ACOUSTIC AND PERCEPTUAL CHARACTERISTIC OF ITALIAN STOP CONSONANTS

Loredana Cerrato, Mauro Falcone

Fondazione Ugo Bordoni
Speech Communications Group

ABSTRACT

We report in this paper the results of a study carried out to analyse the acoustic and perceptual characteristics of Italian stop consonants. The aim of this study is twofold: give an acoustical description of Italian stops and investigate which are the perceptual cues relative to their place of articulation.

From the acoustic point of view we report: the measurements relative to the length of the whole consonant and of its release burst; the F1 and F2 of the following vowel measured at the beginning of it. Moreover we counted the presence of the release burst and we tried to describe its acoustical characteristics in terms of the spectral structure as suggested by Blumstein [1] [2].

From the perceptual point of view we report the results of three perceptual tests that we run with the aim of evaluating whether the release burst or the formant transitions are more relevant for the perception of Italian stop consonants' place of articulation.

1. INTRODUCTION

Although it is commonly agreed that the acoustic cues which make the identification of stop consonants possible lie in the burst portions and in the adjacent transition segments, there is no unanimity as to the relative contribution of each cue [3] [4] [5] [6]. In particular for Italian there are very few studies that investigated the acoustic information of stop consonants [7] [8] [9], our study represents a pioneering work which gives some insight into the relative importance of the different acoustic cues of stop consonants.

2. SPEECH MATERIAL

We had to build an *ad hoc* corpus for our study as all the available databases didn't seem to contain enough material for our scope. We created a kind of "building-sentences task" to elicit semi-spontaneous speech from 10 speakers (5 male and 5 female university students from the area of Rome) who were recorded in our labs while trying to build a series of 12 sentences containing target syllables.

Target syllables were VCV and VC:V, with V=a and C= /b d g p t k/ and C:=/b: d: g: p: t: k:/ that is the complete set of the Italian single and geminated stop consonants. The VCV and VC:V sequences all have the same structure /a/+stop+/a/, in order to minimise the acoustic and phonetic variability due to the coarticulation phenomena.

All the target syllables were inserted in stress position in real words which were embedded in meaningful sentences such as: *Il papà di ADA ha zAPPAto nell'orto con la zAPPA rotta.*

2.1. Corpus for the acoustic analyses

The material used for the preliminary acoustic analysis was extracted from the production of VCV and VC:V sequences by all the 10 speakers in two different speaking styles: semi-spontaneous speech and read speech.

The parameters we analysed are:

1. total duration of consonantal segment (interval between the end of the preceding vowel and the beginning of the next vowel);
2. burst length (if present);
3. frequency of F1 and F2 of the following vowel, measured at the beginning of the vowel;
4. description of the burst in terms of its structure.

The results of the acoustic analysis in the time domain (points 1 and 2) are reported in table 1 and 2.

CONSONANT DURATION	spontaneous	read
Single stop	66 m	71 ms
Geminate stop	134 ms	112 ms

Table 1: Average values of the syllable duration of the Italian stop consonants measured in our corpus, for semi-spontaneous and read speech.

BURST DURATION	spontaneous	read
Single stop	7 ms	8 ms
Geminate stop	11 ms	10 ms

Table 2: Average values of the burst duration of the Italian stops measured in our corpus, for semi-spontaneous and read speech.

No other European languages, among the most common ones, have geminated consonants like Italian. For this reason, no studies have been conducted to analyse and describe geminated stop consonants. Our analysis show that geminated stops have a total length which is almost the double of single stops, and the release burst appears to be longer in geminated stops showing at times a particular "double" realization.

The main results relative to the frequency domain are reported in table 3.

VOWEL	F1 (Hz)	F2 (Hz)
reference /a/	*600*	*1500*
pa	620	1145
ba	620	1320
ta	620	1420
da	620	1545
ka	630	1610
ga	640	1585

Table 3: Average values of the first and second formant of the following vowel, measured at the beginning of it.

The value of F2 represents the transition from the consonant to the vowel, therefore it varies according to the place of articulation of the preceding consonant. For instance in the case of bilabial stops /b, d/, that have a very low place of articulation, as the F2 value is lowered towards their articulation point.

The results relative to the spectral structure of the release burst are quite consistent with those reported in [1]:

- DA diffuse raising pattern with energy around 2.5 kHz for dental consonants;
- DD diffuse falling pattern with energy around 1.5 kHz for labials;
- C compact pattern with energy around 1.7 kHz for velars.

We also counted the presence of the release burst in all the stop consonants. The results, reported in table 4, show that it ranges from 66% to 96% depending on the style of speech (semi-spontaneous vs. read) and on the type of consonant (single vs. geminated, bilabial vs. dental vs. velar).

Percent of burst presence	spontaneous	read
Single stop	66 %	85 %
Geminate stop	95 %	96 %

Table 4: Percent value of the presence of the release burst.

2.2 Corpus for the perceptual tests

From the speech material gathered for this first analysis, we extracted a smaller corpus consisting of 6 VCV e 6 VC:V syllabic segments containing the 12 Italian stop consonants uttered in semi-spontaneous speech by 2 male and 2 female speakers. In total we selected 48 stimuli. With this corpus we created three sets of stimuli for three different perceptual tests.

1. The first set, which we call **syllable**, is made of syllabic stimuli having a slightly variable duration (±1ms) which is around 250 ms.
2. The second set, which we call **burst**, is made of short stimuli, of the length of about 25 ms (±1ms) which we extracted from the syllable stimuli. They consist of the short portion containing the *release burst* relative to the stop consonants.
3. The third set, which we call **transitions**, consist of the same stimuli forming the syllable set, from which, this time we took out the short segment containing the release burst. The length of this stimuli is about 225 ms.

2.3 Editing criteria

The target words were excised from the sentences with the aid of a digital computer program, which displays the waveforms and spectrograms of the syllables to be analysed and plays them if required.

Attention was paid to cut the waveform where the value is as closest to zero as possible.

Syllable set

From the target words we selected the syllables with the following criteria: the beginning of the stimulus was edited from the end of the transitions with the previous sound and the end was marked at the beginning of the transition with the following sound.

Burst set

From the syllable stimuli, we extracted the burst stimuli with the following criterion: we started to edit the signal portion from the end point, that is where on the waveform it is visible the beginning of the final vowel of the VCV syllable and from a perceptual point of view it is hearable the sound of the vowel.

Usually this point corresponds to the second periodic peak visible on the waveform, while the first peak should represents the burst. Starting from this point and going backward on the time axis for about 25ms we selected our stimuli. This segmentation assures that always the same part of the consonant-vowel transition is included in the stimuli, that is only the first two visible peaks on the waveform. Moreover when the burst is not visible on the waveform this criterion it still allows to accomplish the segmentation (see fig.1 and fig 2).

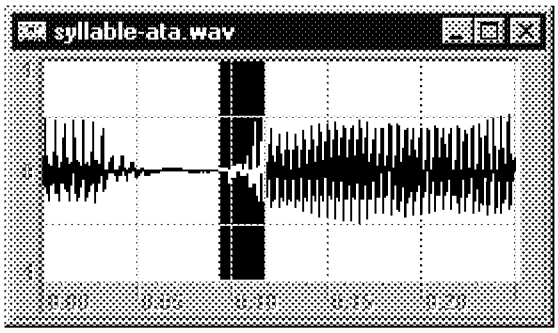


Figure 1: A typical segmentation of a CVC stimulus. The highlighted portion in the waveform is the selected burst stimulus.

Of course, it could be argued that, in order to evaluate the perceptual importance of the release burst we shouldn't have

included any part of the previous or following transition. But this would not be possible as the effective duration of the burst alone is on average 10 ms, which is practically too short to be perceivable.

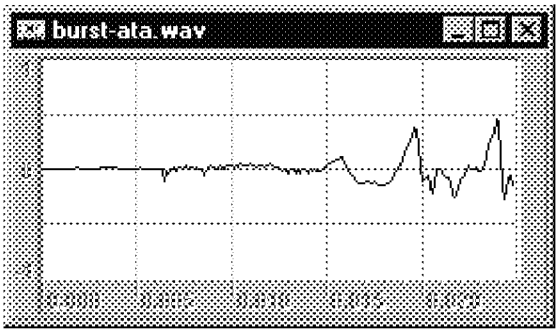


Figure 2: Detail of the release burst waveform.

Transition set

These stimuli consist of the same stimuli forming the syllable set, from which we took out the short segment representing the release burst, as a consequence they have a length of 225 ms.

3. EXPERIMENTS SET UP

Three perceptual test were run, one for each set of stimuli:

1. test syllable
2. test transition
3. test burst

3 different groups of 20 listeners (university students, aged between 21 and 30) served as subjects for each test. They had no hearing pathologies and nobody was an expert in perceptual phonetics. In all the tests they listened to the stimuli which were presented in a random order.

Each stimulus was repeated three times with a short pause of approximately 0.5 seconds and it was introduced by a voice announcing it by a sequential number. At the end of each stimulus presentation the listeners had to make a choice among the twelve possible Italian stops (six singles and six geminates) reported on an answer sheet. Before carrying out each test, the subjects underwent an initial phase of training of the duration of approximately 5 minutes. They listened to the stimuli over headphones, in binary modality with fixed volume at a comfortable level about 73dB SPL.

4. RESULTS

The results of these tests, summarised in table 5, show that the accuracy of identification of the consonantal place of articulation from burst stimuli only is very low (25%). The accuracy of identification increases dramatically when the subjects are presented with stimuli made of the burst and the transition with the following vocalic sound. Moreover the accuracy of identification rises even when VCV syllables deprived of the segment containing the release burst are presented.

In total, without considering the confusion between single and double consonants (see column % correct+ in table 5), the results of the syllable test show a correct perception of the stimuli for 94%. If we consider also the mistakes due to the confusion between single and double consonants (see column % correct in table 5), we still have a high percentage of correct answers 75%, and the 20% of incorrect answers are due to the confusion between singles and geminates and vice versa.

In particular the results of the syllable test show that only two types of mistakes occurred: /t/ perceived as /d/ with 1.46%; /k/ perceived as /g/ 2.71%. This confusion occurred mainly because in semi-spontaneous speech unvoiced stop consonants tend to have an incomplete closure, or no closure at all, sounding more like fricatives than stops; but as we didn't put the fricatives among the possible answers, subjects identified these fricativised unvoiced stops as the relative voiced stop.

TEST TYPE	% correct	% correct+
Syllable set	74.4	94.2
Transition set	51.1	76.8
Burst set	12.8	25.3

Table 5: Summary of the results obtained in the three subjective tests. The value % correct is the percent of stimuli correctly perceived. The % correct + is the percent of percent of stimuli correctly perceived allowing confusion between a single and its geminate or vice-versa.

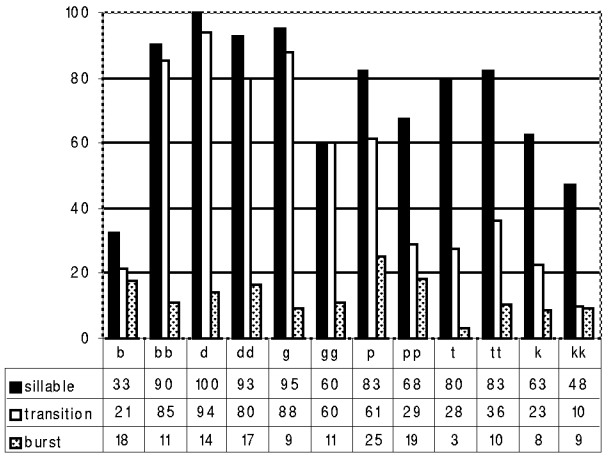
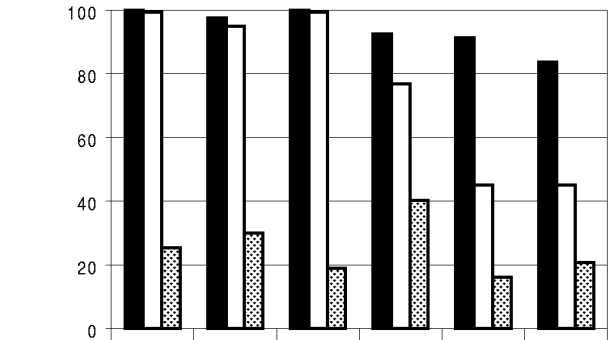


Figure 3: Percent of correctly perceived stimuli in the three subjective tests for all the single and geminate stop consonants of the Italian language.

In figure 3 it is reported a very high degree of confusion between the perception of /b/ and /b:/ in both syllable and transition stimuli. This is due to the fact that the speech material we used is affected by a specific phenomenon of regional pronunciation: the very low correct perceived rate (33%) of /b/ depends on the production of the voiced bilabial /b/ that in the

central Italian regional variation (and in particular in the area of Rome) is usually produced as geminated.

The results for the transition test need further discussion. While the deletion of the burst affects only slightly the performance of listeners, who can still correctly perceive the voiced stop consonants /b,d,g/, on the contrary the perception of the unvoiced stops /p,t,k/ and their geminates, appears to be affected by the deletion of the burst, with the consequent loss of about 10 to 50 percent in correct perceived stimuli. It is important to outline that these errors are not equally distributed in the set of stimuli. There are in fact four stimuli which are always misperceived by all listeners: a /k/ perceived as /g/ (or /g:/), a /k:/ perceived as /t/ (or /t:/), a /p/ perceived as /b/ (or (/b:/), and finally a /t/ perceived as /d/. As all listeners perceive these stimuli in the same wrong manner, we believe that the stimuli really "sound as" a different stop consonant. In this case it is hard to consider this result as an error. Probably the manipulation of the speech signal, i.e. the removal of the burst, deeply affected the nature of these stimuli.



	b+	d+	g+	p+	t+	k+
sillable	100	98	100	93	91	84
transition	99	95	99	77	45	45
burst	25	30	19	40	16	21

Figure 4: Percent of correctly perceived stimuli, allowing confusion between singles and geminates, in the three subjective tests.

The percentage of the correctly perceived stimuli in the three tests is reported in figure 4. In the burst test the result is so low that there are no evidence to support any hypothesis. In fact if we consider that a random choice will have a value around 8%, it is clear that the obtained value of 13% is not indicative of any relation between the perceived stimulus and the subjects' choice. None of the single or geminate stops have a high score of correct perception. This means that the burst itself does not deliver any information about the stop consonant's place of articulation independently of its phonetic characteristic. In other words the judgement given in this test appears to be almost a random choice.

5 CONCLUSIONS

Deriving conclusions from the results of our tests, we have to underline that the limited speech material set we used might have affected our results. Our corpus was constrained by three different factors: the fixed context of the stop consonant in the VCV syllable (V=a), the speech modality that is near to spontaneous speech and the limited speaker class (only young speakers from the same area). For a wider investigation less constraints are of course necessary. Nevertheless the results of this study provide some interesting information on the acoustic and perceptual cues of Italian stops, and in particular that: Italian stop consonants in syllabic context are correctly identified by listeners, while the stimuli representing the release burst seem not to be sufficient for the correct identification of the stop consonants' place of articulation. Moreover with the transition stimuli we obtain a very good percentage of correct identification, which support the hypothesis that the acoustic information relative to the place of articulation of Italian stop consonants doesn't lie in the release burst portion, but, in the transition with the previous and following vowel [6,8].

5. REFERENCES

1. Blumstein, S., Stevens, K., "Acoustic invariance in speech production: evidence from measurements of spectral characteristics of stop consonants", *JASA 66*, pp.1001-1017, 1979

2. Blumstein, S., Stevens, K. "Perceptual invariance and onset spectra for stop consonants in different vowel environments", *JASA. 67*, pp.648-662, 1980

3. Cooper, F., Delattre, P.C., Liberman, A. M., Borst, J. M., Gerstman, L., "Some experiments on the perception of synthetic speech sounds", *JASA. 24*, pp.597-606, 1952

4. Liberman, A.M.,"Some Results of Research on Speech Perception", *JASA. 29*, pp.117-123, 1957

5. Bonneau, A., Djezzar Laprie, Y., "Perception of the place of articulation of French stop burst", *JASA. 100*, pp.555-564, 1996

6. Kewley-Port, D., "Representation of spectral change as cues to place of articulation in stop consonants", *Technical Report n.3 Research on speech perception*, Bloomington Indiana University Press, 1980

7. Cerrato, L., Falcone, M., "Il burst nelle occlusive in sequenze VCV e VC:V dell'italiano: un'analisi acustica", *Atti delle VIII° Giornate di Studio del Gruppo di Fonetica Sperimentale (in press)*, Pisa 1997

8. Albano Leoni, F., Maturi, P., "Forma e sostanza nei suoni del linguaggio", in L'Interfaccia tra fonetica e fonologia E. Magno Caldognetto (a cura di) Studi di Linguistica applicata Unipress, pp.115-126

9. Landi, R., "Le consonanti occlusive in stili differenti di parlato", *Atti delle 7° Giornate di Studio del GFS 1996*, pp. 143-155, Napoli, 1996