be true. Finally, and again in accord with the hypotheses made at the outset, there was a continuous shift towards the 'bad' corner of the scale alongside with the shortening increaments: that is, the shorter the unstressed vowel, the more severe the judgements by both language groups. The factors 'language' (**L**), 'context' (**C**) and 'shortening' (**S**), all individually are highly significant; also the interactions **L x C** and **L x S** are highly significant. However, because of the uneven number of observations in each sample (due to the word *canonical-canonico* having only three shortening steps, instead of four, also in isolation), the interactions **C x S** and **L x C x S** could not be computed. To remedy this flaw, two further computations were performed: one ANOVA procedure in which the levels of shortening were limited to three in all cases, and one GLM (general linear model) procedure. In the first case, the interaction **L x C x S** became highly significant, but **C x S** and even **L x C** did not prove significant. In the second case, **L x C x S** turned out highly significant and **C x S** moderately significant (.05 threshold), although **L x C** was again non-significant.

3.1.2. *Lengthenings.*

The data on lengthenings, still with respect to the 'quality' judgements, are more delicate to interpret (cf. table IV). Here, the only factors to be considered are **L** and **S** (where **S** is to be regarded as negative shortening, i.e. lengthening). The only significant factor is **S**; both **L** and **L x S** are not significant. An inspection of the mean values shows that the judgements are altogether quite severe in all cases: in fact, these stimuli are seldom considered anything better than tolerable (score 3), with the Italian subjects giving again more severe judgements than their English colleagues. Notice further that for the English subjects the worst stimulus was the one with the second step of lengthening, followed by the third and the fourth; for the Italian subjects, on the other hand, the descending order of acceptability was: third, fourth, second and first.

These data show that the reaction to lengthening was very sharp in all cases. Contrary to the initial hypothesis, which attributed more tolerance to lengthening to the Italian subjects (at least for the lower degrees of lengthening) the language factor does not seem to play a role here. However, it must be recalled that these data are only explorative in nature, due to the small number of observations. Interestingly, also Nishinuma & Duez [1987] observe that French speakers reject lengthenings more promptly than shortenings, with both stressed and unstressed vowels.

3.2. *Nature judgements.*

Let us now turn to the judgements concerning the 'nature' of the stimuli ('original/modified'). These were submitted to a preliminary elaboration, together with the 'confidence' ratings strictly connected with them. The 3-point scale of the confidence ratings was multiplied by 1 when the response was 'original', and by −1 when the response was 'modified'. By this means, a continuous scale was built, ranging from 3 ('I am certain that the stimulus is original') to −3 ('I am certain that the stimulus is modified'), which combines together the data of these two types of responses; we shall henceforth refer to this scale as the 'nature' rating *tout court*. We then performed an ANOVA analysis, with the same factors already selected for the 'quality' judgements.

3.2.1. *Shortenings.*

As before, we shall first examine the shortening data (cf. again table III). Altogether, the Italian subjects gave sharper ratings: i.e., they were

Factors and interactions			*Quality*	*Nature*
L	E		2.14	0.45
	I		2.09	0.65
S	1		3.01	−2.66
	2		3.96	−2.98
	3		3.91	−2.92
	4		3.75	−2.89
L x S	E	1	3.07	−2.67
	E	2	3.96	−3.00
	E	3	3.75	−3.00
	E	4	3.50	−2.96
	I	1	2.96	−2.64
	I	2	3.96	−2.96
	I	3	4.07	−2.85
	I	4	4.00	−2.82

Table IV: lengthenings. Mean values of the 'quality' and 'nature' ratings for the lengthened tokens. See table III for the abbreviations; however, **S** stands here for 'lengthening', and 1-4 stand of course for degrees of lengthening.

more determinate in considering 'modified' the stimuli. They switch over to sharp negative values already by the second level of shortening, whereas the English subjects, although producing a negative rating by the very same level of shortening, have lower values, showing thus to be nearer chance. Indeed, with words in context they maintain an overall positive rating up to the third level. Moreover, both language groups show increasing negative values alongside with the increments of shortening. All this is consistent with the 'quality' judgements. Also consistent with these is the ratings of the context variable: words in context are predominantly judged 'original', whereas those in isolation are predominantly felt to be 'modified'. This is encouraging, especially considering that in this case we have to be content with the raw data more than with the ANOVA results, which are slightly less striking than with the 'quality' judgements. In fact, only the factors **C** and **S** are significant, plus the interaction **L x S**; of the remaining factors, the interaction **L x C x S** is only moderately significant (.05 threshold), whereas **L** and **L x C** are non significant. As to **C x S**, this result could not be computed for reasons which have been alluded to before. The further computations carried out by excluding the fourth level of shortening did not improve the picture: **C x S** turned out moderately significant (.05 threshold), but **L x C x S** resulted non significant, all the rest being as before. These outputs were further confirmed by the additional computations performed by means of the GLM procedure. Thus, the interactions **C x S** and **L x C x S** oscillated between being significant and non-significant according to the procedure of computation. However, the general picture is altogether satisfactory: the non-significant value of factor **L** and of the interaction **L x C** might reflect the fact that for both language groups the particular task involved in the confidence ratings was equally hard.

3.2.2. *Lengthenings.*

With the lengthening data, the picture is even more difficult to disentangle (but recall that these data are based on a small set of observations). Here there is a strong 'floor' effect, all the ratings being very near the highest negative value (cf. table IV). The decreasing order of negativity is (for both language groups): second level of lengthening, followed by third, fourth and first respectively. Overall, the English speakers are slightly more determinate in their negative ratings than the Italian ones. However, the factorial analysis points to **S** (to be read here as 'lengthening') as the only significant factor; **L** and **L x S** are not, just as with the 'quality' ratings.

4. *Discussion.*

The data gathered in this experiment were on the whole in accord with the general assumptions that were put forth in the introductory section. Italian subjects seem to be more sensitive to durational changes, as far as these concern the shortening of unstressed vowels. As a matter of fact, this is true for the 'quality' judgements, and not for the 'nature' ratings; however, this should not be too surprising, given the intrinsic difficulty of the second task. In both cases, however, there was a significant effect of the variable 'context'; as predicted, words in isolation are more easy to judge than words in context, which enjoy some sort of protection by the rest of the sentence in which they are embedded. Finally, the variable 'shortening' was significant too, showing that the values chosen for this factor were effective four our purposes.

As to the lengthening of unstressed vowels, the picture is less clear, partly because of the limited number of observations. In both 'quality' and 'nature' ratings, 'lengthening' itself turns out to be the only significant factor, although a 'floor' effect was observed. This suggests that for both language groups the prolonged articulation of an unstressed vowel is felt, in the given circumstances, very unnatural. It should be noticed, though, that the lengthening values used in this experiment might have proved to be much too extreme than the shortening ones, which lend themselves to more articulated ratings.

A word of caution is in order here. The kind of data selected for this experiment forced us to use different base stimuli for the two language groups. Although great care was employed in order to make the English-Italian pairs as similar as possible, it should not be overlooked that the ineliminable diversity of the two members of each pair might have played a role. It is encouraging, however, to note that the language factor, although overall significant according to our predictions, did not turn out to be significant in every single case (cf. the 'nature' ratings). Thus, one is invited to speculate that these diversities were not so extreme as to automatically cause different responses.

It should be further noticed that even within each language group the various words used for the experiment were very likely to behave differently. Recall, in particular, that some words had to undergo preliminary modifications in order to create the 'base' words used for the shortening (and lengthening) procedures (cf. tables I and II). It is therefore intriguing to verify whether the so-called 'derived originals' were perceived as less 'original' than the truely original ones. With this aim, we performed an ANOVA procedure on the 'nature' ratings with **L**, **S** and **W** ('word') as factors. All factors and interactions are significant in this analysis, with the only exception of the factor **L**. This suggests that indeed individual words caused different sets of responses, irrespective of the language group. A closer

examination of the 'derived originals' shows that, with words in context, the two English tokens (*esoteric, canonical*) were perceived as perfectly natural ones, one of them receiving the overall highest score; in contrast, however, the single Italian token (*mesopotamico*) received the lowest score of all. With words in isolation, something similar happens: the three English words (*superfluous, esoteric, stochastic*) shows altogether better than average scores, whereas of the two Italian words (*mesopotamico, canonico*) the first one shows by far the lowest score, exhibiting in fact the only negative value in the whole set of 'base' words. Interestingly, the first shortening step applied to this word yields a mean positive rating. Thus, one of the two following remarks holds: either the Italian subjects were more sensitive to even minor modifications of the originals; or the overall quality of the 'derived originals' happened to be worse in the Italian set. One can observe, however, that the general dynamics of the ratings develops quite coherently for all words, with only marginal divergences. For instance, the Italian word *stocastico*, in isolation, scores better than all the others, but shows nevertheless the expected decrements alongside with the successive shorthening steps.

The results of the experiment reported upon in this paper seem to confirm the hypothesis that speakers of (supposed) stress-timed languages such as English, and speakers of (supposed) syllable- timed languages such as Italian show different degrees of sensitivity to durational (ultimately, articulatory) characteristics concerning unstressed vowels (i.e. syllables). This line of research is therefore worth pursuing. We are currently envisaging further investigations.

Address of the authors:
Pier Marco Bertinetto, Scuola Normale Superiore, I-56100 Pisa
Carol A. Fowler, Dartmouth College, New Hampshire and Haskins Laboratories, 270 Crown Street, New Haven, Conn. 06510

Bibliographical references

Beckman, M. (1982), "Segment duration and the 'mora' in Japanese", *Phonetica*, 39: 113-135.

Benguerel, A.-P. (1986), "Time-warping and the perception of rhythm in speech", *Journal of Phonetics*, 14: 231-246.

Bertinetto, P. M. (1977), " 'Syllabic blood', ovvero l'italiano come lingua ad isocronismo sillabico", *Studi di Grammatica Italiana*, 6: 69-96.

Bertinetto, P. M. (1981), *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*, Firenze.

Bertinetto, P. M. (1988), "Reflexions on the dichotomy 'stress-' vs. 'syllable-timing' ", paper read at the First International Conference on Experimental Phonostylistics & Sociophonology and Speech Acoustic Variability, Florianópolis, S. Catarina, Brazil (the proceedings will be published by the Belgian International Center of Applied Phonetics, Bruxelles).

Bolinger, D. L. (1981), *Two kinds of vowels, two kinds of rhythms*, Indiana University Linguistics Club.

Bond, Z. S. & Fokes, J. (1985), "Non-native patterns of English syllable timing", *Journal of Phonetics*, 13: 407-420.

Brakel, A. (1985), "Towards a morphophonological approach to the study of linguistic rhythm", *Chicago Linguistic Society*, n° 21: 15-25.

Dauer, R. M. (1983), "Stress-timing and syllable-timing reanalyzed", *Journal of Phonetics*, 11: 51-62.

Duez, D. & Nishinuma, Y. (1984-85), "Some evidence on rhythmic patterns of spoken French", *Experiments in Speech Processes*, 4: 30-40.

Essens, P. J. (1986), "Hierarchical organization of temporal patterns", *Perception and Psychophysics*, 40: 69-73.

Farnetani, E. & Kori, S. (1983), "Interaction of syntactic structure and rhythmical constraints on the realization of word prosody", *Quaderni del Centro di Studio per le Ricerche di Fonetica*, 2: 287-318.

Fletcher, J. (1985), "Some effects of tempo change on timing in French", *Phonetics Laboratory Univ. Reading Work in Progress*, 5: 45-64.

Fowler, C. A. (1977), *Timing Control in Speech Production*, Indiana University Linguistics Club.

Fowler, C. A. & Smith, M. (1986), "Speech perception as 'vector analysis': an approach to the problems of invariance and variability", in S. J. Perkell & D. H. Klatt (eds.), *Invariance and Variability in Speech Processes*, Hillsdale: 123-139.

Hill, D. R., Jassem, W. & Witten, I. H. (1978), "A statistical approach to the problem of isochrony in spoken British English", *Man-Machine Systems Laboratory Report*, 78/12/6, Univ. of Calgary.

Hoequist, C. (1983*a*), "Durational correlates of linguistic rhythm categories", *Phonetica*, 40: 19-31.

Hoequist, C. (1983*b*), "Syllable duration in stress-, syllable- and mora-timed languages", *Phonetica*, 40: 203-237.

Huggins, A. W. F. (1972), "On the perception of temporal phenomena in speech", *Journal of the Acoustical Society of America*, 51: 1279-1290.

Kohler, K. J. (1986), "Invariance and variability in speech timing: from utterance to segment in German", in Perkell, J. S. & Klatt, D. H. (eds.), *Invariance and variability in Speech Processes*, Hillsdale.

Major, R. C. (1981), "Stress-timing in Brazilian Portuguese", *Journal of Phonetics*, 9: 343-351.

Major, R. C. (1985), "Stress and rhythm in Brazilian Portuguese", *Language*, 61: 259-282.

Marotta, G. (1985), *Modelli e misure ritmiche. La durata vocalica in italiano*, Bologna.

Martin, J. C. (1972), "Rhythmic (hierarchical) versus serial structure in speech and other behavior", *Psychological Review*, 79: 487-509.

Miller, M. (1984), "On the perception of rhythm", *Journal of Phonetics*, 12: 75-83.

Nishinuma, Y. & Duez, D. (1987), "Etude perceptive de l'organization temporelle de l'énoncé en français", *Travaux de l'Inst. de Phonétique d'Aix*, 11: 181-201.

Osika, B. T., Zue, V. W., Weeks, R. V., Neu, H. & Auerbach, J. (1975), "The role of phonological rules in speech understanding", *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP, 23: 104-112.

Pointon, G. E. (1980), "Is Spanish really syllable-timed?", *Journal of Phonetics*, 8: 293-304.

Price, P. J. (1980), "Sonority and syllabicity: acoustic correlates of perception", *Phonetica*, 37: 327-343.

Roach, P. (1982), "On the distinction between 'stress-timed' and 'syllable-timed' languages", in Crystal, D. (ed.), *Linguistic Controversies*, London: 73-79.

Scott, D. R., Isard, S. D., Boysson-Bardies. B. de (1985), "Perceptual isochrony in English and in French", *Journal of Phonetics*, 13: 155-162.

Vayra, M., Avesani, C., Fowler, C. A. (1984), "Patterns of temporal compression in spoken Italian", *Proceedings of the Tenth Int. Congress of Phonetic Sciences*, Utrecht.