Speech Communication 11 (1992) 369–378
North-Holland

What is the input to the speech production mechanism?[1]

John J. Ohala

Department of Linguistics, University of Alberta, Edmonton, AB, Canada T6G 2E7 and
Department of Linguistics, University of California, Berkeley, CA 94720, USA

Received 30 September 1991
Revised 19 May 1992

Abstract. Our conception of what it is that the speech production mechanism is attempting to implement in speaking comes from linguistics. But linguistics first developed its methods for other purposes. In the early 19th century linguistics produced a method for tracing the family relationship between suspected cognate words and their constituent sounds. This, the *comparative method*, involved establishing an optimal path between these forms via a reconstructed parent form. 20th century structuralist linguistics (including generative phonology), essentially grafted the same method onto the task of finding the underlying phonemic constituents of words. Although the underlying structure found in this way may be a good hypothesis as to the mental elements determining actual spoken utterances, there are reasons to suspect that it is too simple. Too much emphasis is placed on the simplicity of the system and on the purely lexical (as opposed to the demarcative and attitudinal) function of the elements in speech. This paper presents some initial attempts to differentiate between phonetic variants in speech which stem from single underlying forms as opposed to those which arise from separate underlying forms (though they may have had a common source historically).

Zusammenfassung. Unsere Auffassung über das was der Sprachproduktionsmechanismus versucht auszuführen beim Sprechen stammt her von der Linguistik. Die Linguistik aber entwickelte ihre Methoden für andere Ziele. Im frühen 19. Jahrhundert produzierte die Linguistik eine Methode um die Familienverhältnisse herauszuarbeiten zwischen verwandten Worten und den sie zusammensetzenden Lauten. Diese vergleichende Methode umfasste das Auffinden eines optimalen Wegs zwischen diesen Formen via einer rekonstruierten Stammform. Die strukturalistiche Linguistik des 20. Jahrhunderts (inklusiv der generativen Phonologie) übertrug dieselbe Methode auf die Aufgabe die zugrundeliegenden phonemischen Bestandteile von Wörtern aufzufinden. Obwohl die zugrundeliegende Struktur welche so aufgefunden wird möglicherweise eine gute Hypothese darstellt was die mentalen Elemente angeht welche sprachliche Äusserungen ausmachen, bestehen gute Gründe anzunehmen, dass diese Hypothese zu einfach ist. Es wird zuviel Wert gelegt auf die Einfachheit des Systems und auf die reinen lexikalen Funktionen der Sprachelemente (im Gegensatz zu demarkativen Funktionen und zum Ausdruck der Einstellung des Sprechers zum Gesagten). Dieser Beitrag stellt erste Versuche vor zu unterscheiden zwischen phonetischen Sprachvarianten welche von gleichen und verschiedenen zugrundeliegenden Formen stammen (und die möglicherweise die gleiche historische Quelle besitzen).

Résumé. Notre conception de ce que le mécanisme de production de parole essaye d'implémenter lors de l'acte de parole vient de la linguistique. Mais la linguistique a d'abord développé ses méthodes pour d'autres buts. Au début du 19èmes siècle, elle a mis au point une méthode pour retracer les relations de parenté entre des mots de même origine et les sons qui les constituent. Cette méthode, la méthode comparative, impliquait d'établir une filiation optimale entre ces formes via une forme-parent reconstruite. La linguistique structuraliste du 20ème siècle (y compris la phonologie générative) a appliqué essentiellement la même méthode pour tâcher de découvrir les constituants phonémiques sous-jacents des mots. Bien que la structure sous-jacente découverte de cette manière puisse constituer une bonne hypothèse de ce que sont les éléments mentaux qui déterminent les phrases effectivement produites, il y a des raisons de suspecter que cela est trop simple. Trop d'importance est attachée à la simplicité du système et à la fonction purement lexicale (en opposition avec les fonctions démarcative et stylistiqué) des éléments dans la parole. Cet article présente quelques essais préliminaires pour différencier les variantes phonétiques de la parole qui prennent leur origine dans une seule forme sous-jacente de celles qui proviennent de formes sous-jacentes séparées (bien qu'elles puissent avoir eu historiquement une source commune).

Keywords. Speech production; linguistics; phoneme; epenthetic stops; dissimilation; comparative method.

[1] This paper was originally presented at the Workshop on Speech Perception and Speech Production sponsored by the ATR Auditory & Visual Perception Research Laboratories, Kyoto, 15–16 November 1990, and has also appeared in *Speech Perception, Production, and Linguistic Structure*, ed. by Y. Tohkura, E. Vatikiotis-Bateson and Y. Sagisaka (Ohmsma, Tokyo), pp. 297–311.

Elsevier Science Publishers B.V.

1. Introduction

In the fascinating history of the development of the different forms of the Roman alphabet there is a period between the 3rd and 10th centuries when the small letters, the so-called "lower-case" or uncials, developed gradually from the angular, largely straight-line Roman capitals. The curves and new lines joining parts of the letters came about when writing was done more on paper than in stone or wax, more with pen or reed than with chisel or stylus, such that it was done more rapidly and without lifting the writing instrument off the inscribed surface. Here we can see how task constraints – writing rapidly, and physical properties and limitations of the hand and of the material used – led to a modification of letter shape. Yet when we use lower-case letters today, whether with a typewriter, word processor or even in handwriting, it is not true that they are present due to these original constraints of the task, the writer and the instruments. Lower-case letters as opposed to capital letters have been fossilized or conventionalized. Even so, contemporary handwriting may still show many of the effects from the same forces that originally led to the development of the uncials: parts of letters not previously joined – capitals or uncials – are often linked by a curve by virtue of the writer not lifting the pen from the paper.

Variant forms of pronunciation are similar to this. Some of the variants observed in speech are fossilized remnants of earlier modifications of speech due to speaking rapidly or softly and others are modifications due to the same factors acting at present – on-line, so to speak. The first type of variants are selected by the speaker as capitals and lower-case are: they are drawn from an inventory of separate norms. The second type of variants are not so much selected by the speaker as they are just consequences of physical and physiological constraints of the speaking mechanism which accompany choice of a different speech style (e.g., the more open jaw position of shouting, the lesser amplitude of articulatory movement when speaking rapidly). In the second case, the variants may spring from a single norm.

In studies of speech production it is important to know what the input to the speech production mechanism is, but unfortunately it is not as easy to differentiate between these two types of variants in speech as it is in writing. For one thing, speech norms have no existence outside of the mouths (or more properly, the brains) of speakers; norms for letter shapes have the advantage that they can be printed and consulted in books. Second, in some cases the discipline studying such variants, phonology, has never succeeded in developing a viable method to differentiate these two types of variant. It is only a slight oversimplification to say that there is only *one* fully developed method in phonology (so far) when it comes to giving an account of variation: this is the method of comparative reconstruction plus two more restricted methods which are based on it: internal reconstruction and phonemic analysis.

Many of those working on spoken language may think that the concerns and methods of the historical linguist are irrelevant to their interests. This is not the case, though, because virtually all the "linguistic" information as to the underlying structure of speech (i.e., the "norms" according to which speech is generated) comes from linguists who, whether they know it or not, base their pronouncements on an application of the comparative method. Understanding the strengths and weaknesses of the method therefore is vitally important to an informed study of the variations in pronunciation. In what follows I propose first to give a brief sketch of the comparative method and to point out some of its successes and limitations as they impact on the study of speech variation. Then I will suggest very programmatically how new methods may be developed which would better serve these new interests.

2. Linguistic methodology

2.1. The comparative method

The comparative method, developed in the 19th century, takes suspected cognates in different languages, attempts to establish proof of family relationship between the languages, and, if successful, then attempts to reconstruct as much as possible of the parent language based on the data from the daughter languages. Historical linguists approach these tasks using principles and methods

that have a quasi-mathematical character even though actual quantification is rarely if ever used. Some of the techniques of reconstruction were first used in philology in the late 18th century to reconstruct "parent" texts from variant daughter texts, e.g., by Friedrich Wolf (1759–1824).

2.1.1. Showing that languages and their parts are related

To establish that two or more languages and their parts are genetically related, i.e., that they spring from a common source, a kind of "folk probability theory" is used[2]. The linguist must be familiar with the structure of many different languages and know intuitively the "degrees of freedom" within which all languages' structures may vary. Then, when so many points of similarity in structure are found (in the phonetic shape of words and their meaning and in grammatical structure) which could not plausibly arise by chance or direct borrowing, it is concluded that the languages are genetically related, i.e., that they have a common parent.

The case for a family relationship is strengthened if the points of similarity are regular and systematic, e.g., every (or most cases of) speech sound *X* in one language (perhaps in a specific environment) corresponds to a *Y* in the other or at a higher level of phonetic abstraction, every sound with feature *X'* in one language is manifested as feature *Y'* in the other. The data in (1) lend themselves to just such an analysis.

(1) Cognate forms in three Indo-European dialects (taken from (Cowan and Rakusan, 1980)), see Table 1.

Focusing just on the beginning of these words, the first step is to establish what are called correspondence sets. Items a, b, c suggest the correspondence set in (2)[3].

[2] Thomas Young (1773–1829), who demonstrated the wave nature of light and was the first to figure out how to read Egyptian hieroglyphics, provided a mathematical account (1819) of the need for large numbers of points of similarity in order to establish convincingly that the resemblance of languages is not due to chance.

[3] Many more examples could be given and to make a full case for genetic relationship *must* be given.

Table 1

	Sanskrit	Greek	Germanic	Translation
a.	$b^hr\bar{a}ter$	$p^hr\bar{e}t\bar{e}r$	brōθar	brother
b.	b^harati	$p^her\bar{o}$	bāira	carry
c.	$b^h\bar{u}tih$	$p^hu\bar{o}$	būa	be
d.	d^hanus	t^henar	denn	flat
e.	$d^h\bar{u}mah$	$t^h\bar{u}mos$	dōmian	smoke
f.	$band^hati$	$pent^heros$	bindan	bind
g.	$bard^hakah$	$pert^h\bar{o}$	bord	cut
h.	$b\bar{o}d^hati$	$peut^homai$	−biudan	wake up

(2) Sanskrit $\mathbf{b^h}$ = Greek $\mathbf{p^h}$ = Germanic **b**

This conclusion is reinforced by the words in the next two rows, d, e, which suggest the correspondence in (3).

(3) Sanskrit $\mathbf{d^h}$ = Greek $\mathbf{t^h}$ = Germanic **d**

In fact, items f, g, h also exhibit this correspondence in non-initial position.

These, along with similar patterns for velars, permit a generalization which abstracts away from the particular sounds involved, at least their place of articulation, and instead is stated in terms of the natural classes of segments defined by their manners of articulation, as in (4).

(4) Sanskrit voiced aspirated stop
=Greek voiceless aspirated stop
=Germanic voice stop

These words also display other similarities, e.g., the medial **r**, **n**, **m**, in rows b, d, e, respectively.

2.1.2. Reconstructing the parent forms

The next step in historical reconstruction is finding an optimal path between all these correspondence sets via a common parent form. There are two basic elements of this reconstruction, then: figuring out a plausible parent form (and optionally, intermediate forms) and the sound changes which changed it into the daughter forms. These are obviously closely related since one can only posit given parent forms if there are plausible sound changes which would transform them into the daughter forms. Judgments of what's "plausible" come from wide experience with how other languages are structured and from experience with other previously-established sound changes. "Majority

vote" and other quasi-probabilistic considerations also enter into the process of reconstruction.

For the parent forms of the correspondence sets in (2) and (3) voiced aspirates were reconstructed since 2 out of these 3 show voicing and 2 out of 3 show aspiration. Treating the sounds and the features more or less as algebraic entities, it was judged that a simpler set of sound changes would be required to derive the daughter forms from the parent form[4].

But rows f, g, h in (1) present an apparent difficulty. They would suggest the correspondence set in (5).

(5) Sanskrit **b** = Greek **p** = Germanic **b**

Here the Germanic **b** has different correspondences in Sanskrit and Greek from those in (2) and (4). There were three options to resolve this apparent inconsistency:

(a) If we assume the same parent form (for (5) as for (2), a voiced aspirate, we would have to justify how in Sanskrit it could sometimes become $\mathbf{b}^{h}$ and sometimes **b** and in Greek, sometimes $\mathbf{p}^{h}$ and sometimes **p**.

(b) A second option of assuming a parent form other than a voiced aspirate is theoretically possible but is not desirable since other correspondence sets (not discussed here) were already established for a full complement of labial stops, namely **b** and **p**. An added labial stop would be unwelcome. (Neither **b** nor **p** as the parent form for the set in (5) would avoid the difficulty since then we would have to explain how in Germanic a **b** resulted sometimes from $\mathbf{b}^{h}$ and sometimes from this other stop, i.e., a problem similar to (a), above.)

(c) A third option, that sound change is sporadic and unpredictable was accepted initially then strongly denied during the neogrammarian revolution (c. 1876). It remains a controversial topic to this day (Wang and Cheng, 1977; Labov, 1981).

[4] This is a gross oversimplification, of course, because the daughter forms in other Indo-European languages have to be taken into account, too, e.g., Italic, Celtic, Balto-Slavic, Armenian, Anatolian, etc.

Herman Grassmann (1863), who made his mark in mathematics and psychology (of color vision) and was also a leading Sanskrit scholar, gave a satisfactory solution to this problem in 1863, taking option (a), by pointing out that an original aspirated stop in Sanskrit and Greek lost its aspiration if the word contained a second aspirated consonant. In other words, Sanskrit and Greek showed dissimilation of aspiration. This solution was celebrated not only because it avoided the awkwardness of the other two options but also because it maintained the overall simplicity and thus optimality of the paths between the daughter forms.

2.2. *Internal reconstruction*

The method called "internal reconstruction" is the comparative method applied to data exclusively from a single language. For example, given the morphologically related words in (6) which show variants with a [k] ~ [tʃ] alternation (on what is apparently a verb/verb nominalization paradigm), one may reconstruct a verb stem for *thatch*: *θæk.

(6)		
	wake	watch
	bake	batch
	dike	ditch
	make	match
	wreak, wreck	wretch
	?	thatch

In fact, Anglo-Saxon texts contain this word, meaning "roof" (plus the related verb *theccan*, "to thatch"); it is cognate with *deck*.

The predominant method of generative phonology and its offshoots, e.g., lexical phonology, autosegmental phonology, is internal reconstruction (although it is not called that). For example, this permits the reconstruction of an "underlying" (=parent) form for the stems of the morphological alternants *profound/profundity* which has the vowel [u] where the present-day variants have [aw] and [ʌ]. The rules which derive the "surface" forms (the pronounced form) from the underlying forms are the equivalent of sound changes. Reinforcing this analysis is the existence of hundreds of other word pairs whose vowel alternations can also be

accounted for by a generalized version of the same rules; see (7).

(7)

south	southern
house	husband
divine	divinity
extreme	extremity
serene	serenity
sane	sanity
suppose	suppository
etc.	

2.3. *Phonemic analysis*

The method of phonemic analysis establishes the functional equivalence of sounds by finding variant sounds in mutually exclusive environments just as the comparative method does. Phonemes (=parent forms) for sound are established in order to create optimal or simplest paths between the allophones (daughter forms). It is quite plausible, in fact, that some of the sound variants considered by traditional phonemic analysis to be allophones of a single phoneme are actually the product of completed sound changes. If so, they would be manifestations of distinct underlying norms, like the upper case and lower case letters. Plausible examples of this in English are the plain and velarized lateral approximant /l/ at the beginning and end of syllables, respectively, and the aspirated and unaspirated allophones of the voiceless stops in word pairs such as *pit/spit*, *tack/stack*. This is not to say that all allophonic differences discovered by phonemic analysis are due to sound change; some may quite legitimately represent phonetically-caused, "online" variation. A probable example of this is the differing degrees of aspiration on voiceless stops as a function of the degree of closeness (and thus resistance to the outgoing airflow) of the following vowel (Klatt, 1975; Ohala, 1981a). Wang and Fillmore (1961) called those allophones that are not phonetically conditioned "extrinsic allophones", while those that are "intrinsic allophones". The problem is that phonemic analysis by itself cannot always tell the difference between these two because they often look the same: extrinsic allophones represent present-day phonetic variation whereas most sound change is fossilized phonetic variation.

Most of those using the methods of internal reconstruction and phonemic analysis in modern phonology claim to be discovering not historical parent forms for variants but underlying psychological forms. Of course, they can claim anything, but perhaps it is not unwarranted to be skeptical of such claims since we do not see them using psychological techniques in their investigations and because of the isomorphism of their procedures to the method of historical reconstruction.

2.4. *Summary*

All these methods, the comparative method and its variants, have this in common: they establish the functional equivalence of phonetically distinct forms by positing a series of intermediate states, one of which is designated the parent form; both the intermediate states and the changes required to transform one state into another must adhere to certain inductively-derived norms. When two or more alternate analyses can be posited, the simplest and most general one is preferred, "general" in the sense of applying to more forms.

I present in Table 2 a brief comparison of the three methods which highlights their fundamental similarities and superficial differences.

3. Implications

3.1. *Disadvantages*

If it is accepted that modern phonological analytical techniques are based on the methods optimized for the study of reconstruction of the linguistic past, what are the consequences for the study of speech variants? First, the "function" of the functionally-equivalent entities that are grouped together in phonemes is the *lexical function*, i.e., the power of different sounds to create different words. But there are other functions of speech sound variation besides those making lexical differences. There is the function of *demarcating* words or other units, e.g., morphemes, phrases, sentences, in the stream of speech. All digital communications incorporate headers, block and sector markers, checksums, and other bits of information which are not part of the message per se but rather

Table 2
Comparison of the comparative method, internal reconstruction, phonemic analysis

Method	Daughter forms	Parent form	Conditioning factors	Time depth
Comparative method	Functionally-equivalent morphemes, phonemes in different languages	Original morpheme, phoneme common to daughter forms	Different language morphological, or phonological environment	Deep (several K years)
Internal reconstruction	Funtionally-equivalent morphemes, phonemes in one language	Original morpheme, phoneme common to daughter forms	Different morphological or phonological environment	Intermediate (1 or 2 K years)
Phonemic analysis	Functionally-equivalent allophones	Phoneme	Different phonological environment	Shallow (a few centuries?)

divide the message up into independent chunks. All communication systems – and speech is no exception – need such segmentation or articulation in order to minimize or restrict errors of transmission, see (Mandelbrot, 1954). Use of Roman capital letters at the beginning of sentences is an example of this. In Arabic writing the presence or absence of the long hook-like descender on the *sin* or *shin* provides some clues about whether it is at the end of a word or not. Using the techniques of phonemic analysis such contextual variants would be considered non-distinctive "allographs" of single "graphemes". Yet we clearly recognize that they have a function in marking boundaries.

Besides a demarcative function there are numerous *stylistic* functions of pronunciation variation, e.g., expressing the attitude of speakers, including the degree to which they are accommodating to listeners. When giving instructions or orders, the degree of precision taken in the articulation is probably related to the speakers' estimate of the importance of the instructions and the probability of the listeners misinterpreting the content. To the linguist the release of word final stops in English is non-distinctive; it exhibits so-called "free variation". But this is undoubtedly something a speaker can choose to do or not as a function of the style of speaking. The neglect of variant sounds that serve functions other than differentiating words would be analogous in typography to ignoring the "stylistic" use of bold or italic face, or even different styles of type faces, e.g., Bodoni, Caslon, Baskerville.

A second limitation of traditional structuralist analytic methods in phonology is that there is as yet no deep understanding of the nature of sound change, i.e., how and why it occurs and how to know the difference between variation that is the result of accomplished sound change on the one hand and on the other hand variation that is the stuff out of which sound change arises but which is not sound change itself. (See, however, (Ohala 1989, in press).) For example, in English the difference in vowel length (ΔD_v) before voiced and voiceless obstruents (House and Fairbanks, 1953) is traditionally regarded as allophonic, i.e., non-distinctive. But all the evidence we have suggests that a ΔD_v of the magnitude found in English is extremely unusual in comparison with ΔD_v in other languages (Lehiste, 1970); this would suggest that the extreme value found in English is intended by the speakers, i.e., an invariable characteristic of English speech, probably the result of sound change. Similarly, traditional analyses of American English regard the nasalization of vowels before nasal consonants as a purely allophonic process – and so it may be in some languages – but again all the evidence we have suggests that the degree of anticipatory nasalization on vowels in North American English is extreme in comparison to that in other languages of the world (Clumeck, 1976; Solé and Ohala, 1991). Once more we are forced to conclude that this is probably an integral characteristic of the language emanating from different underlying psychological norms and not simply a function of physiological constraints of the speech organs. A speech researcher who uncritically accepted the linguists' pronouncements on the nature of the input to the speech production mechanism would thus probably be misled.

3.2. Advantages

The preceding discussion focussed on some of the limitations of traditional linguistic analysis as this impacts on studies of speech styles, but we should not overlook one of its great advantages. One thing historical linguistics is demonstrably successful at is reconstructing the history of the languages, including the making of vast catalogs of the sound changes leading from parent to daughter languages. If one accepts the widely-held view that many sound changes originate from synchronic phonetic variation of the type seen in different speech styles, then an informed study of sound change can give us some insight into these on-line mechanisms of variation (Ohala, 1975, 1985).

4. Towards new methods of studying variation in speech

4.1. Introduction

The purpose of the preceding discussion was to call attention to a fundamental problem in the study of speech variation: differentiating phonetic variation of a single underlying norm from that due to sound change and which therefore stems from separate underlying norms, just as upper-case and lower-case letters do. The problem is difficult because sound change is largely just fossilized phonetic variation and thus the two may often resemble each other. We cannot rely entirely on traditional linguistic methods for this purpose, although we certainly could and should take these methods as starting points for a study of variation. But new methods will be necessary. I will suggest a few.

4.2. Example 1. Establishing the causal basis of variation

As mentioned, present-day phonetic variation and sound change will resemble each other. For example, vowels will be somewhat nasalized adjacent to nasal consonants – but only some 50 to 70 msec of the portion of the vowel immediately adjacent to the nasal needs to be nasalized (due to inertial properties of the velic valve). A sound change can occur when a listener reinterprets this nasalization as an intended feature of pronunciation and incorporates it and exaggerates it in his own speech (this is what I suspect has happened in English). But there is this difference: in the former case the phonetic effect will be mechanically linked, as it were, to the presence of the conditioning environment, the nasal consonant. Solé and Ohala (1991) have attempted to demonstrate such a difference in the behavior of nasalization on vowels in American English versus Continental Spanish.

Another example comes from the influence of consonants on the quality of adjacent vowels. In a study of Swedish vowels Lindblom (1963) demonstrated that as syllables are pronounced more quickly and their duration shrinks, the formant frequencies of the vowels move closer to those characteristic of the adjacent consonant. For example, back rounded vowels like /u, o/ which have low F_2s get higher F_2s when pronounced next to a high F_2 consonant like /d/. The change in formant frequency as a function of the consonant's formants and the duration of the vowel is sufficiently lawful to be expressible by equations. Nevertheless, Nord (1974), also working with Swedish, showed that the changes in vowel quality characteristic of vowels in some prosodic environments could not be explained simply by the effects of duration and consonantal environment and thus must stem from norms that are distinct.

It is widely observed that low vowels are longer than high vowels, other things being equal, and many have hypothesized that this effect is due to the greater distance the jaw has to travel for low, open vowels and thus the greater time taken in the articulation. But Nooteboom and Slis (1970) found that these durational differences persist even when speech is produced with a clenched jaw. Although follow-up experiments are necessary, this tends to suggest that differences in vowel duration as a function of vowel height are centrally programmed.

4.3. Example 2. Duration

Consider a small problem: In certain words in English (and other languages), one finds intrusive

or epenthetic stops, e.g. in the English words in (8).

(8) warm[p]th
some[p]thing
team[p]ster
prin[t]ce (homophonous with "prints")
Lan[t]sing
young[k]ster

Are these epenthetic stops just the result of low-level phonetic events – transitional elements – caused by anticipatory denasalization of the latter half of the nasals (before the following oral consonants)? Or are these purposeful, i.e., underlying?

Through sound change such stops can and have become fossilized, e.g., in the pronunciation of the English words in (9a) and the French in (9b).

(9) (a) glimpse (<gleam)
Thompson (< Thom + son)
thunder (<Thunor)

(b) chambre (<cam(e)re) "room"
sembler (<sim(u)lare) "to seem"
vendredi (<ven(i)ris-die) "Friday"

I conducted a study (Ohala, 1981b) to see if durational measures can differentiate phonetic, transitional stops from underlying phonological stops. Specifically I investigated whether the VN sequences in such words would be different if they were underlying VN# versus VNC#. It was expected that VN in a syllable not closed by a stop would be longer than one closed by a stop. I recruited 24 linguistically naive native speakers of English to derive and pronounce the novel form *clam + ster*. Then much later in the same session they were asked to do the same with *clamp + ster*. (In some cases the speakers spontaneously repeated these words.) Since these are novel derivations, if an intrusive [p] did appear in "clamster" it would have to be a transitional, purely phonetic event, not lexical or underlying. The [p] in "clampster", on the other hand, must be underlying or lexical since it is inherent in the stem "clamp".

The results of measuring the VN sequence in both words are given in Figure 1. The mean duration in 24 tokens of *clamp + ster* was 255 msec (standard deviation = 33), whereas for 25 tokens of *clam + ster* which did not have any detectable epenthetic stop it was 351 msec (standard deviation = 48). For 8 tokens of *clam + ster* that had an epenthetic [p], the mean duration was 353 msec (standard deviation = 54). Thus the durational difference of the VN sequence can be used to differentiate between underlying and phonetic epenthetic stops. In (Ohala, 1981b) this result is used to determine that in some speakers' pronunciation an epenthetic [p] in the existing word "teamster" is not underlying.

4.4. *Example 3. Ask listeners*

The case of dissimilation of aspiration in Sanskrit and Greek, cited above, along with numerous other cases of dissimilation found in the literature of historical phonology may provide some insight into how the listener deals with the vast amount of variation present in speech. I have proposed a theory of dissimilation which characterizes it as the result of listeners attempting to correct or normalize (inappropriately) the speech signal which they think shows purely phonetic variation (Ohala, 1981c, 1986, 1989). Briefly, the theory assumes that faced with phonetic variation in the speech signal, e.g., the migration of aspiration, nasalization, lip rounding, etc. from the site where it is distinctive

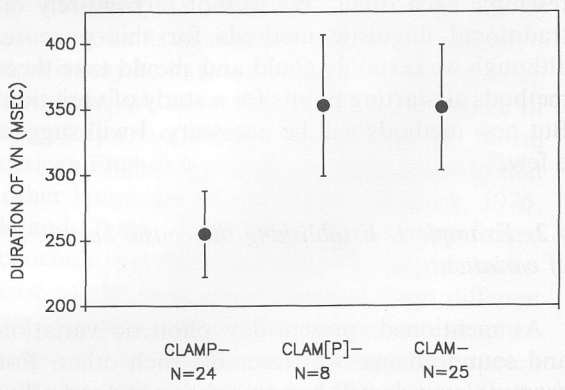

Fig. 1. Mean duration of VN sequence in novel derivations *clamp + ster* and *clam + ster*, in the latter case both those with and those without an epenthetic stop. Bars indicate standard deviation.

to adjacent segments where it is not, that listeners can "normalize" or "correct" the signal by somehow factoring out these features from the non-distinctive sites. E.g., given non-distinctive lip rounding on the fricative /s/ before the rounded vowel /u/, the listener should be able to factor out the consequences of this anticipatory assimilation of lip configuration: a lower center frequency of the fricative noise. Dissimilation comes about when the same feature is distinctive on *two* nearby sites within a word and the listener guesses (incorrectly) that it is distinctive on one of those but its presence on the other is due to assimilation. In constructing the underlying lexical form of a word (on which the listener bases his own pronunciation), the listener therefore factors out this feature on one of these two sites. Supporting this hypothesis is the fact that dissimilation is found almost exclusively with features such as aspiration, nasalization, lip-rounding, palatalization, etc., which are known to be subject to "long-distance" assimilation, but not to features such as "stop" which rarely if ever show assimilation.

What is of interest here for our purpose is not dissimilation itself which is, in a sense, an error of perception, but rather that the few cases of erroneous correction of the speech signal imply that there is a process of correction which applies normally in the vast majority of cases. In fact, the speech perception literature contains abundant evidence that listeners can normalize a variable speech signal, e.g., Mann and Repp (1980) showed that listeners accept as /s/ a fricative with a lower center frequency when it appears before the vowel /u/, but not before the unrounded vowel /a/. (See also (Beddor et al., 1986; Ohala and Feder, 1987; Ọhala and Shriberg, 1990.) The implication of such studies is that somehow listeners know what is distinctive (=underlying, lexical) and what is not in the speech signal; careful perception studies that tap this aspect of listeners' knowledge can help us to discover the input to the speech production mechanism.

5. Conclusion

There is no easy way to discover the input to the speech mechanism; we will have to be clever and invent our own techniques as we progress. The linguists' claims may be good starting points, but since they are most suitable for historical reconstruction, not for psychological accounts of speech production, they have to be independently verified.

References

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

H. Clumeck (1976), "Patterns of soft palate movements in six languages", *J. Phonetics*, Vol. 4, pp. 337–351.

W. Cowan and J. Rakusan (1980), *Source Book for Linguistics*, Language Resource Center, Carleton Univ., Ottawa.

H. Grassmann (1863), "Über die Aspiraten und ihr gleichzeitiges Vorhandensein im An – und Auslate der Wurzels", *Z.f.v. Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen*, Vol. 12, No. 2, pp. 81–138.

A.S. House and G. Fairbanks (1953), "The influence of consonant environment upon the secondary acoustical characteristics of vowels", *J. Acoust. Soc. Amer.*, Vol. 25, pp. 105–113.

D.H. Klatt (1975), "Voice onset time, frication, and aspiration in word-initial consonant clusters", *J. Speech Hearing Res.*, Vol. 18, pp. 686–706.

W. Labov (1981), "Resolving the neo-grammarian controversy", *Language*, Vol. 57, pp. 267–308.

I. Lehiste (1970), *Suprasegmentals* (MIT Press, Cambridge, MA).

B. Lindblom (1963), "Spectrographic study of vowel reduction", *J. Acoust. Soc. Amer.*, Vol. 35, pp. 1773–1781.

B. Mandelbrot (1954), "Structure formelle des textes et communication", *Word*, Vol. 10, pp. 1–27.

V.A. Mann and B.H. Repp (1980), "Influence of vocalic context on perception of the [š] vs [s] distinction", *Perception & Psychophysics*, Vol. 28, pp. 213–228.

S.G. Nooteboom and I. Slis (1970), "A note on the degree of opening and the duration of vowels in normal and "pipe" speech", *IPO Ann. Rep.*, Vol. 5, pp. 55–58.

L. Nord (1974), "Vowel reduction. Centralization or contextual assimilation?", *Preprints of the Speech Communication Seminar*, Stockholm, 1–3 August 1974, Vol. 2, pp. 149–154.

J.J. Ohala (1975), "How a study of sound change can aid in automatic speech recognition", in *Speech Communication. Vol. 3: Speech Perception and Automatic Recognition*, ed. by G. Fant (Almqvist & Wiksell, Stockholm), pp. 299–302.

J.J. Ohala (1981a), "Articulatory constraints on the cognitive representation of speech", in *The Cognitive Representation of Speech*, ed. by T. Myers, J. Laver and J. Anderson (North-Holland, Amsterdam), pp. 111–122.

J.J. Ohala (1981b), "Speech timing as tool in phonology", *Phonetica*, Vol. 38, pp. 204–212.

J.J. Ohala (1981c), "The listener as a source of sound change", in *Papers from the Parasession on Language and Behavior*, ed. by C.S. Masek, R.A. Hendrick and M.F. Miller (Chicago Ling. Soc., Chicago), pp. 178–203.

J.J. Ohala (1985), "Linguistics and automatic speech processing", in *New Systems and Architectures for Automatic Speech Recognition and Synthesis* (NATO ASI Series, Series F: Computer and System Sciences, Vol. 16), ed. by R. de Mori and C.-Y. Suen (Springer, Berlin), pp. 447–475.

J.J. Ohala (1986), "Phonological evidence for top-down processing in speech perception", in *Invariance and Variability in Speech Processes*, ed. by J.S. Perkell and D.H. Klatt (Lawrence Erlbaum, Hillsdale, NJ), pp. 386–397.

J.J. Ohala (1989), "Sound change is drawn from a pool of synchronic variation", in *Language Change: Contributions to the Study of its Causes* (Series: Trends in Linguistics, Studies and Monographs No. 43), ed. by L.E. Breivik and E.H. Jahr (Mouton, Berlin), pp. 173–198.

J.J. Ohala (in press), "What's cognitive, what's not, in sound change", *Diachrony within Synchrony* (Duisburger Arbeiten zur Sprach- und Kulturwissenschaft), ed. by M. Morrissey and G. Kellermann (Peter Lang, Frankfurt/M).

J.J. Ohala and D. Feder (1987), "Listeners' identification of speech sounds is influenced by adjacent 'restored' phonemes", *Proc. 11th Internat. Congress of Phonetic Sciences*, Tallinn, Estonia, USSR, Vol. 4, pp. 120–123.

J.J. Ohala and E.E. Shriberg (1990), "Hyper-corrections in speech perception", *Proc. ICSLP 90 (Internat. Conf. on Spoken Language Processing*, Kobe, 18–22 November 1990), Vol. 1, pp. 405–408.

M.J. Solé and J.J. Ohala (1991), "Differentiating between phonetic and phonological processes: The case of nasalization", *Proc. 12th Internat. Congress of Phonetic Sciences*, Aix-en-Provence, 19–24 August 1991, Vol. 2, pp. 110–113.

W.S.-Y. Wang and C.-C. Cheng (1977), "Implementation of phonological change: The Shuang-Feng Chinese case", in *The Lexicon in Phonological Change*, ed. by W.S.-Y. Wang (Mouton, The Hague), pp. 148–158.

W.S.-Y. Wang and C.J. Fillmore (1961), Intrinsic cues and consonant perception", *J. Speech Hearing Res.*, Vol. 4, pp. 130–136.

T. Young (1819), "Remarks on the probabilities of error in physical observations", *Trans. Cambridge Philos. Soc.* (Reprinted in *Miscellaneous Works of the late Thomas Young, M.D., F.R.S., etc.* (John Murray, London, 1855), Vol. 2, pp. 8–28.