Journal of Phonetics (1986) **14**, 197–221

Stop Epenthesis in English

Marios Fourakis

Central Institute for the Deaf, 818 South Euclid, St. Louis, Missouri 63110, U.S.A.

and

Robert Port

Indiana University, Bloomington, Indiana 47405, U.S.A.

Received 14*th May* 1985, *and in revised form* 11*th January* 1986

Some phonologists have claimed that the insertion of a stop between a sonorant and a fricative consonant in syllable-final sonorant–fricative clusters follows from universal constraints on the human speech perception and production mechanism. Others have claimed that the intrusive stops are products of language or dialect specific phonological rules that are stated in the grammar. In this experiment we examined the production of sonorant–fricative and sonorant–stop–fricative clusters by two groups of English speakers. One spoke a South African dialect and the other an American mid-western dialect. The words tested ended in clusters of [n] or [l] plus [s] or [ts] and their voiced counterparts. Spectrographic analysis revealed that the South African speakers maintained a clear contrast between sonorant–fricative and sonorant–stop–fricative clusters. The American speakers always inserted stops after the sonorant if the fricative was voiceless, but when the fricative was voiced, they more often omitted the stop in underlying clusters containing a stop (/ldz/ or /ndz/) but sometimes inserted a stop in clusters such as /nz/ or /lz/.

Measurements of the durations of the vowels, sonorants, stops and the final fricatives were made from the spectrograms. The inserted stop in the American productions was significantly shorter than the underlying one and its presence also affected the duration of the preceding nasal. This incomplete neutralization is similar to other cases reported in the literature. It is proposed here that cases of incomplete neutralization result from the application of language-specific rules that we call *phase rules* which govern articulatory timing. These rules sometimes appear to yield results very similar to those of segmental phonological rules.

1. Introduction

1.1. *The theoretical issue*

One of the most important issues in the theory of language is how cognitive grammar makes contact with such realms of the non-linguistic world as muscles and acoustics. In the case of speech production, that amounts to the question of how phonological

0095–4470/86/020127 + 25 $03.00/0

segments control muscles to generate speech gestures. This paper is a study of phenomena that bear on this interface between phonology, phonetics and physiology. We investigated the case of the apparent epenthesis of a stop between the sonorant and fricative segments in the American pronunciation of words like *dense*, *prince* and *else*. That is, the word *dense* seems to be pronounced like *dents* and *prince* like *prints*. At the same time, because this effect depends on the voicing of the final cluster (since *fines* and *finds* remain distinct in careful speech), we examined differences in the voicing of the final obstruent or obstruent cluster: thus, clusters like *tens* vs. *tense* and *bends* vs. *dents* were compared.

There are basically two approaches to analysis of this problem found in the literature. One approach is to treat the epenthesis as a change in segmental phonological structure: that is, a rule of $\emptyset \rightarrow$ [t] applies in these contexts (Zwicky, 1972; Dinnsen, 1980). The second approach suggests that there is a universal tendency for the articulatory system to spontaneously introduce a silent gap in such segmental sequences so as to create an acoustic effect like that of a stop (Ohala, 1974). Thus one interpretation argues that this is a standard phonological process, while the other says that it is a low-level articulatory effect that is external to the language altogether.

But there is a third possibility that we would like to pursue in this paper. This is the possibility that there is a quite different type of rule in evidence here that is neither clearly phonetic nor phonological as the terms are customarily interpreted. It seems that many phenomena near the boundary of phonetics and phonology need a special type of linguistic rule that has two specific propterties. First, the rules employ *linguistically defined contexts* for their domain of application. This implies that the rules have access to linguistic structure, and that they are language specific, not universal. Second, the rules deal directly in the *dimensions of speech production*, such as the timing of articulatory gestures.

There are several classes of evidence that support the existence of such rules. One important class is the many cases of *incomplete neutralization* (Dinnsen & Charles-Luce, 1984; Port & O'Dell, 1985; see Dinnsen, 1985 for a review). There are now many attested cases of phonological contrasts that are said to be neutralized in a certain context and which native speakers assert are neutralized, but which are not actually identical. Such cases appear to require control over details of the timing of production without a corresponding perceptual effect. A second class of evidence is the case of linguistic contrasts which a linguist may hear but which are denied by the speakers themselves (Labov, Yaeger & Steiner, 1972; Labov, 1981). Apparently speakers have control over many subtleties of timing of which they have little perceptual awareness.

We postulate the existence of rules having the property that they employ linguistically definable environments but have temporal effects on a variety of articulatory gestures. We will call them *phase rules* in order to suggest their most prominent property which is to modify the timing of phonetic events relative to each other. Phase rules are thus somewhat like temporal implementation rules (Klatt, 1976; Port, 1981), but they deal directly with underlying gestures rather than with the acoustic events that happen to be prominent in the acoustic record.

A problem with phonetic implementation rules such as those proposed in the literature is that the timing units controlled by them are arbitrarily determined by the kind of data the experimentalist happens to find convenient to measure. Thus, for example, Port (1981) proposed timing rules that would account for the vowel and consonant duration data that were measured in the experiments. But we have no assurance that the acoustic intervals that were measured (even though they were chosen because of their acoustic prominence) are the ones directly controlled by speakers. Obviously it is implausible that *all* of the intervals that might be measured could be *directly* controlled by speakers. The computational burden would be enormous and, besides, there is great variability within and between subjects in this sort of timing (Port, 1981). It is more plausible that the speech production system is arranged so that a fairly small number of parameters may be directly controlled by linguistic factors. This is to propose that there may be an *intermediate structure* between the language (that is, the phonological representation) and actual gestures. Phonological features may determine simple control parameters of this intermediate structure yet have very complex time-distributed effects on the outputs of this intermediate structure. But if this is so, then we must admit that we have little idea what this structure is like or what parameters are directly controlled by speakers.

Thus, if there is this possibility of another level between abstract rule systems, it may be premature to attempt to make computational models of speech timing. Before we start formalizing the arithmetic, we should carefully define the formal properties of the *kind of rule* we need. Thus, for the time being, we shall make do with qualitative descriptions of how phase rules operate.

In this paper, we explore some of the timing details of several phonological sequences of English where such rules might apply: first, the epenthesis of stops in American English words such as *tense*, *Chomsky* and *strength*, and, second, the well-known timing effects of the voicing contrast in the coda of syllables. Basically, our results show that there must be a language-specific "rule" for American English that modifies speech timing sufficiently to create an interval that resembles a [t] but which is clearly distinct from the underlying lexical /t/ in *tents* and *prints*. In addition, we replicate earlier results showing that the effects of a change in the voicing feature in a syllable coda results in a complex set of changes in the timing of speech articulation. We argue that neither a model based on the segment as the unit of articulatory control nor physiological necessity provide plausible accounts of these effects.

1.2. *The problem*

Many dialects of American English have no surface contrast between clusters of sonorant–fricative (henceforth S–F) and sonorant–stop–fricative (S–T–F) when all the segments are in the same syllable. One example is the pair *prince* and *prints*. The underlying contrast between the two kinds of cluster seems to be neutralized in favor of the segmentally more complex one as is shown by the transcriptions below.

(a)	[tεnts]	[tεnsiti]	'tense, tensity'[1]	(1)
	[tεnts]	[tεnt]	'tents, tent'	
(b)	[tεndz]	[tεndin]	'tends, tending'	
	[tεndz]	[tεn]	'tens, ten'	
	[fαlts]	[fαlsiti]	'false, falsity'	(2)
	[fαlts]	[fαlt]	'faults, fault'	

[1]In this text, we use [ε] to represent IPA [ɛ] and [α] to represent IPA [ɔ].

The underlying representations of *tense*, *ten*, and *false* contain no stop after the sonorant /n/ or /l/ but are said to exhibit one in the phonetic output. Of course, /n/ and /l/ are not quite the same, as *tense* and *tents* seem to be perceptually neutralized while *false* and *faults* are probably distinct most of the time and only rarely confusable.

This process of epenthesis appears to be restricted to American speech. Jones (1966, p. 73) pointed out that the occurrence of a stop between a coronal sonorant and a coronal fricative is not characteristic of British English. If this is so, then it makes clear that there is nothing inevitable about the presence of a stop in such a position.

Another significant property of this epenthesis effect is that speakers are not completely consistent about their speech patterns. Javkin (1979) examined the temporal programming of the tongue in clusters with [l] as the sonorant and found that the probability of occurrence of an epenthetic stop in words like *Wells* or *bells* was dependent upon many factors, including the morphemic makeup of the word. He found that when a word ends in /l/ in the singular, then the plural usually does *not* show a phonetic stop between the [l] and [z]. On the other hand, despite this evidence for phonological influence, investigations of velum movements during the production of such clusters by speakers of British English revealed that the raising gesture for the velum lags behind (rather than precedes) the gesture of partial occlusion for the fricative (Ali, 1979; Ali, Daniloff & Hammarberg, 1979). Thus, what experimental evidence is available does not support the position that the gap between the nasal and the fricative must needs be produced due to articulatory constraints. In this respect, the American stop epenthesis rule resembles the rules that delete vowels in unstressed syllables in such words as *parade* and *Carolina* (Dalby, 1984). Recent results show that at least 10% of such vowels are deleted in unattended careful speech and as many as 30% disappear at faster tempos (Dalby, 1984). Thus the variability in occurrence of epenthesis resembles the variability found in rules of deletion as well as of insertion. The general problem, then, is what the nature of this epenthesis is and what its relationship is to the phonological descriptions on one hand, and to articulatory descriptions on the other.

1.3. *Analyses*

As noted above, there are basically two different claims in the literature regarding the origin of these non-underlying stops. One is that these stops are the result of a phonological insertion rule which applies to the systematic phonemic representation (UR) and derives the observed surface forms containing stops. The other is that the occurrence of the stops is the product of constraints applicable to the articulators themselves during the phonetic implementation of the systematic phonetic representation (SR) which itself contains no stop.

Dinnsen (1980) formulates a phonological insertion rule that would apply to underlying tautosyllabic nasal-fricative clusters to derive the observed forms containing stops. This rule would be specific to the dialects exemplifying surface forms as in (1) and (2) and can be stated in a transformational format as:

$$\underset{1}{\begin{bmatrix} +\text{con} \\ +\text{nas} \\ \end{bmatrix}} \underset{2}{\begin{bmatrix} -\text{son} \\ +\text{cont} \\ \alpha\text{voi} \end{bmatrix}} \underset{3}{\$} \quad \overset{1}{} \rightarrow \quad 1 \begin{bmatrix} -\text{nas} \\ \alpha\text{voi} \end{bmatrix} 2 \quad 3 \tag{3}$$

This rule inserts a stop that is homorganic with the preceding nasal and shares voicing with the following fricative at the end of a syllable. This rule would account for the data in (1) by applying to the simpler S–F clusters and deriving S–T–F clusters. The rule proposed by Zwicky, however, attempts to account for the data in (2) as well as in (1) above because it is not restricted to apply only to [−nas] contexts. Thus his rule would apply to any S–F cluster, regardless of the type of sonorant, whether nasal or liquid, present in the cluster.

The other interpretation of the occurrence of these epenthetic stops (Harms 1973; Ohala, 1974; Donegan & Stampe, 1979) is that it results not from the application of a phonological rule, but from universal constraints on the articulators at the production level. This constraint dictates that if the phonetic string produced as output by the phonological component includes tautosyllabic S–F clusters, then a stop *must* occur in the articulatory output. This is due, according to Harms, to "the disparity in timing between the neural commands and motor events". In a similar vein, Ohala (1974) claims that the occurrence of stops in the nasal fricative clusters is the result of some kind of mistiming. He proposes, quite reasonably, that closing the velum before the release of the occlusion for the nasal will produce a configuration of the articulators similar to that of a homorganic stop. No reason is given for why this mistiming should be expected to occur, but presumably it could be attributed either to carelessness or to inability to finely control the articulators in coordinated movements. The possibility of carelessness cannot easily be refuted but the absolute inability to finely control the articulators and coordinate them in space and time is a claim that can be tested either by physiological or acoustical experiments.

Furthermore, Ohala (1974, p. 359) offers a similar explanation for the occurrence of stops in *l*-fricative clusters. Specifically, he suggests that "the contact areas for [l] and [s] are to a certain extent complementary" and that "in moving from an [l] to an [s], contact and release of contact must be made simultaneously in these complimentary areas". Thus an inadvertent [t] frequently results. One must again presume that the ultimate reason must be either carelessness or lack of fine control. Thus the effect is not due to any linguistic factors at all, only to characteristics of the speech output system.

We suggest that another possibility exists: that epenthetic stops might be language-specific and yet distinct from the underlying stops. We shall propose on the basis of our data that there is a language-specific rule of articulation that applies in the phonological context of a sequence of nasal and voiceless fricative segments. This articulatory rule modifies the relative timing of articulatory gestures in a way that results in the presence of an articulatory and acoustic "stop" interval. A parsimonious interpretation of our data does not support the view that a /t/ segment, similar to those of the lexicon, is inserted here.

In the following eperiment, we examine the speech of two dialects of English, only one of which we expect to exhibit the epenthesis effect that is our primary interest. What we are looking for is evidence regarding the nature of epenthesis rules. In particular, if only one dialect exhibits epenthesis or if both do but do so in different ways, then the effect would seem to be "under linguistic control". If it turns out, in addition, that the epenthetic stops are clearly different from the underlying stops in their phonetic detail, we would have to view that as evidence for a low level articulatory source for the effect rather than a segmental account. This follows since a /t/ should be a /t/ whether it is underlying or inserted. Thus, a result showing a difference between underlying and inserted /t/s would suggest either that there is a wider variety of segmental allophones than had previously been expected, or else that there is an epenthesis rule that directly controls articulators below the level of segments.

2. Method

A list of 18 words was constructed containing two groups, a test group and a control-group. The eight test words, some of which were nonsense words, contained minimally distinct words contrasting in the type of sonorant (nasal or lateral), the presence or absence of an underlying stop and the voicing of the final fricative. In order to verify effects found in our nonsense test words, a control group containing only real words was selected to exemplify the major points of contrast present in the test group. It must be noted here that in tautosyllabic final clusters in English, the value of voicing is common across the sequence stop + fricative. It is not independent for the final fricative.

Test words		Control words			
dense	dens	tense	tens		
dents	dends*	tents	tends		
delse*	dells		coals	false	falls
delts	delds*	colts	colds	faults	

The words were placed in the carrier sentence: "Bob said ___ today". Four lists were then constructed by randomizing the words and four more by inverting the order of the words in the first four lists. The lists were typed onto single pages and presented to the subjects. The subjects read the eight lists and were recorded in single sessions per subject. The subjects comprised two groups, one group of Americans and one of South Africans. The five speakers of American English were all male, aged 25–35 years, and either attending or teaching at Indiana University. The four speakers of South African English were three males and one female, aged 22–30 years, attending Indiana University. The subjects were instructed to speak at a normal rate, to place the intonation peak for each utterance on the test word and to keep the intonation pattern and tempo as constant as possible.

The recordings were made in a soundproof booth, using a high impedance microphone, a custom-made preamplifier, and a Revox A700 tape recorder. Wide-band spectrograms were made of the first six readings of each word by each speaker (using a Voice Identification Series 700 sound spectrograph). The other two samples were used only if the speaker had not kept the intonation contour constant or had paused during the reading.

The test words were then segmented on the spectrograms according to the following general principles. The vowel was measured from the apparent release of the initial consonant closure to the onset of the following consonant. Any interval between the sonorant and final fricative with no acoustic energy visible on the spectrogram was considered to be a voiceless stop if its duration exceeded 10 ms; otherwise it was measured as part of the proceding resonant segment. Similarly, following the criteria employed by Raphael, Dorman, Freeman & Tobin (1975), in interval with nothing but a voice bar beginning at the cessation of an apparent resonant was measured as a voiced stop if its duration exceeded 10 ms. Otherwise it was measured as part of the preceding segment. The interval with visible high frequency noise and no formant structure was considered to be the final fricative interval. The detailed segmentation criteria were the following.

The *vowel* [ɛ] was measured from the release burst of the stop to the onset of the [n] as indicated by an abrupt drop in the first formant coinciding with the abrupt cessation of the second and third formants of the vowel. The vowel, whether [ɛ] or [ɑ], could not always be separated from a following [l]. Thus the syllabic nucleus of all words containing [l] was measured from the release of the closure of the initial consonant, the burst for [d,k] and the cessation of noise for [f], to the point where all the [l] formants disappeared before [t,d] or the onset of high frequency noise for [s,z], as a single unit.

The *nasal* segment was measured from the end of the preceding vowel to the cessation of all formants before [t] or the weakening of the first formant before [d], or to the onset of high frequency noise for [s,z].

The *voiceless stop* [t] was measured from the end of the preceding sonorant to the abrupt onset of high frequency noise for [s].

The *voiced stop* [d] was measured from the offset of the preceding sonorant to the onset of frication for [z].

The *fricatives* [s,z] were measured as the interval containing high frequency noise in the range of 4–5 kHz.

All the data collected were then statistically analysed using Version 6 of SPSS (Nie, Hull, Jenkins, Steinbrenner & Bent, 1975). Analysis of variance tests were run on the durations of all intervals to determine the effects of the independent variables. All null hypotheses were rejected at $p < 0.01$ and accepted at $p > 0.05$. For marginal cases the exact probabilities are given in the tables and figures

3. Results

In this section we will discuss the results of the experiment under three separate headings. First we will discuss the production of the clusters by the two groups of subjects in terms of the presence or absence of acoustically evident stops between the sonorant and fricative. Second, we will discuss the timing of the measured intervals in terms of the presence or absence of an underlying stop and in terms of the voicing of the obstruent cluster.

3.1. *Presence of stop in the acoustic signal*

One of the aims of this experiment was to determine whether or not the two groups of subjects differed in the frequency of occurrence of some acoustic event that could be identified as a stop. Table I summarizes the results for both groups. In Table I we see that the South African speakers invariably produced S–F clusters (as in *dense*) with no acoustic stop present in the signal, whereas they invariably produced S–T–F clusters (as in *dents*) with an acoustic stop.

The contrasts between word final /ns/, /ls/, /nz/, /lz/ on the one hand and /nts/, /lts/, /ndz/, /ldz/ on the other were always maintained on the surface in terms of the presence or absence of an acoustic stop that was at least 10 ms long. Figure 1 shows sound spectrograms for one token of each word illustrating the obvious difference between *dense* vs. *dents* and *false* vs. *faults* in the speech of South African speakers.

TABLE I. Percentage occurrence of acoustic stops in the speech of South Africans and Americans. A stop was defined as an interval of silence longer than 10 ms visible on a spectrogram between the vowel and the fricative

		Test words	Control words
South Africans			
Voiceless			
n-clusters	*ns*	0	0
	nts	100	100
l-clusters	*ls*	0	0
	lts	100	100
Voiced			
n-clusters	*nz*	0	0
	ndz	100	100
l-clusters	*lz*	0	0*
	ldz	100	100
Americans			
Voiceless			
n-clusters	*ns*	100	100
	nts	100	100
l-clusters	*ls*	100	100†
	lts	100	100
Voiced			
n-clusters	*nz*	7	10
	ndz	27	41
l-clusters	*lz*	7	2†
	ldz	97	100

*$n = 48$.
†$n = 60$.

The American speakers' results are shown in the lower half of Table I. Looking first at the clusters ending in voiceless /s/ we see that the American speakers always exhibited some sort of phonetic stop between the sonorant and the voiceless fricative regardless of whether the stop was present underlyingly. Thus the underlying contrasts between /ns/, /ls/ on the one hand and /nts/, /lts/ on the other hand were always neutralized in favor of the segmentally more complex cluster. Figure 2 shows sound spectrograms for one token of each kind of cluster for the American speakers.

However, as can be seen at the bottom of the table the situation was rather different when the clusters ended in voiced /z/. In the case of the underlying *ndz* clusters, the American speakers evidenced an acoustic stop only about a third of the time. This was contrary to the expected 100% stop occurrence since there had been no stop deletion rules posited for these clusters. In the productions of the *lz* cluster, the underlying stop surfaced in the great majority of cases (97% for test words and 100% for control words). In the S–F clusters, for both the nasal and the liquid sonorant, the American speakers

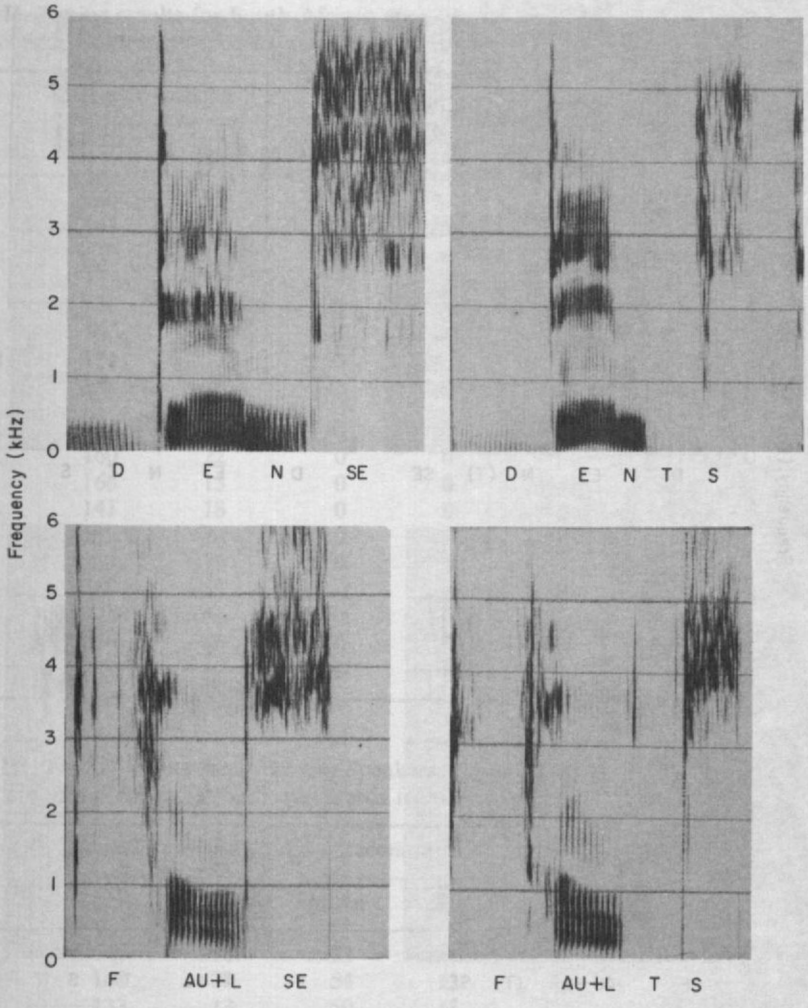

Figure 1. South African productions of *dense–dents* and *false–faults*. Sound spectrograms of typical tokens.

exhibited an (epenthetic) acoustic stop in very few cases (the highest being 10% for control words of *nz* type).

3.2. *Temporal patterns*

In this section we present the timing patterns exhibited by S–F and S–T–F clusters. The relevant intervals that were segmented and measured were, in general, the vocalic nucleus, the stop and the fricative. The vocalic nucleus, in turn, was divided into the vowel and the nasal part in words containing nasals, since we found that no reliable division was possible for the *Vl* sequences. Here we will present the effects of the presence (vs. absence) of an underlying stop and of the voicing of the final cluster on all the measured intervals and on the word duration. Word duration was computed for each token by finding the sum Vocalic Nucleus + (Stop) + Fricative. Tables II and III show

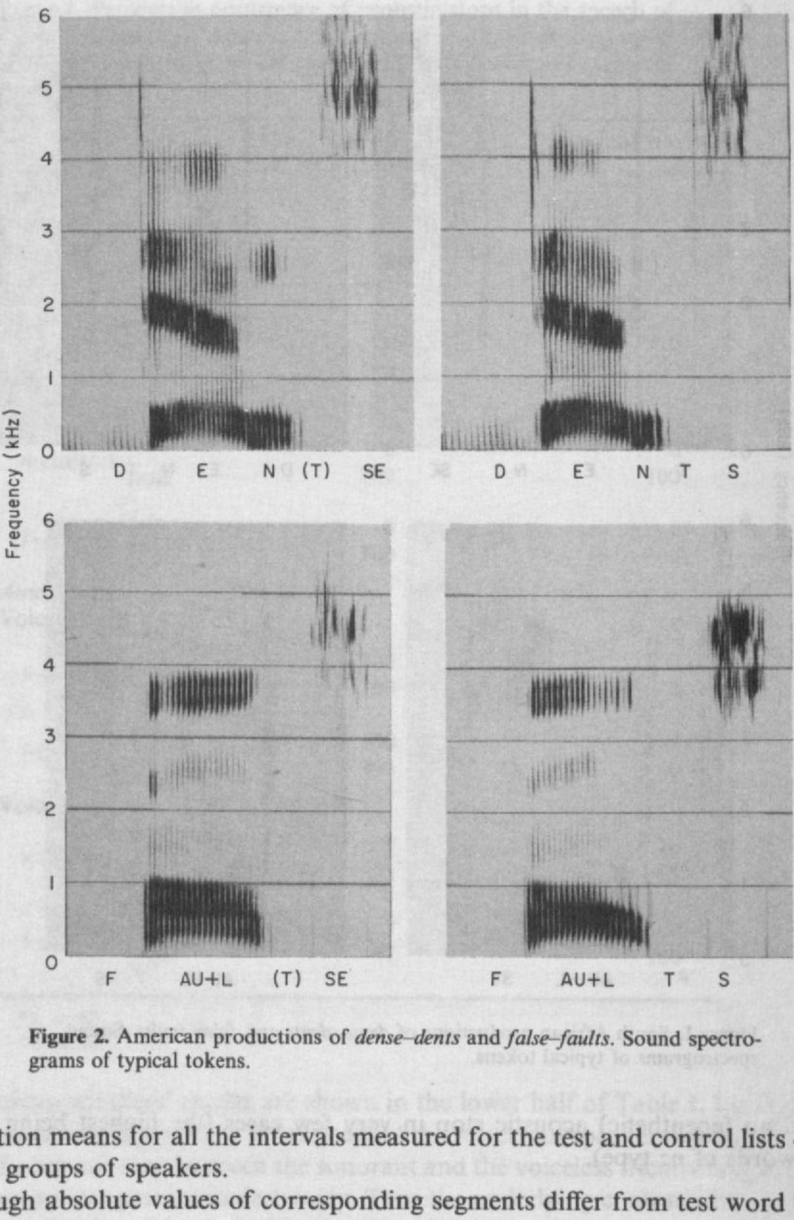

Figure 2. American productions of *dense–dents* and *false–faults*. Sound spectrograms of typical tokens.

the duration means for all the intervals measured for the test and control lists of words for both groups of speakers.

Although absolute values of corresponding segments differ from test word to corresponding control word, the patterns exemplified by the two groups are remarkably similar and in many cases identical. The upper half of Table IV presents the results of analysis of variance for both groups of speakers to determine the effects of the presence or absence of an underlying stop on each interval and on the total word duration. For each comparison it is stated whether the *F*-value had probability less than 0.01 (**), less than 0.05 (*) or greater than 0.05 (ns). One can see, first, that the results of the test list correspond remarkably well with those of the control list. Second, the results for the Americans and South Africans are remarkably similar, with a few prominent exceptions.

Since the test and control words display nearly identical patterns, it was decided to pool the durations of all words that were similar in their segmental makeup in order to

TABLE II. Timing results for South African speakers. For all means $n = 24$. Values are in milliseconds; * indicates words from the control group of real English words

Word	Vocalic nucleus		Resonant		Stop closure		Frication	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
dense	120	12	87	14	0	0	141	33
*tense	147	15	88	10	0	0	131	23
dents	110	8	61	9	67	26	114	20
*tents	132	11	61	8	66	24	112	19
dens	136	20	159	40	0	0	90	15
*tens	166	27	165	26	0	0	82	11
dends	134	13	155	40	49	21	97	20
*tends	158	13	146	30	50	19	86	12
delse	194	19	0	0	0	0	152	35
*false	160	22	0	0	0	0	147	32
delts	166	13	0	0	95	19	119	16
*faults	141	18	0	0	85	20	121	23
*colts	193	24	0	0	85	18	117	17
dells	287	75	0	0	0	0	90	15
*falls	268	50	0	0	0	0	88	20
*coals	321	55	0	0	0	0	85	17
delds	256	40	0	0	77	37	100	18
*colds	289	47	0	0	66	29	92	19

TABLE III. Timing results for American Speakers. For all means $n = 30$, except where noted. Values are in milliseconds; * indicates words from the control group of real English words

Word	Vocalic nucleus		Resonant		Stop closure		Frication	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
dense	140	18	58	14	31	12	103	24
*tense	168	20	54	11	36	12	98	26
dents	133	13	50	13	42	24	97	30
*tents	162	17	48	9	42	18	94	25
dens	166	21	131	28	0	0	88	24
*tens	198	25	127	29	0	0	89	26
dends†	156	19	133	32	0	0	89	30
dends‡	168	14	151	35	42	18	83	18
*tends§	182	19	114	20	0	0	88	29
*tends‖	193	17	134	30	39	18	82	20
delse	201	27	0	0	23	18	117	28
*false	190	14	0	0	39	24	103	19
delts	183	21	0	0	74	17	93	26
*faults	183	25	0	0	73	20	97	27
*colts	223	30	0	0	73	20	100	26
dells	287	49	0	0	0	0	89	19
*falls	297	42	0	0	0	0	88	17
*coals	335	55	0	0	0	0	92	12
delds	240	36	0	0	71	27	90	18
*colds	293	47	0	0	65	24	87	27

†$n = 22$. ‡$n = 8$. §$n = 17$. ‖$n = 12$.

TABLE IV. Results of analysis of variance for the test and control words separately, and for both groups of speakers: effects of presence vs. absence of an underlying stop; effects of a change in obstruent voicing

Word	Vocalic nucleus	Resonant	Stop	Fricative	Word duration
Effects of presence of a stop					
South Africans					
dense–dents	**	**		**	ns†
dens–dends	ns	ns		ns	**
tense–tents	**	**		**	ns
tens–tends	ns	ns		ns	*
delse–delts	**			**	**
dells–delds	ns			ns	**
false–faults	**			**	**
coals–colds	*			ns	*
Americans					
dense–dents	ns	*	*	ns	ns
dens–dends	ns	ns		ns	ns
dens–dends	ns	ns		ns	ns
tense–tents	ns	ns	ns	ns	*
tens–tends‡	ns	ns		ns	ns
tens–tends§	ns	ns		ns	ns
delse–delts	**		**	**	ns
dells–delds	**			ns	ns
false–faults	ns		**	ns	ns
coals–colds	**			ns	ns
Effect of change in voicing					
South Africans					
dense–dens	**	**		**	**
dents–dends	**	**	*	**	**
tense–tens	**	**		**	**
tents–tends	**	**	*	**	**
delse–dells	**			**	*
delts–delds	**		ns	**	**
false–falls	**			**	**
colts–colds	**		*	**	*
Americans					
dense–dens	**	**		*	**
dents–dends	**	**	ns	ns	**
tense–tens	**	**		ns	**
tents–tends	**	**	ns	ns	**
delse–dells	**			**	**
delts–delds	**		ns	ns	**
false–falls	**			**	**
colts–colds	**		**	ns	**

* = $p < 0.05$.
** = $p < 0.01$.
†ns = not significant.
‡$n = 27$ for tens, $n = 17$ for tends.
§$n = 12$ for tends.

TABLE V. Results of analysis of variance on the test and control words combined for the set of words with a nasal sonorant: effects of a change in the presence of a stop; effects of a change in voicing

Word-final sequence	Vocalic nucleus	Resonant	Stop	Fricative	Word duration
Effects of presence of stop					
South Africans					
-ense, -ents	**	**	—	**	ns†
-ens, -ends	ns	ns	—	ns	**
Americans					
-ense, -ents	ns	**	**	ns	ns
-ens, ends $n = 55$, $n = 20$	ns	ns	—	ns	**
-ens, -en(d)s $n = 55$, $n = 39$	*	ns	—	ns	ns
Effect of change in voicing					
South Africans					
-ense, -ens	**	**	—	**	**
-ents, -ends	**	**	*	**	**
Americans					
-ense, -ens	**	**	—	*	**
-ents, -ends	**	**	ns	ns	**

** = $p < 0.01$.
† ns = not significant.

increase the degrees of freedom. The analysis of variance results for the pooled data shown here are summarized in Table V. The discussion that follows and the figures that accompany it, refer then, to these pooled results for the clusters with nasal sonorants and to the control words for the clusters with a liquid sonorant.

3.2.1. *Presence of an underlying stop—South African speakers*

Figure 3 shows pooled duration means for the South African speakers. The words are paired in terms of whether they contained an underlying stop or not. The *n*-group is pooled over both control and test words. The plots for the *l*-group, on the other hand, represent only the test words (since the differences in segmental makeup were too great to allow pooling). The duration means for each interval and for the total word, either Vowel + Nasal + (Stop) + Fricative or Voc.Nucleus + (Stop) + Fricative, are plotted cumulatively.

Looking first the results for the *n*-group we notice two distinct cases. For words ending in voiceless /s/, the total word duration was not affected by the presence of the stop [$F(1.93) = 0.259$, ns]. The stop, however, significantly affected all the other segments by shortening them. That is, it resulted in a shorter vowel [$F(1,93) = 12.445$, $p < 0.001$], nasal [$F(1.93) = 164$, $p < 0.001$], and fricative [$F(1,93) = 21.3$, $p < 0.001$]. It appears as though all the intervals shortened in order to accommodate the stop while still keeping word duration constant.

For the voiced pairs, though, the results were markedly different. The presence of the underlying stop significantly increased the total word duration [$F(1,92) = 16.531$, $p < 0.001$] but did not affect any of the other intervals, neither the vowel [$F(1,92) = 1.008$, ns], nasal [$F(1,92) = 2.535$, ns] nor fricative [$F(1,92) = 2.7$, ns]. This is exactly the

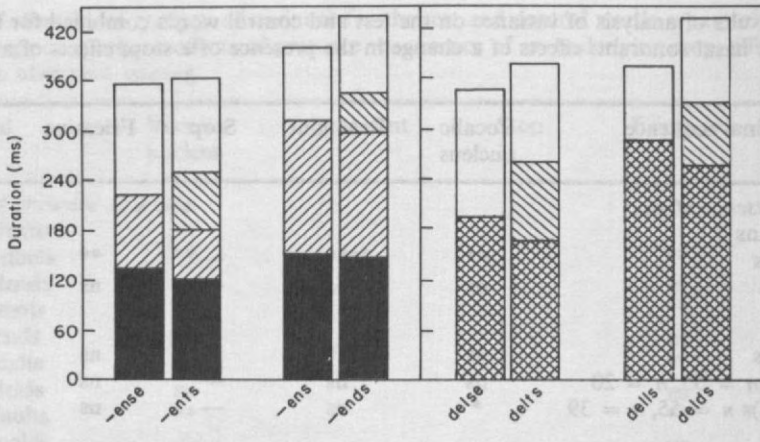

Figure 3. South African results: Mean duration of vowel, sonorant, stop, and fricative, paired by presence vs. absence of underlying stop. Data are pooled across speakers and across the test words and control words. □, Fricative; ▨, nasal; ▨, stop; ■, vowel; ▩, vowel + l.

opposite of the case for the voiceless clusters since here the word duration was affected but not the individual segmental intervals.

Looking next at the right hand panel of Fig. 3, for words with /l/ in voiceless clusters, it can be seen that the presence of an underlying stop had a significant shortening effect on both the vocalic nucleus [$F(1,46) = 36, p < 0.001$] and the fricative [$F(1,46) = 17.22, p < 0.001$]. However, the word duration was still lengthened significantly by the presence of the stop [$F(1,46) = 9.882, p < 0.003$]. We see, then, a trend toward keeping the word duration constant, as in the voiceless clusters with /n/, but one that does not quite succeed. On the other hand, when the final cluster is voiced, the results of the *l*-group show a pattern exactly like that of the *n*-group. The presence of the underlying stop has no effect on the segmental intervals of vocalic nucleus [$F(1,46) = 5.8$, ns], or the fricative [$F(1,46) = 4.9$, ns], yet it did significantly increase the word duration [$F(1,46) = 16{,}118, p < 0.003$]. Again it appears as though the /d/ is simply concatenated into the word.

It is clear that the temporal effects of an underlying stop depend critically on its voicing (and thus, of course, on the voicing of the fricative). When the stop is voiceless there appears to be an effort to compensate for the increase in total duration caused by the presence of the stop by decreasing the durations of all the other intervals. This effort is successful with the nasal but only partially so with the liquid. When the stop is voiced, however, there is no effort to compensate regardless of the sonorant type. The words containing the voiced stop are longer than the words without it by approximately the duration of the stop.

3.2.2. *Presence of underlying stop, American speakers*

Figure 4 shows duration means for the American speakers. Again the words are paired in terms of whether or not they contain an underlying stop.

As mentioned above in Section 3.1, all voiceless clusters wer produced by the Americans with an acoustic stop whether the resonant was /n/ or /l/. Thus, in these cases, to talk about the effect of the presence or absence of an underlying stop on the stop duration is equivalent to talking at the phonetic level about the effect of an underlying

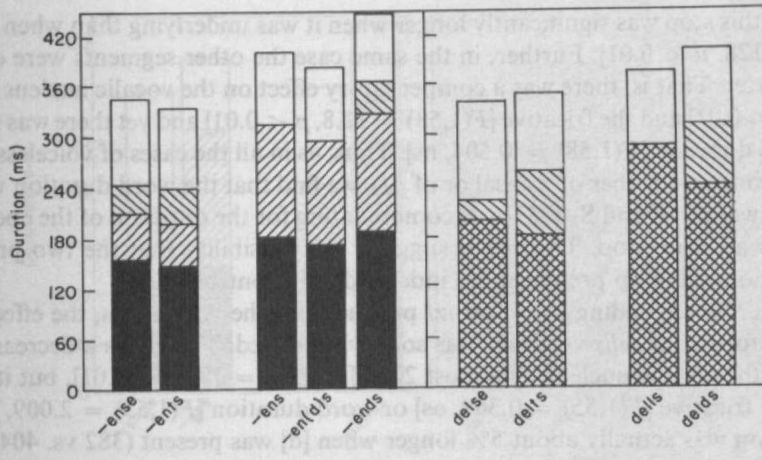

Figure 4. American results: Mean durations of vowel, sonorant, stop, and fricative paired, by presence vs. absence of underlying stop. Data and pooled across speakers and across the test words and control words. □, Fricative; ▨, nasal; ▨, stop; ■, vowel; ▩, vowel + l.

vs. an inserted stop. In the discussion that follows the two descriptions will be used interchangeably.

Analysis of variance on the combined *dents–dense* and *tents–tense* results showed highly significant effects of the nature of the stop on both the sonorant, *n* [$F(1,117) = 9.671$, $p < 0.002$], and on the stop [$F(1,117) = 8.574$, $p < 0.004$]. The stop was longer when it was an underlying one and the nasal that preceded an underlying stop significantly shorter. The nature of the stop did not affect the duration of the vowel [$F(1,117) = 3.118$, ns], the fricative [$F(1,117) = 1.094$, ns], or the word [$F(1,117) = 1.318$, ns]. Thus, the total word duration was kept constant regardless of whether or not the stop present in the cluster was an underlying one or an inserted one.

Looking next at *n*-words ending in /z/, Fig. 4 shows the results of the comparisons in two different cases. First, we examined the effects of an underlying stop when that stop did not actually surface in the acoustic output. Thus we compared /nz/ which surfaces as [nz] with the instances of /ndz/ that surfaced as [nz], spelled -EN(D)S in Fig. 4. There was no significant effect of the presence of absence of an underlying stop on any of the intervals [vowel: $F(1,92) = 6.938$, $p = 0.011$; sonorant $F(1,92) = 0.439$, ns; fricative $F(1,92) = 0.002$, ns; word duration $F(1,92) = 2.091$, ns]. Thus, it seems one might propose deletion of the underlying stop segment for these tokens, since some productions of *dends* were homophonous with *dens* as far as we can observe with these data.

Secondly, we examined the effect of the underlying voiced stop when it actually occurs in the acoustic output. We compared, that is, /nz/ with cases of /ndz/ surfacing as [ndz]. The presence of the stop did not significantly affect the durations of either the vowel [$F(1,73) = 0.050$, ns], the nasal [$F(1,73) = 2.205$, ns] or the fricative [$F(1,73) = 0.977$, ns]. It did, however, significantly increase the duration of the word [$F(1,73) = 8.968$, $p < 0.004$]. Thus, the Americans here behave the same as the South Africans in apparently concatenating a /d/ segment onto the string of segments.

In the *l*-group, whose results are plotted in the right-hand panel of Fig. 4, the effect of the underlying stop in the words that ended in a voiceless cluster was varied. As mentioned above, an acoustic stop was always present between the /l/ and /s/. But the duration of this stop was significantly longer when it was underlying than when inserted [$F(1,58) = 128, p < 0.01$]. Further, in the same case the other segments were compensatorily shorter. That is, there was a compensatory effect on the vocalic nucleus [$F(1,58) = 8.671, p < 0.01$] and the fricative [$F(1,58) = 11.8, p < 0.01$] and yet there was no effect on the word duration [$F(1,58) = 0.504$, ns]. Thus, as in all the cases of voiceless clusters in this experiment (whether of a nasal or of /l/), we find that the word duration was kept constant between S–F and S–T–F by accommodating for the duration of the epenthetically inserted acoustic stop. This result suggests the possibility that the two properties (word duration and stop presence) are independently controlled.

Finally, in *l*-words ending in voiced /z/ produced by the Americans, the effect of the underlying stop, as in *dells* vs. *delds*, was somewhat mixed. There was a decrease in the duration of the vocalic nucleus by almost 20% [$F(1,55) = 19, p < 0.01$], but it had no effect on the fricative [$F(1,55) = 0.369$, ns] or word duration [$F(1,55) = 2.009$, ns]. The word duration was actually about 5% longer when [d] was present (382 vs. 404 ms) but not significantly so.

In general, we see again two distinct patterns. For words ending in voiceless /s/, segmental durations were clearly adjusted compensatorily in order to accommodate the presence of an acoustic stop with minimal lengthening of the word. Still it is found that underlying and inserted stops have different durations. In the voiced clusters the optional deletion of underlying /d/ made the /ndz/ identical to an underlying /nz/ cluster. When the /d/ was not deleted the word duration tended to increase for S–T–F clusters as opposed to S–F clusters.

For the American speakers, then, the duration of the acoustic stop followed by a voiceless final fricative depends on the origin of the stop, i.e. on whether it is present or not in the underlying representation. In general, the results for the *l*-cluster and the *n*-clusters show a tendency to reflect the underlying segmental makeup in the surface temporal characteristics of the cluster although in a complex way.

3.2.3. *Voicing of final fricative*

The voicing of the word final fricative and the preceding stop had a prominent effect on the durations of most intervals. This is true regardless of the type of sonorant present in the cluster.

3.2.3.1. *South African speakers*. Figure 5 shows duration means for the South African speakers. In the *n*-group of test words, for words without stops (as in *dense* and *dens*), the voicing of the final fricative had a significant effect on all the intervals: on the vowel [$F(1,45) = 12.002, p < 0.01$], on the sonorant [$F(1,45) = 67, p < 0.01$], on the fricative [$F(1,45) = 45, p < 0.01$], and on the word [$F(1,45) = 7.662, p < 0.01$]. When stop was acoustically present, as in *dents* vs. *dends*, the duration of all intervals were significantly affected by the voicing of the final cluster except those of the stops. The effect of voicing on the duration of the stops was marginal: [$F(1,44) = 6.282, = 6.282, p = 0.016$]. The other intervals were significantly longer when the final cluster was voiced than when it was voiceless: the vowel [$F(1,44) = 131, p < 0.01$], the sonorant [$F(1,44) = 131, p < 0.01$] the fricative [$F(1,44) = 9, p < 0.01$] and the word [$F(1,44) = 37, p < 0.01$].

Turning next to the *l*-group, we see that the vocing of the fricative had a significant effect on the vocalic nucleus [$F(1,46) = 75, p < 0.01$] and on the fricative [$F(1,46) = 63, p < 0.01$] when there were no stops in the clusters as in *delse* vs. *dells*. But the combined result of lengthening the vocalic nucleus and shortening the fricative was to reduce

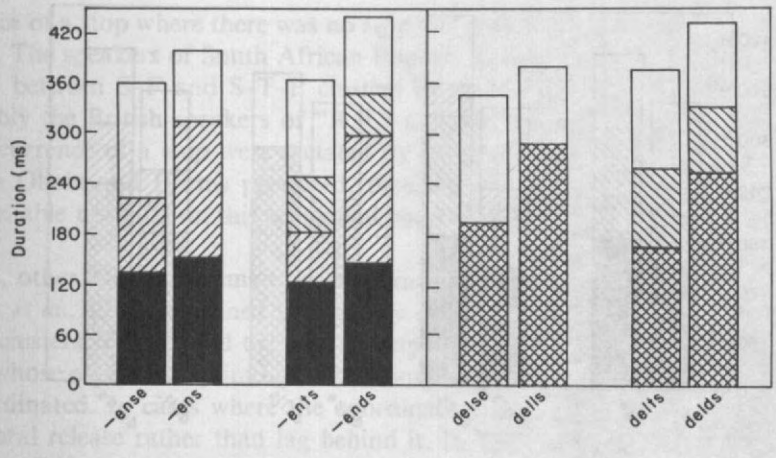

Figure 5. Mean durations of vowel, sonorant, stop, and fricative for South Africans, paired by the voicing of the final obstruent (voiceless–voiced). □, Fricative; ▨, nasal; ▨, stop; ■, vowel; ▩, vowel + l.

the effect on the increase of the total word duration to only marginal significance [$F(1,46) = 6.021$, $p = 0.018$]. On the other hand when a stop was acoustically present, all intervals except for the stop were affected [for the vocalic nucleus $F(1,46) = 109$, $p < 0.01$; for the fricative $F(1,46) = 14.2$, $p < 0.01$; and for the word $F(1,46) = 18$, $p < 0.01$]. But the effect on the stop was marginal [$F(1,46) = 4.018$, $p = 0.051$]. In general, the voicing had an effect on all durations for the South Africans. It is surprising that the effect on the stop itself was only marginal in both the *n*- and the *l*-groups of words, since many researchers have reported significant differences between [d] and [t] for American English in syllable-final positions.

3.2.3.2. *American speakers.* Looking at the left panel of Fig. 6, *n*-group of words show that the voicing of the final fricative significantly affects the durations of most intervals. When there was no underlying stop present (*dense–dens*) the words with a voiced fricative had a significantly longer vowel [$F(1,55) = 24$, $p < 0.01$], sonorant [$F(1,55) = 155$, $p < 0.01$] and word duration [$F(1,55) = 15.284$, $p < 0.01$]. The effect on the fricative was marginal [$F(1,55) = 5.364$, $p = 0.024$]. When there was an underlying stop present which also occurred on the surface, then the effect of a voiced cluster was to significantly increase the duration of the vowel [$F(1,36) = 44$, $p < 0.01$], the sonorant [$F(1,36) = 171$, $p < 0.01$], and the word [$F(1,36) = 33$, $p < 0.01$]. But there was no effect on the stop [$F(1,36) = 0.003$, ns] or the fricative [$F(1,36) = 1.584$, ns].

In the *l*-group of words the clusters with a voiced fricative had significantly longer durations for all the intervals except the stop. Thus when no underlying stop was present the effect was significant on the vocalic nucleus [$F(1,52) = 75$, $p < 0.01$], on the fricative [$F(1,52) = 12.534$, $p < 0.01$] and the word duration [$F(1,52) = 12.555$, $p < 0.01$]. When there was an underlying stop present, only the vocalic nucleus and the word durations were significantly affected. They were both longer [for the vocalic nucleus $F(1,57) = 61$, $p < 0.01$]; and the word duration [$F(1,57) = 13.68$, $p < 0.01$]. But the fricative was not affected [$F(1,57) = 0.552$, ns] and neither was the stop [$F(1,57) = 0.001$, ns].

For both South Africans and Americans, then, the difference in acoustic duration between voiced and voiceless underlying stops was either marginally significant or not

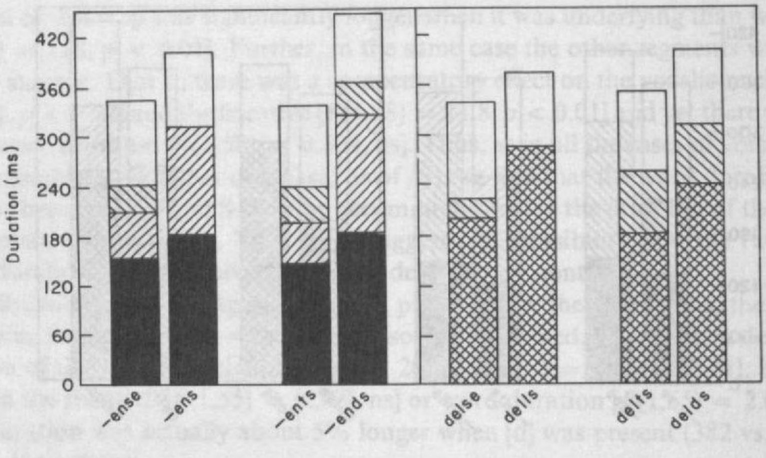

Figure 6. Mean durations of vowel, sonorant, stop, and fricative for Americans, paired by the voicing of the final obstruent (voiceless–voiced). □, Fricative; ▨, nasal; ▨, stop; ■, vowel; ▩, vowel + l.

significant at all. This is quite different from previously reported results that have shown postvocalic voiced stops to be shorter than voiceless ones (e.g. Peterson & Lehiste, 1960; Port, 1981). It may be that since much of the temporal information relevant to the perception of voicing in this position is carried by the preceding vocalic nucleus (cf. Raphael *et al.*, 1975, for nasals; Lehiste, 1972, for sonorants in general), the gesture for the production of /t/ and /d/ is allowed to vary widely (as suggested by Zue & Laferriere, 1979). Whenever the duration of the two stops is not significantly different, the difference in the durations of the vocalic nuclei still carries the distinctive timing information either in absolute terms or in the ratio of the vocalic nucleus to the consonant closure, as argued by Port (1981). Thus, it is clear that the effects of voicing on segmental durations are not confined to the segment whose voicing is being varied, and the timing of several other segments carries the information necessary to distinguish the two voicing classes.

4. Discussion

In this experiment we attempted to determine the segmental make-up and temporal patterning of English words ending in S–F and S–T–F clusters. The discussion is divided into several parts: the first having to do with the apparent segmental make-up of these clusters, the second with the detailed timing characteristics relevant to both the epenthesis effect and the voicing effect. The third section presents our own interpretation of the epenthesis found in American English.

4.1. *Segmental make-up in underlying and surface representation*

This research was partly stimulated by differing claims concerning the occurrence of stops in S–F clusters by two groups of linguists (Dinnsen, Zwicky vs. Ohala, Harms) and by the fact that Jones (1966) reserved his comment about epenthesis to American English. The results of the experiment clearly support Jones. They showed that the occurrence of a stop where there was no stop underlyingly was limited to the American speakers. The speakers of South African English maintained the underlying segmental contrasts between S–F and S–T–F clusters in both voiced and voiceless conditions. Presumably the British speakers of "R.P." analysed by Jones also share this feature. If the occurrence of a stop were dictated by intrinsic timing constaints on the articulators, as Ohala and Harms proposed, then the South African speakers should not have been able to maintain this segmental contrast. It seems then that the effect is not universal.

In fact, other evidence seems to corroborate this point. Ali and his colleagues (Ali, 1979; Ali *et al.*, 1979) examined the activity of the velum in the production of nasal-fricative clusters. They found that the closing of the velum and the release of the oral closure, whose expected mistiming was the source of Ohala's interpretation, seem to be well coordinated. In cases where the coordination was poor, velar closing tended to precede oral release rather than lag behind it. In their experiment, nasal airflow was observed well into the temporal interval assigned to the fricative. Interestingly enough, the subjects in their study were not Americans.

It seems then that the occurrence of these stops is not universally predictable or required. This seems to leave us with no explanation for the occurrence of these stops in American English, other than one along the lines of Dinnsen and Zwicky. That is, these stops must be the product of a language-specific rule of stop epenthesis. The rule, in either Dinnsen's or Zwicky's versions (Dinnsen restricted it to nasal-fricative clusters), predicted the insertion of stops regardless of the voicing of the final fricative, the only constraint being that the inserted stop agree in voicing with that of the fricative. The American speakers however, produced the words ending in voiced and voiceless fricatives in two markedly different ways as discussed in Section 3.1. In review:

(1) When the fricative was voiceless they produced both S–F and S–T–F clusters as S–T–F clusters but with different timing patterns which nevertheless left word duration the same;

(2) When the fricative was voiced, then two distinct cases surfaced, being (a) that in *n*-clusters, an underlying voiced stop was usually dropped and epenthesis was infrequent, and (b) that in *l*-clusters an underlying voiced stop was almost never dropped and epenthesis was infrequent.

Thus, the rule given in the introduction must be reformulated. In fact, if all the data are to be accounted for, three rules are necessary. The first, an epenthesis rule, would insert a voiceless stop between a sonorant and a voiceless fricative, obligatorily. The second rule would insert a voiced stop between a sonorant and a voiced fricative, optionally. Despite their obvious similarity it is difficult to see how the two rules could be collapsed since the first is obligatory and the second optional. A third rule would be necessary to optionally delete a voiced stop after a nasal and before a fricative in a syllable coda. These three rules, then, would have to be part of the phonological component of American English, and thus specific to this group of dialects.

Nevertheless, there is a very deep problem with a segmental analysis of this sort. Our detailed study of the timing of these segments revealed that the epenthetic stops were different from those that derive from underlying forms. Yet the differences observed are sufficiently small that it is implausible to argue that they represent different "allophones". In the next section, we review these results.

4.2. *Timing patterns*

The second rationale for this project was to obtain experimental data on timing in S–F clusters and to examine the effects of the voicing of syllable final consonants on preceding segments. Some of these results are at odds with other published data while other aspects of the timing raise difficult problems for a phonological analysis of the epenthesis effect.

4.2.1. *Vowel duration effects*

What is the effect of the voice feature of a final cluster on the preceding vowel? It is typically found that both the nasal (or other sonorant) and the vowel are shorter when the following obstruent is voiceless and longer when the following consonant is voiced (Chen, 1970; Lehiste, 1972; Raphael *et al.*, 1975; Lovins, 1978). Still, Zue & Laferriere (1979) reported no difference in vowel durations between words containing /Vnt/ and /Vnd/. These authors ascribe the differences between their results and those of Raphael *et al.* to the larger size of their corpus (608 vs. 85 tokens), their use of a carrier sentence instead of words in isolation, and to the fact that their sample included polysyllabic words instead of only monosyllabic ones. It is well-known now that many timing effects due to segmental features can be seen clearly only by looking at minimally contrastive word sets under conditions that maximize overall durations (such as monosyllables pronounced either in isolation or in carrier sentences where they receive sentence stress) (see, for example Klatt, 1976). Of course,the fact that the effects are small and that they become difficult to observe in connected speech does not make them less interesting or relevant for a theory of the production of language.

4.2.2. *Consonant duration effects*

We found as others have that with respect to the voicing contrast, a nasal segment in a VNC behaves more like the vowel; it is longer before a voiced segment. The nasal consonant was much shorter before a voiceless consonant than before a voiced one. In our experiment, however, the duration of the nasal for the American speakers was also longer when followed by an inserted stop than by an underlying one. Conversely, the voiceless stop that followed the nasal or liquid was shorter when it was epenthetic than when it was present in the underlying representation of a word.[1]

[1]This result bears on some claims made by Fujimura & Lovins (1978) who propose a syllable production model that makes predictions about such timing effects. In their model, syllables are constructed by concatenating *core elements* and phonetic *affixes*. The core is roughly the onset and nucleus of the syllable (Hayes, 1981) but the affix is a certain fraction of the final consonant cluster. For example, a word like *dends* is generated by concatenating a core element [den] and the affixes [d] and [z]. The [d] and [z] are considered affixes here because they are a strong of apical obstruents that agree in voicing with each other (Fujimura & Lovins, 1978, p. 111). On the other hand, a word like *strength* is generated by concatenating a core element /strenk/ and an affix /θ/. The /k/ is not an affix because it differs in voicing from the nasal and is not apical.

Our data bear on this model since particular cores and affixes are assumed to have characteristic durations at a given tempo so that a concatenation process should produce roughly additive results without any compression in the core (cf. Klatt, 1973; Port, 1981). We found that in words with voiced fricatives, the existence of a stop lengthened the word duration while having no effect on the other intervals in the word. Thus *delds* was longer than *dells* and *dends* longer than *dens* while other segmental intervals were unaffected by the addition of the /d/ segment.

For the voiced consonant sequences, then, the model made very nearly the right predictions concerning both the segmental and overall durations. In the case of the clusters with voiceless obstruents, however, the model does not seem to help at all. This is because for the American speakers the presence of the "inserted" stop and the readjustments of the adjacent segments kept the word duration constant.

This phenomenon—in which an apparent neutralization of a phonological contrast turns out to be incomplete—is not an isolated case. There are a number of examples in the literature which we shall discuss in the next section.

4.3. *Implications*

How could the epenthetic /t/ be different from an underlying /t/? This puzzling effect raises some important problems for current views of the relation between phonetics and phonology. Most theories of language and speech assume that morphemes are spelled from phoneme-like units and that, during speech production, phonological rules translate these into a far larger alphabet of allophones employed in that language. Since allophones are segmental units, it falls to "implementation rules", the next lower stage, to generate specifications for articulatory gestures in time (e.g. Chomsky & Halle, 1968; Klatt, 1976).

In such a model, epenthesis should be performed at the phonological rule stage for two reasons: first, because it is clearly a segment-sized unit that is inserted and, second, because the pattern is invariant over the language community. Thus the effect cannot be interpreted as a low-level speaker idiosyncrasy. But, as noted above, if a phonological /t/ is what is inserted, there should be no way that the implementation rules which are responsible for timing could discriminate an underlying /t/ from a derived one. Yet our results show a difference between them.

Three interpretations seem possible at this point:

(a) *Temporal implementation rules have access to the derivational history of a string*

This renders linguistic formalization almost absurdly powerful and implies carrying a staggering amount of information along as part of the derivational structure. This solution seems to violate a fundamental assumption of formal linguistic theory, the assumption that a stretch of speech has only one structure at a time (at a given level) within a derivation.

(b) *Some timing rules* (such as American stop epenthesis), *are assigned before the application of some other phonological rules* (considered in Dinnsen, 1985).

Thus, we might try to imagine a way for some timing specifications to be attached to segments before application of certain segment changing rules, that is, relational timing properties, such as "longer by k%", attached as diacritic properties to the segments. These eventually would be manifested as particular timing patterns and might offer a way of having the timing rules be sensitive to derivation. In any case, however, this solution provides two different allophones of /t/, one derived from an underlying segment and the other produced only by the epenthesis rule. Yet because these allophones are not auditorily distinguishable to native speakers (at least not very well), the plan violates a basic assumption about what the systematic phonetic segments are. Systematic phonetic segments should be auditorily distinctive to native speakers of a dialect since that is the only way they could be learned (Chomsky & Halle, 1968).

A third proposal could be made as an alternative to (b). It may have been noted that most cases of the underlying stops in nasal clusters have more than one morpheme (viz., *tents*, *tends*, etc.). Thus it might be that:

(c) *morphological layers in the sense of Kiparsky (1982) have their own phonetic timing rules.*

We might imagine that the stem *tent* is pronounced with a certain set of articulatory control parameters and that the inflectional suffix, perhaps because it is on a different layer, has its own articulatory (hence acoustic) structure. This implies that at this level of the speech production system a morphological /s/ is distinct from an ordinary /s/ at the end of a syllable in the layer-one lexicon. In *tense* all the segments lie on the same morphological level and are executed with their own articulatory commands—commands which result in the insertion of an acoustic stop in these clusters in certain dialects of English but not in other dialects. This suggestion can be viewed as an alternative way of "mixing" timing rules with the phonological rule systems.

We shall propose in the next section another hypothesis—one that is not necessarily incompatible with the above notions but which is intended to relate this effect more closely to a model of the motor control of speech production.

It should be noted that there are now a considerable number of well documented cases of incomplete neutralization between phonological entities (see Dinnsen, 1985, for a review). For example, the American English flap comes close to naturalizing /t/ and /d/ but there remains a statistical difference for many dialects (Fox & Terbeek, 1977; Huff, 1980). Word-final devoicing in German is supposed to produce complete neutralization of voicing in word-final position yet there remain certain differences in the timing of production (Fourakis and Iverson, 1984; Port & O'Dell, 1985). Similar incomplete neutralization of phonological contrasts has been found in Catalan (Dinnsen, 1985; Dinnsen & Charles-Luce, 1984) and Polish (Slowiaczek & Dinnsen, 1985). These cases required experimental analysis to detect the differences but other more mundane examples abound when speaking tempo is changed. For example, pairs like *parade* and *prayed* are very distinct in slow or careful speech in the U.S.A., but become gradually neutralized at faster tempos. Such gradual degradation cannot be accounted for with a rule that deletes the schwa probabilistically (Zwicky, 1972). Effects of this sort, whether traditionally viewed as phonological rules or as phenomena of casual or sloppy speech, seem to require rules with many similar properties. They are described in our fourth hypothesis presented in the next section.

4.4. *Phase rules and implementation rules*

We propose that there exist a set of low-level rules governing articulatory transitions between neighboring segments which we will call *phase rules*. They have the following characteristics:

(1) They govern local articulatory gestures including their timing. They are perhaps confined within the domain of a one or two syllabic cycles. This would account for the tendency for syllable durations to be unaffected by them and for syllable structure to play a prominent role in their contexts (Kahn, 1976).

(2) They are controlled partially by phonological contextual features. This is why statements in terms of phonological features do such an effective job of describing where they occur.

(3) They are learned. Hence, they are language specific and variable in the details of articulatory output from speaker to speaker.

(4) They are very sensitive to pragmatic communicative needs that change rapidly from moment to moment during speech. Hence, speakers may adjust the degree of application from word to word, thus accounting for the fairly continuous control over

parameters like voice-onset time, flapping and German voicing neutralization (Fourakis & Iverson, 1984).

We think of the input to these rules as a sequence of complex phonological symbols, that is, of either distinctive feature bundles (in the sense of Chomsky & Halle, 1968) or multidimensional structures (viz. Halle & Vergnaux, 1980). The output of a phase rule seems to be either a specification for a gesture in time or perhaps the actual gesture itself (Fowler, 1980; Kelso & Tuller, 1983). Thus phase rules appear very similar to (and possible indistinguishable from) temporal implementation rules (Klatt, 1976; Port, 1981) since both rule types specify which articulatory gesture to make and its timing. Clearly any model of speech production that takes linguistic structures as input must have implementation rules of some sort to mediate between linguistic segments and gestures. What is new in our proposal is the notion that such rules may also perform many of the tasks previously assigned to segmental phonological rules, such as the stop epenthesis discussed here, aspiration of English voiceless stops, German syllable-final devoicing and so forth. We suspect that those phonological rules that turn out to be phase rules will coincide rather closely with the ones called "post-lexical" by Kiparsky (1982) or called "P-rules" by some natural phonologists (Hooper, 1974).

Such rules could be responsible for the timing patterns that have been observed in the flapping environment of some American English dialects. The longer vowel durations before /d/-derived flaps than before /t/-derived flaps could be the result of a phase rule that adjusts vowel and closure durations of the underlying /t/ and /d/ in a way that happens to leave an acoustic trace of the underlying difference. Similarly, for German, it could be proposed that the devoicing of final voiced obstruents is achieved by a phase rule that has an output similar to that of the underlying voiceless obstruents.

In the case of the results in the present experiment, we might propose that the implementation of S–F clusters in syllable codas of American speech reflects a phase rule that controls the timing of such clusters in such a way that they become nearly neutralized with S–T–F clusters. This phase rule happens to be highly invariant across the American community. Other rules are more variable (such as vowel reduction/deletion or flapping) and are fairly sensitive to pragmatic factors in their degree of application.

The differing effects of phase rules are sometimes highly perceptible but sometimes not. It is easy to tell a flap from a [t] but frequently very hard to tell a *d*-flap from a *t*-flap (at least in many mid-western dialects). But this becomes a theoretical problem only if one assumes that production-based grammars and perception-based grammars must be identical. However, as Dinnsen (1985) has argued, there is continually mounting evidence that this is not a supportable position, In our view, the basic cause of these timing effects lies with phase rules which operate as coercive rules only in production. They are restricted to governing the transitions between articulatory gestures relative to the syllabic cycle. Each syllable may be thought of as representing a passage around the "syllable wheel" and each is modulated by various possible linguistic warpings.

5. Conclusions

An investigation of the presence and the timing associated with epenthetic stops in American English and South African English showed that only the Americans epenthesized a stop-like interval in words like *tense* and further that the insertion does not yield a stop identical to the underlying stop found in words like *tents*. This result demonstrates clearly that the analysis of this effect that claims there is a universally required epenthesis is clearly incorrect. Apparently speakers are able to learn to produce the insertion or not in a completely consistent fashion. It also suggests, however, that an interpretation that relies on a rule of segmental insertion at the phonological level in implausible since it implies that the inserted unit is a simple phonological segment. Yet our evidence shows that the underlying and inserted segments are articulatorily distinct—even if there is only a very small and nearly imperceptible difference.

Although a variety of models could be invoked to account for this difference, all of them seem to require radical revision of the relationship between phonetics and phonology. In particular, it appears that this "phonological rule", like many others, has many of the properties usually associated with "implementation rules"—such as the fact that rules make reference to phonological properties in specification of their environment. Yet other properties of these rules seem more non-linguistic; they control graded parameters of articulation and exhibit differences in production that are probably only marginally perceptible. We propose that there are a large number of rules that have these properties and we suggest the term phase rule to describe them.

We would like to express our thanks to Daniel A. Dinnsen, and Osamu Fujimura for comments on various drafts of this manuscript. This research was supported by the National Institutes of Health, Grant HD-12511. Some of this material was presented to the Acoustical Society of America Annual Meeting in 1979 in Cambridge, Massachusetts, U.S.A.

References

Ali, L. H. (1979) Observations on English nasal fricatives. *Speech communication papers presented at the 97th meeting of the Acoustical Society of America* (J. J. Wolf & D. H. Klatt, editors), pp. 119–122. New York: Acoustical Society of America.

Ali, L. H. Daniloff, R. G. & Hammarberg, R. (1979) Intrusive stops in nasal-fricative clusters: An aerodynamic and acoustic investigation, *Phonetica*, **36**, 85–97.

Chen, M. (1970) Vowel length variation as a function of the voicing of the consonant environment, *Phonetica*, **22**, 129–159.

Chomsky, N. & Halle, M. (1968) *The sound pattern of English*. New York: Harper and Row.

Dalby, J. (1984) *The phonetic structure of fast speech*. Doctoral dissertation, Department of Linguistics, Indiana University, U.S.A.

Dinnsen, D. A. (1980) Phonological rules and phonetic explanation, *Journal of Linguistics*, **16**, 171–191.

Dinnsen, D. A. (1985). *Journal of Linguistics*, **22**.

Dinnsen, D. A. & Charles-Luce, J. (1984). Phonological neutralization, phonetic implementation and individual differences, *Journal of Phonetics*, **12**, 49–60.

Donegan, P. J. & Stampe, D. (1979) The study of natural phonology. In *Current Approaches to Phonological Theory* (D. Dinnsen, editor), pp. 126–173. Bloomington: Indiana University Press.

Fourakis, M. & Iverson, G. K. (1984) On the 'incomplete neutralization' of German final obstruents, *Phonetica*, **41**, 140–149.

Fowler, C. (1980) Coarticulation and theories of extrinsic timing control. *Journal of Phonetics*, **8**, 113–133.

Fox, R. A. & Terbeek, D. (1977) Dental flaps, vowel duration and rule ordering in American English, *Journal of Phonetics*, **5**, 27–34.

Fujimura, O. & Lovins, J. (1978) Syllables as concatenative phonetic units. In *Syllables and segments* (A. Bell & J. B. Hooper, editors), pp. 107–120. New York: North-Holland Publishing Company.

Halle, M. & Vergnaud, J.-R. (1980) Three dimensional phonology, *Journal of Linguistic Research*, **1**, 83–105.

Harms, R. T. (1973) *Some nonrules of English*. Bloomington: Indiana University Linguistics Club.

Hayes, B. (1981) *A metrical theory of stress rules*. Bloomington, Indiana; Indiana University Linguistic Club.

Hooper, J. (1974) *An introduction to natural generative phonology*. New York: Academic Press.

Huff, C. (1980) Voicing and flap neutralization in New York City English, In *Research in phonetics, report No. 1*, pp. 233–256. Bloomington: Indiana University Department of Linguistics.

Javkin, H. (1979) Phonetic universals and phonetic change. In *Report of the phonology laboratory no. 4*. Berkeley: Department of Linguistics, University of California.

Jones, D. A. (1966) *The pronunciation of English*. Cambridge: Cambridge University Press.

Kahn, D. (1976) *Syllable-based generalizations in English phonology*. New York: Garland Publishing Co.
Kelso, J. A. & Tuller, B. (1984) A dynamical basis for action systems. In *Handbook of cognitive neuroscience* (M. S. Gazzaniga, editor).
Kiparsky, P. (1982) From cyclic phonology to lexical phonology, In *The structure of phonological representations* (H. van der Hulst & N. Smith, editors). Dordrecht Holland: Foris Publications.
Klatt, D. H. (1973) Interaction between two factors that influence vowel duration, *Journal of the Acoustical Society of America*, **54**, 1102–1104.
Klatt, D. H. (1976) The linguistic uses of segmental duration in English, *Journal of the Acoustical Society of America*, **59**, 1208–1221.
Labov, W. (1981) Resolving the neogrammarian controversy, *Language*, **57**, 267–308.
Labov, W., Yaeger, M. & Steiner, R. (1972) *A Quantitative Study of Sound Change in Progress*. Philadelphia: Report to the National Science Foundation.
Lehiste, I. (1972) Manner of articulation, parallel processing, and the perception of duration, *Ohio State Working Papers in Linguistics*,**12**, 33–52.
Lovins, J. B. (1978) A study of "nasal reduction" in English syllable codas (unpublished). Murray Hill, New Jersey: Bell Laboratories.
Nie, N. H., Hadlai Hull, C., Jenkins, J. G., Steinbrenner, K. & Bent, D. H. (1975) *Statistical package for the social sciences*. New York: McGraw-Hill.
Ohala, J. (1974) Experimental historical phonology. In *Historical linguistics II. Theory and description in phonology* (J. M Anderson & C. Jones, editors), pp. 353–389. Amsterdam: North-Holland Publishing Co.
Peterson, G. & Lehiste, I. (1960) Duration of syllabic nuclei in English, *Journal of the Acoustical Society of America*, **32**, 693–703.
Port, R. (1981) Linguistic timing factors in combination, *Journal of the Acoustical Society of America*, **69**, 262-274.
Port, R. & O'Dell, M. L. (1985) Neutralization of syllable final voicing in German, *Journal of Phonetics*, **13**, 455–471.
Slowiaczek, L. M. & Dinnsen, D. A. (1985) On the neutralizing status of Polish word-final devoicing, *Journal of Phonetics*, **13**, 325–341.
Raphael, L. J., Dorman, M. F., Freeman, F. & Tobin, C. (1975) Vowel and nasal duration as cues to voicing in word final stop consonants: spectrographic and perceptual studies, *Journal of Speech and Hearing Research*, **18**, 389–400.
Zue, V. & Laferriere, M. (1979) Acoustic study of medial /t,d/ in American English *Journal of the Acoustical Society of America*, **66**, 1039–1050.
Zwicky, A. M. (1972) Note on a phonological hierarchy in English. In *Linguistic change and generative theory* (R. P. Stockwell & R. K. S. Macauley) 275–301. Bloomington: Indiana University Press.

Volume 14 Number 2 April 1986

journal of phonetics

ACADEMIC PRESS
Harcourt Brace Jovanovich, Publishers
London Orlando San Diego New York
Austin Boston Sydney Tokyo Toronto

JOURNAL OF PHONETICS

Volume 14 (4 issues) January–October 1986. Published quarterly, Inland £73.00 including postage; abroad $151.50 including postage. Subscription orders should be sent to Academic Press, High Street, Foots Cray, Sidcup, Kent DA14 5HP, U.K. *In 1986 only, back volumes can be purchased at the special rate of £36.50 (U.K. only), $75.75 (abroad).* Reduced rate for personal subscribers £36.50 (U.K.), $75.75 (abroad). Send notices of change of address to the office of the Publishers at least 6–8 weeks in advance. Please include both old and new addresses.

Second class postage is paid at Jamaica, New York 11431, U.S.A. Air freight and mailing in the U.S.A. by Publications Expediting Inc., 200 Meacham Avenue, Elmont, New York, 11003.

U.S. Postmaster: Send address changes to *Journal of Phonetics*, c/o Publications Expediting Inc., 200 Meacham Avenue, Elmont, NY 11003, U.S.A.

ISSN 0095–4470