133
134

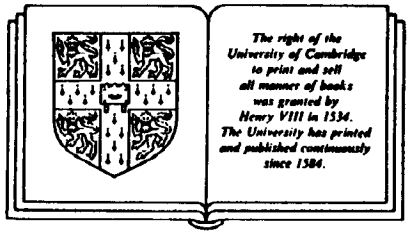
Papers in Laboratory Phonology I
Between the Grammar and Physics of Speech

EDITED BY JOHN KINGSTON

Department of Modern Languages and Linguistics, Cornell University

AND MARY E. BECKMAN

Department of Linguistics, Ohio State University



CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE

NEW YORK PORT CHESTER MELBOURNE SYDNEY

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1990

Printed in Great Britain by the
University Press, Cambridge

British Library cataloguing in publication data

Between the grammar and physics of speech. – (Papers in laboratory phonology; 1)
1. Phonology
I. Kingston, John II. Beckman, Mary E. III. Series 414

Library of Congress cataloguing in publication data

Between the grammar and physics of speech/edited by John Kingston and Mary E. Beckman.
p. cm. – (Papers in laboratory phonology: 1)
Largely papers of the First Conference in Laboratory Phonology,
held in 1988 in Columbus, Ohio.
Includes indexes.
ISBN 0-521-36238-5
Grammar, Comparative and general – Phonology – Congresses.
2. Phonetics–Congresses. I. Kingston, John (John C.)
II. Beckman, Mary E. III. Conference in Laboratory Phonology (1st
1986 Columbus, Ohio) IV. Series.
P217.847 1990
414–dc20 89-36131 CIP

ISBN 0 521 36238 5 hard covers
ISBN 0 521 36808 1 paperback

14

The phonetics and phonology of aspects of assimilation

JOHN J. OHALA

Some [scholars of language]... have allowed themselves... to be led astray by paying more attention to the symbols of sound than to sounds themselves.

KEY 1851

...on paper almost everything is possible.

OSTHOFF AND BRUGMANN 1878 [1967]

Real languages are not minimal redundancy codes invented by scholars fascinated by the powers of algebra, but social institutions serving fundamental needs of living people in a real world. [In trying to understand] how human beings communicate by means of language, it is impossible for us to discount physical considerations, [i.e.,] the facts of physics and physiology.

HALLE 1954: 79–80. [Parts of this quote have been rearranged from the original without, I think, distorting the sense.]

14.1 Introduction

Assimilations of the type given in (1) are extremely common where, when two stops of different place of articulation abut, the first (C1) assimilates totally to the second (C2).[1]

(1)	L. Latin	scriptu	>	Italian	scritto
		nocte	>		notte
	Sanskrit	bhaktum	>	Pali	bhattum
		praptum	>		pattum
		labdha	>		laddha
	Old Irish	fret− (frith−) + cor	>	freccor ~ frecur	
		*ad-gladam	>	ac(c)aldam	
		ad + bongid	>	apaig	
				(Thurneysen 1961)	

Even more common are cases where a nasal assimilates to the place of articulation of a following stop, as exemplified in (2).

(2)					
	L. Latin	primu tempus	>	French	printemps
		amita	>	Old French	ante (Mod. French tante)
	Shona	N+tuta	>	*nt^{h}uta >	nɦuta
		N+bato	>	mɓato	
				(Doke 1931)	

There is no particular difficulty in representing such variation, whether linear or nonlinear notation is used, as in (3a) and (3b) (where the assimilation of nasals to stops is given).

(3) a. [nasal] → [α place]/ ___ [stop, α place]

b. +nasal C (α place) −nasal C (β place) > +nasal C (α place) −nasal C (β place)

Nevertheless, there is something profoundly unsatisfactory in these representations. As noted by Chomsky and Halle (1968:400ff), there is nothing in them which would show that the reverse process whereby a stop C2 assimilates to the place of articulation of stop or nasal C1, i.e. as represented in (4a) and (4b), is rarely found and is quite unnatural.

(4) a. [stop] → [α place]/ [nasal, α place] ___

b. +nasal C (α place) −nasal C (β place) > +nasal C (α place) −nasal C (β place)

Chomsky and Halle's solution to this was to invoke a marking convention which when linked to an assimilation rule would provide the "natural" or unmarked feature values. However marking conventions are clearly no solution at all; they are just a patch for a defective notation system, like an added epicycle in the Ptolemaic model of planetary orbits. If marking conventions could solve all defects of notation, then there would never be any motivation to adopt newer, "improved" notations – and yet this has occurred frequently in the past 30 years: feature notation in the early 60s, articulatory features in the 60s, autosegmental

notation in the mid 70s, not to mention elaborations such as mirror image rules, alpha variables, etc. It is evident that phonologists want their notations to be revealing, to explain or render self-evident the behavior of speech sounds and to do this by incorporating some degree of isomorphism between the elements of the representation and the phenomena it stands for.

14.2 Past work relevant to assimilation

If we want to represent the processes in (1) and (2) we must first understand them. The most common explanation given for them is that they come about due to "ease of articulation," i.e. that the speaker opts for an articulation that is easier or simpler than the original (cf. Zipf 1935: 96–97). Unfortunately, the notion of "ease" or "simplicity" has never been satisfactorily defined. It is true that when a heterorganic cluster becomes a geminate (and necessarily homorganic) or when a nasal assimilates in place to the following consonant, there is one less articulator involved but it does not follow so straightforwardly that this yields an easier task. No one knows how to quantify articulatory effort but certainly the neurological control operations should be counted, too, not just the energy required to move the speech organs. For all we know it may "cost" more to have the velum execute a closing gesture in the middle of a consonantal closure (as in Shona *mɓato*, above) rather than to synchronize this gesture with the offset of one segment and onset of another. Likewise, it may very well cost more to hold a consonantal closure for an "extra" amount of time – as required in geminates – rather than to give it a more normal duration. Finally, and this is the crucial defect, the notion of ease of articulation fails to explain why, in the above cases, it is typically C1 which assimilates to C2 and not vice-versa. *A priori*, it seems more plausible that if degree of effort really mattered, C1 is the consonant that should prevail in these assimilations, i.e. after supposedly "lazy" speakers adopt a given articulatory posture, one would expect them to maintain it during C2. That the opposite happens is sufficient reason to be highly suspicious of such accounts.

There are also accounts of assimilation that do not rely on the notion of ease of articulation, e.g. Kent's (1936) "the speaker's thoughts are inevitably somewhat ahead of his actual utterance," but their relevance to cases of assimilation have not been demonstrated. Kent's basic notion is not implausible: speech errors do exhibit anticipation of sounds, e.g. *teep a cape* < *keep a tape*; but the character of such speech errors does not resemble in detail what one finds in (1) and (2). Often the anticipated sound is itself replaced by the sound it supplants and, moreover, the anticipated sound almost invariably is one occupying a similar position in another syllable, i.e. in onset, nucleus, or coda position. Also, it is universally accepted that all articulations (all voluntary movement, in fact) must be preceded by the "thoughts" that control them, but how, exactly, does "thought" get transduced into movement? It is not true that by just thinking or intending to say

an utterance we actually say it. So Kent's account is not a sufficient explanation for anticipatory assimilation.

There is reason to believe, though, that such assimilations owe very little to the speaker for their initiation (cf. Ohala 1974a, 1974b). Malécot (1958) and Wang (1959) showed that final heterorganic stop clusters created by tape splicing, e.g. /epk/ from joining /ep/ (minus the stop release) to the release of the final stop of the syllable /ek/, are identified overwhelmingly as single consonants having the place of articulation of C2. Malécot concluded:

> Voiceless p t k releases and voiced b d g releases contain sufficient cues for conveying both place and manner of articulation of American English plosives in final position. These cues are powerful enough, in most instances, to override all other place and manner cues present in the vowel-plus-closing transitions segment of those plosives... (380)

In addition, there is abundant evidence that the place of unreleased final stops – i.e. where only the stop onset cues are present – is frequently misidentified (Householder 1956), suggesting that place cues are relatively less salient in this environment. In contrast, place cues for stops in pre-vocalic position are generally very strong, so much so that just the burst is generally sufficient to cue place (Winitz, Scheib, and Reeds 1971). Place cues are even weaker in the case of nasals, which, although as a class are highly distinct from non-nasals, are often confused among themselves (House 1957; Malécot 1956); thus when joined to a following stop it is not surprising that the listener has relatively less trouble hearing the nasal consonant as such but takes the place cue from the more salient stop release.

The relative perceptual value of VC vs. CV transitions for intervocalic stops have been investigated in numerous studies (Repp 1976, 1977a, 1977b, 1978; Fujimura, Macchi, and Streeter 1978; Dorman, Raphael, and Liberman 1979; Streeter and Nigro 1979). The results of these studies show consistently (and, in the case of Fujimura et al., cross-linguistically) that when spliced VC and CV transitions differ, e.g. /eb/ spliced onto /de/ with a gap equal to that typical of a single stop, listeners generally "hear" only the consonant cued by the CV transitions, that is /eb + de/ is heard as /ede/.

The cause of this effect is still debated. One view is that recent cues dominate over earlier cues. Another view (e.g. Malécot's, quoted above) is that the VC and CV transitions have inherently different quality of place cues: especially in the case of stops, the VC cues reside almost totally in the formant transitions whereas the CV cues include transitions and the stop burst, the latter of which has been demonstrated to carry more reliable information for place than transitions alone (Schouten and Pols 1983). Against this view, however, is the evidence of Fujimura et al. that when the mismatched VC-CV utterances are played backwards, it is still the CV cues, that is, those originally from the VC portion, that dominate the

percept. Moreover, there were slight but significant differences in the reaction of native speakers of English and of Japanese in their reaction to these stimuli which were attributed to the differing syllable structures of the two languages. These arguments support the notion that in addition to any physical differences between VC and CV cues, listeners' experience, including their native language background, dictates which cues they pay most attention to. These three hypotheses do not conflict; they could all be right.

In order to explore further the influence of the listeners' prior experience on the interpretation of mismatched VC1 + C2V utterances, the following two studies were done. These experiments were done as class projects by students in one of my graduate seminars.[2]

14.3 Experiment 1

14.3.1 *Method*

An adult male native speaker of American English (from Southern California) recorded VCV and VC1C2V utterances where V was always [a] and the single C was any of the six stops of English, /p, t, k, b, d, g/, and the clusters consisted of the same six stops in C2 position and a homorganic nasal in C1 position, e.g. apa, ata, aba, anta, amba, etc. In addition, stress was placed in one reading on the initial vowel and, in a second reading, on the final vowel. The list is given in (5).

(5)

'apa	'aba	'ampa	'amba
'ata	'ada	'anta	'anda
'aka	'aga	'aŋka	'aŋga
a'pa	a'ba	am'pa	am'ba
a'ta	a'da	an'ta	an'da
a'ka	a'ga	aŋ'ka	aŋ'ga

The recording was done in a sound-treated room using high-quality recording equipment. These utterances were filtered at 5 kHz and digitized at a 10 kHz sampling rate and subjected to splicing such that within each of the eight groups in (5) the initial VC was spliced onto the final CV. Where there is only one intervocalic C, VC is the interval from initial vowel onset to the middle of the consonant closure and CV is the remainder. Thus the first group yielded the spliced utterances in (6).

(6)

'ap-pa	'ap-ta	'ap-ka
'at-pa	'at-ta	'at-ka
'ak-pa	'ak-ta	'ak-ka

In the case of the homorganic nasal + stop clusters, the cut was also made at the middle of the nasal + stop closure. Thus 72 stimuli were created.

These spliced tokens were copied, randomized, onto an audio tape with a 250 ms 1000 Hz sine wave "warning signal" 500 ms. before each stimulus and 4·6 seconds between each stimulus (with extra time between every sixth stimulus). Six tokens randomly chosen were also prefixed to the 72 test stimuli to serve as examples which would familiarize listeners to the type of stimuli they would be hearing. The entire tape, examples plus test stimuli, lasted about 9 minutes.

Twelve volunteer listeners, native speakers of American English, were recruited from an elementary linguistics course at University of California, Berkeley, and were presented with an answer sheet and instructions which indicated that we were evaluating the intelligibility of synthetic speech and wanted them to identify some isolated utterances. The answer sheet gave three choices for each stimulus, one which represented the medial consonant or sequence as having the place of C1, another C2, and a third as "other." For example, for the stimulus /ap-ka/ the possible answers were "apa, aka, other;" for the stimulus /an-pa/ the choices were "anta, ampa, other," etc.

14.3.2 *Results*

Of the 576 responses (those where the place of articulation of C1 ≠ C2), 93% were such as to show that the place of articulation of C2 dominated the percept. Differences in stress placement did not significantly influence the proportion of C2 responses nor was there any significant difference between stop + stop clusters and nasal + stop clusters. However, there was a significantly lower proportion of C2 responses in the case of voiced clusters (89%) vis-à-vis voiceless clusters (97%).

14.4 Experiment 2

14.4.1 *Method*

For experiment 2 the same VC1 + C2V utterances were used except that tokens where C1 = C2 were eliminated as were all tokens where C1 = nasal. The rest were modified by incrementing and decrementing the closure interval in 10 ms. chunks: 120 to 170 ms. for the voiceless clusters and 70 to 140 ms. for the voiced clusters. In the voiced series some cuts did not coincide with zero crossings but since the amplitude of the signal was so low this did not introduce any noticeable discontinuities into the stimuli. These tokens were randomized and presented to 18 listeners (young adult native speakers of English) in a way similar to that of experiment 1, except that the answer sheet gave the options of VC1V, VC2V, and VC1C2V.

14.4.2 *Results*

Figure 14.1 presents graphically the pattern of listeners' responses in terms of the percentage of -CC- judgements (vertical axis) as a function of the closure duration (horizontal axis); the solid line gives the results for the voiced series and the

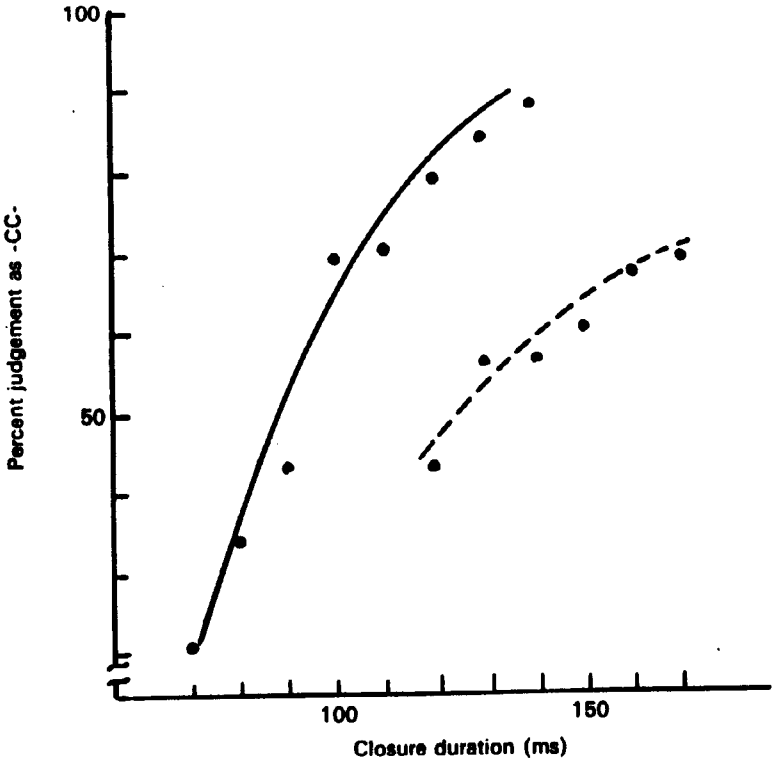


Figure 14.1 Listeners' responses to stimuli in experiment 2 (see text). Ordinate: percent of 18 listeners' pooled responses identifying the stimulus as –C1C2–; abscissa: duration of stop closure (ms). Solid line: voiced C's; dashed line: voiceless C's (curves fitted by eye).

dashed line, the voiceless series. These data confirm previously published results: when the gap between the VC and CV transitions is short, listeners report only a single consonant, whereas when the interval is longer, more clusters are reported. However, the results extend earlier results by showing that the identification functions are different for the voiced and voiceless clusters: voiced stops were heard as clusters at shorter durations (down to 95 ms.) than were voiceless stops (only down to 150 ms.). One may speculate plausibly that this difference is a reflection of the intrinsic difference between the durations of voiced and voiceless intervocalic consonants: single voiced stops are typically shorter than single voiceless stops (see Westbury 1979: 98 for a review),[3] in comparison with voiceless stop closures, listeners therefore must hear a shorter voiced stop closure before they relinquish the -CC- percept and embrace -C-.

These results suggest that it is experience with the natural structure of speech which guides the listener in evaluating and integrating the cues present in speech:

since there is a richer, more reliable set of place cues in the CV transition than the VC transition, listeners weight the former more heavily than the latter in deciding what they've heard. Since intervocalic clusters have a longer closure than single consonants, a longer closure duration is necessary for listeners to hear clusters. Since voiceless stops have longer closure durations than voiced stops, a shorter closure duration is necessary for the latter to be heard. This interpretation is compatible with all the experimental results reviewed earlier and supports the interpretation that the lack of salience of the VC transitions is mediated by linguistic experience and is not – or is not simply – a general psychological constraint. This interpretation also suggests that the sound changes shown above in (1) and (2) could have occurred due to less experienced listeners lacking the perceptual ability to integrate the weaker place cues in the VC transitions.

14.5 Discussion

These results have implications for a number of points.

14.5.1 *A general theory of assimilation*

The first experiment reported here and the earlier ones reviewed have, in effect, reproduced in the laboratory one aspect of the sound changes responsible for the data in (1) and (2) and can, therefore, be said to have (partially) explained it (see also Ohala 1987). They have shown that its source is to be found in the acoustic–auditory domain, not in the articulatory. The source of variation–change in pronunciation–happened not in the mouth of the speaker who uttered the test tokens in the experiment but in the ears of the listeners. Of course, the factors which give rise to the acoustic–auditory factors behind this asymmetry in direction of assimilation are ultimately articulatory and we can speculate fairly confidently about their nature. During a stop closure there is, by its very nature, a continuous increase in the air pressure behind the constriction. Thus at implosion there is low pressure but at release there is high pressure and consequently a high rate of airflow past the constriction. This high airflow creates audible turbulence, the burst. The burst has been shown to be a more reliable and robust cue to place of articulation (since the spectrum of the noise generated is largely determined by the resonating cavity forward of the constriction) than the formant transitions which occur as the articulators move towards or away from a constriction, where the pattern is determined as much by the cavity in back of the constriction as in front and is thus subject to more overlaid and thus obscuring influences (Öhman 1966; see also Ohala and Kawasaki 1984, and references therein).

In the case of nasal + stop clusters we again have the highly reliable stop burst plus formant transitions cuing place vs. the less reliable place cues in the nasal and its formant transitions thus leading to the stop dominating the cues for place.

I do not claim that all assimilations would be subject to the same principles. In fact I believe that many of the phenomena labeled "assimilation" are likely to be governed by very different principles (see note 1). Although I am not prepared to defend these beliefs here I think that voicing assimilation will exhibit very different tendencies (it would probably have a greater incidence of perseverative assimilation of voicelessness), as will assimilation that, unlike those in (1) and (2) above, involves only one articulator, e.g. velum, tongue tip, lips.

14.5.2 *Sound change in general*

These results add to the growing body of evidence pointing to the crucial role of the listener in initiating certain sound changes (Jonasson 1971; Ohala 1981, 1985a). This is not to deny that much of the synchronic variation in speech – from which diachronic variation arises – can be traced to the speaker or the physical principles which map articulation to sound (see e.g. Weymouth 1856; Ohala 1979, 1983a; Goldstein 1983); nevertheless, the role of the speaker has been greatly overemphasized in previous speculation on this point.

Furthermore, these results reinforce a non-teleological view of sound change, that is, that neither speaker nor hearer chooses – consciously or not – to change pronunciation (Ohala 1975, 1985a; Hombert, Ohala, and Ewan 1979). Rather, variation occurs due to "innocent" misapprehensions about the interpretation of the speech signal or, as suggested above, due to listeners' inexperience. In this respect sound change is not unlike the transmission of scribal errors in the copying of manuscripts. It does not occur to "optimize" speech in any way: it does not make it easier to pronounce, easier to detect, or easier to learn. I acknowledge that this is a complex issue and that anyone holding an opposite view would not likely be convinced by these few remarks. Perhaps some of the works cited in the references at the end of this paper would serve that purpose.

14.5.3 *Representation of sound patterns in a way that facilitates their explanation*

How would one go about representing these factors in a way which would make them self-evident, that is, to fall out naturally from the representation and not to require propping up by external declarations like markedness (or other) conventions? The answer, I maintain, is models which incorporate known aerodynamic principles (Rothenberg 1968; Stevens 1971; Müller and Brown 1980; Ohala 1975a, 1976, 1983b; Keating 1984), known principles relating vocal tract shape and acoustic output (Fant 1960; Stevens 1972), and some of the principles (to the extent that we known them) of how our auditory system extracts information from the acoustic signal (e.g. see Bladon and Lindblom 1981; Bladon 1986; Lindblom 1986). I will refer to these as "phonetic" models or representations.[4] There has even been considerable success in developing models

which incorporate two or more of these links in the speech chain (Flanagan, Ishizaka, and Shipley 1975; Lindblom 1986). It is possible to "wind up" such models and make them "go" and see natural speech sound behaviour happen.

For the sake of explaining natural sound patterns there are advantages to representations using phonetic primitives – advantages not found in other currently popular phonological representations. One advantage is that *a few well-chosen primitives go a long way*. The basic anatomical, aerodynamic, acoustic, and perceptual constraints of speech can be and have been invoked to explain many quite specific forms of speech sound behavior. Table 14.1 lists just a few of these. This partial list constitutes by itself a "critical mass" of successful explanatory studies of sound patterns (even allowing that some may require revision) such that it would deserve serious attention. That they are all based on the same minimum machinery and are experimentally supported makes them doubly worthy of attention. In contrast, with currently fashionable phonological representations, each new fact considered as often as not requires *ad hoc* patches to the existing framework, e.g. "geminate integrity," "inalterability," "the obligatory contour principle," "the shared feature convention" (see Kingston, this volume). These patches must be established by decree; they do not "fall out" from the primitives assumed for the basic nonlinear mechanisms. How impressive would it have been if Newton, after presenting his theory which unifies free fall of terrestrial objects and planetary orbits, had added "and, oh yes, on top of this we also have to recognize that tides exist and that they are related somehow to the relative positions of the sun and moon"?

A second advantage – and this follows from the first, just mentioned – is that *none of the terms of the explanation are unfamiliar, other-worldly entities*. If Boyle-Mariotte's Law has to be invoked to explain the devoicing of stops, it is the same Boyle-Mariotte's Law that applies in other parts of the familiar universe we live in (bicycle pumps, automobile pistons, party balloons, barometers, etc.). If the principle of "camouflage" is invoked in the auditory domain to explain dissimilation (Ohala 1981), it is the same camouflage that applies in the visual domain. In contrast, currently popular phonological representations "explain" sound patterns by conjuring up a vast array of devices and conventions that seem to apply exclusively to speech.[5]

In sum, phonetic accounts of natural sound patterns adhere to the constraint of Occam's razor, expressed by Newton as "more is in vain when less will serve."

Physical and physiological representations may represent unfamiliar territory for most phonologists. Nevertheless, the one who asks the question presumably is responsible for providing the answer. It cannot be the case that inferior answers to questions are accorded any status in science because the asker shows no interest or ability in the domain where the answers lie. The answer to questions such as why C2 is favored in C1C2 assimilations of stops is to be found in these phonetic domains – not in spider-web networks of phonetic labels such as one finds in

Table 14.1

1. Treatments of sound patterns due to production constraints:
   a. Devoicing of obstruents (Ohala 1983b), especially those with back articulations (Javkin 1977) and especially those with long closure duration;
   b. (Af)frication of stops before high close vowels and glides (Ohala 1983b);
   c. Devoicing and/or frication of high close vowels or glides (Ohala 1983b);
   d. Nasalization inhibits devoicing and/or frication of vowels and glides (Ohala 1978a, 1983b);
   e. Segments which do and do not block spreading nasalization (Ohala 1975b, 1983b);
   f. Stop epenthesis (stop preservation) in nasal+oral sequences, [s]+[l] (and vice versa) (Weymouth 1856; Phelps 1937; Ohala 1974b).

2. Treatments of sound patterns by reference to acoustic auditory constraints:
   a. Labiovelars behave like labials when interacting with fricatives but behave like velars when interacting with nasals and coloring the quality of adjacent vowels (Ohala and Lorentz 1977; Ohala 1979);
   b. Palatalized labials and velars change to apicals (Ohala 1978b, 1983a, 1985a);
   c. Labialized apicals and velars change to labials (Durand 1955);
   d. Cross-language prohibitions against labials+w, apicals+l, and apicals and palatals+j (Kawasaki 1982; Ohala and Kawasaki 1984);
   e. The above changes (and many others) are asymmetrical in their directionality (Ohala 1983a, 1985a).
   f. An account of dissimilation which explains:
      i. what features do not assimilate.
      ii. why dissimilation tends not to introduce new segments to a language whereas assimilation may do so,
      iii. why dissimilation requires the conditioning environment to remain in the process of the change, whereas assimilation does not (Ohala 1981, 1985a)
   g. Why nasalization tends not to be distinctive near nasal consonants (Kawasaki 1986);
   h. Spontaneous nasalization (Ohala 1983a);
   i. How nasalization affects vowel quality (Wright 1986; Beddor, Krakow, and Goldstein 1986).

autosegmental notation. Drawing a line between two primitive entities, however valid they might be as primitives, e.g. "obstruent," "labial," and "oral," or simply grouping them together inside square brackets, will not show how this combination will create an increase in oral pressure which, when released, gives rise to a rich set of place cues. Somewhere in the representation there will have to appear equations such as that in (7),[6]

(7) air pressure = air mass × (1/volume) × constant,

reflecting that the more air one stuffs into a cavity, the more pressure rises (see

Table 14.2 *Comparison of physical phonetic vs. autosegmental representations of sound patterns*

Goal	Physical Phonetic	Autosegmental	Comments
Explaining 'naturalness'	Excellent	Poor	These two are related
Reflecting history of languages	Excellent for presumed phonetic initiation; not for subsequent transmission	Undemonstrated	
Taxonomic descriptive	Poor to Fair; too cumbersome	Possibly very good; especially when a 'melody' needs to be abstracted from words manifesting it	There is no measure of success in description
Reflecting psychological structure	Makes no claim to this (but see Ohala 1986)	Largely unproven except for certain, speech error and word game data	This has nothing to do with naturalness
Representing phonology in a pedagogically effective way	Makes no claim to this	Makes no claim to this	
Representing phonology in a computationally efficient way (e.g. for word parsing)	Makes no claim to this	Makes no claim to this	

Ohala 1976, 1983b). Granted, it would be possible in principle to represent this fact through an *ad hoc* rule such as (8).

(8) $\emptyset \rightarrow$ burst / $\left[\begin{matrix}\text{obstruent} \\ \text{etc.}\end{matrix}\right]$ ___

But, as discussed above, we would soon find ourselves having to add more such rules – "epicycles" – each time a new consequence of the interactions of parameters was discovered.

I do not regard autosegmental notation as useless for all purposes nor do I think phonetic representations are suited for all tasks which are legitimate concerns of the phonologist. It may be helpful if I present this evaluation in the form of the scorecard given in table 14.2.

I presume there is no need to elaborate further on the reasons why I think that in comparison with phonetic representations autosegmental notation fails to explain the naturalness of common sound patterns. A closely related task, of course, is to give an account of the history of phonological events which gave rise to current sound patterns. Naturally, phonetic representations would excel at representing that stage at which phonetic factors played a role: the initiation of sound change; they would probably be of relatively little value in explaining the transmission of sound change.

Nevertheless, there is a sense – perhaps an unintended sense – in which traditional, supposedly synchronic, phonological accounts are actually good descriptions (but still not explanations) of the historical sequences leading up to various sound patterns. This is so because in spite of modern phonology's goal of discovering the "mental processes" underlying language (Chomsky and Halle 1968: viii), the chief method used is still that of internal reconstruction, which is appropriate for historical but not psychological study.

What autosegmental notation seems to be good at and for which physical phonetic representation is less suited is *describing* sound patterns and therefore *classifying* them, especially when certain phonological "melodies" need to be abstracted out of the forms which manifest them, e.g. tone and other prosodic properties of words such as quantity, Semitic-like separations of vowels and consonants in inflectional paradigms. These are the kind of phenomena autosegmental notation was originally developed to serve (Mattingly 1966; Goldsmith 1976). (I am not convinced that other "prosodies" such as vowel harmony or spreading nasalization, benefit from an autosegmental description – as opposed to, say standard linear notation. And one of the more popular uses of autosegmental notation – that of being able to posit abstract segments without having to commit oneself to an arbitrary declaration about their phonetic character – strikes me as a clever solution to a pseudo-problem, or more accurately, a problem necessitated by arbitrary constraints on the form of linguistic description.[7]) But if nonlinear representation is useful as a descriptive device – that is, simply as a notation – it must be remembered that there are no absolute criteria for evaluating descriptions. Different modes of description may be suitable for different purposes. A hawk may be variously described as a vertebrate, a warm-blooded creature, a predator, a creature found at the top of the food chain, one which has altricial (as opposed to precocious) young, etc. These descriptions are not mutually exclusive; they are not right or wrong; all have some usefulness; it makes little sense to try to "prove" that a hawk is a "predator." These are informal descriptors. We get a poor return (in terms of greater understanding) from any efforts to make them formal.

As for representing the psychological structure of speech, physical phonetic representation makes no claim to this and the value of autosegmental notation for this goal has scarcely begun to be seriously investigated. The mere possibility of representing speech in a certain way – even in an allegedly economical way (but who is keeping the accounts?) – is insufficient evidence of the psychological reality of that representation.

As for the other goals listed in table 14.2, neither representation has made any claim to meeting them. I list them simply to make the point that there are many jobs the phonologist takes on and a given representation may not be suitable for all of them. Ultimately we seek the best tool for each job. It is no disparagement of a given tool to point out that it is unsuitable for a certain purpose. Equally,

however, there is no advantage to "selling" a tool for a function to which it is ill-adapted.

Notes

Pat Keating, Ilse Lehiste, Manjari Ohala, Janet Pierrehumbert, and Rich Rhodes provided helpful comments on earlier versions on this paper, for which I thank them.

1 This pattern is statistically frequent but not exceptionless; Tauli (1956) reports the following stop + stop assimilations in Estonian dialects:

pikk (cf. Finnish pitkä)
käk'ki ~ kätki
'settember < september

but: 'retseppe ~ retsepte

where in the last case, C2 assimilates to C1.

I am not concerned in this paper with assimilations between segments differing in manner and/or voicing. Nevertheless, it is worth noting that Murray (1982) has argued convincingly that cases such as Sanskrit to Pali *patra-* > *patta-* are not cases of assimilation of C2 to C1 but rather of gemination of C1 and subsequent loss of (original) C2.

2 The first one, a replication of the kind of studies just cited, was done by all members of the seminar: Eugene Buckley, Jeff Chan, Elizabeth Chien, Clarke Cooper, David Costa, Amy Dolcourt, Hana Filip, Laura Michaelis, Richard Shapiro, and Charles Wooters and is cited here as Buckley et al. (1987). The second experiment, which explored the influence of varying time intervals between the spliced segments, was done by Jeff Chan and Amy Dolcourt and is cited here as Chan and Dolcourt (1987).

3 Curiously, the duration of inter-vocalic voiced stop *clusters* is virtually the same as that for voiceless stop clusters, about 185 ms. in post-stress environment (Westbury 1979: 75ff.).

4 Some "natural" sound patterns, e.g. dissimilation, may require for their explanation reference to phonetic *and* cognitive principles.

5 Cf. Isaac Newton's rejection, in his *Optics*, of "occult qualities" to explain natural phenomena.

6 To a physicist such equations have all the faults I have attributed to the notations in (3): they are arbitrary and do not show in a self-explanatory way why they may not be expressed differently. The physicist demands – and does possess – a yet more primitive model from which such equations may be derived. These equations nevertheless serve the phonologist as adequate primitives because they represent reliable general principles from which the target phenomena may be derived and yet which have been established independently of those phenomena.

7 The situation is not unlike that in Christian philosophy which has come up with clever explanations for the existence of evil in spite of the initial assumptions that God was omnipotent and beneficent.

References

Beddor, P. S., R. A. Krakow, and L. M. Goldstein. 1986. Perceptual constraints and phonological change: a study of nasal vowel height. *Phonology Yearbook* 3: 197–217.

Bladon, R. A. W. 1986. Phonetics for hearers. In G. McGregor (ed.) *Language for Hearers*. Oxford: Pergamon, 1–24.

Bladon, R. A. W. and B. Lindblom, 1981. Modeling the judgment of vowel quality differences. *Journal of the Acoustical Society of America* 69: 1414–1422.

Buckley, E. *et al.* 1987. Perception of place of articulation in fused medial consonant clusters. MS, University of California, Berkeley.

Chan, J. and A. Dolcourt. 1987. Investigation of temporal factors in the perception of intervocalic heterorganic stop clusters. MS, University of California, Berkeley.

Chomsky, N. and M. Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Doke, C. M. 1931. *A Comparative Study in Shona Phonetics*. Johannesburg: The University of the Witwatersrand Press.

Dorman, M. F., L. J. Raphael, and A. M. Liberman. 1979. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America* 65: 1518–1532.

Durand, M. 1955. Du rôle de l'auditeur dans la formation des sons du langage. *Journal de psychologie* 52: 347–355.

Fant, G. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.

Flanagan, J. L., K. Ishizaka and K. L. Shipley. 1975. Synthesis of speech from a dynamic model of the vocal cords and vocal tract. *Bell System Technical Journal* 54: 485–506.

Fujimura, O., M. J. Macchi, and L. A. Streeter. 1978. Perception of stop consonants with conflicting transitional cues: a cross-linguistic study. *Language and Speech* 21: 337–346.

Goldsmith, J. A. 1976. *Autosegmental Phonology*. Bloomington: Indiana University Linguistics Club.

Goldstein, L. 1983. Vowel shifts and articulatory-acoustic relations. In A. Cohen and M. P. R. v. d. Broecke (eds.) *Abstracts of the Tenth International Congress of Phonetic Sciences*. Dordrecht: Foris, 267–273.

Halle, M. 1954. Why and how do we study the sounds of speech? In H. J. Mueller (ed.) *Report of the 5th Annual Round Table Meeting on Linguistics and Language Teaching*. Washington DC: Georgetown University, 73–83.

Hombert, J.-M., J. J. Ohala, and W. G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55: 37–58.

Householder, F. W. 1956. Unreleased /ptk/ in American English. In M. Halle (ed.) *For Roman Jakobson*. The Hague: Mouton, 235–244.

House, A. S. 1957. Analog studies of nasal consonants. *Journal of Speech and Hearing Disorders* 22: 190–204.

Javkin, H. R. 1977. Towards a phonetic explanation for universal preferences in implosives and ejectives. *Proceedings of the Annual Meeting of the Berkeley Linguistic Society* 3: 559–565.

Jonasson, J. 1971. Perceptual similarity and articulatory re-interpretation as a source of phonological innovation. *Quarterly Progress and Status Report*, Speech Transmission Laboratory, Stockholm, 11971: 30–41.

Kawasaki, H. 1982. An acoustical basis for universal constraints on sound sequences. Ph.D. dissertation, University of California, Berkeley.

1986. Phonetic explanation for phonological universals: the case of distinctive vowel nasalization. In J. J. Ohala and J. J. Jaegar (eds.) *Experimental Phonology*. Orlando, FL: Academic Press, 81–103.

Keating, P. 1984. Aerodynamic modeling at UCLA. *UCLA Working Papers in Phonetics* 59: 18–28.

Kent, R. 1936. Assimilation and dissimilation. *Language* 12: 245–258.
Key, T. H. 1852. On vowel-assimilations, especially in relation to Professor Willis's experiment on vowel-sounds. *Transactions of the Philological Society* 5: 192–204.
Kohler, K. J. 1984. Phonetic explanation in phonology. The feature fortis/lenis. *Phonetica* 41: 150–174.
Lindblom, B. 1986. Phonetic universals in vowel systems. In J. J. Ohala and J. J. Jaeger (eds.) *Experimental Phonology*. Orlando, FL: Academic Press, 13–44.
Malécot, A. 1956. Acoustic cues for nasal consonants: an experimental study involving tape-splicing technique. *Language* 32: 274–284.
1958. The role of releases in the identification of released final stops. *Language* 34: 370–380.
Mattingly, I. G. 1966. Synthesis by rule of prosodic features. *Language and Speech* 9: 1–13.
Müller, E. M. and W. S. Brown Jr. 1980. Variations in the supraglottal air pressure waveform and their articulatory interpretations. In N. J. Lass (ed.) *Speech and Language: Advances in Basic Research and Practice*. Vol. 4. New York: Academic Press, 317–389.
Murray, R. W. 1982. Consonant cluster developments in Pali. *Folia Linguistica Historica* 3: 163–184.
Ohala, J. J. 1974a. Phonetic explanation in phonology. In A. Bruck, R. A. Fox, and M. W. LaGaly (eds.) *Papers from the Parasession on Natural Phonology*. Chicago: Chicago Linguistic Society, 251–274.
1974b. Experimental historical phonology. In J. M. Anderson and C. Jones (eds.) *Historical Linguistics II. Theory and Description in Phonology*. Amsterdam: North Holland, 353–389.
1975a. A mathematical model of speech aerodynamics. In G. Fant (ed.) *Speech Communication*. [*Proceedings of the Speech Communication Seminar, Stockholm, 1–3 Aug, 1974.*] *Vol. 2: Speech Production and Synthesis by Rule*. Stockholm: Almqvist & Wiksell, 65–72.
1975b. Phonetic explanations for nasal sound patterns. In C. A. Ferguson, L. M. Hyman, and J. J. Ohala (eds.) *Nasalfest: Papers from a Symposium on Nasals and Nasalization*. Stanford: Language Universals Project, 289–316.
1976. A model of speech aerodynamics. *Report of the Phonology Laboratory* (Berkeley) 1: 93–107.
1978a. Phonological notations as models. In W. U. Dressler and W. Meid (eds.) *Proceedings of the 12th International Congress of Linguists*. Innsbruck: Innsbrucker Beitrage zur Sprachwissenschaft, 811–816.
1978b. Southern Bantu vs. the world: the case of palatalization of labials. *Proceedings of the Annual Meeting of the Berkeley Linguistic Society* 4: 370–386.
1979. Universals of labial velars and de Saussure's chess analogy. *Proceedings of the 9th International Congress of Phonetic Sciences*. Vol. 2. Copenhagen: Institute of Phonetics, 41–47.
1980. The application of phonological universals in speech pathology. In N. J. Lass (ed.) *Speech and Language. Advances in Basic Research and Practice*. Vol. 3. New York: Academic Press, 75–97.
1981. The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, and M. F. Miller (eds.) *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society, 178–203.
1983a. The phonological end justifies any means. In S. Hattori and K. Inoue (eds.) *Proceedings of the 13th International Congress of Linguists*. Tokyo, 232–243. [Distributed by Sanseido Shoten.]

1983b. The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (ed.) *The Production of Speech*. New York: Springer Verlag, 189–216.
1985a. Linguistics and automatic speech processing. In R. De Mori and C.-Y. Suen (eds.) *New Systems and Architectures for Automatic Speech Recognition and Synthesis*. [NATO ASI Series, Series F: Computer and System Sciences, Vol. 16.] Berlin: Springer Verlag, 447–475.
1985b. Around *flat*. In V. Fromkin (ed.) *Phonetic Linguistics. Essays in Honor of Peter Ladefoged*. Orlando, FL: Academic Press, 223–241.
1986. Consumer's guide to evidence in phonology. *Phonology Yearbook* 3: 3–26.
1987. Explanation, evidence, and experiment in phonology. In W. U. Dressler, H. C. Luschützky, O. E. Pfeiffer, and J. R. Rennison (eds.) *Phonologica 1984: Proceedings of 5th International Phonology Meeting*, Eisenstadt, Austria. Cambridge: Cambridge University Press, 215–225.
Ohala, J. J. and H. Kawasaki. 1984. Prosodic phonology and phonetics. *Phonology Yearbook* 1: 113–127.
Ohala, J. J. and J. Lorentz. 1977. The story of [w]: an exercise in the phonetic explanation for sound patterns. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 3: 577–599.
Öhman, S. E. G. 1966. Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 39: 151–168.
Osthoff, H. and K. Brugmann, 1967. Preface to morphological investigations in the sphere of the Indo-European languages I. In W. P. Lehmann (ed. and trans.) *A Reader in Nineteenth Century Historical Indo-European Linguistics*. Bloomington: Indiana University Press. 197–209. [Originally published as Preface. *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen I*. Leipzig: S. Hizrel, 1878, iii-xx.]
Phelps, J. 1937. Indo-European *sl*. *Language* 13: 279–284.
Repp. B. H. 1976. Perception of implosive transitions in VCV utterances. *Status Report on Speech Research* (Haskins Laboratories) 48: 209–233.
1977a. Perceptual integration and selective attention in speech perception: further experiments on intervocalic stop consonants. *Status Report on Speech Research* (Haskins Laboratories) 49: 37–69.
1977b. Perceptual integration and differentiation of spectral information across intervocalic stop closure intervals. *Status Report on Speech Research* (Haskins Laboratories) 51/52: 131–138.
1978. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics* 24: 471–485.
Rothenberg, M. 1968. The breath-stream dynamics of simple-released-plosive production. *Bibliotheca Phonetica* 6.
Schouten, M. E. H. and L. C. W. Pols. 1983. Perception of plosive consonants. In M. van den Broecke, V. van Heuven, and W. Zonneveld (eds.) *Sound Structures. Studies for Antonie Cohen*. Dordrecht: Foris, 227–243.
Stevens, K. N. 1971. Airflow and turbulence noise for fricative and stop consonants: static considerations. *Journal of the Acoustical Society of America* 50: 1180–1192.
1972. The quantal nature of speech: evidence from articulatory-acoustic data. In E. E. David and P. B. Denes (eds.) *Human Communication: A Unified View*. New York: McGraw Hill, 51–66.
Streeter, L. A. and G. N. Nigro. 1979. The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America* 65: 1533–1541.
Thurneysen, R. 1961. *A Grammar of Old Irish*. Dublin: The Dublin Institute for Advanced Studies.

Wang, W. S.-Y. 1959. Transition and release as perceptual cues for final plosives. *Journal of Speech and Hearing Research* 2: 66–73. [Reprinted in: I. Lehiste (ed.). 1967. *Readings in Acoustic Phonetics*. Cambridge, MA: MIT Press, 343–350.]

Westbury, J. R. 1979. Aspects of the temporal control of voicing in consonant clusters in English. Ph.D. dissertation, University of Texas, Austin.

Weymouth, R. F. 1856. On the liquids, especially in relation to certain mutes. *Transactions of the Philological Society* [London]. 18–32.

Winitz, H., M. E. Scheib, and J. A. Reeds. 1972. Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech. *Journal of the Acoustical Society of America* 51: 1309–1317.

Wright, J. T. 1986. The behavior of nasalized vowels in the perceptual vowel space. In J. J. Ohala and J. J. Jaeger (eds.) *Experimental Phonology*. Orlando, FL: Academic Press, 45–67.

Zipf, G. K. 1935. *The Psycho-biology of Language. An Introduction to Dynamic Philology*. Boston: Houghton Mifflin.

15

On the value of reductionism and formal explicitness in phonological models: comments on Ohala's paper

JANET B. PIERREHUMBERT

In this paper, Ohala provides a nice case study in the style that he has become so well known for. He presents experimental data indicating that a regular phonological process, the direction of assimilation, is grounded in facts about speech production and perception. Such results are significant not only because of the light they shed on particular phenomena, but also as examples of research methodology. They point out the importance of seeking the right sphere of explanation for observed patterns in sound structure.

The type of explanation which is featured in this paper is phonetic explanation, or explanation based on the physics and physiology of speech. Phonetic explanations are especially attractive because of their reductionist character; it is very satisfying to reduce psychology to biology, and biology to physics. Ohala's comparison of phonetic and nonlinear phonological accounts of assimilation links reductionism ("None of the terms of the explanation are unfamiliar, other-worldly entities") with generality ("A few primitives go a long way"). It is not clear to me that this link is well-founded, especially with respect to ongoing research. Nonlinear phonology has identified a number of principles which have great generality, although their physical basis is unclear. In particular, the principle of hierarchical organization has been shown to be a factor in the lexical inventory, phrasal intonation, and allophony rules of many languages. On the other hand, some parts of phonetics are extremely particular, from a scientific point of view. For example, there is no reason to suppose that the specific nonlinear oscillator responsible for vocal fold vibration has any generality from the point of view of physics. At this point, physics has no general theory of nonlinear oscillators; each new system that has been studied has given rise to new analysis techniques and interpretations. Vocal fold vibration is more interesting than other systems of similar mathematical complexity chiefly because of its role in human language. A researcher who is trying to decide how to use his time often has a choice of whether to aim for reduction or for generality. In some cases, paradoxically, aiming for generality is best even if reduction is the aim; making the description more

comprehensive and exact can narrow the class of possible underlying mechanisms. The field advances best if the judgement is made on a case-by-case basis, by assessing the feasibility and informativeness of the various methods that might be applied.

I think we need to consider, also, the possibility that higher-level domain-specific theories may incorporate scientific insights which are lost (to the human mind, at least) in the theories which "explain" them. I am reminded of a talk I heard a few years ago in which Julian Schwinger discussed his experiences designing microwave guides during the war. He at first viewed this assignment as a trivial and uninteresting one, since the physics of microwave guides is completely specified by Maxwell's equations. However, although the behavior of microwave guides is indeed a solution to Maxwell's equations, their phenomenology proved to be so complex that an additional higher-level theory was needed to make it comprehensible. The interest of the assignment emerged in constructing this theory. In this case, the higher-level theory was desirable even though the explanation was already known. Such theories are doubly desirable when the explanation is being sought.

In his summary comparison of phonetic and nonlinear phonological representations of sound patterns, Ohala says that neither claims to represent sound structure in a computationally efficient manner. However, computational models, whether efficient or not have been very important in our progress towards explaining speech. So a brief review may be worthwhile.

On the phonetic side, the acoustic theory of speech production relied on calculations made using the first available digital computers. It demonstrated an approximation to speech production which can be computed efficiently, something recently emphasized in Fant (1985). This enabled it to support work on synthesis models, which have been so important to our understanding of speech perception, prosody, and allophony in continuous speech. It has also provided the basis for rigorous work on the limitations of the theory (cf. Fant, Lin, and Gobl 1985).

On the other hand, the formalism for phonological rules developed in Chomsky and Halle (1968) was grounded in earlier work by Chomsky and others on the theory of computation. The *Sound Pattern of English* (SPE) showed considerable descriptive flair, but so did much earlier work. It advanced over earlier work chiefly by its algorithmic approach, and the SPE formalism was successfully applied in implementing the phonological rules of text-to-speech systems, both for English and for other languages (see Allen, Hunnicutt and Klatt 1987; Carlson and Granström 1976; Carlson and Granström 1986, and references given there). Work in metrical and autosegmental phonology has built on the theory of trees and connected graphs, which are also computationally tractable. This tractability has made it possible to move quickly to implementations which incorporate theoretical advances in nonlinear phonology, for example Church's (1982) syllable parser and work by Pierrehumbert (1979), Anderson, Pierrehumbert and Liberman (1984)

and Beckman and Pierrehumbert (1986) on synthesis of fundamental frequency contours. Such implementations of nonlinear representations have themselves led both to new observations and to theoretical innovations (see Pierrehumbert and Beckman 1988). Nonlinear phonology still presents one serious obstacle to computational implementations; it is not explicit about the interaction between derivational rules and well-formedness conditions. Compared to *The Sound Pattern of English*, the theory remains underformulated. This has been an obstacle to theoretical progress, too, since it has led to confusion about what claims are being made and what their consequences are for new data.

One lesson which emerges from reviewing computational work on sound structure is the value of formalization. Unlike Ohala, I feel that formalization pays off for all types of descriptions. Formalizing nonlinear phonological descriptions pays off because it clarifies the issues and assists systematic evaluation, in part by supporting the construction of computer programs. Some descriptions may be just as good as others, but not all are equally good; some are just plain wrong. There is no point in seeking a phonetic or cognitive basis for spurious generalizations. It is important to keep in mind, also, the importance of formalizing phonetic descriptions. If we observe a parallel between some facts about speech and some physical law, we don't have an explanation, we have a conjecture. To have an explanation, it is necessary to write down the equations and determine the quantitative correspondence to the data. Only in this way can we find out if additional physical or cognitive mechanisms are crucially involved. Exact modeling will be especially important for determining the interplay of phonetic and cognitive factors, since the cognitive system can apparently exaggerate and extend tendencies which arise in the phonetics.

A second lesson is the value of distinguishing levels of representation. For instance, one level in a synthesis system will have the job of specifying what linguistic contrasts in sound structure are possible. Another will specify how sounds are pronounced, in terms of the time course of acoustic or articulatory parameters. Yet another will specify a speech waveform. For the most part, different kinds of work are done at different levels, and so they are complementary rather than competing. However, competition does arise when the division of labor between levels is unclear. This is how most of the competition between nonlinear phonology and phonetics arises, in my opinion. For example, before Poser (1984), it was unclear whether Japanese had a phonological rule changing High to Mid after a pitch accent, or whether it had a phonetic rule reducing the pitch range after a pitch accent. Poser's experiments resolved this question. I feel optimistic that such issues will in general be resolvable by empirical investigation, and will not become mired in debates about philosophy and taste.

A third observation, suggested especially by work on speech synthesis, is that different degrees of explanation are possible at all levels of representation. Nonlinear phonology explains the sound patterns in the English lexicon better

than a word list does. This nonlinear explanation in turn requires further explanation; since (as Ohala points out) it is based on internal reconstruction, its cognitive basis is problematic and needs to be worked out. A similar gradation can be found within phonetics proper. The acoustic theory of speech production explains spectra of speech sounds by deriving them from the configurations of the articulators. But it too requires further explanation. Why does the linear approximation work as well as it does? Which articulatory configurations are possible in general? And why does one configuration rather than another occur in any particular case?

References

Allen, J., M. Sharon Hunnicutt, and D. Klatt. 1987. *From Text to Speech: The MITalk System*. Cambridge Studies in Speech Science and Communication. Cambridge: Cambridge University Press.

Anderson, M., J. Pierrehumbert, and M. Y. Liberman. 1984. Synthesis by rule of English intonation patterns. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* 1: 2.8.1–2.8.4.

Beckman, M. and J. Pierrehumbert. 1986. Japanese prosodic phrasing and intonation synthesis. *Proceedings of the 24th Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 136–144.

Carlson, R. and B. Granström, 1976. A text-to-speech system based entirely on rules. *Conference Record, 1976 IEEE International Conference on Acoustics, Speech, and Signal Processing* 686–688.

Carlson, R. and B. Granström. 1986. Linguistic processing in the KTH multi-lingual text-to-speech system. *Proceedings of the 1986 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2403–2406.

Chomsky, N. and M. Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Church, K. 1982. Phrase-structure parsing: a method for taking advantage of allophonic constraints. Ph.D. dissertation, MIT.

Fant, G. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton (2nd edition 1970).

1985. The vocal tract in your pocket calculator. In V. Fromkin (ed.) *Phonetic Linguistics*. London: Academic Press. 55–77.

Fant, G., Q. Lin, and C. Gobl. 1985. Notes on glottal flow interaction. *STL-QPSR* 2-3/85, Department of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, 21–41.

Hertz, S. 1982. From text to speech with SRS. *Journal of the Accoustical Society of America* 72: 1155–1170.

Pierrehumbert, J. 1979. Intonation synthesis based on metrical grids. *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, Acoustical Society of America, New York, 523–526.

Pierrehumbert, J. and M. Beckman. 1988. *Japanese Tone Structure*. Linguistic Inquiry Monograph Series. Cambridge, MA: MIT Press.

Poser, W. 1984. The phonetics and phonology of tone and intonation in Japanese. Ph.D dissertation. MIT.

16

A response to Pierrehumbert's commentary

JOHN J. OHALA

I thank both Pat Keating and Janet Pierrehumbert for their thoughtful and constructive oral comments on my paper during the conference. I especially thank Janet Pierrehumbert for offering the above written commentary.

Pierrehumbert raises the issue of the connection between generality and reductionism ("A researcher... has a choice whether to aim for reduction or for generality"). The connection or lack of it is, I think, a very simple matter. In *explanations* the link is obligatory. To explain the unfamiliar by reducing it to the familiar means to bring the unknown into the fold of the known and therefore to enlarge the domain to which the explicanda apply, thus achieving greater generality. An example is Watson and Crick's explanation of the genetic code and the mechanism of inheritance by reducing it to previously known chemical facts, e.g. how adenine bonds only with thymine and cytocine only with guanine in such a way as to guarantee the construction of exact copies of molecular chains (DNA) consisting of those substances. Needless to say, in cases like this it may take genius and inspiration to figure out which facts to bring together into an explanation.

One can achieve generality without reduction but then this is a form of description. Explanation is deductive generalization; a systematic description of a sufficiently wide range of phenomena is inductive generalization. One of the points in my paper was that nonlinear phonology is a good description of certain sound patterns. Pierrehumbert seems to agree with me, then. Presumably we also agree that although at any given stage of the development of a scientific discipline, inductively-based generalizations are important and necessary, ultimately all disciplines strive to deduce their facts from first principles. Where we may disagree – and here we get "mired in debates about... taste" – is whether our discipline is at a stage where deduction of natural phonological processes is possible. I say that it is and have offered table 14.1 (and the accompanying references) in support of this. Everyone should follow their own hunches on this matter; I only hope that their choice will be based on a thorough understanding of what the phonetic explanations have to offer.

I am not sure I see the relevance of Pierrehumbert's example of the larynx as a nonlinear oscillator. Primitives – the things to which we try to reduce more complex behavior – are, so to speak, the building blocks of the universe. Within physics the larynx as an oscillator can be understood or constructed for the most part out of basic physical building blocks – this, at least, is what vocal-cord modelers such as Flanagan, Ishizaka, and Titze have done. Pierrehumbert's point that "there is no reason to suppose that the specific nonlinear oscillator responsible for vocal fold vibration has any general application in physics" may be true but uncontroversially so because as far as I know, no one has claimed otherwise. The claim – my claim – is the reverse: principles of physics apply to certain aspects of speech and language, including the behavior of the vocal cords.

As for the microwave guide example, it is instructive but I do not think it helps to resolve the present point of contention. Although domain-specific investigations are invariably necessary for any specific task (in vocal cord modeling one must produce or assume a value for the elasticity of the vocal cords in order to make the rest of the model work), it can still be a matter of judgement or of contention which features of a model call for domain-specific treatment and which can be reduced to primitives of a more general sort. My position is that, as opposed to current practice, more of the problems that occupy phonologists today can and should be reduced to primitives from, say, physics, physiology, or psychology.

To avoid misunderstanding let me say that I do not advocate reduction of phonological phenomena unless (a) the opportunity to do so presents itself (i.e. someone has a bright idea which makes the reduction possible) and (b) this reduction results in an increase in our understanding of the behavior in question. It is possible to give some account of articulator movements in terms of muscle contractions which in turn can be accounted for to some degree in terms of physical and chemical processes in the neuromotor system. But this may be neither necessary nor helpful to our understanding of how certain articulations are made. More to the point, there will always be gaps in our knowledge which prevent our understanding of some things. In other words, reduction will not be an option until someone is inspired to propose and test a hypothesis which specifies candidate primitives underlying the objects of our curiosity. I believe in opportunistic reductionism not obligatory reductionism. (See also Ohala 1986, 1987a, 1987b.)

Pierrehumbert's review of computational models in linguistics is useful though it seems to have little to do with my paper. I agree with most of her remarks. The purpose of my table 2 was simply to suggest that there are a variety of criteria by which we might want to evaluate phonological models, including, perhaps, computational efficiency. It is doubtful that one model would be optimal for all tasks but whatever the task we should seek out the best methods to accomplish it. It would not contradict my claim that existing phonetic models do better than nonlinear notations at the task of accounting for natural sound patterns if it turned

out that nonlinear phonology was better at the task of computational efficiency. In any case, Pierrehumbert does not seem to insist on this latter point.

References

Ohala, J. J. 1986. Consumer's guide to evidence in phonology. *Phonology Yearbook* 3: 3–26.

1987a. Experimental phonology. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 13: 207–222.

1987b. Explanation, evidence, and experiment in phonology. In W. U. Dressler, H. C. Luschützky, O. E. Pfeiffer, and J. R. Rennison (eds.) *Phonologica 1984*. Cambridge: Cambridge University Press, 215–225.