Acoustical and perceptual study of gemination in Italian stops

Anna Esposito
International Institute for Advanced Scientific Studies ''E. R. Caianiello'' (IIASS), Via G. Pellegrino 19, I84019 Vietri sul mare (SA), Italy, INFM Salerno University

Maria Gabriella Di Benedetto
INFOCOM Department, Rome University ''La Sapienza,'' Via Eudossiana 18, I00184 Rome, Italy

(Received 10 October 1996; accepted for publication 13 May 1999)

On the basis of theoretical considerations and results from acoustic and perceptual analyses, it is hypothesized that closure duration is the primary cue for gemination in Italian. Results of an acoustic analysis of a large number of single and geminate Italian utterances show two acoustic correlates of gemination: the length of the closure and the length of the vowel preceding the consonant. Other acoustic parameters were not systematically related to gemination. These results were validated perceptually. At the perceptual level, the above cues were used by the listeners in the geminate/nongeminate discrimination; however, closure duration played a major role. Moreover, it was found that the significant lengthening of consonant was only partially compensated by the shortening of the previous vowel and by a small lengthening of the geminate utterance with respect to the nongeminate one. This result suggests that speakers follow a sort of timing (rhythm) which is fixed in duration and depends on the number of syllables in the word: words with equal numbers of syllables do not change in utterance length, an elongated segment being partly compensated by the shortening of another. This process seems to be applied also perceptually suggesting that the timing (rhythm) of a language is also an auditory attitude. © *1999 Acoustical Society of America.* [S0001-4966(99)04309-X]

PACS numbers: 43.71.Es, 43.71.Hw, 43.70.Fg, 43.70.Hs [WS]

INTRODUCTION

Some languages allow the clustering of the same consonant in vowel contexts and this phenomenon is known as ''consonant gemination.'' Gemination plays a particular role in the phonology of such languages because several words change meaning as a function of singleton versus geminate consonants (minimal pairs). Phonetic theories agree in considering the gemination of a phoneme as a particular realization of the original one (Muljacic, 1972) which is modified in some of the acoustic parameters. Recent papers (Rochet and Rochet, 1995; Shrotriya *et al.*, 1995) report that there is an acoustic relationship between consonant closure duration and gemination as well as between the length of the vowel preceding the consonant and gemination. Moreover, these studies also report that there is a perceptual relationship between closure duration and gemination, whereas variation in the length of the vowel preceding the consonant does not seem to be perceptually relevant. Other acoustic parameters which appear to be related with gemination (Shrotriya *et al.*, 1995) are the burst energy and the $F0$ values at the offset of the vowel preceding the consonant. However, there is not extensive work on this phenomenon and the results reported are based on a small number of data.

Our interest was to examine the acoustic parameters which play a role in the production of geminate consonants and to validate their perceptual importance. To this end, we set up a series of experiments for collecting the acoustic data and for synthesizing the stimuli for the perceptual tests. Our aim was to try to give an answer to the following questions: (1) Which acoustic parameters are related with gemination? Specifically, does a feature of tension associated with gemination cause more extreme vowel formants [or some other spectral shape effects as for Hindi (Shrotriya *et al.*, 1995)] for Italian geminate stops? To this end, a data base of geminate and nongeminate utterances with no semantic meaning was built up and acoustic analysis of such data was carried out. (2) Are the significant acoustic attributes of a geminate phoneme production also perceptually significant?

I. SPEECH MATERIALS AND MEASUREMENTS

A. Data

A set of vowel–consonant–vowel (the nongeminate case) and vowel–consonant–consonant–vowel (the geminate case) utterances in which the consonant was [b,d,g,p,t,k] (the complete set of stop consonants in Italian) in the environment of the vowels [i,a,u] was recorded. These bisyllabic utterances were chosen because, in Italian, many minimal pairs, such as *papa* (pope) and *pappa* (baby food), *fato* (fate) and *fatto* (fact), *rito* (rite) and *ritto* (stand up), *pala* (shovel) and *palla* (ball), are bisyllabic words; therefore, the use of a bisyllabic structure is justified by the natural attitude of the native speakers in producing it. The use of nonsense words was necessary for having in all cases a symmetrical context and the same stress pattern. Since different acoustic parameters, including durational parameters, were measured, the utterances were not included in a carrier phrase; the stress and intonation pattern of the whole sentence would obviously influence such durational parameters in a way which would be difficult to control. The utterances were produced by six Italian speakers (three male and three female). Each utterance was repeated three times for a total of 324 utterances in single form and 324 utterances in geminate form.

B. Subjects

Six Italian adult speakers with no known articulatory impairment served as subjects. Four of them were students and two of them professors at Rome University ''La Sapienza.'' All subjects were native speakers of standard Italian that they learned during childhood. Their speech was characterized by the accent of the Italian region where they spent most of their life or by the accent of the closest relatives (parents and grandparents). Some of them showed no dialectal accent, whereas some others showed a Roman dialect accent.

C. Recording procedures

The speech materials were produced by the speakers in a sound-treated room and recorded on a high-quality magnetic tape recording system. Care was taken to ensure that vocal effort and patterns of stress and intonation were reasonably natural and consistent from word to word. Before measurements were performed, the recorded utterances were evaluated and the utterances which were judged by the experimenter to be unacceptable samples of the phonemes in question were re-recorded. For example, a sample was judged to be unacceptable when mistakes happened in the recording procedure, or the speaker made pronunciation errors.

The recordings were made in the Speech Laboratory, INFOCOM Department, Rome University ''La Sapienza'' (Italy). The measurements were performed using the UNICE version 1.6 by VECSYS speech analysis program, which accepts user commands to read in waveform files and generates spectral displays of various types. The spectral representation used for the analysis of our data was the DFT (Discrete Fourier Transform) magnitude spectrum. The analysis window (Hamming window) duration was set to a default of 256 samples which corresponds roughly to 26 ms at a sampling rate of 10 kHz. The first step in the analysis procedure was to process the speech signals by a low-pass digital filter with a cutoff frequency of 5 kHz.[1]

The output of the filter was sampled at 10 kHz and stored both on a PC computer memory and on floppy disks. Sound spectrograms of all utterances and visual displays of the corresponding waveforms were also made. For these records, the criteria to perform the different measurements were established by visual inspection of the spectrogram, the waveform, and the spectrum to obtain a coherent set of measures.

D. Measurements in the frequency domain

1. V1 formant frequencies

Using a DFT spectrum, measurements of the formant frequencies, $F1$, $F2$, and $F3$, at the offset and in the middle of the vowel preceding the consonant were made. In order to avoid incorrect DFT estimations due to windowing and other side effects, these measurements were also checked for errors by visual inspection of the spectrograms, which can give a coarse idea of each formant frequency range.

2. Parameters

Once the temporal sampling points corresponding to the consonant release were located, the burst offset, and the onset of the vowel following the consonant, the following parameters were examined, using software programs developed *ad hoc* for these tasks:

(1) The burst energy: the energy in the temporal interval from the consonant release to the burst offset was computed by squaring and adding the samples contained in it. The result was divided by $V1$ energy which was computed on three vowel frames located around the middle of $V1$.
(2) The VOT energy: the energy in the temporal interval from the consonant release to the onset of the vowel following the consonant was computed. The result was divided by $V1$ energy.
(3) The burst power: the ratio between the burst energy and the burst duration;
(4) The VOT power: the ratio between the VOT energy and the VOT duration.

These computations were only made for voiceless consonants because of the difficulty in defining the temporal sampling point corresponding to the burst offset for voiced consonants. For voiceless consonants, this temporal sampling point was identified as the time (after the consonant release) at which no energy below 1.5 kHz was visible in the spectrogram.

3. DFT spectra at consonant release

All the spectra were pre-emphasized. The 256-point analysis window was placed on the waveform in order to have the burst onset fall in the zone corresponding to the maximum value of the analysis window. A visual examination of the spectra was first performed. Then, the ratio of the 0–0.3 kHz to the 0–5 kHz frequency range signal energy was computed. Finally, the spectrum energy was quantized using a vector quantization framework (Vannucci, 1994; Rossetti, 1994) by dividing the 0.3–5 kHz frequency range into frequency bands. Since low-frequency samples are influenced by the voicing, the 0–0.3 kHz band was excluded from the centroid's calculations. The total energy of the signal in the 0.3–5 kHz frequency range was computed and equally distributed over the frequency bands. Therefore, the frequency bands had different widths. Each band B_i, $i=1,\dots,N=10$ was then assigned to a centroid C_i defined by the pair of values C_{ix} and C_{iy}:

$$C_{ix}=\sum_{f_n\in B_i} f_n A(f_n) \Big/ \sum_{f_n\in B_i} A(f_n),$$
$$C_{iy}=\sum_{f_n\in B_i} A(f_n)/|B_i|, \tag{1}$$

where $A(f_n)$ is the magnitude of the nth DFT sample in the B_i frequency band, $|B_i|$ is the number of DFT samples in the

B_i frequency band, and f_n is the nth harmonic component in the B_i frequency band. The connected N code-points gave a smoothed representation of the spectrum energy distribution in the 0.3–5 kHz frequency range. Such representation was computed for each consonant in geminate and nongeminate forms and was considered as a template of the consonant burst spectrum.

E. Measurements in the time domain

1. Duration of the vowel preceding the consonant (V1 duration) and the vowel following the consonant (V2 duration)

The temporal sampling point defining the vowel onset was identified as the temporal instant at which, in the waveform, a glottal pulse appears, followed by other regular glottal pulses. The vowel onset time was set by the experimenter by placing a cursor on the waveform display. In some cases, mainly for the vowel [a], a glottal excitation was visible before the regular vowel voicing. This glottal excitation was discarded in the vowel onset measurement. The temporal sampling point defining the vowel offset was identified as the temporal instant in which, in the spectrogram, the frequency energy was lower than 1 kHz. This criterion was also used to define the offset of the vowel following the consonant. The temporal sampling point defining the onset of the vowel following the consonant was identified as the temporal instant in which a frequency energy greater than 1 kHz appeared in the spectrogram.

2. Consonant closure duration

This measure was defined as the time interval from the offset of $V1$ to the consonantal release. To identify the temporal sampling point corresponding to the consonant release, the waveform, and the spectrogram were examined in parallel. The oral release is marked in the spectrogram by an abrupt onset of energy. An abrupt release does not always occur in the case of voiced consonants, and it is not always possible to identify the release by only looking at the spectrogram. In such cases, the examination of the waveform was useful because the amplitude of the consonantal voicing is lower than the amplitude of the vowel voicing. The consonant release was identified as the instant at which there was an abrupt onset of energy in the spectrogram and/or an amplitude change in the waveform. Other durational measures, such as VOT and complete utterance duration, were obtained as difference or sum of the measurements made in Secs. I E 1 and 2.

II. RESULTS OF ACOUSTIC ANALYSES

The values of the twelve acoustic attributes, (1) $V1$ formant frequencies, (2) burst energy, (3) VOT energy, (4) burst power, (5) VOT power, (6) DFT spectra at (12) consonant release, (7) VOT duration, (8) burst duration, (9) utterance duration, (10) $V1$ duration, (11) $V2$ duration, and (12) consonant closure duration were computed for the 648 utterances (324 single and 324 geminate).

The aim of the production experiment was to try to understand whether the above parameters would show any significant difference in the single versus the geminate consonants. To this purpose, each parameter was analyzed separately as described below. An ANOVA statistical analysis was performed on all the data. Gemination was treated as a between subjects factor, whereas other parameters, such as vowel category, formant values, and consonants, were treated as within subjects factors. Test of the main effect were performed when interactions were present, that is, the effect of one factor was explored at each level of the other factor. Furthermore, on some of the duration data was applied the Maximum Likelihood Criterion.

A. Results in the frequency domain

Formant frequency values were averaged over all speakers and repetitions, keeping separate sentences differing in vowel identity, and single versus geminate consonants forms. An ANOVA analysis was performed on formant frequency values. Gemination was treated as a between subject factor, and vowel category [i,a,u], and formant values (measured in the middle and at the offset of the vowels) were treated as a within subjects factors. The ANOVA analysis showed that gemination had no effect on formant frequency values [$F(1,10)=0.485$, $p>0.1$ for $F1$, $F(1,10)=0.028$, $p>0.1$ for $F2$, $F(1,10)=0.650$, $p>0.1$ for $F3$]. Furthermore, no interaction was found between gemination and vowel category [$F(2,20)=0.113$, $p>0.1$ for $F1$, $F(2,20)=0.006$, $p>0.1$ for $F2$, $F(2,20)=0.179$, $p>0.1$ for $F3$], and between formant values measured in the middle and at the offset of the vowels [$F(1,10)=0.485$, $p>0.1$ for $F1$, $F(1,10)=0.018$, $p>0.1$ for $F2$, $F(1,10)=0.872$, $p>0.1$ for $F3$]. Although, obviously, the formant frequency values change when the vowel or the consonant context were varied, results were very stable in terms of comparison between geminate and nongeminate consonants, suggesting that there was no relationship between formant frequencies and gemination.

The ANOVA analysis performed on the burst energy values showed that gemination had no effect on burst energy [$F(1,16)=0.480$, $p>0.1$ for female; $F(1,16)=0.597$, $p>0.1$ for male]. No interaction was found between gemination and vowels and between gemination and consonants both for female [$F(2,32)=0.088$, $p>0.1$; $F(2,32)=0.259$, $p>0.1$] and male [$F(2,32)=2.956$, $p>0.05$; $F(2,32)=1.331$, $p>0.1$] speakers. Burst energy was significantly affected by consonant category [$F(2,32)=30.279$, $p<0.0001$ for female; $F(2,32)=28.988$, $p<0.0001$ for male] and to a less extent by vowel category [$F(2,32)=7.418$, $p<0.01$ for female; $F(2,32)=5.263$, $p<0.05$ for male].

The ANOVA analysis performed on the VOT energy values showed that gemination had no effect on VOT energy [$F(1,16)=0.117$, $p>0.1$ for female; $F(1,16)=0.428$, $p>0.1$ for male]. No interaction was found between gemination and vowels and gemination and consonants both for female [$F(2,32)=0.388$, $p>0.1$; $F(2,32)=0.437$, $p>0.1$] and male [$F(2,32)=4.115$, $p>0.01$; $F(2,32)=0.831$, $p>0.1$] speakers. VOT energy was significantly affected by consonant category [$F(2,32)=31.094$, $p<0.0001$ for female; $F(2,32)=29.597$, $p<0.0001$ for male] and to a less extent by vowel category [$F(2,32)=8.174$, $p<0.01$ for female; $F(2,32)=4.411$, $p<0.05$ for male]. Figure 1(a) and

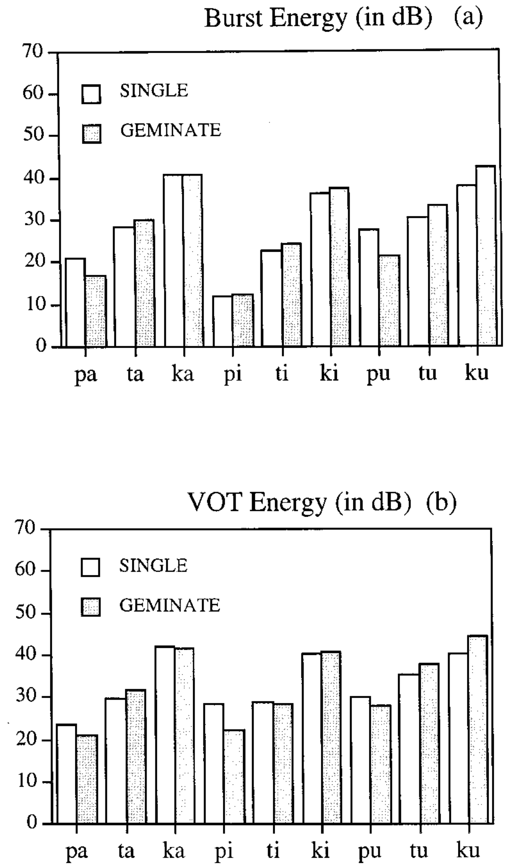

FIG. 1. Averaged values of burst energy (a) and averaged values of VOT energy (b) for single and geminate utterances. The data are averaged over speakers and repetitions.

(b) reports, respectively, the burst and the VOT energy values averaged over all the repetitions, keeping separate sentences differing in consonant place of articulation, vowel identity, and single versus geminate forms.

The ANOVA analysis performed on the burst power values showed that gemination had no effect on burst power [$F(1,16)=0.834$, $p>0.1$ for female; $F(1,16)=0.384$, $p>0.1$ for male]. No interaction was found between gemination and vowels and between gemination and consonants both for female [$F(2,32)=0.074$, $p>0.1$; $F(2,32)=1.543$, $p>0.1$] and male [$F(2,32)=3.065$, $p>0.1$; $F(2,32)=.438$, $p>0.1$] speakers. Burst power was significantly affected by consonant category [$F(2,32)=14.811$, $p<0.0001$ for female; $F(2,32)=14.117$, $p<0.0001$ for male] and to a lesser extent by vowel category [$F(2,32)=14.065$, $p<0.0001$ for female; $F(2,32)=7.526$, $p<0.01$ for male].

The ANOVA analysis performed on the VOT power values showed that gemination had no effect on VOT power [$F(1,16)=1.093$, $p>0.1$ for female; $F(1,16)=0.007$, $p>0.1$ for male]. No interaction was found between gemination and vowels [$F(2,32)=1.372$, $p>0.1$, for female; $F(2,32)=2.425$, $p>0.1$, for male] and gemination and consonants [$F(2,32)=2.339$, $p>0.1$, for female; $F(2,32)=0.378$, $p>0.1$ for male]. VOT power was significantly affected by consonant category [$F(2,32)=12.132$, $p<0.001$ for female; $F(2,32)=9.388$, $p<0.001$ for male] and by vowel category [$F(2,32)=16.409$, $p<0.0001$ for female; $F(2,32)=9.000$, $p<0.001$ for male].

Both burst energy and VOT energy values showed great variability among speakers even for the same place of articulation and the same vowel context. Burst power values showed great variability between the geminate and nongeminate case which were depending on the place of articulation, the vowel context, and the speakers. Geminate labial consonants showed lower burst power values than single ones, whereas the opposite was true for geminate dental consonants. Geminate velar consonants showed lower burst power values than single ones in the context of the vowels [i,a] and higher values in the context of the vowel [u]. However, this was not true for all speakers. For example, labial consonants in [a] context showed for some speakers higher burst power values in the nongeminate case, while for some others the opposite was true. No systematic difference between single and geminate forms was observed to be present across place of articulation and vowel context.

VOT power values showed a behavior similar to the burst power (such measurements were computed only for voiceless consonants) with no significant difference between single and geminate forms. This result was somewhat unexpected, because of the general feeling that geminates are produced with greater effort than nongeminates, resulting in a greater energy at the release.

The ANOVA analysis performed on the ratio of the signal energy in the 0–0.3 kHz to the 0–5 kHz frequency range showed that gemination had no effect on the signal-energy ratio both for female [$F(1,16)=0.029$, $p>0.1$] and male [$F(1,16)=0.041$, $p>0.1$] speakers. No interaction was found between gemination and vowels [$F(2,32)=0.153$, $p>0.1$ for female; $F(2,32)=0.635$, $p>0.1$, for male]. No interaction was found between gemination and consonants for female speakers [$F(2,32)=1.125$, $p>0.1$]. For male speakers it was found an interaction between gemination and consonants [$F(2,32)=3.965$, $p<0.01$]. We explored, in this case, the effect of gemination for each consonant and we found that gemination was not significant for all the consonants [$F(1,16)=0.587$, $p>0.1$ for [t]; $F(1,16)=1.722$, $p>0.1$ for [k]; $F(1,16)=1.006$, $p>0.1$ for [b]; $F(1,16)=0.343$, $p>0.1$ for [d]; $F(1,16)=2.837$, $p>0.1$ for [g]] except for [p] [$F(1,16)=11.609$, $p<0.01$]. Since, the effect of gemination on consonant category was not systematic, we concluded that this statistical significance was not of practical importance. However, consonant category was significant both for single [$F(5,80)=140.541$, $p<0.001$] and geminate utterances [$F(5,80)=135.587$, $p<0.001$]. Figure 2 reports the ratio of the signal energy in the 0–0.3 kHz to the 0–5 kHz frequency range for each vowel and consonant in the single and geminate form.

The ratio of the signal energy was significantly affected by consonant category [$F(5,80)=61.108$, $p<0.0001$ for female; $F(5,80)=272.162$, $p<0.0001$ for male] and by vowel category [$F(2,32)=16.021$, $p<0.0001$ for female; $F(2,32)=77.957$, $p<0.0001$ for male]. An interaction was found between vowels and consonants both for female [$F(10,160)=22.014$, $p<0.0001$] and male [$F(10,160)=35.660$, $p<0.0001$] speakers. We explored, in this case,

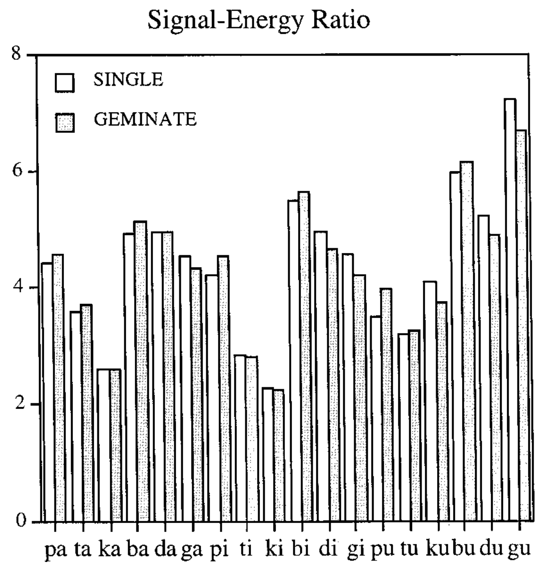

FIG. 2. Percentage of the ratio of the signal energy in the 0–0.3 kHz to 0–5 kHz frequency ranges $[En(0-0.3\ \mathrm{kHz})/En(0-5\ \mathrm{kHz})]$. The reported values were averaged over three repetitions for each speaker.

the effect of vowels for each consonant and viceversa. We found that, for male speakers, the ratio of the signal-energy values were significantly affected by vowel category in the context of [p] [$F(2,32)=7.386$, $p<0.01$], [k] [$F(2,32)=77.677$, $p<0.001$], [b] [$F(2,32)=9.340$, $p<0.01$], [g] [$F(2,32)=119.917$, $p<0.001$], but not in the context of [t] [$F(2,32)=3.850$, $p>0.01$], and [d] [$F(2,32)=5.121$, $p>0.01$]. Moreover, the effect of the consonants was significant for each vowel [$F(5,80)=105.865$, $p<0.01$ for [a]; $F(5,80)=169.329$, $p<0.01$ for [i]; $F(5,80)=116.040$, $p<0.01$ for [u]].

For female speakers, the ratio of the signal-energy values were significantly affected by vowel category in the context of [t] [$F(2,32)=33.279$, $p<0.001$], [k] [$F(2,32)=51.656$, $p<0.001$], [b] [$F(2,32)=12.799$, $p<0.001$], [g] [$F(2,32)=50.591$, $p<0.001$], but not in the context of [p] [$F(2,32)=4.714$, $p>0.01$], and [d] [$F(2,32)=0.606$, $p>0.01$]. Moreover, the effect of the consonants was significant for each vowel [$F(5,80)=29.369$, $p<0.01$, for [a]; $F(5,80)=43.321$, $p<0.01$, for [i]; $F(5,80)=40.632$, $p<0.01$, for [u]]. It is hard from these results to determine which of the two features (vowel or consonant category) play a major role in determining the signal-energy ratio values, even though the effect of consonant appear to be more systematic. Further research is necessary to evaluate this possibility, and at the moment it is beyond the aim of this paper.

DFT spectra of all consonants in single and geminate form were visually examined. The smoothed energy distribution (for the description of this measure see Sec. D 3) did not show any difference between the geminates and nongeminates. Figure 3 shows the energy distribution for vowel [a] (squares), [i] (circles), and [u] (triangles) in labial [Fig. 3(a)], dental [Fig. 3(b)], and velar context [Fig. 3(c)] (respectively) in geminate and nongeminate forms. As shown, the plots of the geminate and nongeminate energy distribution overlapped quite closely. In conclusion, no relationship between

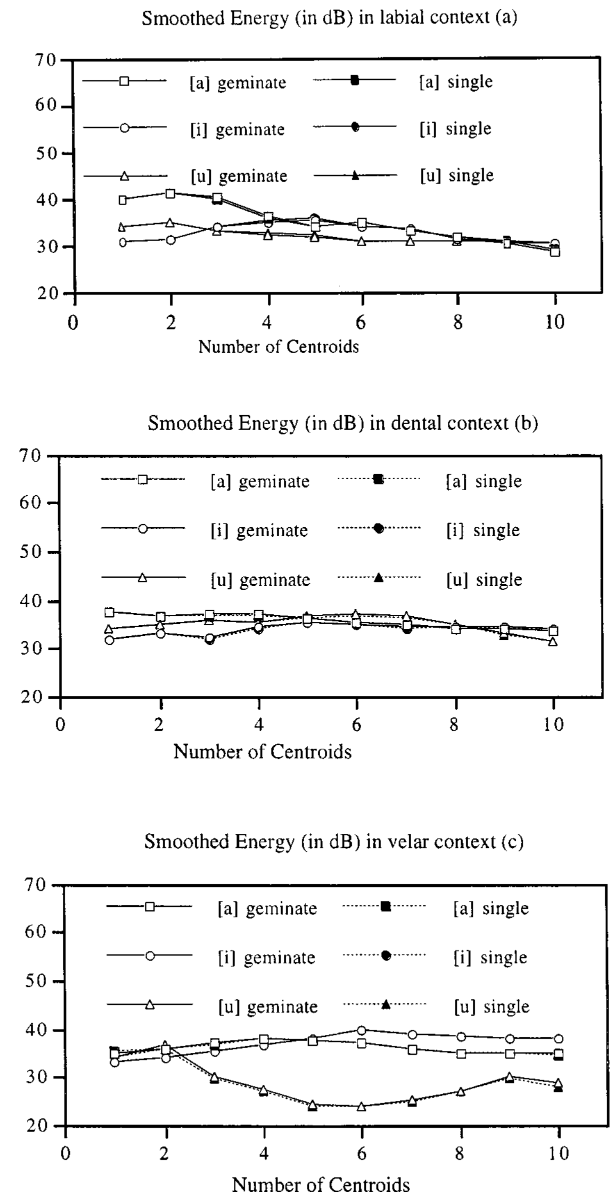

FIG. 3. Smoothed energy distribution of labial consonants in the [a] environment in geminate (empty squares) and nongeminate (filled squares) forms, in the [i] environment in geminate (empty circles) and nongeminate (filled circles) forms, and in the [u] environment in geminate (empty triangles) and nongeminate (filled triangles) forms. Labial (a), dental (b), and velar (c) consonants are plotted separately.

the energy representations and the acoustics of the gemination was observed.

B. Results in the time domain

The ANOVA analyses in the time domain were performed first by taking into account separately female and male data. The reason for that was that we expect that gender could play a role on duration measures. Successively, as suggested by the reviewers, the ANOVA analyses were performed putting together the male and female data. This was done for all the duration measures reported below (VOT, burst duration, $V2$ duration, utterance duration, $V1$ duration, and closure duration). The results obtained (collapsing together male and female data) were consistent with those obtained considering the male and female data separately, except for a little effect of gemination on utterance duration which will be discussed below. To be consistent with all the other data reported above, below are reported the ANOVA analyses performed separately on male and female data, where gemination has been considered as a between subject variable and vowels and consonants have been considered as within subject variables.

First, results which did not show any relationship with gemination are reported. The ANOVA analysis performed on VOT duration showed that gemination had no effect on VOT duration both for female $[F(1,16)=.249,\ p>0.1]$ and male $[F(1,16)=.005,\ p>0.1]$ speakers. No interaction was found between gemination and vowels $[F(2,32)=.251,\ p>0.1$ for female; $F(2,32)=1.048,\ p>0.1$ for male], and between gemination and consonants $[F(2,32)=3.173,\ p>0.01$ for female; $F(2,32)=5.326,\ p>0.01$ for male]. As we expected, VOT duration was significantly affected by vowel category $[F(2,32)=52.820,\ p<0.0001$ for female; $F(2,32)=73.254,\ p<0.0001$ for male], and by consonant category $[F(2,32)=99.734,\ p<0.0001$ for female; $F(2,32)=179.673,\ p<0.0001$ for male].

The ANOVA analysis performed on burst duration showed that gemination had no effect on burst duration both for female $[F(1,16)=0.049,\ p>0.1]$ and male $[F(1,16)=0.735,\ p>0.1]$ speakers. No interaction was found between gemination and vowels $[F(2,32)=0.349,\ p>0.1$ for female; $F(2,32)=0.423,\ p>0.1$ for male]. No interaction was found between gemination and consonants for female speakers $[F(2,32)=0.051,\ p>0.1]$. For male speakers an interaction was found between gemination and consonants $[F(2,32)=7.246,\ p<0.01]$. We explored, in this case, the effect of gemination for each consonant and found that gemination was not significant for all the consonants $[F(1,16)=4.862,\ p>0.01$ for [p]; $F(1,16)=0.152,\ p>0.01$ for [t]; $F(1,16)=6.612,\ p>0.01$ for [k]]. However, consonant category was significant both for single $[F(2,32)=16.117,\ p<0.001]$ and geminate utterances $[F(2,32)=60.375,\ p<0.001]$. Hence, we concluded that also for male speakers, gemination had no effect on burst duration.

As we expected, burst duration was significantly affected by consonant category $[F(2,32)=33.661,\ p<0.0001$ for female; $F(2,32)=69.245,\ p<0.0001$ for male], whereas vowel category did not play a significant role because we found a small effect for male speakers $[F(2,32)=5.802,\ p<0.01]$ but no effect for female speakers $[F(2,32)=1.301,\ p>0.1]$.

The ANOVA analysis performed on $V2$ duration showed that gemination had no effect on $V2$ duration both for female $[F(1,16)=2.029,\ p>0.1]$ and male $[F(1,16)=1.545,\ p>0.1]$ speakers. No interaction was found between gemination and vowels $[F(2,32)=1.943,\ p>0.1$ for female; $F(2,32)=0.958,\ p>0.1$ for male], and between gemination and consonants $[F(2,32)=1.397,\ p>0.1$ for female; $F(2,32)=0.731,\ p>0.1$ for male]. $V2$ duration was significantly affected by vowel category $[F(2,32)=46.141,\ p<0.0001$ for female; $F(2,32)=26.357,\ p<0.0001$ for male], and by consonant category $[F(2,32)=5.692,\ p<0.001$ for female; $F(2,32)=11.088,\ p<0.0001$ for male]. Figure 4 reports the averaged $V2$ durations in the single and geminate form, for the vowels [a,i,u] in the consonant context [p,t,k,b,d,g].

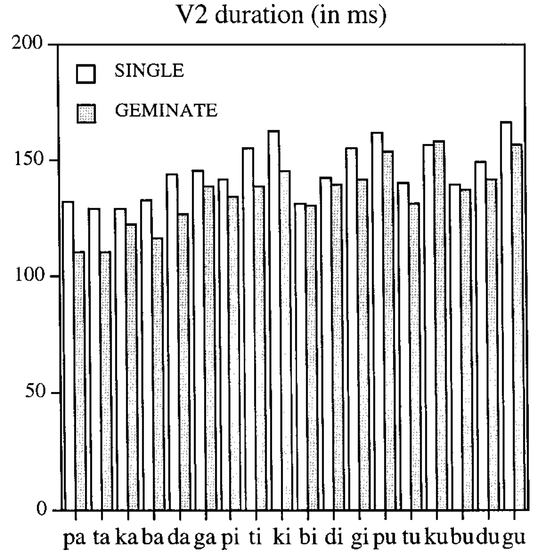

FIG. 4. $V2$ duration for single and geminate utterances. The reported values are averaged over speakers and repetitions.

The ANOVA analysis performed on utterance duration showed that gemination had no effect on utterance duration both for female $[F(1,16)=7.952,\ p>0.01]$ and male $[F(1,16)=7.065,\ p>0.01]$ speakers. No interaction was found between gemination and vowels $[F(2,32)=0.772,\ p>0.1$ for female; $F(2,32)=0.521,\ p>0.1$ for male]. No interaction was found between gemination and consonants for female speakers $[F(5,80)=1.604,\ p>0.1]$. For male speakers an interaction was found between gemination and consonants $[F(5,80)=3.749,\ p<0.01]$. We explored, in this case, the effect of gemination for each consonant and we found that gemination was not significant for [p] $[F(1,16)=3.985,\ p>0.01]$, [b] $[F(1,16)=4.602,\ p>0.01]$, [d] $[F(1,16)=4.059,\ p>0.01]$, and [g] $[F(1,16)=0.690,\ p>0.01]$, whereas it was significant for [t] $[F(1,16)=10.473,\ p<0.01]$ and for [k] $[F(1,16)=16.220,\ p<0.01]$. However, the consonant category was significant both for single $[F(5,80)=3.686,\ p<0.01]$ and geminate utterances $[F(2,32)=7.080,\ p<0.001]$. Utterance duration was significantly affected by consonant category $[F(5,80)=7.882,\ p<0.001$ for female; $F(5,80)=7.017,\ p<0.0001$ for male], whereas vowel category do not play a significant role both for female $[F(2,32)=3.530,\ p>0.01]$ and male $[F(2,32)=0.112,\ p>0.01]$ speakers. When the data for male and female were collapsed together, an effect of gemination was found on utterance duration $[F(1,32)=14.936\ p=0.0005]$ and an interaction between gemination and consonants $[F(5,160)=5.098,\ p=0.0002]$.

We explored, in this case, the effect of gemination for each consonant and found that gemination was not significant for [g] $[F(1,32)=3.720,\ p>0.01]$, whereas it was significant for [b] $[F(1,32)=8.485,\ p<0.01]$, [d] $[F(1,32)=8.485,\ p<0.01]$, [p] $[F(1,32)=11.206,\ p<0.01]$, [t] $[F(1,32)=23.921,\ p<0.001]$, and for [k] $[F(1,32)=23.596,\ p<0.001]$. The consonant category also was significant both for single $[F(5,160)=5.276,\ p<0.001]$ and

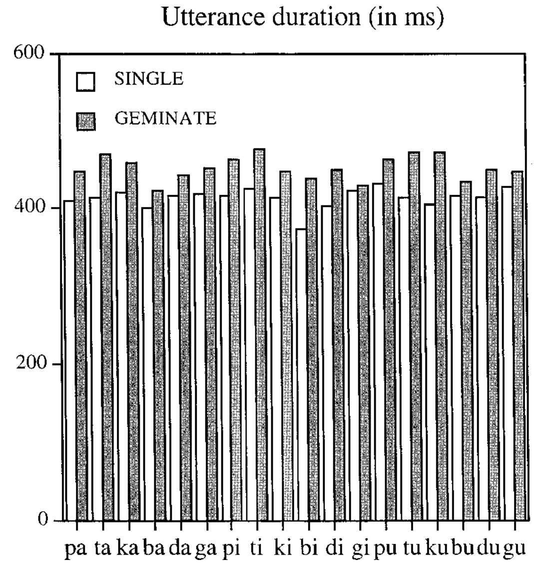

FIG. 5. Utterance duration for single and geminate utterances. The reported values are averaged over speakers and repetitions.

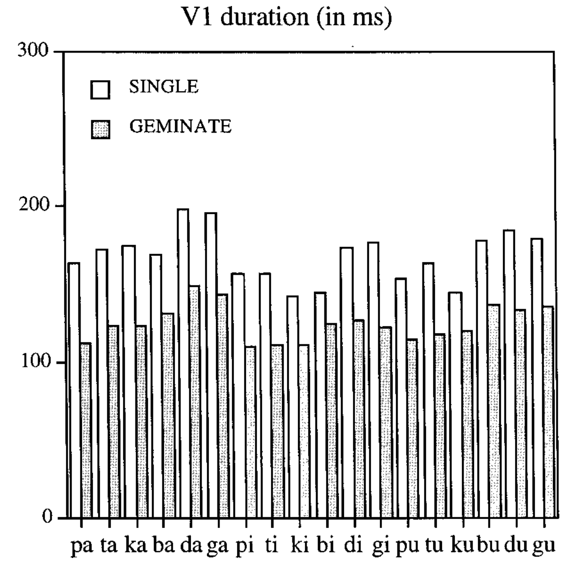

FIG. 6. $V1$ duration for single and geminate utterances. The reported values are averaged over speakers and repetitions.

geminate utterances [$F(5,160)=14.147$, $p<0.001$].

From these results, we concluded that there was an effect of gemination on consonant category. However, since the consonant category also has a significant effect on gemination, it becomes difficult to define the role played by gemination on utterance duration. Further research is necessary to evaluate this effect, and to determine whether the present result generalizes to other consonant categories.

Figure 5 reports the averaged utterance durations in the single and geminate form, for the vowels [a,i,u] in the consonant context [p,t,k,b,d,g].

It is worth noting that the standard deviation values for VOT, burst duration, $V2$ duration, and utterance duration were very high in comparison with the differences among their values in the geminate and nongeminate case.

Moreover, for these parameters, there was a great deal of variability among speakers: Some speakers did not show any durational difference, whereas some others did. Some vowel contexts showed larger differences than others. Consonantal place and voicing also played a role for such variability. We can conclude from these data that $V2$ duration, burst duration, and VOT do not play a role in the geminate/nongeminate distinction, whereas further research is necessary to define the role played by gemination on utterance duration.

The only durational parameters which showed significant differences between geminate and nongeminate consonants were the $V1$ duration and the closure duration.

For $V1$ duration, the results of the ANOVA analysis showed that gemination plays a significant role both for female [$F(1,16)=12.531$, $p<0.01$] and male [$F(1,16)=59.871$, $p<0.0001$] speakers. No interaction was found between gemination and vowels both for female [$F(2,32)=3.915$, $p>0.01$] and male [$F(2,32)=0.365$, $p>0.1$] speakers. No interaction was found between gemination and consonants for male speakers [$F(5,80)=2.680$, $p>0.01$]. For female speakers an interaction was found between gemination and consonants [$F(5,80)=6.605$, $p<0.0001$]. We explored, in this case, the effect of gemination for each consonant and found that gemination was significant for [p] [$F(1,16)=18.107$, $p<0.01$], [t] [$F(1,16)=9.704$, $p<0.01$], [d] [$F(1,16)=24.685$, $p<0.001$], and [g] [$F(1,16)=11.158$, $p<0.01$], whereas it was not significant for [k] [$F(1,16)=7.222$, $p>0.01$] and for [b] [$F(1,16)=6.731$, $p>0.01$]. Moreover, the consonant category was significant for both single [$F(5,80)=31.051$, $p<0.001$] and geminate utterances [$F(5,80)=11.715$, $p<0.001$]. From these results we concluded that gemination was of practical importance for the parameter we are considering ($V1$ duration). $V1$ duration was also significantly affected by the consonant category both for male [$F(5,80)=24.653$, $p<0.0001$] and female [$F(5,80)=36.161$, $p<0.0001$] speakers, and by vowel category both for female [$F(2,32)=17.488$, $p<0.0001$] and male [$F(2,32)=28.989$, $p<0.0001$] speakers. Averaged $V1$ durations (for geminate and nongeminate cases) as a function of vowels context and consonantal place are reported in Fig. 6.

The results of the ANOVA analysis on closure duration showed that gemination plays a significant role both for female [$F(1,16)=99.110$, $p<0.0001$] and male [$F(1,16)=258.114$, $p<0.0001$] speakers. No interaction was found between vowels and gemination for male [$F(2,32)=3.471$, $p>0.01$] speakers. For female speakers an interaction was found between gemination and vowels [$F(2,32)=9.086$, $p<0.001$]. We explored, in this case, the effect of gemination for each vowel and found that gemination was significant for all the vowels [$F(1,16)=100.819$, $p<0.001$ for [a]; $F(1,16)=81.534$, $p<0.001$ for [i]; $F(1,16)=97.257$, $p<0.001$ for [u]], whereas the effect of vowel category was significant for geminate utterances [$F(2,32)=28.620$, $p<0.001$] but not for single utterances [$F(2,32)=1.322$, $p>0.1$]. An interaction was found between gemination and consonants both for male [$F(5,80)=12.403$, $p<0.0001$] and female [$F(5,80)=10.935$, $p<0.0001$] speakers. We explored, in this case, the effect of gemination for each consonant and found that gemination was significant for all con-

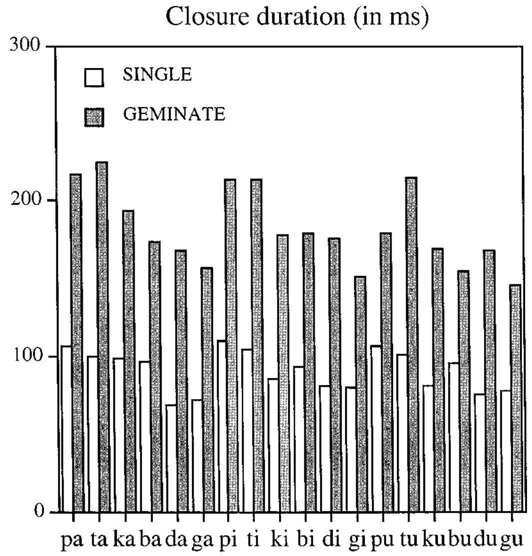

FIG. 7. Closure duration for single and geminate utterances. The reported values are averaged over speakers and repetitions.

sonants both for male [$F(1,16) = 205.908$, $p < 0.001$ for [p]; $F(1,16) = 254.019$, $p < 0.001$ for [t]; $F(1,16) = 208.164$, $p < 0.001$ for [k]; $F(1,16) = 105.225$, $p < 0.001$ for [b]; $F(1,16) = 175.347$, $p < 0.001$ for [d]; $F(1,16) = 157.184$, $p < 0.001$ for [g]] and female [$F(1,16) = 69.345$, $p < 0.001$ for [p]; $F(1,16) = 72.764$, $p < 0.001$ for [t]; $F(1,16) = 82.609$, $p < 0.001$ for [k]; $F(1,16) = 69.284$, $p < 0.001$ for [b]; $F(1,16) = 97.300$, $p < 0.001$ for [d]; $F(1,16) = 92.591$, $p < 0.001$ for [g]] speakers. Consonant category also was significant both for single [$F(5,80) = 14.893$, $p < 0.001$ for female; $F(5,80) = 19.180$, $p < 0.001$ for male] and geminate utterances [$F(5,80) = 37.786$, $p < 0.001$ for female; $F(5,80) = 77.204$, $p < 0.001$ for male]. From these results, we concluded that gemination plays a significant role on closure duration. Closure duration was significantly affected by consonant category both for male [$F(5,80) = 12.403$, $p < 0.0001$] and female [$F(5,80) = 41.743$, $p < 0.0001$] speakers, whereas vowel category showed a significant effect for female [$F(2,32) = 20.856$, $p < 0.0001$] but not for male [$F(2,32) = 4.953$, $p > 0.01$] speakers. Averaged closure durations (for geminate and nongeminate cases) as a function of vowels context and consonantal place are reported in Fig. 7.

$V1$ duration in the geminate case was observed to be reduced by about 25% with respect to its duration in the nongeminate case. Closure duration in the geminate case was significantly elongated, by about the 100%, with respect to the nongeminate case. This result was present for all speakers, vowel contexts, consonant place of articulation, and consonantal voicing.

Classification based on the Maximum Likelihood Criterion (Dillon and Goldstein, 1984) when applied to all the measured acoustic parameters confirmed that gemination was significant only for $V1$ and closure durations. The basic idea behind MLC is that the parameters of a set of data can be described through a Gaussian with mean m and a covariance matrix S. Generally this hypothesis is applied to all kinds of natural phenomena and is not a limitation. The mean m and the covariance matrix S are computed from the data. Once we compute m and S from two (or more) different set of data, the two Gaussians which describe the two sets of data are compared (to see how much they differ) using different criteria. One criterion is the MLC criterion. The method uses, as a measure of separability (separability scores) between the two Gaussians, the percentage of errors made, making, through a maximum likelihood criterion, an ''*a posteriori*'' classification of each data in the two groups.

The Maximum Likelihood Criterion applied on $V1$ alone gives a percentage of error equal to 20.4% (Error% in Table I) and as Errors/Utterances 132/648. When the MLC criterion was applied on different subsets of all the utterances, the misclassifying percentage for $V1$ remained stable (about 20%), suggesting that this parameter is not influenced by the context.

Table I shows the separability scores obtained by applying the Maximum Likelihood Criterion to closure duration, and to both closure and $V1$ duration (for different speakers, vowel contexts, and places of articulation) and to $V1$ duration alone.

The separability scores were good in both cases, although they did not improve when closure and vowel durations were considered together. These results suggested that closure duration can be considered a primary acoustic cue for the geminate/nongeminate distinction, whereas the role played by $V1$ duration must still be investigated.

III. DISCUSSION ON THE ACOUSTIC DATA

The results obtained from the acoustic measurements showed that:

(1) Formant frequency values of the vowel preceding the consonant showed no relationship with gemination, suggesting that no extra vocal effort is needed in a geminate production.

(2) There is no relationship between any representation of the energy at consonant release and gemination, in contrast with the general feeling that geminate consonants must show, at the release, greater energy than singletons. We measured the burst energy, the burst power, the VOT energy, the VOT power, the spectrum shapes, the energy distribution, and other energy measures (see Sec. II A). None of these acoustic measurements showed any significant relationship with gemination. Shrotriya *et al.* reported that:

> ''...the burst of geminate consonant is stronger (by about 10 dB) as compared to the burst of nongeminate'' (cf. Shrotriya *et al.*, 1995, ''Acoustic and perceptual characteristics of geminated Hindi stop consonants,'' ICPhS95 4, pp. 134).

However, they did not give a precise definition of the measurements made and moreover, this result could be attributed to a language specific effect (Hindi vs Italian).

(3) The acoustic parameters which appeared strongly related with gemination were the durations of the intervocalic consonant (which is close to twice the duration of the singletons) and of those of the vowel preceding it. The above parameters were found to be significant also in

TABLE I. Separability scores obtained by applying the Maximum Likelihood Criterion (MLC) to Closure Duration (CLd), to both Closure and $V1$ Duration ($V1d$), and to $V1$ duration alone. Percentage of errors and number of errors over number of utterances are reported. (The voiced/voiceless feature, the gender of speakers, the vowel context, the place of articulation, and each consonant are considered separately.)

	MLC on CLd		MLC on CLd and $V1d$	
Context	Errors%	Errors/Utterances	Errors%	Errors/Utterances
Overall	3.9%	25/648	3.9%	25/648
Voiceless	2.5%	8/324	1.9%	6/324
Voiced	2.8%	9/324	2.5%	8/324
Male	3.1%	10/324	2.8%	9/324
Female	5.6%	18/324	4.3%	14/324
[a]	2.8%	6/216	3.2%	7/216
[i]	2.8%	6/216	4.2%	9/216
[u]	4.6%	10/216	5.1%	11/216
Labials	2.8%	6/216	3.7%	8/216
Dentals	2.3%	5/216	2.8%	6/216
Velars	3.2%	7/216	2.8%	6/216
[b]	3.7%	4/108	2.8%	3/108
[d]	0.9%	1/108	0.9%	1/108
[g]	0.9%	1/108	0.9%	1/108
[p]	1.9%	2/108	1.9%	2/108
[t]	0.0%	0/108	0.0%	0/108
[k]	2.8%	3/108	2.8%	3/108
	MLC on $V1$ alone			
	Errors%	Errors/Utterances		
$V1d$	20.4%	132/648		

Hindi geminate consonants (Shrotriya *et al.*, 1995), suggesting that this effect is language independent.

Results obtained from the measurements in the time domain needed a more accurate discussion. Closure and $V1$ durations were found to be the only acoustic parameters which indicate the presence or absence of a geminate consonant, hence, it appeared natural to face the problem of what this implies at a higher level.

We observed that the length of $V2$ was shorter in the geminate case, but this shortening was not significant. VOT and burst lengths remained roughly unchanged in both the single and geminate case. In the geminate case, the length of the closure (on the average) was twice the length of the singleton closure and the length of $V1$ was reduced by about 25%. The length of utterance was, on the average, longer in the geminate case. However (even though the results from the ANOVA analysis showed that gemination has a little effect on utterance duration) for this parameter, there was a great deal of variability among speakers: Some speakers did not show any durational difference, whereas some others did; some vowel contexts and consonant contexts showed larger differences than others. Considering that there was a great lengthening of the consonant, a partial reduction of $V1$, and that the other durational parameters as VOT, burst duration, and $V2$ duration were not significantly affected by gemination, we would expect a great lengthening of the geminate utterance with respect to the single one. However, the lengthening of the geminate utterance was considerably lower than expected.

This effect cannot be explained on the basis of a hypothesis of ''anticipatory and backward compensation'' suggested by Lindblom and Rapp (1973) for Swedish. The above theory attributes the segmental reduction of a phoneme to the number of syllables that precede (anticipatory compensation) and follow (backward compensation) it. Surely, our data showed an effect of anticipatory compensation which is evident in the reduction of $V1$. The effects of the backward compensation were negligible even though they can play a role in the similar timing of the single and geminate utterances. However, this compensation did not depend on the number of syllables in the word (which in our experiment are the same for the single and geminate utterances). Thus it could only be attributed to the acoustic attributes of the phonetic segments composing the utterance. Therefore, all other factors being equal, if the only significant acoustic attribute to discriminate a segment from another is its duration, and if it is necessary to maintain a constant timing because of the well-defined stress conditions, then speakers perform on the other segments a phonetic shortening. This could suggest that there are temporal constraints on the rhythmic structure of a word (for well-defined conditions of lexical stress and number of syllables) that speakers unconsciously tend to maintain, by balancing the durational change of some of the phonetic segments (which are distinctive only through a durational feature) with the durational change of others (which are distinctive through other features not related to segmental duration).

IV. PERCEPTUAL EXPERIMENT

In order to test the relevance of the acoustic parameters which distinguish geminate versus single consonants, a perceptual experiment was carried out. The aim of this experiment was to define a closure length which works as a threshold in the perception of gemination (that is, below such average closure duration consonants are perceived as singleton; above it consonants are perceived as geminate). Moreover, this experiment was also devoted to investigate the perceptual role played by $V1$ duration in geminate and nongeminate contexts.

The experiment was carried out using /apa/ and /appa/ stimuli. The use of such stimuli (which are not representative of all consonants and vowel contexts) was aimed to evaluate the perceptual relevance of the durational difference observed in the acoustic data (see Sec. II). The perceptual experiment was intended as a validation of a general effect. In order to obtain quantitative estimations of this effect, the above experiment should be duplicated with others VCV and VCCV utterances.

A symmetrical context was used because it removes a spectral cue: the rate of change of formants through the closure. Huggins (1972a) showed that, in medial stops, a symmetrical vowel context had no effect on the perception of different closure durations, except for an increased sensitivity of the subjects to the durational changes. Since we are investigating the perceptual relevance of durational changes, a symmetrical context was selected for our experiment.

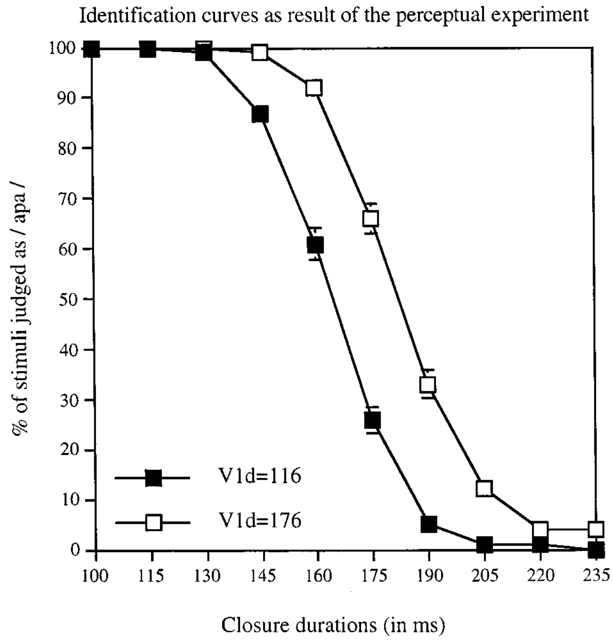

FIG. 8. Identification functions of /apa/ stimuli for $V1=116$ ms (filled squares) and $V1=176$ ms (empty squares) versus different consonant closure durations. The identification curves are average functions, where each point plots the mean of 10 observations for each of 20 subjects (also reported is the Standard Deviation around the data points).

A. Stimuli

The stimuli were synthesized as follows: a natural /apa/ token, spoken by one subject, whose durational behavior was close to the average durational behavior of all the subjects, was extracted from the database. The vowel and closure durations in the original stimulus were 176 ms and 99 ms, respectively.

The digitized signal was then modified by means of a waveform editor (UNICE editor) to produce two stimuli, by decreasing the length of $V1$ from 176 ms ($V1$ duration in the original token) to 116 ms, in a step corresponding to 5 $F0$ periods (60 ms). The reason for this choice was that, in voiced sounds, as shown by Huggins (1968), the minimum step size which could be used without introducing an abnormal pitch period was one period of the fundamental. The $F0$ periods were removed from the middle portion of the vowel to leave the VC and CV transitions intact. For each stimulus obtained, ten new stimuli were produced by increasing the length of the silent portion of the intervocalic consonant ([p]) from 100 to 235 ms in steps of 15 ms. This yielded to a total of 20 stimuli (2 vowel durations$\times$10 consonantal durations).

B. Subjects

Subjects were 20 native speakers (12 males and 8 females) of standard Italian enrolled at Rome University, with no known hearing impairment. Their age ranged from 21 to 29 years.

C. Experimental tasks

Subjects were asked to identify the stimulus words as geminate or singleton. The experiment was run separately for the two vowel durations, 176 ms and 116 ms, with the order of presentation balanced across subjects. For each vowel duration, the ten closure durations were presented such that each stimulus was preceded and followed once, by every other. Subjects listened to a total of 101 stimuli in each stage. Between two subsequent stimuli, there was a 2-s silent interval to allow the subjects to answer, a 4-s music interval to minimize any psychoacoustic effect due to the previous stimulus, and a 2-s silent interval to let subjects concentrate on the next stimulus. The subject's task was to identify each stimulus as either /apa/ or /appa/ by typing an appropriate key on a computer keyboard. Subjects listened to a total of 202 stimuli played in random order via a computer program and delivered through good quality headphones. Some stimuli (about 25) were played, for practice, before the experiment start. Subjects were run individually, in sessions lasting about 32 min each.

D. Results

Results are presented below (Fig. 8) by two identification functions where the varying closure durations are plotted for a fixed $V1$ duration [$V1$duration$=116$ ms (filled squares), $V1$ duration$=176$ ms (empty squares)]. The two identification functions in Fig. 8 are average functions, where each point plots the mean of 10 observations on each of 20 subjects. The behavior of the identification functions of each speaker was similar to the above identification functions (Neter and Wassermann, 1974).

The identification functions reported in Fig. 8 show two different closure duration thresholds at which the perception of a consonant as geminate overcomes the chance. A shorter $V1$ duration (filled squares) requires a shorter closure duration to allow the perception of a geminate consonant; in this

TABLE II. Summary of the perceptual experiment results. Mean threshold values for the two different $V1$ durations (116 and 176 ms). The mean threshold differences (positive for each listener), the Standard Deviation (SD), and the result of a two-tailed student t-test applied over all the listeners (last row in the table) are also reported.

$V1d=116$ ms	$V1d=176$ ms	Differences in ms
Mean 165.8	182.7	16.9
SD 9.5	16.1	9.6
$t(19)=7.882$, $p<0.001$		

case, the average closure duration at which the perception of a geminate overcomes the chance was about 165 ms. However, a longer closure duration (about 183 ms) was required when $V1$ duration was longer (empty squares). The displacement (17 ms on the average) in the perceptual threshold for gemination, when $V1$ duration was longer, was present for all listeners. Table II summarizes the results of the perceptual test. The difference among the threshold values for the two different $V1$ duration is positive for all listeners, which reinforces the observation that when $V1$ is lengthened the closure duration must be longer for perceiving a geminate consonant. Moreover, a two-tailed student t-test was applied to the data which show that the differences are statistically significant [$t(19)=7.882$ at level $p<0.001$].

V. DISCUSSION ON THE PERCEPTUAL DATA

The first result of our perceptual experiment was that stimuli with shorter $V1$ duration needed a shorter closure duration to allow geminate perception. This leads back to the discussion on the acoustic data, and particularly on the different lengths of the geminate and nongeminate utterances. In the previous discussion, we reported that the differences between the length of geminate and nongeminate utterances were not statistically significant, even though the geminate utterances were, on the average, longer than the single ones. That is not a contradiction with the fact that the closure length of a geminate is almost twice the length of a singleton. Our explanation was that speakers follow a sort of timing (rhythm) which is fixed in duration and depends on the number of syllables in the word: substituting a singleton to a geminate in words (leaving unchanged the number of syllables) does not significantly change the utterance length, an elongated segment (the geminate consonant) being partly compensated by the shortening of another (the vowels). The data from the perceptual experiment showed a similar effect in the listening process; i.e., changing the duration of the vowel in the utterance results, for a fixed consonant length, in a different discrimination threshold between singleton and geminate. Generally, if the vowel is shortened, a singleton could be perceived as a geminate. This confirms our hypothesis of a constant word timing for well-defined conditions of lexical stress and number of syllables. Speakers follow this timing even though the nature of the segments composing the word changes (as in the single versus the geminate case); listeners are sensitive to this timing by balancing changes in the utterance duration with changes of their perceptual durational thresholds (as in the shorter versus longer utterances).

The parameters ($V1$ and closure duration) which showed acoustic relationships with gemination were the object of a further analysis to define their perceptual role. We found that after a given threshold value which depends on the $V1$ duration, a consonant is always perceived as geminate. This means that the length of closure duration was used by listeners, as a perceptual cue, to distinguish between geminate and nongeminate consonants. However, it was also an acoustic attribute of the signal (see results on the acoustic data), supporting our hypothesis that closure duration is a distinctive feature for gemination. Since all segment durations which are to act as a cue must be perceived as reference to some baseline, this durational feature was relative rather than absolute. This is a very interesting finding because only durational parameters are found to play a role—whereas duration is often found to be relevant in connection to other acoustic attributes. Several authors pointed out that the pattern of duration of individual phonetic segments conveys information about the linguistic content of an utterance. Miller (1956), for example, has argued that duration should be classified as a distinctive feature for the recognition of vowels. Similarly, Lisker (1957) has shown that, in American English, if the stop-closure is lengthened in the [b] in **ruby** the word is heard as **rupee**, suggesting that it was a durational parameter (the VOT length) that made the distinction between voiced and voiceless stops. Bastian *et al.* (1961), and Liberman *et al.* (1961) reported similar results in studies on other consonants. Finally Klatt (1976) emphasized the role of durational features as a primary perceptual cue in the distinction of long versus short vowels, voiced versus voiceless fricatives, phrase final versus nonfinal syllables, stressed versus unstressed vowels, and presence or absence of emphasis. Moreover, results on the perceptual role of the closure length in the geminate distinction are in agreement with our finding. Pickett and Decker (1960) measured the phoneme boundary between a single and a double /p/ in the pair **topic** and **top pick**. At a speaking rate of six syllables per second, they found the phoneme boundary between **topic** and **top pick** to correspond to a closure duration of about 160 ms. Furthermore, Rochet and Rochet (1995), and Shrotriya *et al.* (1995) showed that for native speakers of different languages, closure duration is perceptually relevant in the geminate versus singleton distinction.

What about the reduction in length of $V1$? From an acoustic point of view, we expect that $V1$ before a geminate be shorter than before a nongeminate because in the first case it is part of a closed syllable and in the latter case it is part of an open syllable. However, our perceptual data show that the longer the $V1$ duration, the greater the crossover value (measured in ms) in perceiving a consonant as geminate. This result was present for all listeners, suggesting that the shortening of $V1$ could not be attributed only to syllable structure. Our interpretation is that there might be two superimposed effects: one due to syllable structure and one due to the presence of a geminate consonant. This conclusion is not in disagreement with the findings reported by Rochet and Rochet who showed that Italian listeners distinguished between *fato* and *fatto* on the basis of consonant duration but not on the basis of vowel duration. We concluded that $V1$ is shortened to balance the abnormal lengthening of the closure in order to keep the rhythm constant, and make the utterance sound natural. This interpretation is also supported by Huggins (1972a) who found that subjects were much more sensitive to changes in vowel duration than to changes in closure duration, and that these changes were perceived as changes in the sentence rhythm, when the duration of the other segments in the utterance remained unchanged.

Informal listening of synthetic stimuli (unpublished data), in which the vowel duration was kept constant while the closure was lengthened up to a value typical of closure of a geminate showed that, in this case, the utterance was heard as broken into two syllables, a long [a] followed by a [pa]. This result, although only preliminary, support [together with the results reported by Kozhevnikov and Chistovich (1965) and by Huggins (1972b)] our hypothesis that the perception of timing in natural speech is based on rhythm rather than on sound segments, and explains why compensation occurs between vowels and consonants if the sentence is to remain temporally fluent.

ACKNOWLEDGMENTS

Acknowledgment goes to Carmela Marangi, Maria Marinaro, Gian Antonio Mian, and Maurizio Ricciardi. The authors thank Riccardo Rossetti and Armando Vannucci who contributed to the acoustic measurement analyses and to the setup of the perceptual experiment. The authors would like to thank the editor Winifred Strange and two anonymous reviewers for their thorough reviews. This work was funded by Rome University ''La Sapienza.''

[1]The software UNICE which was used for digitizing the data allows the use of appropriate oversampling factors in order to obtain a correct A/D conversion. Thus if the filter passbands are 12 kHz, one cannot use a sampling rate lower than 40 kHz. To overcome this limitation, UNICE can perform a sampling rate conversion by a factor of two or four (the user sampling rate can be twice or four times lower than the real sampling rate). By using this capability, the available sampling frequency range starts from the half of the filter passband.

Bastian, J., Eimas, P. D., and Liberman, A. M. (**1961**). ''Identification and discrimination of a phonemic contrast induced by a silent interval,'' J. Acoust. Soc. Am. **33**, 842(A).

Dillon, W. R., and Goldstein, M. (**1984**). *Multivariate Analysis* (Wiley, New York).

Huggins, A. W. F. (**1968**). ''How accurately must a speaker time his articulation,'' IEEE Trans. Audio Electroacoust. **AU-16**, 112–117.

Huggins, A. W. F. (**1972a**). ''Just noticeable differences for segment duration in natural speech,'' J. Acoust. Soc. Am. **51**, 1270–1278.

Huggins, A. W. F. (**1972b**). ''On the perception of temporal phenomena in speech,'' J. Acoust. Soc. Am. **51**, 1279–1290.

Klatt, D. H. (**1976**). ''Linguistic uses of segmental duration in English: Acoustic and perceptual evidence,'' J. Acoust. Soc. Am. **59**, 1208–1221.

Kozhevnikov, V. A., and Chistovich, L. A. (**1965**). ''Speech: Articulation and perception,'' (Moscov-Leningrad), (English Translation: J.P.R.S., Washington D.C., No. JPRS 30543).

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., and Bastian, J. (**1961**). ''An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance,'' Lang. Speech **4**, 175–195.

Lindblom, B., and Rapp, K. (**1973**). ''Some temporal regularities of spoken Swedish,'' Papers from the Institute of Linguistics, Stockholm University, **21**, pp. 1–62.

Lisker, L. (**1957**). ''Closure duration and the intervocalic voiced–voiceless distinction in English,'' Language **33**, 42–49.

Miller, G. A. (**1956**). ''The perception of speech,'' in *For Roman Jakobson* (Mouton, The Hague), pp. 353–360.

Muljacic, Z. (**1972**). *Fonologia della lingua Italiana* (II Mulino, Bologna).

Neter, J., and Wasserman, W. (**1974**). *Applied Linear Statistical Models* (Irwin, Momewood).

Pickett, J. M., and Decker, L. R. (**1960**). ''Time factors in perception of a double consonant,'' Language and Speech **3**, 11–17.

Rochet, L. B., and Rochet, A. P. (**1995**). ''The perception of the single-geminate consonant contrast by native speakers of Italian and Anglophones,'' in *Proceedings of ICPhS95*, edited by K. Elenius and P. Branderud, Vol. 3 (Arne Strömbergs Grafiska, Stockholm), pp. 616–619.

Rossetti, R. (**1994**). ''Gemination of Italian stops,'' J. Acoust. Soc. Am. **95**, 2874.

Shrotriya, N., Siva Sarma, A. S., Verma, R., and Agrawal, S. S. (**1995**). ''Acoustic and perceptual characteristics of geminate Hindi stop consonants,'' in *Proceedings of ICPhS95*, edited by K. Elenius and P. Branderud, Vol. 4 (Arne Strömbergs Grafiska, Stockholm), pp. 132–135.

Vannucci, A. (**1994**). ''Acoustic correlates of distinctive features of Italian stops,'' J. Acoust. Soc. Am. **95**, 2874.