THE LIMITS OF SEGMENTAL DESCRIPTION

FRANCIS NOLAN

Department of Linguistics
Cambridge University
Cambridge CB3 9DA, UK

ABSTRACT

Evidence is discussed which perturbs the segmental, categorial foundation of descriptive phonetics. EPG studies showed that in cases which would be treated in auditory phonetic analysis and in phonological description as place assimilation, there is often a residual gesture towards the 'underlying' segment. Such results underline that the performance of segmental contrasts is neither discretely segmental in time, nor categorial in the sense of involving an inventory of discretely different elements. Segmentalised phonetic description is further challenged by instrumental evidence that neutralisation may be phonetically incomplete; and that segmental contrast may be cued over domains as large as the stress foot. Phonetics needs a more explicit statement of the relation of segments to articulation and to perception.

INTRODUCTION

Throughout the history of modern phonetics the phone-sized segment has been crucial. True, other elements such as syllables and features have periodically competed for attention; but the centrality of the phone is such that even these alternative elements are often talked of as combinations of, or decompositions of, phones.

The phone-sized segment remains at the heart of phonetic description despite decades of instrumental research into articulation and acoustics demonstrating beyond doubt that discrete phones do not exist in a straightforward way in the speech event, at least as superficially observable. An x-ray film of speech, or a dynamic palatographic record, shows gestures for different segments overlapping and blending. And from the earliest speech synthesis it has been known, for instance, that the perceptual cues to a consonant are distributed at least over the adjacent vowels.

But the survival of the phone is not hard to explain. It is the basis of our only extensive model of general phonetic description, as embodied for instance in the alphabet of the International Phonetic Association. This in turn reflects the fact that phone- or phoneme-sized units provide the most generally applicable and revealing descriptions of the phonologies of languages.

Thus the phonetic sciences have proceeded in a somewhat schizophrenic state of mind, knowing that phones aren't really there, but at the same time they have to be there. The hope is generally that at some stage the relationship between segmentalised descriptions and the continuum of speech performance will become clear and well specified.

This paper draws together a number of cases where it seems that the tension between the discretely segmental description and the observable speech event is high enough to make the resolution of their relationship a priority.

DISCRETENESS ON TWO AXES

The traditional phone-based model of phonetic description implies discreteness on two axes.

Firstly, the phone symbols from left to right in a transcription imply a temporal sequence of discrete phonetic events. The strongest interpretation of this, with for instance all acoustic cues to a segment ending simultaneously and abruptly at a boundary with a following segment, is clearly falsified even by casual observation of spectrograms. Perhaps the weakest interpretation is one which allows overlapping in the realisation of phones, but still expects their implied sequencing to be respected in that the realisation of phone n will not extend later in time than that of phone n+1 nor earlier than that of phone n-1 (see Fig.1). For instance, if in

the utterances [ski] and [sku] the friction of the [s] contains from its onset cues to the velar, and to the tongue+lip configuration of the vowel, the implied sequencing has been respected. If however the velar or is not cued until late in the friction, whilst the vowel configuration is cued from the start, the implied sequencing has been violated (because cues to segment 3 begin before cues to segment 2). The issue of 'proper sequencing' has probably not been addressed in quite this form in existing work on coarticulation.

Secondly, the phone symbols imply selection of phonetic events from a paradigm of discrete phonetic units. Thus [d] is either present in an utterance or it is not, and if it is, it is whole-heartedly [d] and not something which vacillates between being a [d] and being a [b]. Speech performance is thus implied to be categorical.

CONNECTED SPEECH PROCESSES

It is reasonable to expect that phone-segmental phonetic description would be under greatest stress with fluent connected speech. Carefully produced citation forms yield maximal phonological contrast, and come nearest to exhibiting a simple relationship between segmental representations and the physical event. In connected speech the explicitness of the realisation of phonological contrasts may be reduced in a variety of ways, including assimilation and deletion. These various reductions in explicitness have been termed connected speech processes (e.g. Barry 1985).

Linguistic phonetics has, perforce, described connected speech processes (CSPs) in phone-segmental terms: a phone is deleted, or changes into another phone (which, in the case of assimilation, more closely matches an adjacent phone in one or more phonetic dimension). It is unclear, however, whether the categorical change implied by this type of description reflects the facts of speech performance (production, or perception) since, as pointed out above, a phone-segmental representation inherently implies discreteness sequentially and paradigmatically.

To find out if assimilation involves a discrete change in production we can compare forms eligible for assimilation with forms which underlyingly contain the segment potentially created by the assimilation. For instance, when a speaker assimilates the place of articulation of the final stop in road to the following velar in the road collapsed, is the utterance then phonetically identical in every respect to the realisation of the rogue collapsed?

This question has recently begun to be studied using electropalatography (EPG). For instance, Barry (1985) shows that where a word-final alveolar precedes a word-initial velar there are three possibilities (see Fig. 2). The EPG display may show complete alveolar closure (no assimilation); it may be identical to the display for a matched utterance with an underlying velar word-finally (complete assimilation); or, crucially, in many tokens it shows that no closure is completed across the alveolar ridge, but nevertheless the sides of the tongue make contact far forward along the sides of the palate in a 'residual' gesture towards the alveolar ridge (partial assimilation). The occurrence of these types is influenced, though not directly determined, by speech rate.

The existence of partially assimilated forms is supported in a similar experiment by Kerswill (1985). The gradual nature of assimilation in production is at variance with the paradigmatic discreteness of phone based representations. In principle, articulation could be categorical in that a speaker either made a gesture sufficient to create a given configuration of the vocal tract, or did not make it. Instead, articulation appears to be gradual – in that some gestures are allowed to be present, but inadequate (from the point of view of the phonetic target, and probably from that of perception – as discussed later). Note that it is not simply the case that a gesture is being curtailed by increased rate; Kerswill (1985) shows that a speaker can speak fast but with relatively few reductions when asked to speak 'carefully'.

NEUTRALISATION

It appears that it is not only connected speech processes which put phone-segmental descriptions under strain. Recent instrumental work has suggested that in many long-accepted instances of phonological neutralisation there is, contrary to the traditional view of neutralisation, some phonetic realisation of the underlying (morpho-) phonological contrast.

Thus it has been argued that the underlying word-final voicing contrast is reflected in small, but measurable phonetic differences in languages where it is normally considered to be neutralised on the surface, such as German (Mitleb 1981, Charles-Luce 1985), Russian (Chen 1970), Polish (Slowiaczek and Dinnsen 1985), and Catalan (Dinnsen and Charles-Luce 1984). The dimensions of the realisation include the duration of the vowel preceding the stop, the duration of the stop, and the way in which these durations are affected by the class of sound at the beginning of a following word. Such

evidence is not uncontroversial (e.g. Iverson and Fourakis 1984), but may at least be seen as raising the possibility that neutralisation in these languages is incomplete, rather in the way that the assimilations discussed above may be partial.

To the extent that the residue of the underlying contrast is in the preceding vowel, sequential discreteness of segments is violated, rather in the way argued by Chomsky (e.g. 1964) to make linearity an unfeasible condition on the relationship between phonemic and phonetic representations. It has recently been suggested, however, that such violations may be more extensive than implied by Chomsky's discussion of adjacent segments. Scott (1984) claims that American English listeners can differentiate potentially neutralised pairs of the writer-rider type, and that they do so on the basis of 'cues other than preceding vowel duration or the acoustic properties of the flap'. These cues include overall durational properties of the words and global differences in phonetic quality – e.g. that rider is more 'open mouthed' in its articulation than writer. Kelly and Local (1986) suggest, too, that the spectral cues to /l/ versus /r/ in English extend over appreciably larger domains than usually considered - perhaps as extensive as the stress foot.

PRODUCTION, PERCEPTION, AND PHONETICS

Might it be the case that phenomena which hit the limits of segmental description are of no interest to phonetics because they are not perceptible, and therefore of no communicative value? On the contrary, their perceptual status forces consideration of one of the major ambiguities of phonetic analysis.

The ambiguity is whether a transcription is a record of what is said, or what is heard. As long as these coincide, the ambiguity is unobtrusive. But if, for instance, it were the case that German speakers reliably produced a measurable difference in Rad-Rat but neither native speakers nor phoneticians could perceive it, what would the correct phonetic transcription for the pair be?

The evidence as yet is inconclusive. Port and O'Dell (1985) report, for German, 59% correct identification of (incompletely) neutralised lexical items, compared with 50% as chance. Experiments are proceeding in Cambridge to test whether listeners are able to exploit perceptually the residual articulations of partial assimilations. And in a case of a phonological merger in progress, Costa and Mattingley (1981) show

that subjects exibit a residual vowel duration difference in New England cod versus card, but are unable to exploit it perceptually.

On the whole it seems probable that at least some cases will emerge where reliable production differences realising phonological contrasts are not perceived. The following table sets out some of the logical possibilities. In the three columns + or - indicates whether or not a distinction is (A) articulatorily realised by a speaker, (N) perceived by a native speaker, and (P) perceived by a well-trained phonetician in 'analytic' mode.

	A	N	P
(a)	+	+	+
(b)	+	+	-
(c)	-	+	-
(d)	+	-	+
(e)	+	-	-

(a) represents the unproblematic ideal.
(b) the native speaker and listener are coping fine with the distinction; the phonetician must try harder.
(c) the native listener hears a distinction which isn't there; this makes sense within a view such as that of Chomsky and Halle (1968) where the phonetic percept is partially determined by higher level linguistic knowledge.
(d) native users produce a distinction without reliable perception, while it may be salient enough for phoneticians to identify; some 'merger in progress' cases appear to fit here.
(e) in other cases the measured effect may be too small for the phonetician.

Hitherto it has been convenient to regard a phonetic representation as a linguistic construct, independent of articulatory and perceptual domains, but with definable (if as yet undefined) and equivalent relationships to each. The emergence of evidence of a lack of congruence between what a speaker produces and what he perceives may force a reappraisal of precisely what a phonetic representation should account for.

CONCLUSION

Phonetic description has revolved around the phone-sized segment. This construct is essentially discrete both sequentially and paradigmatically.

Sequential discreteness has long been recognised not to characterise any aspect (acoustic, articulatory) of the speech signal. The questions which seem currently worth pursuing are how extensive the influence of a segment is in time; and,

perhaps, as summarised in Fig. 1, whether even proper sequencing is preserved in the speech signal.

Do speakers behave as if segments represent categorial choices? Apparently not; in environments with the potential for place of articulation assimilation a gradation of assimilation occurs.

Categories may be a function of hearing, rather than speaking. The continuum of behaviour from no place assimilation through partial to complete assimilation may turn out to yield a categorial perceptual boundary somewhere in the 'partial' region. But on the other hand it is possible that no perceptual boundary will emerge because, as with the cod – card case, listeners can't exploit the acoustic details.

A consideration of the limits of segmental description, then, inevitably leads to consideration of the status of the categories which phone-segments imply, and of the representations which they comprise. If the disparity between production and perception which is hinted at by work cited here is confirmed, the general conception of phonetic analysis will have to be radically revised and its relation to aspects of speech performance made explicit.

REFERENCES

BARRY, M. 1985 A palatographic study of connected speech processes. Cambridge Papers in Phonetics & Exp'l Linguistics 4.
CHARLES-LUCE, J. 1985 Word-final devoicing in German: effects of phonetic and sentential contexts. J.Phon. 13, 309-24.
CHEN, M. 1970 Vowel length variation as a function of the consonantal environment. Phonetica 22, 129-59.
CHOMSKY, N. 1964 Current issues in linguistic theory. In: J.A. Fodor & J.J. Katz (eds), The Structure of Language.
COSTA, P.J. & MATTINGLY, I.G. 1981 Production and perception of phonetic contrast during phonetic change. Status Rep. on Sp. Res. SR-67/68, 191-6. Haskins Labs.
DINNSEN, D.A. & CHARLES-LUCE, J. 1984 Phonological neutralisation, phonetic implementation and individual differences. J.Phon. 12, 49-60.
FOURAKIS, M. & IVERSON, G.K. 1984 On the 'incomplete neutralization' of German final obstruents. Phonetica 41, 140-9.
KELLEY, J. & LOCAL, J.K. 1986 Long-domain resonance patterns in English. To appear in IEE Proceedings of Conference on Speech Input/Output, London.
KERSWILL, P.E. 1985 A sociophonetic study of CSPs in Cambridge English: an outline and some results. Cambridge Papers in Phonetics & Exp'l Linguistics 4.

PORT, R.F. & O'DELL, M.L. 1985 Neutralization of syllable-final voicing in German. J.Phon. 13, 455-71.
SLOWIACZEK, L.M. & DINNSEN, D.A. 1985 On the neutralizing status of Polish word-final devoicing. J.Phon. 13, 325-41.
SCOTT, D.R. 1984 More on the /t/:/d/ distinction in American alveolar flaps. JASA 75 (Suppl. 1), S66.
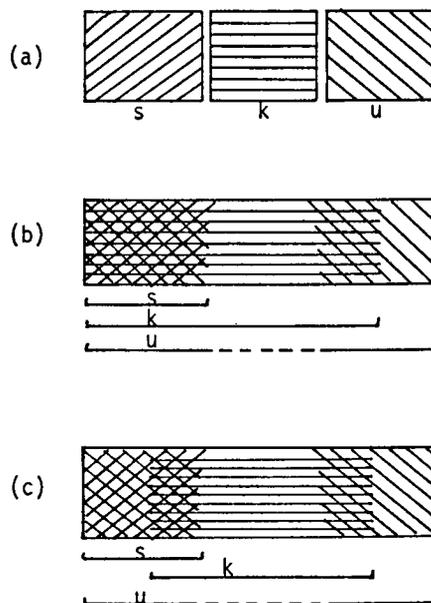
Fig 1 (a) discrete phones, as not found in the speech event; (b) implied sequence of phones in [sku] preserved; and (c) violated, since cues to the vowel precede those to the stop.
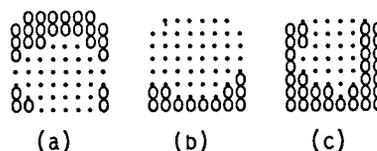

Fig 2 EPG displays taken from (a) unassimilated alveolar; (b) alveolar completely assimilated to following velar; (c) alveolar partially assimilated to following velar, showing maximum 'residual' alveolar gesture.