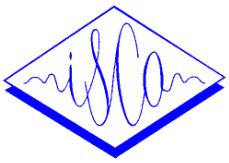

4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

THE SPATIAL AND THE TEMPORAL DIMENSIONS OF CONSONANT REDUCTION IN CONVERSATIONAL ITALIAN

Edda Farnetani
e-mail: Farnetani@csrf00.csrf.pd.cnr.it
Centro di Fonetica del C.N.R. - Padova, Italy

ABSTRACT

This study is an electropalatographic investigation of the spatiotemporal characteristics of consonants /t/,/d/,/n/,/l/, /s/ in Italian, produced by two subjects in spontaneous, conversational speech. The data show that consonant reduction is a continuous process, inversely related to duration, in agreement with Lindblom's model [1,2]. Duration-dependent target undershoot is manifested as a progressive decrease in tongue tip/blade contact and a retraction of the constriction location, to extreme cases where no sign of front contact is detectable (and the consonant is no longer perceived). An important finding is that the degree and the frequency of undershoot is consonant-specific and that the reduction patterns lawfully mirror the tongue-body coarticulation patterns observed in previous research on isolated words. This indicates the consonant-specific production constraints are preserved across speaking styles.

1. INTRODUCTION

1.1. Reduction and style

The acoustic study on Swedish vowels by Lindblom [1] is the first investigation showing that reduction is a duration-dependent and contextually-determined continuous process, and is the first attempt to uncover its underlying mechanism. Lindblom's "duration dependent undershoot" model assumes that reduction, and the inverse relation between duration and undershoot, is the automatic response of the speech system to invariant vocalic and consonantal commands, which, in rapid speech overlap in time. In the revised version of the model [2], based on comparison between two speech styles, vowel reduction is still dependent on duration but its degree can vary as a function of actively controlled style-related factors, which affect the velocity and the amplitude of the articulatory movements. This account implies that different styles entail different target specifications along a continuum from hyper- to hypo-articulation: consequently the observed speech variability reflects in part underlying invariants, and in part a reorganization of speech gestures.
According to articulatory phonology [3], vowel reductions, consonant lenitions (e.g. incomplete closures in stops), and other casual speech processes like deletions, are all accounted for by an increase in gestural overlap and a decrease in gestural amplitude. The former automatically reduces duration and movement displacement, the letter automatically reduces movement velocity and displacement. So, also within this framework reduction tends to increase as duration decreases. A fundamental difference between Lindblom's theory and the gestural framework is that here all speech variability can be accounted for by assuming an underlying gestural invariance, since different speech styles do not entail gestural changes, but only quantitative changes in the gestural parameters. Although the two different perspectives are hardly testable, one advantage of gestural phonology is the unitary account of reductions and deletions.
Data on reduction in spontaneous speech are scarce. In a pilot study by Engstrand and Krull [4] on selected CVC samples of spontaneous speech, the vowels show clear duration-dependent undershoot effects which parallel the systematicity observed in laboratory speech. Another finding is that within the same range of durational variations the vowels show less reduction in content words than in adverbial phrases, indicating that the pronunciation style tends to be more elaborate in semantically more important words. A subsequent experiment on the diphthong /aI/ in spontaneous speech [5] shows duration-dependent undershoot patterns, and a comparison between spontaneous speech and citation forms indicates that in the former both /a/ and the glide fail to attain the formant values of the corresponding citation-forms.
The EPG study by Shockey and Farnetani [6] on coronal consonants in spontaneous English and Italian, shows that in both languages consonant reduction ranges along a continuum from full to incomplete closures, to extreme cases where no sign of tongue-to-palate gesture is detectable and the key consonant is no longer perceivable. A second finding (concerning Italian) is that a consonantal context is more powerful than any vocalic context in determining the presence or absence of reduction. Sequences of /rt/, /lt/, /tr/, /ld/, /nd/ always favour complete closures in the key consonants, while in /st/ clusters, complete /t/ closure is never achieved, instead in VCV sequences reduction tends to occur in any vocalic context. The present study continues the investigation on Italian, with an extended speech material and new perspectives.

1.2. The present study

This study examines the articulatory characteristics of /t/, /d/,/n/,/l/, and of the two allophones of the alveolar fricative, the word initial [s] and the intervocalic [z], produced by two speakers of standard Northern Italian in spontaneous speech.
The aims of this research are:
a) to determine whether and to what extent the degree of articulatory reduction is predictable from consonant duration;
b) to examine whether the frequency and degree of reduction depends on consonant identity. A reasonable hypothesis is that consonant-specific production constraints should be preserved to some extent in spontaneous speech: accordingly, the degree of reduction should have to vary across consonants, and mirror the tongue body coarticulation patterns which are strictly related to consonant specific constraint (as shown in [7], and in a number of other studies).

c) the third aim is to compare the production characteristics of these consonants with those observed in citation forms. The study also explores the possible effects of position in the word, initial vs medial (also final for /n/) and of word category (function vs content) on the frequency and the degree of reduction.

2. METHODS

Speech materials were two-second samples of spontaneous speech, 60 for Sub. GB, 75 for Sub. FC, randomly selected from one hour conversation with another native-language speaker. The recording system was the Reading EPG-3, which simultaneously records EPG data (s.r.100 Hz) and the speech signal (s.r. 10.000 Hz). The artificial palate contains 62 electrodes distributed in eight rows along the front/back dimension. Phonemic and narrow phonetic transcription were made through careful auditory analyses from large to small speech samples to the key VCVs. Cs are *intervocalic* single and gemimated consonants occurring within and across word boundaries. A preliminary EPG analysis was carried out for determining frequency of full, reduced, and deleted closures, and, for the fricative, to determine the number of deletions. In a subsequent analysis of non deleted consonants, the following EPG parameters were measured:
DUR = duration of the constriction movement; the onset was taken at the point in time where the first EPG evidence of the consonant appears (usually at the third or fourth row location), the offset time was the point where the last consonant-related contact disappears;
MAX-CNT = The total number of on-electrodes in the first four rows, taken at the point of maximum contact;
MEAN-CNT = The average of on-electrodes during the 40 or 50 ms of maximum contact;
FP = Front Profile: calculated by summing the row order number where the frontmost contact appears (Row 1, 2, etc.), to the percentage of non-activated electrodes in that row, so that the index increases as the frontmost contact shifts back.
Although MEAN-CNT (or MAX-CNT) and FP are related to each other, and together reflect the amplitude of the constriction movement, the separate use of the two is an attempt to distinguish between variations in tongue tip/blade height (relatable to the amount of contact) and those along the front/back dimension (relatable to FP).
For fricatives one more parameter was used: ROWN = the row number(s) where maximum constriction occurs.
Each consonant in spontaneous speech was compared with 15 to 20 repetitions of the same consonant in initial and medial position of isolated words in the contexts of /a/ and /i/. The comparisons were made using MAX-CNT (or MEAN-CNT) and FP.
For all statistical analyses the significance level was set at $p<0.02$.

3. RESULTS

3.1. Unreduced, reduced and deleted consonants

There were on the average 50 occurrences of each consonant per subject; /s/ had the highest frequency (72 for GB and 56 for FC); and /t/ the lowest (36 for GB and 41 for FC). Figure 1 shows the frequency of occurrence of full, incomplete, and deleted closures for stops, nasals and laterals. Fricatives exhibit two deletions for GB and none for FC. As shown in the figure, this preliminary analysis

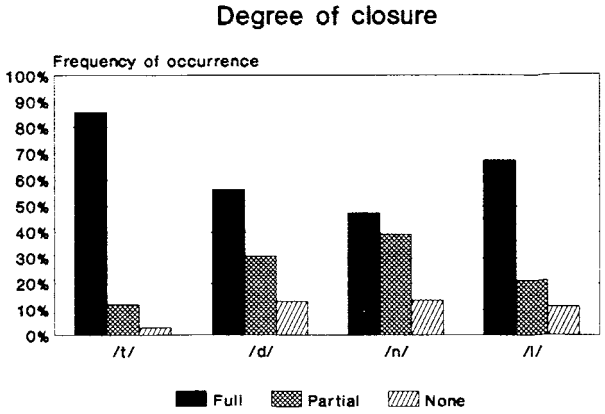

Fig.1. Degrees of closures (data are averaged across subjects)

indicates that fricatives and voiceless stops tend to be more resistant to reduction than the other consonants, and that /n/ appears to be the least resistant of all.

3.2. Spatiotemporal aspects

The analyses refer to both full-closure and reduced-closure consonants: this was necessary for determining the effects of duration on the EPG contact patterns, for whatever amount of contact.
The results of correlations indicate that spatial and the temporal parameters are significantly correlated, with the highest *r*s occurring between DUR and MEAN-CNT, (those between DUR and MAX-CNT were highly significant for all consonants except /s/ for GB).
The results of regression analysis with DUR as independent variable and MEAN-CNT as dependent variable show that the regression coefficients (RC) are all significantly different from 0, in line with the predictions of Lindblom's model. Table 1 shows the RC and R^2 values for each consonant and subject. The RCs indicate the rate at which changes in duration induce changes in contact. They are all highly significant ($p<0.001$), except for /s/ in GB ($p<0.02$). Their apparently low values are due the different scales of the variables (DUR (ms) ranges from 0 to 260, n electrodes from 0 to 25). So, RC= 0.106 (/t/ for GB) means that the contact tends to vary of one electrode at durational changes of 10 ms.

	GB		FC	
	RC	R^2	RC	R^2
/t/	0.106	0.540	0.039	0.315
/d/	0.092	0.430	0.135	0.564
/n/	0.171	0.556	0.226	0.784
/l/	0.188	0.539	0.183	0.653
/s/	0.015	0.082	0.039	0.264

Table 1. Values of Regression Coefficients and of R^2 for each consonant and subject, obtained by regressing EPG front contact (MEAN-CNT) against duration (DUR).

The table shows that for both subjects /s/ has the slowest rate of contact change and the lowest R^2; especially for GB, the proportion of variance accounted for by this relation is only 8%. The consonants tending to change most as a function of duration are /n/ and /l/ for both subjects.
The data also show a significant inverse relation between MEAN- or MAX-CNT and FP, and between DUR and FP

for /d/, /l/, /n/, (and /s/ for FC), suggesting that reduction is manifested as a lowering of the tongue and shifting back of the constriction. This is interesting especially for /s/ where negative and significant RCs were observed between the amount of contact and the row with maximum contact (ROWN * MEAN-CNT), indicating that as the constriction becomes less tight, the location of maximum constriction shifts back.

The scatterplots in Fig.2 illustrate the difference between /n/ and /s/ for FC.It can be seen that the slope is steeper and the data points cluster more closely around the regression line for /n/ than for /s/. She /s/ scatterplot also shows the inverse relation between DUR and FP.

Comparisons among the consonants were also made by confronting the ranges of variations of the different param-

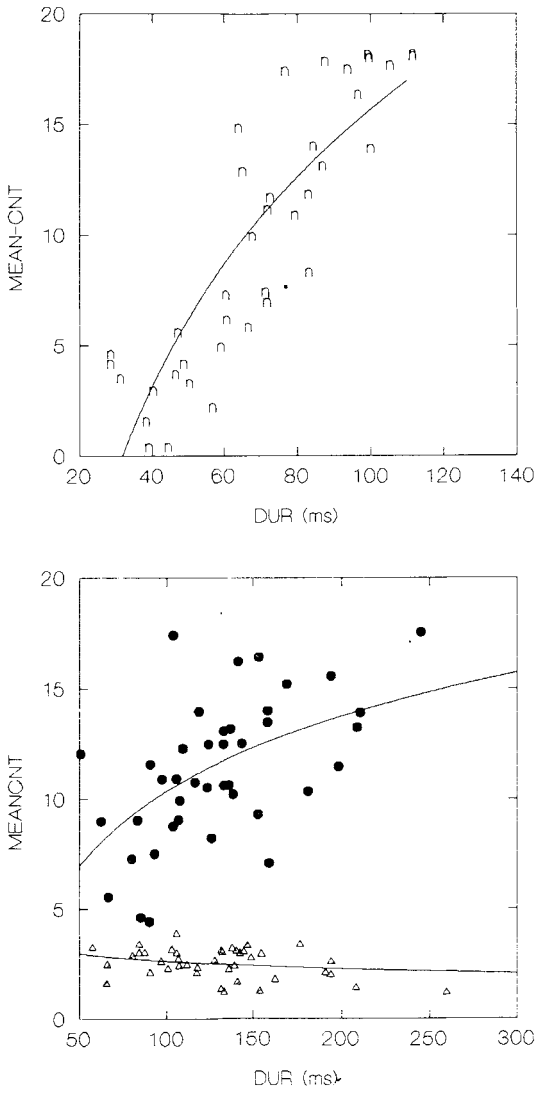

Fig.2. Scatterplot of MEAN-CNT against DUR for /n/ (top), and of MEAN-CNT and FP (Front Profile) against DUR for /s/ (bottom) for Subj. FC.

eters, as shown in Fig.3. Here the two subjects are strikingly similar: it can be seen that the durational ranges are quite small for /n/ and /l/, and rather large for /t/ and especially for /s/; instead, the range of variations of linguopalatal contact show the reverse trend. This mean that much less variation

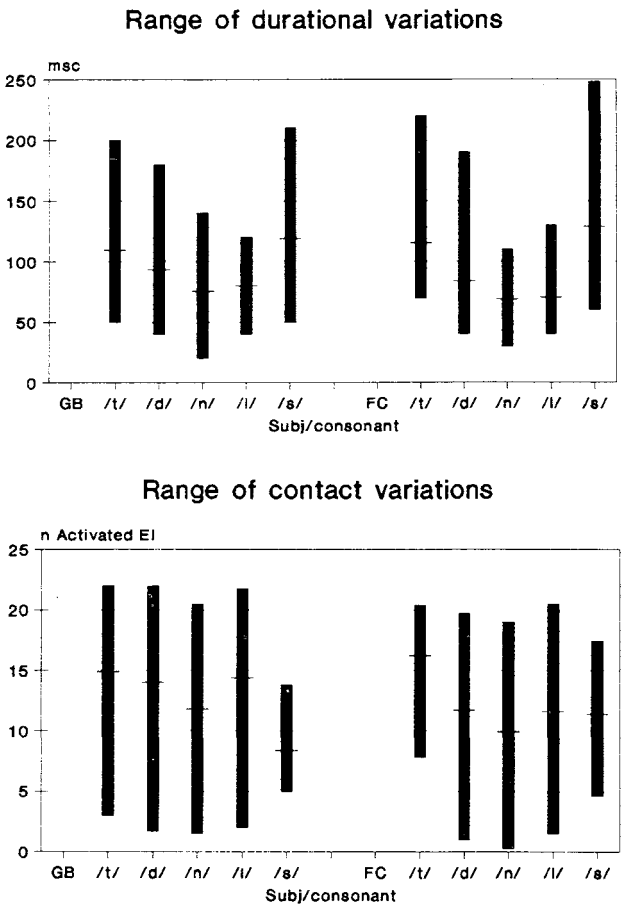

Fig. 3. Ranges of durational variations (top) and of EPG front contact variations (bottom) for each consonant and subject.

in amount of contact is allowed for /t/ and /s/ as a function of duration. The overall data converge in suggesting that the rate of duration dependent undershoot of the tongue tip/blade target is consonant-specific, and tends to decrease from /n/ and /l/ to /d/ to /t/ to /s/. This parallels the trend observed in coarticulation studies for a number of Italian subjects (GB is one of those) where tongue body coarticulaton tends to decrease from /l/ to /d/ to /t/ to /s/ [7].

3.3 Comparisons with citation forms

In the citation forms /t/, /d/, /l/ and /n/ showed all a complete front closure. The results of t-tests indicate, as expected, that consonants in isolated words tends to have a significantly higher EPG contact and a significantly lower FP than in spontaneous speech. A dicrepancy between the subjects concerns /s/, whose production differs substantially, and whose FP differs significantly in the two styles only for GB. The spatial characteristics of the consonants in the two styles are summarized in the plots of MAX- or MEAN-CNT against FP shown in Fig.5, where the data of the two subjects are pooled (except for /s/). The ellipses contain 80% of the data points. The plots show the ranges of variation along the two dimensions, the similarity and the dissimilarities between styles and among consonants. There is a partial overlap between the two styles for both subjects. The stops are the most similar across the subjects, the fricatives the most dissimilar, with a wide and advanced

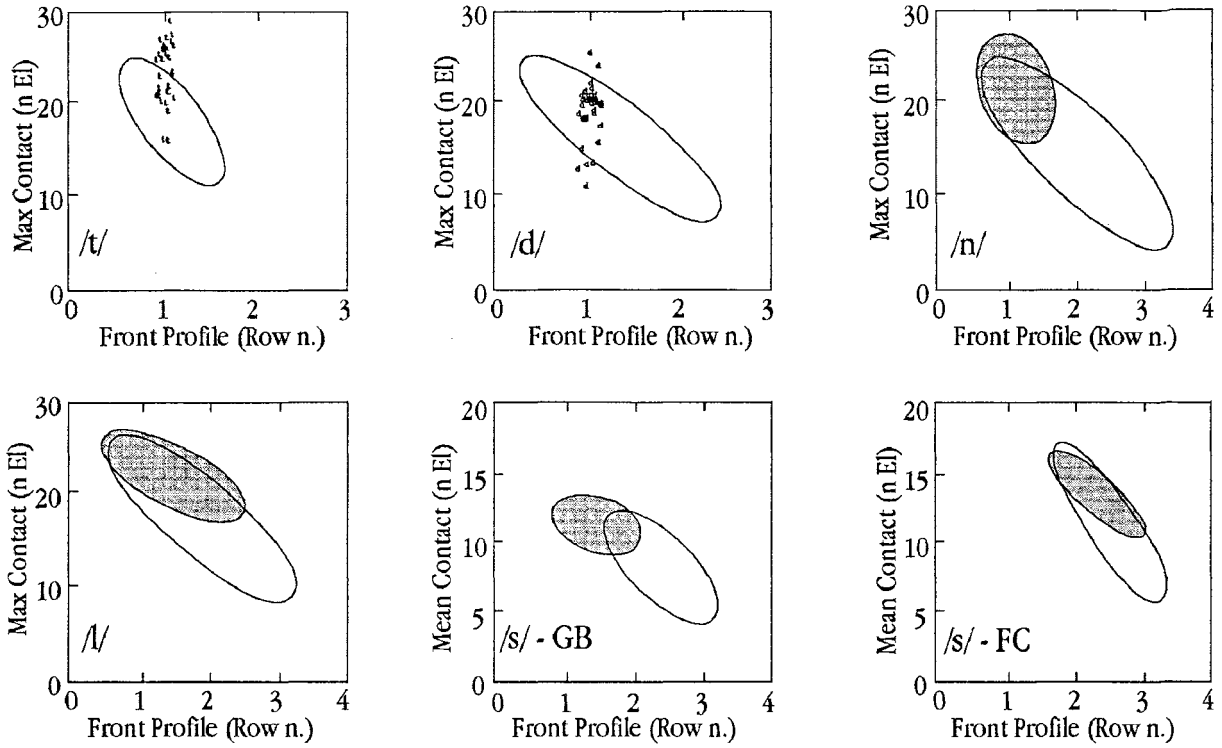

Fig.4. Plots of EPG front contact against Front Profile, for each consonant, comparing citation forms (shaded areas for /n/,/l/,/s/, and data points for /t/ and /d/), with spontaneous speech (unshaded areas). The data of the two subjects are pooled, except for /s/.

constriction for GB, and a narrower and more typically alveolar constriction for FC; for this subject the /s/ data exhibit the largest overlap between the two styles.

3.4 Word boundary and word function

The differences between initial (or final) and medial position in the word, and between content and function words were tested with a series of ANOVAs; the results were integrated with the count of deleted consonants in each condition. The data indicate that the effects of position are significant only for /l/ in FC and for /t/ and /d/ in GB. The deleted consonants are as many at word boundaries as in word medial position for both subjects. The effects of word function for FC are highly significant for /l/ and /n/, but do not reach significance for the other consonants; for GB they are significant for /d/ and /s/. For both subjects, however, the deletions in function words are more than twice the deletions in content words, independently of consonant type. Globally, with all the caveats due to this unconstrained speech material, it seems that consonants in function words (articles, prepositions, negative adverbs), tend to be reduced more than consonants in content words, in agreement with the findings in [4], whilst the effects of position appear to be weaker and less systematic.

4. CONCLUSIONS

This articulatory study shows that the articulation of consonants in spontaneous speech exhibits systematic variations predictable in part from consonant duration and in part from consonant-specific production constraints. The most important result of this study is the lawful relation found between the reduction patterns of tongue tip/ blade and the tongue-body coarticulation patterns observed in previous studies with highly controlled speech material. This indicates that some systematic articulatory features persist in any speech style. The data indicate that spontaneous colloquial speech is far from being a homogeneous sequence of reduced targets, but is rather a composite structure where more or less elaborate forms alternate with extremely casual productions. The differences observed between content and function words, and the failure to find word boundary effects, indicate that the factors shaping articulation in spontaneous speech must be sought at the higher level linguistic, semantic and informational structure.

5. REFERENCES

[1] Lindblom, B. *Spectrographic study of vowel reduction.* J.A.S.A 35, 1963, pp. 1773-1781.

[2] Moon, S.J. Lindblom, B. *Interaction between duration, context, and speaking style in English stressed vowels.* J.A.S.A. 96, 1994, pp. 44-55.

[3] Browman, C. and Goldstein, L. *Articulatory phonology: An overview.* Phonetica 49, 1992, pp. 155-180.

[4] Engstrand, O. and Krull, D. *Determinants of spectral variation in spontaneous speech.* Procs. SPEECH RESEARCH 89', Budapest, 1989, pp. 88-91.

[5] Lindblom, B., Brownlee S., Davis, B. and Moon, S.J. *Speech transforms.* Speech Communication 11, 1992, pp. 357-368.

[6] Shockey L. and Farnetani E. *Spontaneous speech in English and Italian.* Procs. ICSLP 92, 1, 1982, pp. 767-769.

[7] Farnetani E. *Coarticulation and reduction in coronal consonants: comparing isolated words and continuous speech.* PERILUS XIV, 1991, pp.11-15.