

An electropalatographic study of Greek spontaneous speech

Katerina Nicolaidis

Department of Theoretical and Applied Linguistics,
Aristotle University of Thessaloniki
knicol@enl.auth.gr

This paper aims to examine spatio-temporal variability during the production of the lingual consonants /t, d, s, z, k, g l, n, r/ in a sample of Greek spontaneous speech. It provides a description of the range of segmental variability and explores the relationship between articulatory and temporal variability. It also discusses instances of common connected speech processes including assimilations and deletions. Data from two Greek speakers producing a monologue was recorded with the technique of electropalatography (Reading EPG system). The results indicated that segmental variability ranges over a continuum from over-articulated to under-articulated forms, i.e. fully articulated tokens to very open articulations. In addition, duration dependent variability was found to be a function of the production characteristics of the consonant. Evidence of both gradient and categorical changes in spontaneous speech were suggested by the data and are discussed with reference to current theoretical frameworks.

1 Introduction

The study of spontaneous speech provides a challenge to the speech scientist. On the one hand, spontaneous speech is the natural output of everyday communication and as such it should be the obvious object of study. On the other hand, the multiplicity and inter-relationship of factors that shape and influence spontaneous speech make it a difficult task to confront. It is well known that spontaneous speech is characterised by large variability resulting from numerous factors including contextual, prosodic and speaker effects, choice of speaking style, environmental factors, such as presence of noise, and so on. A simultaneous study and evaluation of all these factors is difficult to include in the research design.

A further challenge presents itself, however, when the researcher attempts to explain and model variability within current theories. One of the core issues addressed concerns the question of whether common, so called connected speech processes, e.g. assimilations, elisions, insertions, etc., are categorical in nature involving changes in the organisation of the underlying units or whether they can be explained as gradient processes due to spatio-temporal variation in gestural execution.

Articulatory Phonology (Browman & Goldstein 1989, 1990, 1992) accounts for variations in fluent casual speech in terms of gradient modifications involving (a) a reduction in the magnitude of gestures and/or (b) an increase in gestural overlap. Reduction typically results in weakenings, for instance, incomplete closure for stops or fricative production of stops. Increase in the overlap of gestures may vary in degree and its articulatory and acoustic effects depend on the nature of the gestures involved; it may result in gestures being hidden or blended with other gestures. The postulation that no categorical changes of underlying units are involved in connected speech processes stems from the assumption that gestures are invariant units by definition present in the lexical item.

For the Theory of Adaptive Variability or Hyper-Hypo Speech (H&H; Lindblom 1983, 1990), reduction and coarticulation play a central role in accounting for variability. Fast rate may be characterised by target undershoot and increase in contextual influences. Depending on the communicative situation, however, the speaker can adapt to the particular requirements for, e.g. formal, clear, casual speech, by adopting different strategies and varying the degree of coarticulatory or reduction effects. As such, the speaker can reorganise his/her phonetic gestures and produce forms ranging over a hyper/hypo-speech continuum. The theory places strong emphasis both on production constraints and listener oriented constraints and suggests that ‘the critical conditions that phonetic gestures must meet is that they be *perceptually sufficiently contrastive*’ (Lindblom 1991:14).

Evidence for a reorganisation of gestural units is provided by Kohler for German (1990; 1991a,b; 1992), who claims that connected speech processes imply modification of the basic units of speech, i.e. elimination, addition and changes of gestures. In congruence with the H&H theory, he suggests that a balance needs to be maintained between minimisation of articulatory effort, which results in connected speech processes, and perceptual contrast.

Studies investigating articulatory and acoustic aspects of spontaneous speech have been relatively limited, although there has been an increasing interest in the acquisition and analysis of spontaneous speech data. Larger coarticulatory effects in spontaneous speech than in isolated words were reported in Krull (1989) for Swedish. Variation in the degree of consonant reduction in spontaneous speech was reported in an EPG study by Shockey (1991) for English and Shockey & Farnetani (1992) for English and Italian. Articulation of the alveolar gesture of coronals was reported to range over a continuum from complete closure to incomplete closure to very open articulations. Differences in the reduction characteristics of consonants in the two languages were attributed to language specific phonology. Further work by Farnetani (1995) for Italian showed that consonant reduction is a continuous process relating to duration and depending on the articulatory characteristics of the consonant.

In view of the above, the current study aims to examine articulatory variability in Greek spontaneous speech. The issues it will address include:

- the range of variability in tongue-palate contact patterns during the production of lingual consonants;
- the relationship between the degree of tongue palate contact and consonant duration; while duration dependent undershoot may be expected, the factors that may constrain reduction effects will be explored;
- the articulatory characteristics of some common connected speech processes in Greek spontaneous speech; and
- the overall variability present in the spontaneous speech data with reference to the presence of categorical vs. gradient processes in speech production.

Table 1 Number of tokens of the consonants /t, d, s, z, k, g, l, r/ produced by speakers JM and CN in spontaneous speech.

	t	d	s	z	k	g	n	l	r
JM	74	0	68	4	45	1	43	15	39
CN	46	7	47	6	49	1	40	24	31
Total	120	7	115	10	94	2	83	39	70

2 Methodology

The data reported in this study are from two Greek speakers producing a monologue on a subject of their interest. The speakers were seated in a sound-proof room while the experimenter was listening to their output from an adjoining room. The recording began once the speaker started producing fluent, unselfconscious output. In total, 2.5 min for speaker JM and 1.2 min for CN were recorded. If pauses and filled pauses are excluded, approximately 3 min of speech were analysed.

Data acquisition was carried out with the Reading University-IBM multichannel data acquisition system (Hardcastle et al. 1989), which was used to record simultaneously and display acoustic, electropalatographic and electrolaryngographic data. Acoustic data were sampled at 20.000 Hz, laryngographic data at 10.000 Hz and EPG data at 200 Hz.

A detailed articulatory analysis of the consonants /t, d, s, z, k, g, l, n, r/ was carried out. (For consistency and uniformity in presentation, phonemic bracketing is used for /d, g/ although their phonological status has been debated for Greek.) EPG data were annotated at the frames of onset and end of constriction, the temporal midpoint and the frame of maximum contact. In this study, the results of the qualitative and quantitative analysis are reported for the temporal midpoint frame. In total, 540 tokens were analysed. Table 1 provides the number of tokens analysed per subject for each consonant. The occurrence of each sound in the data varied considerably with the two most frequent consonants being /t/ and /s/.

Duration measurements for all consonants were made from the EPG and acoustic data. Finally, the relationship between degree of EPG contact (at the temporal midpoint) and consonant duration was tested with a series of regression analyses.

3 Results

3.1 EPG contact patterns

The qualitative analysis of the EPG data revealed variability in tongue palate contact patterns suggesting variation in the execution of the lingual gesture for all consonants studied. Variability ranged over a continuum frequently resulting in reduced gestures. Although consonant deletions have been reported in similar studies for other languages (e.g. Shockey 1991 for English), evidence of absence of a consonantal gesture was infrequent in Greek (see table 2); these cases were particularly interesting and will be discussed separately below.

3.1.1 The plosives /t, d/

Realisations of the plosive /t/ varied in terms of the degree of constriction at the alveolar region and the overall degree of contact both at the alveolar region and the

Table 2 Number of consonants produced with incomplete constriction, with no evidence of a lingual gesture ('deleted'), and voiceless consonants produced with full or partial voicing. F indicates the tokens produced with friction; LR stands for lateral release (see text for details).

	Incomplete constriction		Deleted		Voiced	
	<i>JM</i>	<i>CN</i>	<i>JM</i>	<i>CN</i>	<i>JM</i>	<i>CN</i>
t	20, 9 F	13, 3 F	3	2	32	19
d		1				
s				4	4	10
k	14, 10 F, 3 LR	5, 2 F, 25 LR			8	17
n	10	5	2	2		
l		1				
r	34	26				

lateral margins of the palate (figure 1, /t/). There were tokens produced with (a) complete closure (palatograms A–C); (b) incomplete closure with presence of a constricted grooved configuration (palatograms F, G); and (c) incomplete closure with evidence of a very open articulation at the alveolar region (palatograms D, E). Variability in the overall degree of contact is also evident for the tokens displaying complete constriction (palatograms A–C). Palatogram (A) shows tongue palate contact during the production of a word initial [t] produced after a pause. This token showed maximum contact both at the alveolar region and the lateral margins of the palate indicating a considerably raised tongue body during its production. Progressively less anterior and lateral contact may be seen in palatograms (B–C) with the latter token involving complete constriction across the first row of electrodes only, suggesting an apical articulation. In this same token side contact is present only at the outmost lateral electrodes, suggesting a less raised tongue body during its production.

The production of the plosive with a central narrow groove (F, G) was evident in the environment of preceding or following [s], suggesting influence from the grooved configuration required for the production of the fricative. Thus, anticipation and perseveration of the grooved pattern was found in [ts] and [st] sequences respectively (cf. Shockey 1991, Shockey & Farnetani 1992). This grooved configuration for [t] frequently differed from that of preceding/following [s] in that it was generally more anteriorly located and narrower. An example of such differences may be seen in the comparison of palatograms (F) for [t] and [s] in figure 1. More anterior contact and a narrower groove is shown for [t] than [s] in [st] clusters. For some tokens, however, such articulatory differences were not as pronounced (e.g. compare palatograms (G) for [t] and [s] in [ts] sequences). Furthermore, the production of the grooved [t] in [st] and [ts] sequences was characterised by the presence of noise on the acoustic signal; there was, however, a clearly observed difference in the amplitude of noise between preceding/following [s] and reduced [t]. A decrease/increase in amplitude coincided with the onset/end of maximum constriction for [t] in the [st] and [ts] sequences, respectively, providing a boundary between the two sounds. Such articulatory and acoustic differences led to the perception of the alveolar stop segment in these sequences.

Constricted and open EPG patterns occurred in approximately 27–28% of productions for both speakers. In addition, [t] was often partially or fully voiced in intervocalic position (46 tokens) or when flanked by a vowel and a voiced consonant (5 tokens); in

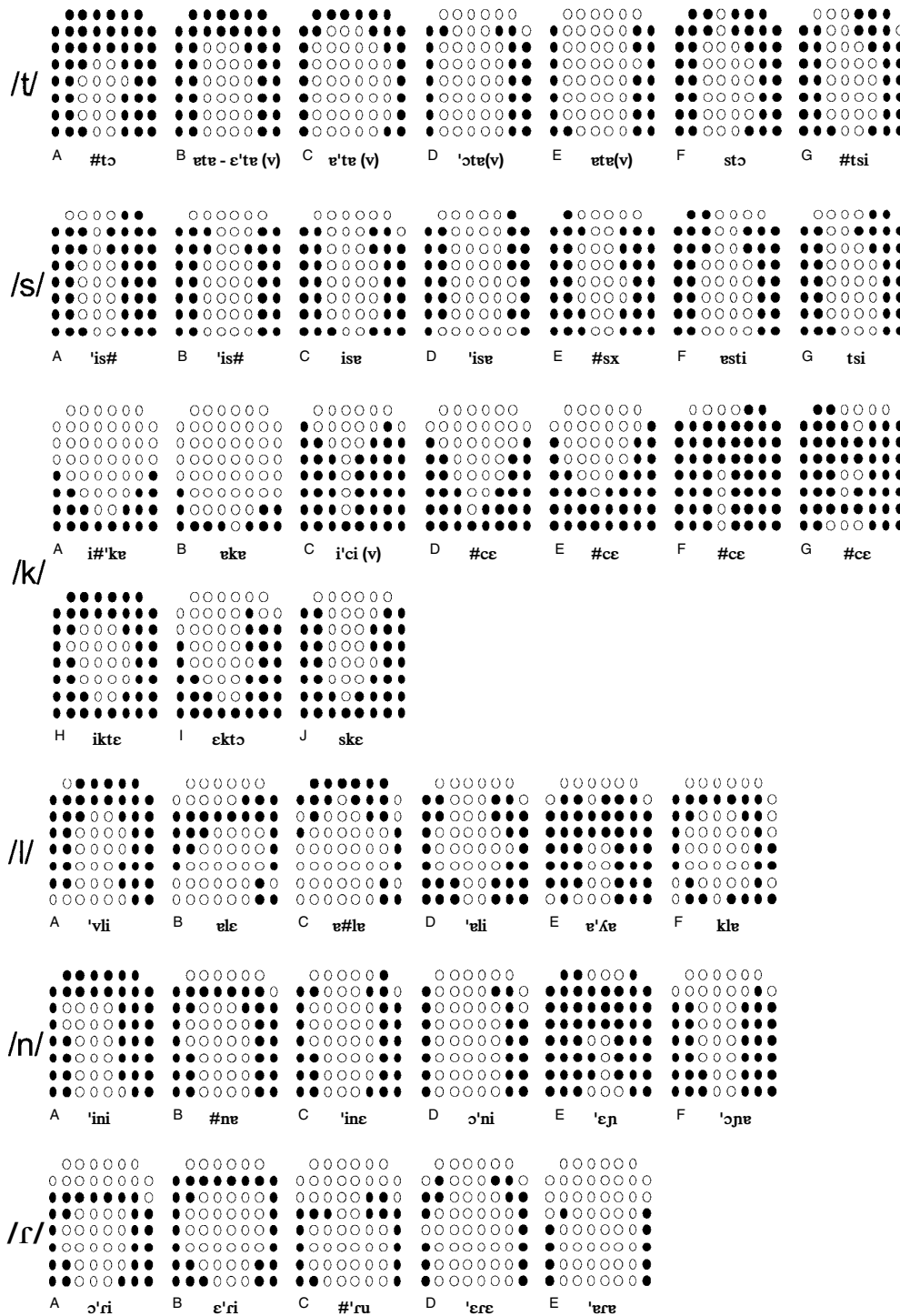


Figure 1 EPG palatograms at the temporal midpoint of the consonants /t, s, k, l, n, r/ produced in several contexts in spontaneous speech. See text for details. ((v) indicates voicing during the consonant).

on the EPG data where there was also evidence of friction on the acoustic data were found and are assumed to involve lateral release of airflow at a location posterior to the last row of EPG contacts. Gibbon, Hardcastle & Nicolaidis (1993) provide a similar interpretation in their analysis of EPG and oral airflow data during [k] in [kl] clusters; they report cases of oral airflow release well before EPG contact release for [k]. In addition, a considerable number of tokens were produced with full or partial voicing when flanked by vowels or a vowel and a voiced consonant, especially for speaker CN (table 2). Contextual variants in the environment of the front vowels /i, ε/ may be seen in palatograms (C–G). Considerable contact along the margins of the palate extending up to the 2nd row of electrodes was frequently present, (e.g. C, E), together with considerable central contact over the palatal area. Two different EPG configurations were evident in the data for both speakers; one involving complete closure across the posterior rows of the palate (retracted palatal/pre-velar region), and the other across the 2nd to 4th rows (retracted alveolar/post-alveolar area), similar to the production of a palatal lateral or nasal (see below). Palatograms (D–F) show such variability in three productions of the conjunction [cε] ‘and’. Palatogram (G) shows evidence of simultaneous constriction at both anterior and posterior regions of the palate, which mainly occurred for speaker JM. The majority of tokens involved posterior constriction while advanced (or simultaneous) constriction occurred in 41% and 20% of the tokens produced by JM and CN, respectively. Similar differences in place of constriction, and consequently in the articulators involved in the making of the constriction, have been noted in the literature with reference to palatals vs. fronted velars or palatalised velars (Recasens 1990, Keating & Lahiri 1993). It is interesting to note that such variability is evident in our data during different productions of the stop in the same context, and in fact, for the same words. Similar variability was evident in the environment of /i/. The data thus suggest different types of gestural organisation for the same underlying units, i.e. variation in articulatory realisation manifested in the involvement of different articulators (tongue regions) and the patterns of overlap. The interesting question that arises is why such variation occurs; depending on the duration of the consonant and the characteristics of the following front vowel, the presence of the fronted pattern may be accounted for as (a) a reduced variant (i.e. involving a change in target) or (b) coarticulation dependent undershoot. Further examination of the data, involving perceptual analysis of the different manifestations, detailed segmental analysis of the articulatory/acoustic characteristics of the consonant and the vowels /i, ε/, as well as an analysis of prosodic structure, is necessary before any conclusive interpretation is made on the nature and factors that influence such variation.

Interesting coarticulatory effects were also evident during the production of [k] in consonantal clusters. Anticipation of the alveolar gesture during the production of the velar plosive in [kt] clusters may be seen in (H, I), with a complete double articulation pattern in (H). A similar double articulation pattern was evident for one token of [g] in a [gl] cluster showing anticipation of the [l] gesture during velar closure. Perseveration of the alveolar gesture in the [sk] cluster may also be seen in (J).

3.1.4 The lateral /l/

Variability in the production of /l/ was manifested in terms of (a) the degree of contact at the region of the constriction (figure 1, /l/, A–D), (b) the degree of contact at the lateral margins (greater lateral opening may be seen in (C) compared to (A, B)), (c) the location of constriction in the front region of the palate (e.g. fronted apical pattern in (C) vs. more retracted constriction in (B)). One token with incomplete alveolar constriction was also found in the data (D). Palatogram E displays one token of the palatal [λ]. This is a typical configuration for the palatal involving more retracted constriction compared to [l] and more contact both at the region of constriction and the lateral margins of the palate, suggesting considerable lingual raising during its

production (cf. Recasens 1990, 1999). Gestural overlap in a [kl] cluster is also shown in (F).

3.1.5 The nasal /n/

The degree of constriction varied from complete closure (figure 1 /n/, A–B) to incomplete constriction (C–D) (cf. Shockey & Farnetani 1992, Shockey 1991). Although the location of constriction was mainly across the 2nd row of electrodes suggesting a clearly alveolar articulation, coarticulatory effects influenced lingual placement with more fronted contact evident in the environment of /i/ (A). The frequency of incomplete constriction varied between 23% for JM and 13% for CN (table 2). These tokens were considerably shorter in duration, especially for speaker JM. Their mean duration was 33.40 ms and 48.20 ms for JM and CN, respectively, compared to 75.36 ms and 64.77 ms for the nasals with complete closure, i.e. 56% and 26% shorter. For several of these tokens nasalisation of neighbouring vowels was noted in the impressionistic analysis; the sharing of nasality with the neighbouring environment may function to enhance the perception of the reduced nasal segment. Palatograms (E–F) show tongue placement during the production of two tokens of the palatal nasal [ɲ] produced with complete and incomplete constriction, respectively, and displaying considerable tongue body involvement during their production.

3.1.6 The tap /ɾ/

For this segment, variability was evident in the degree of constriction and the place of articulation which varied between alveolar and retracted alveolar (figure 1, /ɾ/) Only five tokens for each speaker were produced with complete constriction across the alveolar ridge (e.g. A, B) suggesting that this segment generally involves approximation of the lingual gesture to the palate.

The above data have provided evidence of large variability at the production level. This variability is continuous and ranges from fully articulated tokens to various degrees of weaker articulations and undershoot of front closure. In addition to reduction effects, contextual effects manifested in the spatio-temporal overlap of neighbouring gestures were clearly evident in the data. In what follows, an attempt will be made to explore the relationship between contact and duration, and account for some of the variability present on the basis of duration dependent effects.

3.2 Spatio-temporal variability

Figure 3 presents the averages and ranges of consonantal durations for the two speakers. No data for /g/ are presented because there was only one token per subject; data for /d/ were available only for CN. Interesting speaker differences in the range of durational variability are shown in the graph. Although individual differences in overall speaking rate are to be expected and may account for some of the variation, it is interesting to note that several tokens with very long consonantal durations were produced by speaker JM; his data involved a significant number of pauses or filled pauses which resulted in considerable consonantal lengthening prior to or following the pause. This was especially evident for the consonants /s, k, n/.

The presence of pauses, filled pauses, other dysfluencies including slips of the tongue, corrections, repetitions of syllables, words and so on, resulted in variation in the rate of production of different stretches of speech produced by the two speakers. Slower versus faster, more fluent utterances were produced which are expected to influence consonantal durations differentially. Several additional factors may account for durational variation including position effects in words, phrases, location of word stress, prominence, word function and so on. Although we have not examined such

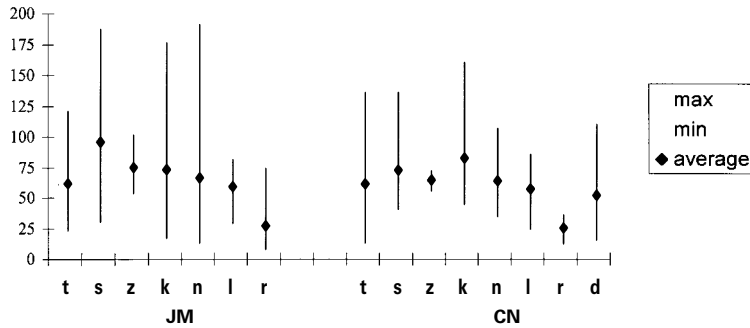


Figure 3 Range and average durations for several lingual consonants produced by speakers JM and CN.

factors systematically, it is worth noting an interesting feature of function words which were frequently lengthened as speakers appeared to be planning the upcoming utterance; in such cases their consonantal segments were produced with considerably long durations. In addition, consonantal segments in word initial and final position (for the latter, /s/ and /n/ only appear in the coda position of word final syllables) were on several occasions observed to be of longer duration than medial ones; this was frequently due to their occurrence after/before (filled) pauses. However, the data provided evidence of large variability, which necessitates a future in-depth statistical investigation in order to examine systematic relationships between duration and several of the parameters mentioned above.

In this study, we focus on quantifying the relationship between consonant duration and degree of contact using multiple regression. The analysis was carried out for the consonants /t/ and /s/ only because there were a good number of tokens for them (over 100). The degree of contact (i.e. absolute number of contacted electrodes) of the first four rows of electrodes (dependent variable) was regressed on the duration of the consonant (DUR), the duration squared (DUR²), and a set of dummy variables describing subjects and context (i.e. different vowel environments and consonantal environment).¹ The contacts of the first four rows were used in order to minimise the effect of coarticulatory influence from neighbouring sounds. Strong articulatory constraints on the tongue tip/blade articulator, which is actively involved in the production of /t/ and /s/, may be expected to inhibit large coarticulatory encroachments

¹ The estimated regression model was:

$$DC = a + b_1DUR + b_2DUR^2 + b_3DUM_1 + b_4DUM_2 + b_5DUM_3 + b_6DUM_4 + \text{Error},$$

where: DC = degree of contact; a, b₁, b₂, b₃, b₄, b₅, b₆ = regression coefficients; DUR = duration of consonant; DUR² = duration of consonant squared; DUM₁ = dummy variable for subject (takes value 1 for speaker JM, 0 for CN); DUM₂ = dummy variable for context (takes 1 for preceding/following consonant, 0 for intervocalic position); DUM₃ = dummy variable for /i/ vowel context (takes value 1 for preceding/following high front vowel /i/, 0 for other vowels); DUM₄ = dummy variable for /e/ vowel context (takes value 1 for preceding/following mid front vowel /e/, 0 for other vowels).

Both the general and resulting nested models were estimated. DUR² was included in the model to test whether the relationship between consonant duration and degree of contact is linear or quadratic. The influence of the vowel context was tested separately for the front vowels /i/ and /e/ because they involve considerable tongue-palate contact (especially /i/) compared to the rest of the Greek vowels /e, ɔ, u/, which generally display little contact in a relatively retracted position on the palate. This design enabled us to know whether consonantal contact was different in the environment of preceding/following /i/ and /e/ as opposed to /e, ɔ, u/. Further specificity in the model, i.e. more dummy variables specifying all five vowels separately, relevant position (preceding or following the consonant) would increase the number of explanatory variables and make estimation problematic, given the number of observations available.

Table 3 Regression results for the contact-duration relationship for /t/.
 (* Dummy variable used to discriminate between cases of consonant and vowel context.)

Dependent Variable: Degree of Contact

<i>Independent Variables:</i>	Intercept	DUR	DUR²	DUM₂[*]
a) The whole sample				
<i>Regression Coefficients</i>	8.7	0.23	-0.0008	2.01
<i>Coefficient t-ratio</i>	4.48	3.76	-1.76	3.60
R ² =0.41 No obs.: 117				
b) Vowel context only				
<i>Regression Coefficients</i>	5.98	0.31	-0.0012	-
<i>Coefficient t-ratio</i>	2.64	4.41	-2.54	-
R ² =0.51 No obs.: 74				
c) Consonant context only				
<i>Regression Coefficients</i>	16.3	0.08	-	-
<i>Coefficient t-ratio</i>	9.4	2.33	-	-
R ² =0.12 No obs.: 43				

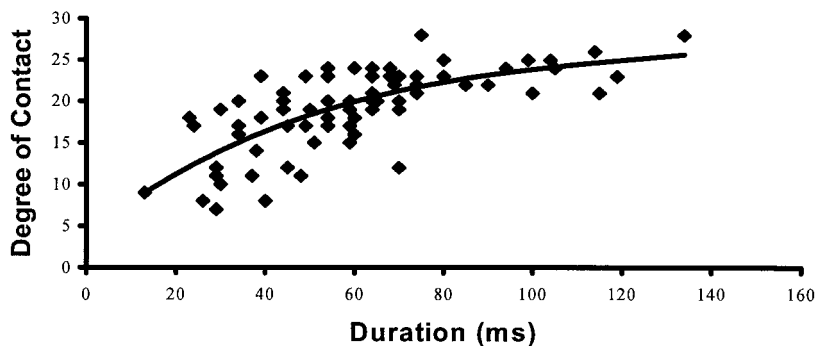


Figure 4 Scatterplot of degree of contact against duration for /t/.

from neighbouring sounds (Farnetani 1990, Recasens 1999). Table 3 and figure 4 present the regression results for the consonant /t/. The figure plots the degree of tongue-palate contact against duration for /t/ and the fitted regression line.

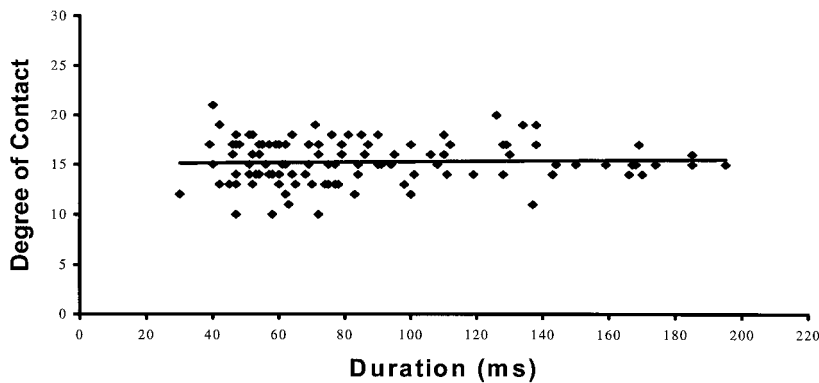
The relationship between the degree of contact and duration for /t/ was found positive and significant at the 5% level (t-ratio = 3.76). Thus a decrease/increase in duration was found to relate to a decrease/increase in the amount of contact at the alveolar region. The analysis revealed a non-linear relationship between the variables (DUR² was significant at the 5% level). This suggests that as duration increases, the rate of contact increase declines (see figure 4). This is to be expected as tongue-palate contact is expected to reach a maximum (maximum configuration for /t/) and not increase any further irrespectively of an increase in duration.

The various dummy variables used to describe the effect of different subjects and

Table 4 Regression results for the contact-duration relationship for /s/.

Dependent Variable: Degree of Contact

<i>Independent Variables:</i>	Intercept	DUR
<i>Regression Coefficients</i>	15.6	0.0043
<i>Coefficient t-ratio</i>	23.3	0.56
$R^2=0.41$ No obs.: 115		

**Figure 5** Scatterplot of degree of contact against duration for /s/.

contexts failed to add significant explanatory power to the model. Only the dummy variable which discriminates between vocalic and consonantal context (DUM_2) showed significance (t-ratio = 3.60). Table 3 shows that the positive relationship between contact and duration is stronger in the vowel environment ($R^2=0.51$) than in the consonant environment ($R^2=0.12$). This may be explained in terms of stronger articulatory constraints on the tongue tip/blade articulator imposed by the coordination requirements with neighbouring consonantal gestures, which appear to influence the nature of duration-dependent effects.

No significant relationship between contact and duration was found for /s/ (t-ratio = 0.56, table 4; figure 5). This is true independently of context, i.e. whether /s/ appears in the environment of vowels or a consonant. Figure 5 plots the degree of tongue-palate contact against duration for /s/ and the fitted regression line. It shows that there is less variation in contact compared to /t/ and this variation does not depend on consonantal duration.

The data indicate that the tongue tip/blade gesture for the /s/ is relatively invariant compared to that of /t/ and that duration dependent variation of the tongue tip/blade gesture is consonant specific. One possible line of interpretation of these results may be with respect to the production characteristics of the two consonants. Stronger articulatory and aerodynamic constraints present for the production of the /s/ limit the degree of variability in tongue placement. A minimal temporal interval may be required for the articulatory configuration needed for the production of friction. Weaker constraints on the same articulator during the production of /t/ may allow for greater duration dependent variability in the lingual gesture. The significant difference due to

contextual environment that was found in the analysis for /t/, i.e. a stronger relationship between duration and degree of contact in the vocalic rather than the consonantal environment (which was frequently /s/), adds further support to a constraint-based interpretation. A similar line of argumentation is presented by Farnetani (1995), who reports that duration dependent undershoot reduces from /n/ to /l/ to /d/ to /t/ to /s/, suggesting that consonant specific constraints determine such behaviour.

3.3 Gradient vs. categorical processes

The articulatory characteristics of consonant weakenings resulting from differential degrees of reduction of the lingual gesture discussed above, such as less fully articulated consonants, approximant or fricative production of stops, together with evidence for partial or full voicing of voiceless obstruents (especially stops), reveal continuous processes suggesting gradient variability over the laryngeal and supralaryngeal gestures during the production of these consonants.

Gradient reduction effects during vowel production were also found. Previous work on Greek vowels has reported high vowel reduction involving devoicing or elision, especially when the high vowels are in the environment of voiceless consonants and in the post-stressed syllable (Dauer 1982). In our data, considerable durational variation was exhibited including vowels of very short durations and tokens where there was no evidence of a vocalic interval on the acoustic data. Some examples of short or ‘deleted’ high vowels include: [ˈlɛrɪsɐ] (i = 17 ms) ‘name of town’, /ˈkrɛtɪsɛ/ ‘lasted’ [ˈkrɛtɪsɛ], /ˈkɛnumɛ/ ‘we make’ [ˈkɛnmɛ], /tɪs/ ‘of’ [ts].

Although we lack the space to report on these cases of ‘deletions’ in detail and examine the acoustic characteristics of the flanking consonants, an interesting example is presented in figure 6, which displays EPG data from the utterance /liˈɣɛki kurɛstiˈkɔ/ (‘a little tiring’) produced as [liˈɣɛckurɛstiˈkɔ], i.e. with no evidence of a vowel between the palatal and velar stop. The EPG data show the production of a palatal sound (onset of palatal closure at frame 930) followed by a gradual retraction of the tongue and production of the velar (velar release at frame 964). This is an interesting case,

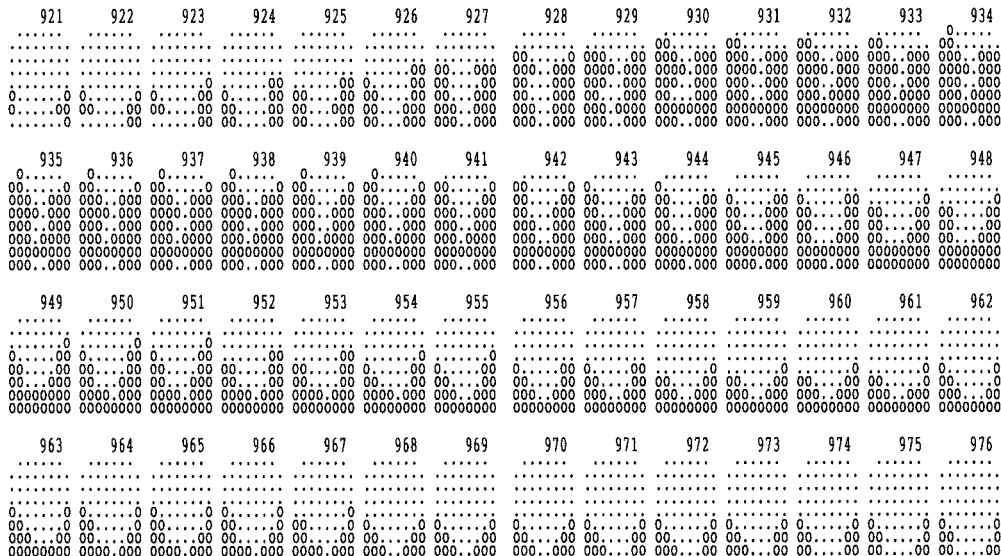


Figure 6 EPG printout of the [ck] sequence in the utterance [liˈɣɛckurɛstiˈkɔ] (see text for details).

where what has been commonly described as a deletion process following a feature spreading operation could be explained in terms of a temporal realignment between the consonant and vowel gestures resulting in completely synchronous gestures. Lack of a discrete vowel interval on the articulatory and acoustic data is compensated for by cues from the preceding palatal.

Tokens of ‘deleted’ consonants were rare in our data (only five tokens for /t/, four for /s/ and /n/) (table 2). Figure 7 (a, b) presents the spectrogram and EPG data of the utterance /'ɛn kɛ tɔ tɐ'ksiði/ ‘even though the trip’ produced as [ʼɛŋçɛʉtɐ'xsiði], i.e. with no evidence of a consonantal gesture for the /t/ of /tɔ/. Similar examples were /ɛkɐ'tɔ tis ɛkɐ'tɔ/ ‘a hundred percent’ produced as [ɛkɐ'ɔtsɛkɐ'tɔ] and /ɐ'pɔ 'pɛdɛ/ ‘from five’ produced as [ɐɔ'pɛdɛ]. These cases were not common but occurred with both speakers when two identical consonantal gestures occurred across a vowel. For the first two examples (i.e. /'ɛn kɛ tɔ tɐ'ksiði/ and /ɛkɐ'tɔ tis ɛkɐ'tɔ/), the EPG data did not reveal any trace of an alveolar gesture for the first /t/ in each utterance, suggesting a deletion of the consonantal gesture. This is illustrated in figure 7 (a, b) for /'ɛn kɛ tɔ tɐ'ksiði/ → [ʼɛŋçɛʉtɐ'xsiði]; the EPG data show that following the release of the palatal stop of [çɛ] (frame 707), there is a gradual decrease in the amount of contact during the vocalic interval till frame 729, where the approach to the alveolar stop [t] of [tɐ'ksiði] starts. The EPG as well as the acoustic data provide no evidence of an alveolar gesture for /tɔ/. Thus a more economical, new gestural organisation appears to occur with one closing gesture.

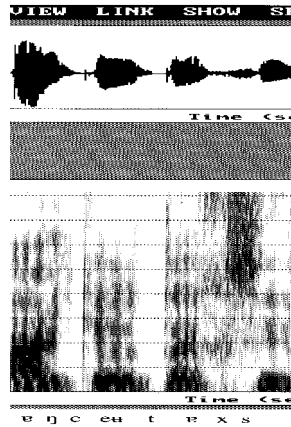
Evidence of processes such as deletion affecting not single segments but whole syllabic components were found in examples /ɛk'θɛsis zɔɣrɛfi'kis/ ‘painting exhibitions’ produced as [ɛk'θɛzɔɣrɛfi'cis] ($z = 60\text{ms}$) and /ɛk'θɛsis vi'vliɔn/ ‘book exhibitions’ produced as [ɛk'θɛzvi'liɔn] ($z = 94\text{ms}$). The EPG data in figure 8 for [ɛk'θɛzvi'liɔn] show the velar configuration for [k] followed by an open anterior pattern for [θ], the vowel [ɛ] and the grooved configuration for the (voiced) fricative before the production of the following word [vi'liɔn].

Such cases of extensive effects involving more than one structural unit may be difficult to explain without appealing to the notion of reorganisation. This new gestural organisation appears to be actively controlled and reflect perceptual constraints manifested in the different organisational patterns of the two examples [ɛk'θɛzɔɣrɛfi'cis] and [ɛk'θɛzvi'liɔn]; i.e. in both examples realisation of at least one alveolar gesture occurs. These examples may also be interpreted to involve a more economical organisation of neighbouring identical gestural units. An explanation based on an increase in gestural overlap due to temporal constraints would assume the active control and complete synchronicity in execution of up to four underlying units, all preplanned and executed well in advance, i.e. during the first unit, and involving both the supralaryngeal and laryngeal gestures.

Gradient and categorical changes are also suggested by data on place and voice assimilation. EPG data on place assimilation of /n/ to the following velar in the /n#k/, /n#ɣ/ context provide evidence of complete place assimilation with no trace of an alveolar gesture. In certain /n#k/ context tokens, a new gestural organisation occurred with no evidence of a lingual gesture for the nasal or nasalisation of the preceding vowel, and realisation of a voiced velar, e.g. /tin ksɛnθi/ ‘the ksɛnθi’ (name of town) produced as [ti'gzɛθi], /ðɛn 'kɛnɛmɛ/ ‘we did not do’ [ðɛ'gɛnɛmɛ] (figure 9).

The EPG palatograms show contact at the posterior rows with no evidence of movement towards a more anterior region. Lack of an oral gesture for the nasal appears to contradict claims that gestures cannot be deleted. More advanced lateral contact was expected and found in the environment of the palatal allophones [j] and [ç] as in /'ɔtɛn 'ɣirisɛ/ ‘when I returned’ [ʼɔtɛŋ'jirisɛ], /'ɛn kɛ/ ‘even though’ [ʼɛŋçɛ] (figure 7(b)). Although it may be possible that a reduced coronal gesture may not be registered with EPG, previous work on English /n#k/ assimilations combining EPG and EMA has shown that when there was no evidence of a residual gesture on contact

(a)



(b)

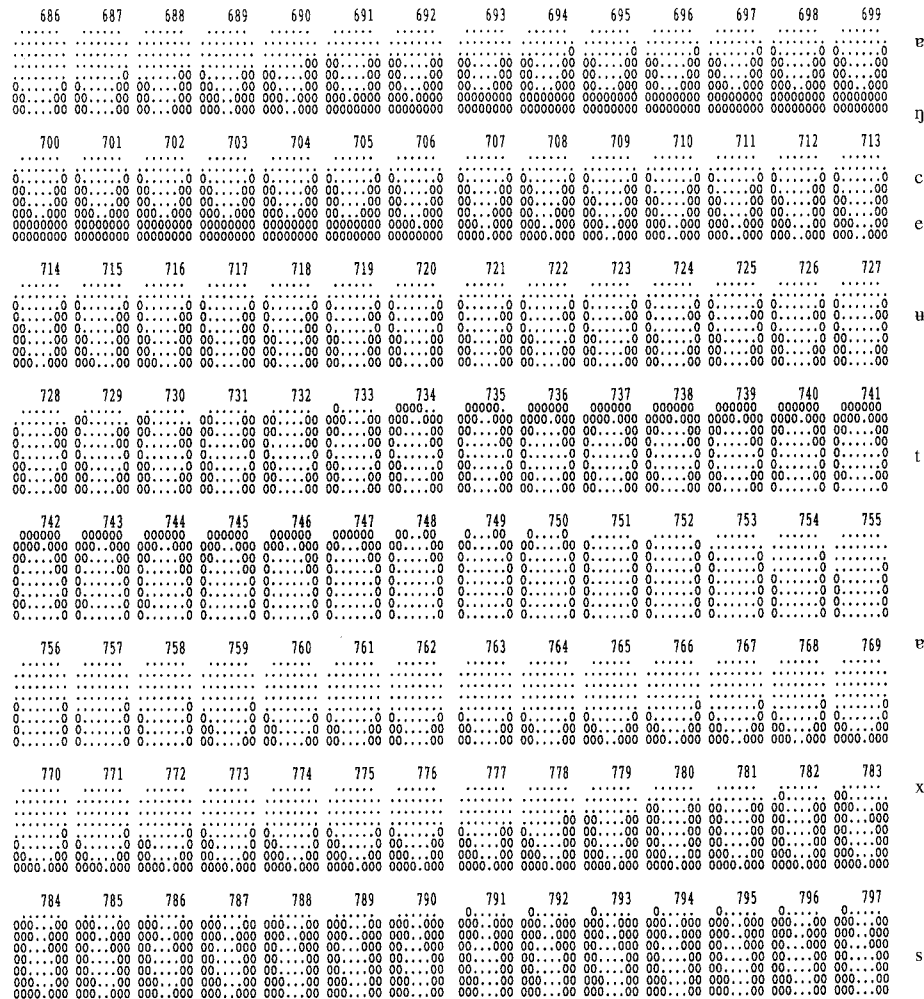


Figure 7 (a, b) Spectrogram and EPG contact patterns for part of the utterance /'en ke to tɛ'ksiði/ produced as [ɛ̃ɲcɛtɛ'ksiði] (underlined part displayed only).

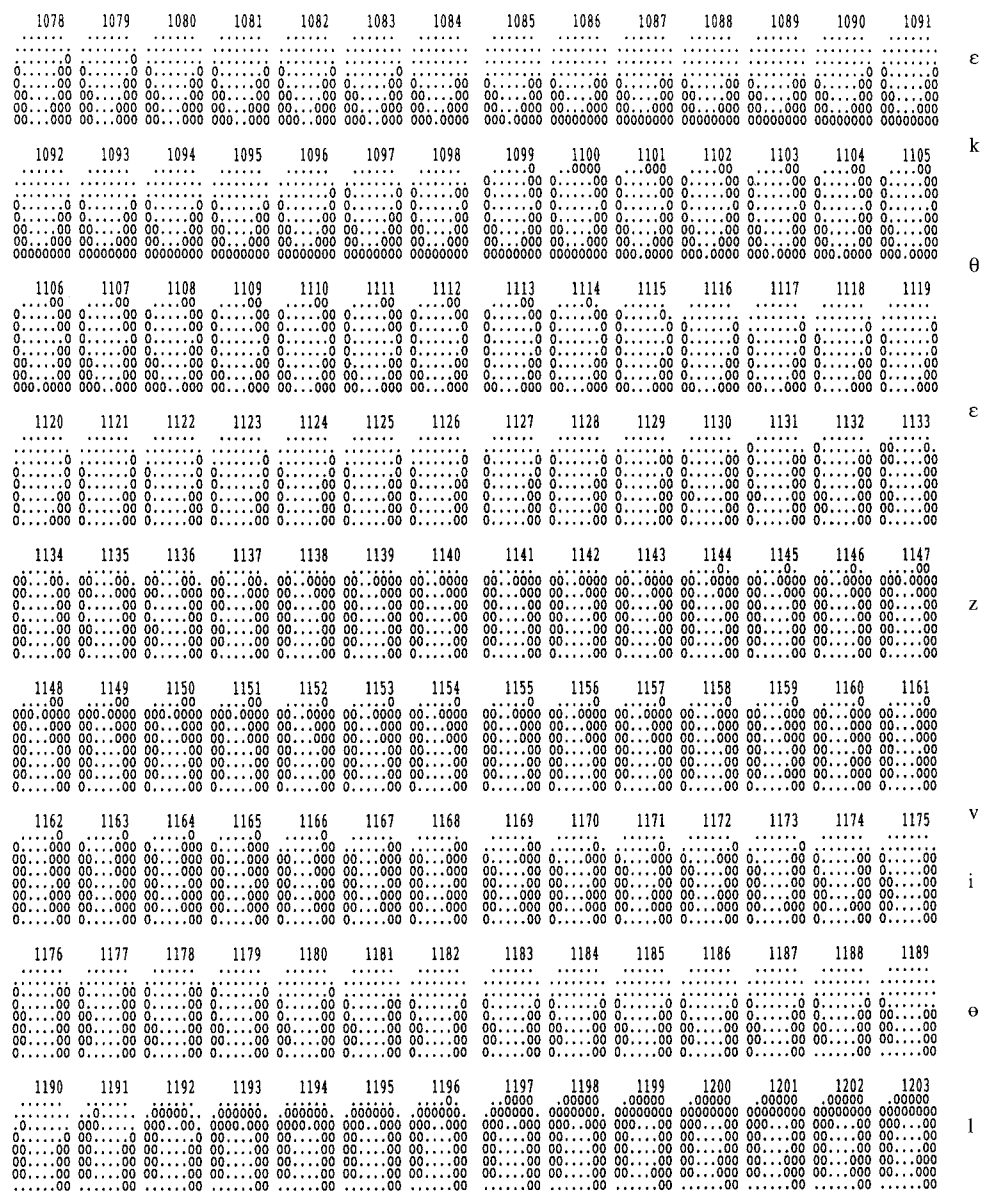


Figure 8 EPG printout of part of the utterance /ɛk'θɛsis vi'vlion/ produced as [ɛk'θɛzvi'liɔn] (underlined part displayed only).

data, a gesture was not registered on EMA either (Ellis & Hardcastle 2000). The data may be interpreted, therefore, to suggest elimination of gestural units indicating categorical changes involved.

Data on voice assimilation provided evidence of gradient effects ranging from no assimilatory influence, to partial or complete voicing. Two possible sites for voice assimilation are the voicing of plosives following nasals, and obstruents assimilating to the voice of the following consonant. Some examples include: /tis γrɛ'mis/ 'of the line'

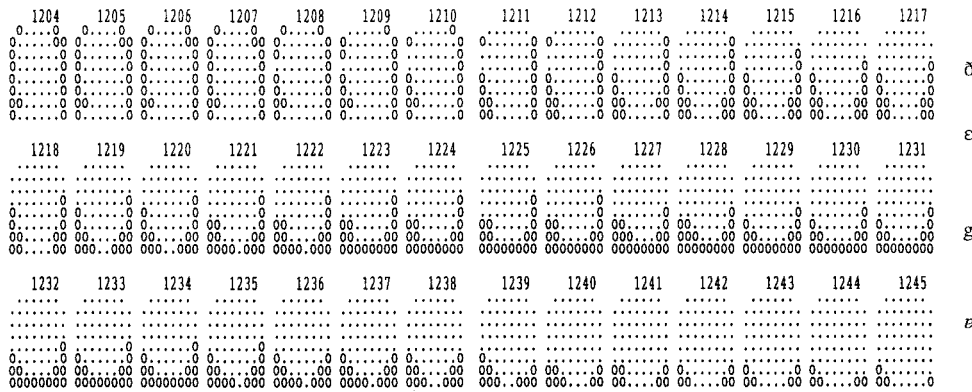


Figure 9 EPG printout of part of the utterance /ðen 'kənɛmɛ/ produced as [ðɛ'gɛnɛmɛ] (underlined part displayed only).

[tizɣrɛ'mis], /tis mɔ'nɛðɛs/ 'of the unit' [tizmɔ'nɛðɛs] (with partial voicing), /'ɔpɔs 'inɛ/ 'as is' ['ɔpɔznɛ], /'ɔpɔs 'inɛ/ ['ɔpɔsnɛ], /'ɛn kɛ/ 'even though' ['ɛnɛ].

Although the data are limited for any strong claims to be made and more data from a number of speakers are needed, they point towards the presence of both phonetic gradient processes and phonological categorical changes in spontaneous speech and as such challenge the claim that connected speech processes such as assimilations and deletions are gradient in nature and do not involve categorical changes of underlying units.

4 Discussion

The ubiquity of variability in speech processes and the elusiveness of the notion of invariance in speech production is best represented in the analysis of spontaneous speech data. At the articulatory level, segmental variability ranges over a continuum from over-articulated to under-articulated forms and as such the feasibility of defining the 'canonical' form is questionable. If we take the plosive /t/ as an example, its production ranged from a very hyperarticulated form, as when after a pause (with increased alveolar contact involving the tongue blade and significant contact over the lateral margins suggesting a raised tongue body), to various degrees of less strongly articulated forms (suggesting apical articulation and a less raised tongue body), to very hypoarticulated forms with evidence of a weak, reduced alveolar gesture and consequent lack of alveolar constriction. Such reduced gestures sometimes resulted in more open articulations, i.e. fricative or approximant production of stops.

Given the variability present at the articulatory as well as acoustic levels (e.g. Crystal & House 1982), two important issues concern the function of variability and its possible constraints. A well-known determinant of variability is language specific phonology. Articulatory data on English spontaneous speech provide evidence of a greater number of consonant deletions or processes such as glottaling of /t/ (Shockey 1991), which are not present in Greek. Vowel deletions or reduced vowels were, however, relatively frequent in our data.

Language differences in the nature of the assimilatory process, i.e. whether continuous or categorical, have also been reported. Various studies on English

alveolar-velar assimilation provide evidence of a gradual process ranging from absence of assimilation, to partial assimilation manifested in residual alveolar gestures, to complete assimilations (Wright & Kerswill 1989, Nolan 1992, Hardcastle 1994, Barry 1991). Further studies have, however, provided evidence of both categorical and gradient processes present within the same language on the basis of assimilation data from English, Italian and Russian (Barry 1991, Farnetani & Busà 1994, Ellis & Hardcastle 2000). Several factors may affect the nature of the process including syllabic structure, type of cluster and speaker specific strategies. Such data, together with evidence from our study suggesting the presence of some categorical processes in Greek, cannot be adequately explained by gestural phonology. Data on total place assimilations and deletions of consonantal gestures and syllables suggest the possibility of reorganisation of gestural units. In addition, our data show that gradient changes involving different degrees of spatio-temporal overlap between gestures at the supralaryngeal and laryngeal levels are present in running speech.

A further determinant of variability that has been well documented is duration (Lindblom 1963, 1983, 1990). While a positive relation between duration and degree of contact has been found for /t/ in our data, lack of such a relationship for /s/ suggests that duration induced articulatory variability is constrained by the production requirements of a segment. Strong articulatory and aerodynamic constraints appear to delimit the range of segmental duration-dependent variability.

Given the overall variability evident in spontaneous speech data, a fundamental issue concerns its function in communication. Lexical access and comprehension of the message rely on a combination of local MICRO and global MACRO effects. The former may be in the form of important segmental cues shared by the neighbouring environment and functioning to enhance the perception of the segment even when this is heavily reduced or apparently deleted (cf. Local 1992). The latter relate to redundancy and predictability of a lexical item due to the overall linguistic structure, i.e. morphological, syntactic or semantic factors which contribute to the reconstruction of the message. Pragmatic knowledge and important background information shared between the speaker and the listener also influence variability and message comprehension. Perception is thus facilitated by redundancy in the speech signal and knowledge of the language and communicative situation. The Theory of Adaptive Variability is attractive because the communicative situation plays a significant role in shaping the organisation of phonetic gestures on the basis of a continuous interplay between production and perception constraints.

The evidence we have in our data for both gradient and categorical processes may be interpreted as suggesting flexibility in gestural organisation and execution, the possibility of new gestural organisations which serve economy in view of the overall redundancy inherent in the speech signal by the speech production system, linguistic structure and the pragmatic, communicative and social context.

Acknowledgements

I would like to thank Edda Farnetani, Klaus Kohler, Adrian Simpson and an anonymous reviewer for their very helpful comments.

References

- BARRY, M. (1991). Temporal modelling of gestures in articulatory assimilation. *Proceedings of the XIIth International Congress of Phonetic Sciences*, Aix-en-Provence, vol. 4, 14–17.
- BROWMAN, C. P. & GOLDSTEIN, L. (1989). Articulatory gestures as phonological units. *Phonology* 6, 201–251.
- BROWMAN, C. P. & GOLDSTEIN, L. (1990). Tiers in Articulatory Phonology with some implications

- for casual speech. In Kingston, J. & Beckman, M. E. (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, 341–376. Cambridge: Cambridge University Press.
- BROWMAN, C. P. & GOLDSTEIN, L. (1992). Articulatory Phonology: an overview. *Phonetica* **49**, 155–180.
- CRYSTAL, T. H. & HOUSE, A. S. (1982). Segmental durations in connected speech signals: preliminary results. *Journal of the Acoustical Society of America* **72**(3), 705–716.
- DAUER, R. M. (1982). The reduction of unstressed high vowels in modern Greek. *Journal of the International Phonetic Association* **10**, 17–27.
- DOCHERTY, G. J. & LADD, D. R. (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press.
- ELLIS, L. & HARDCASTLE, W. J. (2000). Assimilation strategies in the production of alveolar to velar sequences: EPG and EMA data. *Proceedings of the 5th Seminar on Speech Production: Models and Data*, Kloster Seeon, Bavaria, 117–121.
- FARNETANI, E. (1990). V-C-V lingual coarticulation and its spatiotemporal domain. In Hardcastle & Marchal (eds.), 93–130.
- FARNETANI, E. (1995). The spatial and the temporal dimensions of consonant reduction in conversational Italian. *EUROSPEECH 95, European Conference on Speech Communication and Technology*, Madrid, 2255–2258.
- FARNETANI, E. & BUSA, M. G. (1994). Italian clusters in continuous speech. *Proceedings of 1994 International Conference on Spoken Language Processing, Yokohama*, vol. 1, 359–362.
- GIBBON, F., HARDCASTLE, W. J. & NICOLAIDIS, K. (1993). Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: a cross-linguistic investigation. *Language and Speech* **36**(2–3), 261–277.
- GIBBON, F. & NICOLAIDIS, K. (1999). Palatography. In Hardcastle & Hewlett (eds.), 229–245.
- HARDCASTLE, W. J. (1985). Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication* **4**, 247–263.
- HARDCASTLE, W. J. (1994). EPG and acoustic study of some connected speech processes. *Proceedings of 1994 International Conference on Spoken Language Processing, Yokohama*, vol. 2, 515–518.
- HARDCASTLE, W. J. & HEWLETT, N. (eds.) (1999). *Coarticulation: Theory, Data and Techniques*. Cambridge: Cambridge University Press.
- HARDCASTLE, W. J., JONES, W., KNIGHT, C., TRUDGEON, A. & CALDER, G. (1989). New developments in electropalatography: a state-of-the-art report. *Clinical Linguistics and Phonetics* **3**, 1–38.
- HARDCASTLE, W. J. & MARCHAL, A. (eds.) (1990). *Speech Production and Speech Modelling*. Dordrecht: Kluwer.
- KEATING, P. & LAHIRI, A. (1993). Fronted velars, palatalised velars, and palatals. *Phonetica* **50**, 73–101.
- KOHLER, K. J. (1990). Segmental reduction in connected speech in German: phonological facts and phonetic explanation. In Hardcastle & Marchal (eds.), 69–92.
- KOHLER, K. J. (1991a). The phonetics/phonology issue in the study of articulatory reduction. *Phonetica* **48**, 180–192.
- KOHLER, K. J. (1991b). The organisation of speech production: clues from the study of reduction processes. *Proceedings of the XIIth International Congress of Phonetic Sciences, Aix-en-Provence*, vol. 1, 102–106.
- KOHLER, K. J. (1992). Gestural reorganisation in connected speech: a functional viewpoint on ‘Articulatory Phonology’. *Phonetica* **49**, 205–211.
- KRULL, D. (1989). Consonant-vowel coarticulation in spontaneous speech and in reference words. *Speech Transmission Laboratory Quarterly Progress and Status Report* **1**, 101–105.
- LINDBLOM, B. (1963). *On vowel reduction*. FL thesis, University of Uppsala. Royal Institute of Technology, Speech Transmission Laboratory Report No. 29.
- LINDBLOM, B. (1983). Economy of speech gestures. In MacNeilage, P. F. (ed.), *The Production of Speech*, 217–245. New York: Springer-Verlag.

- LINDBLOM, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle & Marchal (eds.), 403–439.
- LINDBLOM, B. (1991). The status of phonetic gestures. In Mattingly, I. G. & Studdert-Kennedy, M. (eds.), *Modularity and the Motor Theory of Speech Perception*, 7–24. Hillsdale, NJ: Lawrence Erlbaum.
- LOCAL, J. (1992). Modeling assimilation in nonsegmental, rule-free synthesis. In Docherty & Ladd (eds.), 190–223.
- MARCHAL, A. (1987). Des clics en français? *Phonetica* **44**, 30–37.
- MARCHAL, A. (1988). Coproduction: evidence from EPG data. *Speech Communication* **7**, 287–295.
- NOLAN, F. (1992). The descriptive role of segments: evidence from assimilation. In Docherty & Ladd (eds.), 261–280.
- RECASENS, D. (1990). The articulatory characteristics of palatal consonants. *Journal of Phonetics* **18**, 267–280.
- RECASENS, D. (1999). Lingual coarticulation. In Hardcastle & Hewlett (eds.), 80–104.
- SHOCKEY, L. (1991). Electropalatography of conversational speech. *Proceedings of the XIIIth International Congress of Phonetic Sciences, Aix-en-Provence*, vol. 3, 10–13.
- SHOCKEY, L. & FARNETANI, E. (1992). Spontaneous speech in English and Italian. *Proceedings of the International Conference on Spoken Language Processing, Banff*, 767–769.
- SILVERMAN, D. & JUN, J. (1994). Aerodynamic evidence for articulatory overlap in Korean. *Phonetica* **51**, 210–220.
- WRIGHT, S. & KERSWILL, P. (1989). Electropalatography in the study of connected speech processes. *Clinical Linguistics and Phonetics* **3**, 49–57.