THE COGNITIVE REPRESENTATION OF SPEECH
T. Myers, J. Laver, J. Anderson (editors)

Articulatory Constraints on the Cognitive Representation of Speech

JOHN J. OHALA

University of California, Berkeley

INTRODUCTION

This is a story in the Sherlock Holmes tradition.[1] As every reader of the Holmes stories knows, the detective's greatest challenge was discovering the activity of the master criminal Professor Moriarty. Professor Moriarty was what we in America would call a "godfather": a behind-the-scenes director of a vast criminal organization—so much behind-the-scenes that only Holmes could deduce his presence and trace a seemingly isolated crime to his doorstep. Holmes' main tools were his knowledge of the world (cultivated by somewhat exotic research), his deductive powers, and his imagination. In spite of his dabblings in chemistry and such, he relied rather little on sophisticated technology; the simple magnifying glass was his principal technical aid.

This is a fairly good model for the task that faces us in trying to discover the

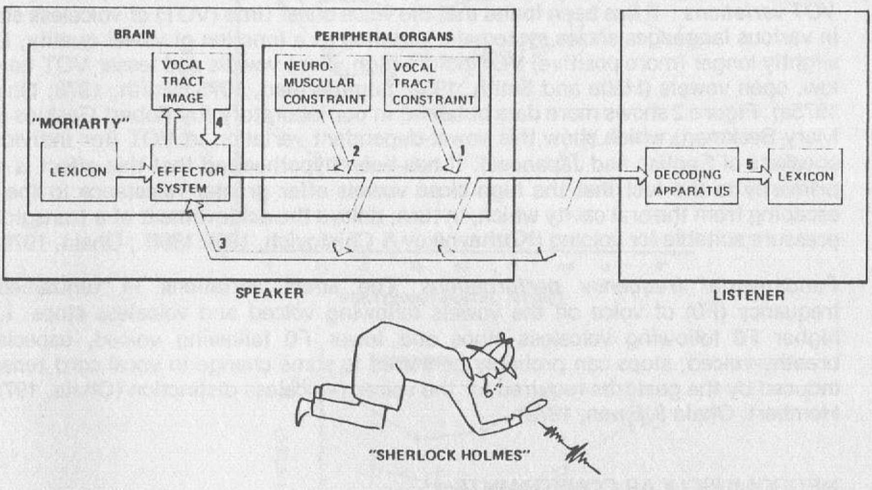

FIGURE 1 *The five routes whereby articulatory constraints can leave their imprint on speech: (1) vocal tract constraints (2) neuromuscular constraints, (3) feedback, (4) feedforward, (5) sound change. In imitation of Sherlock Holmes we must examine the speech signal, the product of these many mechanisms, and isolate the contribution of the lexicon.*

cognitive representation of speech. Somewhere up in the brain there exists the ultimate authority on how to pronounce words; we call this the lexicon. This lexicon—my Moriarty analogue—does its work indirectly through neurological, muscular, and articulatory agents. We are in the position of Holmes in that we can observe the activity of some of these agents and have to deduce the master*mind* behind it all. Like Holmes, I believe that imagination, not engineering, will be our most effective tool in this task.

The particular subquestion I address in this paper is: how is the representation of speech in the brain, i.e., its lexical form, affected by articulatory constraints? We can't really answer this question, however, until we know what the lexical representation of speech is, and we can't know that until we can examine the speech in its acoustical and physiological form and differentiate those features due to lexicon from those added on further downstream. So let's attack that question first.

I think it is possible to identify at least five routes whereby "down-stream" constraints may influence pronunciation. Our task is to examine various features of pronunciation, and assign them to one of these five sources or to the lexicon. I will briefly list and offer examples — in some cases hypothetical — of each of these five. (In the discussion to follow it will be helpful to refer to Figure 1).

VOCAL TRACT MECHANISMS

Some features of the speech signal are probably present due to purely mechanical, i.e., physical, constraints of the vocal tract, e.g., inertial properties of the articulators, the anatomical connections between articulators, the elasticity of articulator tissues, aerodynamic factors, etc. I will discuss two probable examples.

VOT variations It has been found that the voice onset time (VOT) of voiceless stops in various languages shows systematic variation as a function of vowel quality, i.e., slightly longer (more positive) VOT before high, close vowels and lesser VOT before low, open vowels (Halle and Smith, 1952; Summerfield, 1975; Smith, 1978; Ohala, 1975a). Figure 2 shows more data obtained in our laboratory (by Robert Gaskins and Mary Beckman) which show this vowel-dependent variation in VOT (for individual speakers of English and Japanese). It has been hypothesized that this effect is due primarily to the fact that the high close vowels offer greater resistance to the air escaping from the oral cavity which, in turn, delays the achievement of a transglottal pressure suitable for voicing (Kozhevnikov & Chistovich, 1965:186ff.; Ohala, 1976).

Fundamental frequency perturbations The small variations in fundamental frequency (F0) of voice on the vowels following voiced and voiceless stops, i.e., higher F0 following voiceless stops and lower F0 following voiced, especially breathy-voiced, stops can probably be traced to some change in vocal cord tension induced by the gestures required for the voiced/voiceless distinction (Ohala, 1978b; Hombert, Ohala & Ewan, 1979).

NEUROMUSCULAR CONSTRAINTS

Another possible influence on the form of speech is limitations of the neuromuscular system, e.g., the frequency response of the muscles, the transmission time of neural impulses, etc.

The time constant in feedback loops Kozhevnikov & Chistovich (1965:93), following Lashley (1951), suggested that the time it would take to issue a 'command'

to a speech muscle and then receive sensory information on the resulting movement would be something on the order of 100 msec. From this basis they argued that feedback could not be used to regulate the timing of speech since speech gestures follow one another at intervals shorter than 100 msec.

Setting the lower limit of the speed of the efferent-afferent loop to around 100 msec in the case of speech seems to be unjustifiably high, especially when one considers that there are some very fast reflexes served by the cranial nerves: the eye blink reflex (< 10 msec) and the masseter reflex (~ 12-15 msec); see Ohala (1970:124) for a review of this issue.

Inherent phase lag between motor cortex and speech muscles Lenneberg (1967:96) proposed that because of presumed variations in the transit time of nerve impulses over the various nerves serving the speech muscles there could be a 'phase lag' (synchronization error) of some 30 msec between the muscles near the brain (e.g., tensor palatini) and those distant from the brain (e.g., posterior cricoarytenoid) and that this would necessitate a rather complex temporal staggering of the delivery of these motor commands at the center. However, there seems to be little empirical support for this conjecture. Ohala (1970:129ff.) reviewed this issue, made his own estimates of the transit time of neural impulses over the cranial nerves, and arrived at estimated synchronization errors of from 5 to 9 msec — too small to make much of a difference. Subsequent measurement of neural impulse speed over the recurrent nerve by Flisberg and Lindholm (1970) coincided

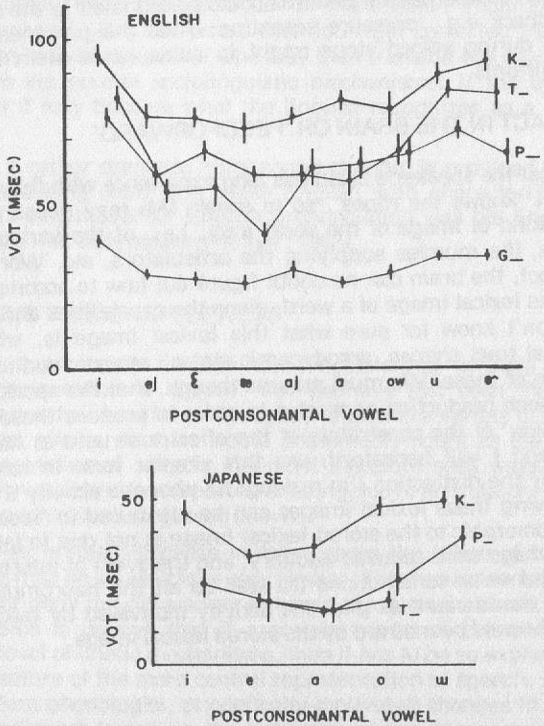

FIGURE 2 ***Voice onset time (VOT) for English (top) and Japanese (bottom) stops in the environment of different vowels.***

remarkably well with Ohala's estimates and tended to support the lower value for synchronization error.

Additional examples could be added, e.g., Lenneberg's hypothesis that all speech articulation is superimposed on a 6Hz oscillation in the brain's electrical state, which claim has also been found to be empirically unsupported (Ohala, 1970:156ff, 1975b).

More recent work in this area has progressed beyond pure speculation, e.g., Hirose, Ushijima, Kobayashi, and Sawashima (1969) and MacNeilage and his associates (MacNeilage, Sussman, Westbury, and Powers, 1979) are contributing important data on the response time and other dynamic properties of specific speech muscles and the motoneurones that drive them.

FEEDBACK

Discrepancies between the centrally intended activity and the activity at the periphery can be sensed via feedback and trigger compensatory activity. The following is a hypothetical example.

Voiced stops are difficult to produce because of the small volume of the oral cavity and limited compliance of its surfaces (Passy, 1890: 161; Ohala and Riordan, 1979). These conditions make it inevitable that the glottal air flow necessary for voicing will lead to the build-up of oral pressure which in turn will lead to the cessation of voicing. The extinguishing of voicing can be avoided though, if the speaker makes more room for the glottal air flow, e.g., by lowering the larynx. Presumably the need to implement such compensatory gestures could be signaled to the speaker's brain by sensory feedback e.g., pressure sensation. Thus the presence of purposeful larynx lowering during voiced stops might in some cases stem indirectly from articulatory constraints.

THE VOCAL TRACT IN THE BRAIN OR 'FEEDFORWARD'

It is obvious that the speaker's brain has long experience with the vocal tract and its constraints; it 'knows the ropes', so to speak. We may guess that the brain possesses some kind of image of the vocal tract, i.e., of the various articulators, their connections, the muscles supplying the articulators, etc. Working with this 'model' of the tract, the brain can no doubt figure out how to accomplish its goals, i.e., reproduce the lexical image of a word, given the capabilities and constraints of the tract. We don't know for sure what this lexical image is, whether muscle contractions, vocal tract shapes, aerodynamic states, acoustic-auditory shapes, or some combination of these. We must assume though, that the speaker knows how to control the speech production apparatus in order to produce these images. This type of 'anticipation' of the constraints of the effector structures has been called 'feed forward', and I will henceforth use this simpler term in referring to this concept. To clarify the distinction I'm making: all phonetic activity that is done for the sake of achieving these lexical images can be attributed to 'feedforward': the action which is isomorphic to the stored lexical image is not due to feedforward. If, for example, the image were acoustic-auditory, and the event of interest were a [w], then the labial and velar constrictions (as well as all the neuromuscular activity underlying them) would count as phonetic activity motivated by feedforward. The lowered F1 and F2 would be motived by the stored lexical image.

SOUND CHANGE

An indirect but extremely interesting mechanism whereby the constraints of the vocal tract may influence even the most central representation of speech is sound change. I have reviewed above some ways that the purely mechanical constraints of the speech production apparatus can leave its imprint on speech. We can think of these as distortions of the speech signal since they are unintended by the speaker. In some cases, the listener may not be able to figure out which features of the speech signal are intended and which are unintentional distortions. When repeating what he heard, he may produce intentionally what was previously unintentional (Ohala, 1974, 1975c, 1978a, 1979; Hombert, et al. 1979; Ohala and Lorentz, 1977). This process is represented schematically in Figure 1, where it is indicated that the speech signal reaching the listener consists of the lexical image plus detectable consequences of all the previously discussed four mechanisms: vocal tract and neuromuscular constraints, feedback and feedforward. This 'distortion'-encrusted signal may form the basis of the listener's lexical image of the word.

The philological literature is filled with examples of this process. As mentioned above, the F0 of vowels varies systematically after voiced and voicless consonants. This mechanically-caused distortion has been reinterpreted centuries ago by the speakers of Chinese, Thai, Vietnamese, and many other languages, as intentional tonal distinctions (Hombert, et al. 1979).

I believe this type of misperception or misjudgment, if you will, on the part of the listener regarding what the intended parts of the speech signal are, goes on all the time — potentially, at least, every time one person speaks and another listens. These I have called 'mini-sound changes'. Most such errors are eventually corrected (because the listener has other sources of information regarding the 'true' pronunciation of words). The few errors that don't get corrected may come to be characteristic of an individual speaker who may then transmit the new pronunciation to other speakers via normal sociolinguistic mechanisms. If the transmission is extensive enough it may become what the linguist recognizes as a sound change proper.

An example of a rather dramatic 'mini-sound change' is provided by a case of a normal Spanish-speaking child who spoke English with characterisitcs of a cleft palate because his only model for English pronunciation was his English-speaking playmate who had a real cleft palate (Klinger, 1962).

DIFFERENTIATING THESE INFLUENCES

To recapitulate: our task is to find out how articulatory constraints may influence the cognitive representation of speech. But we have just seen that articulatory constraints may leave their imprint on the speech signal in a variety of ways and only one, sound change, affects what we would presumably count as the most central, the lexical, representation of speech. We must therefore seek ways to assign any given aspect of pronunciation — suspected of being due to articulatory constraints — to one of the five mechanisms discussed above.

This is a complex problem because in many cases the action of two or more of these mechanisms may be superficially very similar. This has led some writers on the subject to ignore or deny the differences between these mechanisms. For example, Harms (1973) seems to believe that if a certain aspect of pronunciation *can be* attributed to low-level phonetic mechanisms, then it *has to* be so explained; it cannot be counted as a feature of the more central representation of speech. That is, in the jargon of the modern phonologist, phonetically-motivated changes in pronunciation are not to be considered 'rules' in the grammar of the language. Donegan and Stampe (1979) on the other hand, seem to deny that any phonetically-motivated aspects of pronunciation, e.g., stop voicing, vowels nasalizing before nasals, can occur without contribution of the highest levels of the speaker's cognitive system;

for them, presumably, there can be no purely mechanical distortions of the speech signal. I hope to show below that both views oversimplify a complex situation, and, furthermore, that questions about the cognitive/mechanical causes of certain aspects of pronunciation should be decided by empirical evidence, not by decree.

VOCAL TRACT MECHANICS

There are a variety of ways which have been used to reveal whether a certain aspect of pronunciation is present due to vocal tract mechanics.

Control independent variable; observe effect on dependent variable One set of techniques has in common the following principle: one can remove or alter a given vocal tract constraint (the independent variable) and see if the particular feature of interest (the dependent variable) disappears or changes. If it does then it probably owes its existence to the constraint; if there is no change, then it is probably unrelated to the constraint.

Vowel height and vowel duration It has been hypothesized that the often-found direct correlation between vowel duration and the openness of the vowel, that is, open vowels longer than close vowels (other things being equal), stems from the fact that the jaw has to travel a greater distance for open vowels and this takes longer (Lindblom, 1967). This hypothesis was tested by Nooteboom and Slis (1970) by having subjects speak test words normally and with their jaw clenched, i.e., by removing the hypothesized cause of the effect. As it turned out, the variation in vowel duration was the same whether the jaw was closed or not. Thus the vowel duration differences could not be attributed to extent of jaw movement (although it leaves open the questions of whether the movement of other articulators, e.g., the tongue, may be the cause and whether the origins of the effect, via sound change, may not have been the physical properties of the jaw).

Vowel height and F0 As is well known, there is a small but systematic correlation between the height of a vowel and its average F0. There have been at least two hypotheses which have attempted to account for it by peripheral factors. One attributes it to the pull of the tongue on the larynx (Ohala, 1972, 1977, 1978b). Ohala and Eukel (1978) found that by having subjects speak with their jaw propped open by bite blocks (and thus presumably making the tongue pull more forcefully on the larynx during the production of high vowels), the difference in F0 between high and low vowels was enhanced. They interpreted this result as favoring the tongue-pull hypothesis. In this case, a change in the dependent variable, tongue pull, was used to support a claimed causal relationship between a peripheral constraint and an aspect of pronunciation.

Many other studies of this type could be cited (Kozhevnikov & Chistovich, 1965: 186ff; Houde, 1968:88ff; Putnam, Doherty & Shipp, 1976; Ohala and Riordan, 1979).

Looking 'upstream' from the vocal tract Another class of studies uses electromyography (EMG) for the purpose of finding out whether a certain aspect of speech originates in the vocal tract or higher 'upstream'. If the EMG records show that certain muscles actively contribute to a given gesture, then one can conclude that the gesture has a more central origin and is not caused by a purely mechanical factor.

F0 variation in sentence intonation Lieberman (1967) claimed that the fall in F0 at the end of declarative sentences was purely a mechanical consequence of falling subglottal air pressure, itself a consequence of the approaching inspiratory phase of

breathing. The finding, from various EMG studies, that such rapid drops in F0 are invariably accompanied by contraction of one or more of the laryngeal strap muscles (along with other evidence) helped to disprove Lieberman's claim. See Ohala (1970, 1977, 1978b) for a review of this issue.

Velic elevation vs. vowel height It has been hypothesized that the well known correlation between the degree of soft palate elevation and the 'height' of the vowel is due to some sort of mechanical pull of the tongue on the soft palate (or, possibly, vice-versa), presumably via the palatoglossus muscle (Moll, 1962). This hypothesis can be discarded however, since several EMG studies (e.g., Lubker, 1968; Fritzell, 1969) found that the muscles controlling soft palate position are themselves responsible for the vowel-specific variation (cf. Ohala, 1974, 1975c).

Other methods of estimating the mechanical constraints of the articulators Some amount of coarticulation (the values of one or more articulatory parameters of one segment 'spilling over' onto neighboring segments) is probably inevitable due to the inertia of the articulators. How much co-articulation is physically necessary can be estimated either by measuring the frequency response of various articulators or by finding the minimum time required to accomplish specific speech gestures.

In an informal study some years ago, Valerie Mamini and I attempted to estimate the frequency response of my own jaw during speech and non-speech. Jaw

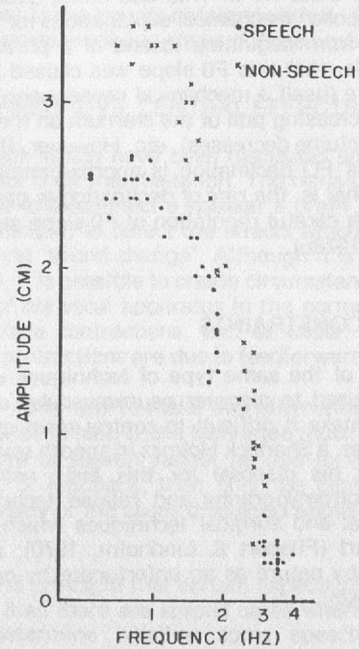

FIGURE 3 *Frequency response of one subject's jaw measured with speech gestures* ([sasasa. . .]) *and non-speech (opening and closing of the jaw). Amplitude scale, which refers to separation between the incisors, is approximate.*

movement was transduced photo-electrically (Ohala, Hiki, Hubler, and Harshman 1968). Figure 3 shows a plot of jaw movement amplitude vs. frequency (1/t, where t = time taken to make one complete opening-closing movement). From this and the results of studies which show that the modal rate of syllable production is about 4/sec (Ohala, 1975b), we can conclude that the jaw must be operating at or near its frequency limit in speech. This would tend to explain the necessity for synergistic action in speech between the jaw and the articulators attached to it, the lower lip and the tongue (Kozhevnikov & Chistovich, 1965: 184; Takahashi and Nigauri, 1962; Lindblom and Sundberg, 1971a; Folkins and Abbs, 1975).

Clumeck (1976), in a cross-language study of velic movement using the nasograph, found that the minimum time required to move from closed to open velopharyngeal port was on the order of 30 to 50 msec. This being the case, we may judge that the extreme anticipatory lowering of the soft palate in American English on vowels before nasal consonants (often 250 msec before the nasal) is *not* caused solely by inertial characteristics of the soft palate.

In a similar vein, the observation that such anticipatory velic opening can be blocked by word or syllable boundary (Ohala, 1971) also argues against such vowel nasalization being due to mechanical characterisitcs. (Cf. similar arguments by Ghazeli (1977:113) regarding coarticulated pharyngealization in Arabic.)

Correlation Analysis If a hypothesis is stated in the form: 'a variation in *x* causes, through mechanical factors, a proportional variation in phonetic feature *y*', then the testing of the hypothesis can be done by seeing whether the stated correlation holds over a wide range of values for *x*. This was done, qualitatively, in testing various hypotheses which proposed mechanical explanations for the 'declination effect', i.e., the slow drop in F0 from beginning to end of a phrase or breath group. It was proposed, for example, that this F0 slope was caused by the gradual lowering of subglottal air pressure (itself a mechanical consequence of respiratory constraints) or by the gradually increasing pull of the sternum on the larynx, via the sternohyoid muscle (as the chest volume decreases), etc. However, there is widespread evidence that the magnitude of FO declination is approximately constant over phrases of different durations. That is, the *rate* of declination is greater for short phrases than for long phrases. Such careful regulation of F0 slope argues for its being centrally controlled (cf. Ohala, 1978b).

NEUROMUSCULAR CONSTRAINTS

In principle, many of the same type of techniques used to identify vocal tract constraints could be used to discover neuromuscular constraints, but ethical and technical limitations make it difficult to control most of the relevant independent variables. Nevertheless, a Sherlock Holmes in speech science would still find enough investigative tools at his disposal for this task: using noninvasive techniques (McClean, 1978), electromyography and related techniques (MacNeilage, et al. 1979), pharmacological and surgical techniques which are a regular part of the physician's healing art (Flisbert & Lindholm, 1970), and finally, exploitation of 'controls' introduced by nature as an unfortunate by-product of disease or injury (Critchley & Kubik, 1925).

FEEDBACK

One obvious way to find out if a given speech gesture is influenced by feedback or not is to interrupt the feedback loop (or reduce the amount of sensory information getting back to the brain) and seeing if the given gesture changes in some way (Ringel and Steer, 1963, Ohala, 1975b). This cannot be used conveniently for all possible feedback channels from all parts of the vocal apparatus, however.

A second method involves the detection of a more or less fixed delay between two events, where it is sensory information from the first event that is suspected of triggering or modulating the second event. This technique was used convincingly by Kozhevnikov & Chistovich (1965:133) in suggesting that the second of two closely spaced consonantal movements must have been triggered by sensory information from the release of the first. Another classic instance of this, although not concerned with speech as such, is the study of Sears and Newsom Davis (1968). They had a subject blow into a tube with an initially high resistance while maintaining a constant pressure which the subject could monitor by seeing the output of a pressure transducer. The activity of the expiratory muscles was monitored via EMG. When the resistance of the tube was suddenly and unexpectedly changed, a corresponding compensatory change in the activity of the breathing muscles was detected within 33 to 80 msec, i.e., a delay time so short that pressure sensing mechanisms within the respiratory tract must have provided the initial triggering of the corrective action.

A third, potentially interesting, way to discover whether feedback was used in the control of the timing of speech was introduced by Kozhevnikov & Chistovich (1965:94ff). They used statistical analysis of the temporal variation of the various speech intervals in a sentence repeated many times. As noted by Ohala (1975b) and Ohala and Lyberg (1976) there are many problems with this type of analysis which call into question most of the results obtained with it. Cooper, Sorensen, and Pacia (1977) have made progress in solving some of these problems.

DIFFERENTIATING PERIPHERAL VS. CENTRAL EFFECTS

Those influences on speech which have been discussed so far require an actual physical human vocal tract for their manifestation. The remaining influences do not. If it were somehow possible to disconnect a speaker's brain from his vocal tract it ought to be possible, in principle, to detect the effects which I have put under the headings of 'feedforward' and 'sound change'. Although it is obviously not possible to decerebrate our speakers, it is possible to create circumstances where the speaker doesn't need to use parts of his vocal apparatus in the normal way; if the speaker nevertheless continues muscle contractions, etc. as usual, this can be taken as evidence that those muscle contractions are due to feedforward (Folkins, 1979).

Elimination of alternatives One way to make this differentiation is the process of elimination of alternatives and, in fact, it is a technique much employed by Holmes. If tests show a given feature of speech cannot be due to vocal tract constraints, neuromuscular constraints, or feedback, then only the central causes remain as viable explanations for it. Many of the cases discussed above have been decided in just this way.

Timing If we can accept the principle that certain aspects of the timing of speech are centrally programmed, then there are special cases where we can use this as a touchstone to help us determine whether some speech event is centrally programmed or simply the product of peripheral constraints.

Consider the case of the intrusive stops in words such as 'warmth' [wɔɹmpθ], 'length' [lɛŋkθ] etc. These stops may be produced unintentionally due to an anticipatory denasalization (and devoicing) of the latter portion of the nasal consonant. Through sound change, however, these unintended stops may come to be intended features of the pronunciation of these words, as the written 'p' suggests in such words as 'Thompson', 'dempster', 'glimpse', (cf. 'Clemson', 'teamster', 'rinse'; Grandgent, 1896). The durational characteristics of such words may be able to reveal whether in any given word, in the speech of any given speaker, such stops are intended (centrally-programmed because of the action of sound change) or unintended (due to physical phonetic constraints). For a concrete example, consider the word 'teamster'[ˈtʰĩmpstəɹ] We know from numerous studies that in English VN sequences are longer in open syllables than in closed syllables, especially those closed with voiceless obstruents (Lehiste, 1970, Lovins, 1978). Therefore, if the [ĩm] sequence in a given speaker's rendering of this word is characteristically short (in comparison to some reference duration), we could conclude the [p] was a 'true' [p] (an 'underlying' /p/, if one prefers such jargon). If, on the other hand, it was characteristically long, we could probably conclude the [p] was unintentional (a 'surface' [p]).

To demonstrate the viability of this approach, the following study was done: twenty-five American English speakers were asked to form, orally, several novel English words by adding suffixes to existing words. In the early part of the interview the form 'clam + ster' was elicited ("Please add the suffix [stəɹ] to 'clam'."). At the end of the interview the form 'clamp + ster' was elicited. Since these were completely new words, I felt that whatever [p] might show up in 'clamster', it would not yet have a chance to be a fossilized [p]; it must of necessity be unintentional. In 'clampster', on the other hand, the [p] must be intentional. Would the durations of the [æ̃m] sequences in the two words be different as predicted? The results, based on the subset of tokens that were measurable, are shown in Figure 4.

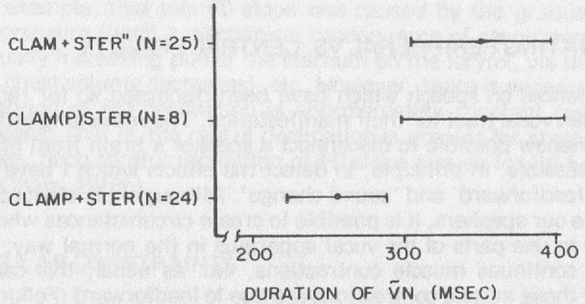

FIGURE 4 *Duration of* [æ̃m] *sequence in 25 utterances of* clamster *(top), in 8 tokens of the previous 25 which exhibited an epenthetic* [p], *i.e.,*[kʰlæ̃mpstəɹ] *(middle), and in 24 tokens of* clampster *(bottom). Horizontal lines indicate the standard deviation associated with each distribution.*

The prediction was borne out. The mean duration of the VN sequence in 'clamster', whether it was said with a clear [p] or not, was about 95 msec longer than the VN sequence in 'clamster', where the [p] was always present because it was intended. In spite of the fact that no normalization of these durations was attempted to correct for inter-subject variations in speaking rate, etc., this difference is highly significant, $p < 0.001$.

Word games and speech errors Evidence relevant to this issue might be provided by word games and speech errors but there are difficulties in interpreting the results.

Chao (1934) reports that a common word game used by speakers of Mandarin separates word initial consonants from their following vowels by inserting a sequence such as [aik] between them. By this process, [pei] becomes [paikei] but [mi] becomes [meitɕi]This does reveal that the phonetically plausible change of [ki] to [tɕi] is an automatic one but it doesn't unambiguously reveal the level — whether central or peripheral — that this process occurs at. Hombert (1976) provides a review and critique of the use of word games in phonology.

Much the same difficulty surrounds the interpretation of speech error data.

DIFFERENTIATING FEEDFORWARD AND SOUND CHANGE

There remains the task of trying to differentiate speech events motivated by feedforward as opposed to those specified in lexical storage, which can itself be influenced by sound change.

It seems to me that this problem can be approached if we first know the form of lexical representations, i.e., whether they are auditory-acoustic, articulatory, etc.

Nature provides some evidence on this issue. Speakers can often overcome rather severe articulatory handicaps and still achieve intelligible speech. The prognosis is not as good for those born deaf, however. This might suggest that the lexical image is acoustic-auditory, not articulatory. Unfortunately, Nature does not bother to run carefully controlled experiments and so the results of her tamperings with human speakers do not unambiguously and unqualifiedly support one view of the other. See Riordan (1978) for a review.

One basic technique of use here is a classic one used by physiologists to study homeostasis in living systems: one creates obstacles for normal behavior and then observes what aspects of behavior remain the same in spite of the obstacle and what aspects change, adapting to the obstacle. Those that remain unchanged are fair candidates for being isomorphic to the ultimate goal of the organism — in our case, the specification of speech in the lexicon (Bernstein, 1967).

A variety of recent experiments, e.g., those by Lindblom and Sundberg (1971a) and Riordan (1977), which show that speakers can quickly overcome artifical articulatory handicaps, seem to rule out conclusively that the lexical image is in the form of an invariant configuration of the articulators.

SOUND CHANGE VS. EVERYTHING ELSE

I think the influence of sound change on speech can, in many cases, be differentiated from other influences by considering the notion of *relatedness* between two or more speech events. In the case of effects due to the purely mechanical or neuromuscular constraints we can say that some speech event, say the F0 perturbation after obstruents, is necessarily related to some other speech event, in the example given, to the voicing distinction on the obstruents. In the case of effects determined by feedback or feedforward, one speech event, say the lowering of the larynx during a voiced obstruent, is necessarily related to another speech event, the continued voicing during the obstruent. Effects due to sound change, however, may lack any necessary relation or connection to other speech events immediately present in the chunk of speech under consideration. For example, the tonal distinctions which developed in Chinese and other languages from the F0 perturbations caused by the voicing distinction of syllable initial obstruents, are now preserved independently of the voicing distinction. In the majority of such cases, in fact, the voicing distinction has been lost from initial obstruents. Likewise, the nasalization of vowels in French, Hindi, Yoruba, and many other languages now exists completely independently of neighboring nasal consonants, which, centuries ago, were the initial cause of the nasalization and which have now been lost.

The investigation of the effects of sound change on speech must start with a knowledge of all the other influences on speech, i.e., vocal tract mechanics, feedback, etc. Before we could make sense of the sound change which led to the development of tones in Chinese, we had to know that the voiced/voiceless distinction on obstruents produces small F0 perturbations even in non-tonal languages such as English and French. In this sense, part of the study of sound change — today, at least — is based in the laboratory. The other part, however, requires the gathering of evidence that the pronunciation of the language has changed. This evidence may come from dialect studies or studies of distinct but genetically related languages, or from language-internal evidence (so-called 'internal reconstruction'), supplemented where possible, by ancient written materials.

CONCLUSION

Although it cannot be said that all aspects of speech can be properly accounted for, i.e., attributed to the lexical representation, to vocal tract or neuromuscular constraints,[2] to feedback or feedforward, we can probably return to the original question 'how can articulatory constraints influence the lexical representation?' and find that it has already been answered. Through *sound change*, any downstream activity — the creakings of the encoding machinery —, that can be detected by the listener may get fixed in his lexicon. Looking at sound patterns cross-linguistically it is difficult to escape the conclusion that this is exactly what has happened to all human languages. To turn our original analogy around, if a Sherlock Holmes were to have presented to him all the structure and details of the lexicon and from this had to discover the properties of the encoding mechanisms, he could probably do fairly well at it. The reason is that our Moriarty figure, the lexicon, keeps getting replaced in generation after generation by a new Moriarty and each new one gets his information only through the agents of the previous one, not directly from the boss himself. Over the centuries then, it is inevitable that the information possessed by these Moriartys should bear the imprint of the agents as much as that of the boss.

Notes

1. This theme was deemed appropriate for a symposium in Edinburgh since Conan Doyle, the author of the Sherlock Holmes stories, was born in Edinburgh and studied there. In fact, the apparent real-life model for Holmes was also a native of Edinburgh, a medical doctor named Joseph Bell who taught medicine at Edinburgh University and amazed his students with his deductive abilities when making diagnoses. May his memory — and the inspiration it provides — live on.

2. Entirely missing from this discussion are auditory constraints. These are covered elsewhere in this volume.

Acknowledgements This work was supported in part by the National Science Foundation and the Committee on Research, University of California, Berkeley.

REFERENCES

Bernstein, N.A. 1967. The coordination and regulation of movements. Oxford: Pergamon Press.

Chao, Y.-R. 1934. The non-uniqueness of phonemic solutions of phonetic systems. Bulletin of the Institute of History and Philology, Academia Sinica 4. 4.363-397.

Clumeck, H. 1976. Patterns of soft palate movements in six languages. J. Phon. 4.337-351.

Cooper, W.E., Sorensen, J.M., and Paccia, J.M. 1977. Correlations of duration for nonadjacent segments in speech: aspects of grammatical coding. JASA 61.1046-1050.

Critchley, M. and Kubik, C.S. 1925. The mechanism of speech and deglutition in progressive bulbar palsy. Brain 48.492-534.

Donegan, P.D. and Stampe, D. 1979. The study of natural phonology. In: D.A. Dinnsen (ed.), Current approaches to phonological theory. Bloomington: Indiana Univ. Press. Pp. 126-173.

Flisberg, K. and Lindholm, T. 1970. Electrical stimulation of the human recurrent laryngeal nerve during thyroid operation. Acta Otolaryngologica 263.63-67.

Folkins, J.W. 1979. Masseter, temporalis, and medial pterygoid activity with the mandible free and fixed. Proc. 9th Int'l. Cong. of Phonetic Sciences, Copenhagen. Vol. 2.337-343.

Folkins, J.W. and Abbs, J.H. 1975. Lip and jaw motor control during speech: response to resistive loading of the jaw. JSHR 18.207-220.

Fritzell, B. 1969. The velopharyngeal muscles in speech. Acta Otolaryngologica. Suppl. 250.

Ghazeli, S. 1977. Back consonants and backing coarticulation in Arabic. Unpublished doctoral dissertation, University of Texas.

Grandgent, C.H. 1896. Warmpth. PMLA 11 (New Series 4).63-75.

Halle, M. and Smith, C.P. 1952. Distinctive features of voiceless stop consonants. Q.P.R., R.L.E. (MIT)27.48-53.

Harms, R.T. 1973. Some non-rules of English. Indiana Linguistics Club.

Hirose, H., Ushijima, T., and Sawashima, M. 1969. An experimental study of the contraction properties of the laryngeal muscles in the cat. Annals of Otolaryngology 78.297-307.

Hombert, J.M. 1976. Word games: some implications for analysis of tone and other phonological processes. Working Papers in Phonetics (UCLA) 33.67-80.

Hombert, J.M., Ohala, J.J., and Ewan, W.G. 1979. Phonetic explanations for the development of tones. Lang. 55.37-58.

Houde, R.A. 1968. A study of tongue body motion during speech sounds. SCRL Monog. 2.

Klinger, H. 1962. Imitated English cleft palate speech in a normal Spanish speaking child. JSHD 27.379-381.

Kozhevnikov, N.A. and Chistovich, L.A. 1965. Speech: articulation and perception. U.S. Dept. of Commerce. JPRS No. 30543.

Lashley, K.S. 1951. The problem of serial order in behavior. In: L.A. Jeffress (ed.), Cerebral mechanisms in behavior. New York: Wiley. Pp. 112-136.

Lehiste, I. 1970. Suprasegmentals. M.I.T. Press.

Lenneberg, P. 1967. Biological foundations of language. New York: Wiley.

Lieberman, P. 1967. Intonation, perception and language. Cambridge: M.I.T. Press.

Lindblom, B. Vowel duration and a model of lip-mandible coordination. STL-QPSR 4.1-29.

Lindblom, B. and Sundberg, J. 1971. Neurophysiological representation of speech sounds. Paper presented at 15th World Congress of Logopedics and Phoniatrics, Buenos Aires.

Lovins, J.B. 1978. A study of "nasal reduction" in English syllable codas. Revised version of paper in: Chicago Linguistic Society Proc. 14.

Lubker, J. 1968. An electromyographic-cinefluorographic investigation of velar function during normal speech production. Cleft. Pal. J. 5.1-19.

MacNeilage, P.F., Sussman, H.M., Westbury, J.R., and Powers, R.K. 1979. Mechanical properties of single motor units in speech musculature. JASA 65.1047-1052.

McClean, M. 1978. Variation in perioral reflex amplitude prior to lip muscle contraction for speech. JSHR 21.176-284.

Moll, K.L. 1962. Velopharyngeal closure on vowels. JSHR 5.30-37.

Nooteboom, S.G. and Slis, I. 1970. A note on the degree of opening and the duration of vowels in normal and "pipe" speech. IPO Ann. Rep. 5.55-58.

Ohala, J.J. 1970. Aspects of the control and production of speech. Working Papers in Phonetics (UCLA) 15.

Ohala, J.J. 1971. Monitoring soft palate activity in speech. POLA (Berkeley) 13.J01-J015.

Ohala, J.J. 1972. How is pitch lowered? JASA 52.124.

Ohala, J.J. 1974. Experimental historical phonology. In: J.M. Anderson and C. Jones (eds.), Historical linguistics II: theory and description in phonology. Amsterdam: North Holland. Pp. 353-389.

Ohala, J.J. 1975a. Conditions for vowel devoicing and frication. JASA 58.S39.

Ohala, J.J. 1975b. The temporal regulation of speech. In: G. Fant and M.A.A. Tatham (eds.), Auditory analysis and perception of speech. London: Academic Press. Pp. 431-453.

Ohala, J.J. 1975c. Phonetic explanations for nasal sound patterns. In: C.A. Ferguson, L.M. Hyman, and J.J. Ohala (eds.), Nasálfest. Stanford: Language Universals Project. Pp. 289-316.

Ohala, J.J. 1976. A model of speech aerodynamics. Report of the Phonology Laboratory (Berkeley) 1.93-107.

Ohala, J.J. 1977. The physiology of stress. In: L.M. Hyman (ed.), Studies in stress and accent. So. Cal. Occasional Papers in Ling. 4.145-168.

Ohala, J.J. 1978a. Southern Bantu vs. the world: the case of palatalization of labials. Berkeley Linguistic Soc., Proc. 4.370-386.

Ohala, J.J. 1978b. Production of tone. In: V.A. Fromkin (ed.), Tone: a linguistic survey. New York: Academic Press. Pp. 5-39.

Ohala, J.J. and Eukel, B.W. 1978. Explaining the intrinsic pitch of vowels. RPL 2.118-125.

Ohala, J.J., Hiki, S., Hubler, S. and Harshman, R. 1968. Photoelectric methods of transducing lip and jaw movements in speech. WPP (UCLA) 10.135-144.

Ohala, J.J. and Lorentz, J. 1977. The story of [w]. An exercise in the phonetic explanation for sound patterns. Berkeley Linguistic Soc., Proc. 3.577-599.

Ohala, J.J. and Lyberg, B. 1976. Comments on 'Temporal interactions within a phrase and sentence context' [Journal of the Acoustical Society of America 56, 1258-1265 (1974)]. JASA 59.990-992.

Ohala, J.J. and Riordan, C.J. 1979. Passive vocal tract enlargement during voiced stops. In: J.J. Wolf and D.H. Klatt (eds.), Speech Communication Papers, 97th ASA meeting. Pp. 89-92.

Passy, P. 1890. Etudes sur les changements phonétiques. Paris: Librarie Firmin-Didot.

Putnam, A.H.B., Doherty, E.T., and Shipp, T. 1976. Lip larynx interaction models for intervocalic /p/ production. JASA 60.S63.

Ringel, R.L. and Steer, M.D. 1963. Some effects of tactile and auditory alterations on speech output. JSHR 6.369-378.

Riordan, C.J. 1977. Control of vocal tract length in speech. JASA 62.998-1002.

Riordan, C.J. 1978. Acoustic aspects of speech production. Unpublished doctoral dissertation, Univ. of Essex.

Sears, T.A. and Newsom Davis, J. 1968. The control of respiratory muscles during voluntary breathing. In: A. Bouhuys (ed.), Sound Production in Man. Annals of the New York Academy of Sciences 155.1.183-190.

Smith, B.L. 1978. Effects of place of articulation and vowel environment on "voiced" stop consonant production. Glossa 12.163-175.

Summerfield, Q. 1975. Aerodynamics versus mechanics in the control of voicing onset in consonant-vowel syllables. Speech Perception (Belfast), Series 2, No. 4. 61-72.

Takahashi, H. and Nigauri, Y. 1962. Relation of vowel articulation to the shape of vocal tract, with special reference to the mouth opening and the tongue. Studia Phonologica 2.71-82.

(following papers by Bybee and Nearey)

John J. Ohala The concept of psychological reality and how to discover what is psychologically real has been made out to be a very subtle and/or complicated issue. I think this has been overdone and would like to attempt to simplify it by means of an extended analogy. Many of us are teachers. One of the tasks required of teachers is certification that students know certain academic material in order that they may receive a grade, a diploma, or a degree. In these cases, the teacher must, in effect, assess the psychological reality of students' knowledge or mental ability. How is this done? By obtaining some behavioral evidence, e.g., performance on a test, presentation of an original scholarly paper, etc., which evidence is consistent with the students having the knowledge or ability attributed to them. Conscientious teachers do not usually take the mere fact that the students have been exposed to the academic material as sufficient evidence that they know and understand the material. Of course, it is not easy to devise good tests and most of us therefore spend much time refining our tests. Nevertheless, I am not aware thay anyone in academia rejects, in principle, the need for such overt behavioral evidence in assessing students' knowledge and mental ability. The same practice should apply when we seek to verify the psychological reality of posited phonological constructs.

This analogy also permits us to evaluate a claim made by Fromkin at this symposium: that it is only necessary to use external evidence, including experimental evidence, to ascertain what a possible psychologically real phonological construct might be and that it is therefore not necessary to justify positing such constructs for each and every language whose sound pattern suggests the need for such constructs. If we were to apply this practice in education, which Fromkin advocates for phonology, it would mean that all we would have to do is find *one* student who demonstrated via test performance or papers that he knew the academic material and we could then assume without the necessity of behavioral evidence that all other students exposed to the same kind of material had likewise mastered it!

Discussion

(of papers by E. Selkirk and R. Ladd)

John J. Ohala Selkirk reports on a very interesting attempt to unify prosodic (sentence-level) and syllabic phonology using rather abstract, almost mathematical constructs. But if this is to be taken as something real, cognitively or otherwise, I would ask what recommends it over the dozens of other schemes offered to account for the same or similar data? Have the other schemes been disproved? In fact, none of them, including Selkirk's, have been the subject of empirical verification. As far as I can determine, none of the data cited, including that mentioned by Cutler in her remarks, unambiguously supports this model over any other.

I grant that the attempt to unify some of the phonological processes at the word and the phrase level is ingenious, but I wonder if all this machinery at the word level is necessary? A lexicon containing the pronunciation (including stress) of all existing words, both stems and derivations, can account for the data as well. The pronunciation of new words can be done by analogy (Ohala, 1974).

One of the examples mentioned in Selkirk's (pre-circulated) paper deserves careful examination. She notes that the word *rhythm* [rɪðm̩] with a syllabic [m̩] yields the derived forms *rhythmy* [rɪðm̩i] with the syllabic [m̩] retained, and *rhythmic* [rɪðmɪk] with the syllabicity of the [m] lost. The explanation offered for the different treatment of [m] is that the suffix -y [i] is a morphologically neutral suffix, and thus does not interact phonologically with the stem, whereas *-ic* [ɪk], being morphologically active, can affect the syllable structure of the stem. This explanation could be tested. We could examine speakers' pronunciation of new derivations of e.g., *prism* [prɪzm]: *prismy* and *prismic*. I can't say what the results will be but I feel I would render both with syllabic [m̩], i.e., [prɪzm̩i] and [prɪzm̩ɪk]. The pronunciation that would be predicted according to Selkirk, [prɪzmɪk], I would reject because I would doubt that my listeners could recover the stem *prism* from it. Only after extensive usage and long familiarity with such a form would I risk the pronunciation [prɪzmɪk] and then by the process of analogy, the models being *rhythmic, orgasmic, cataclysmic, logarithmic,* etc. There being no existing models (that I know of) for a pronunciation such as *[prɪzmi], i.e., with non-syllabic [m]. I would expect that *prismy* would always remain [prɪzm̩i] in spite of familiarity and long usage.

I find Ladd's paper well reasoned and, on the whole, quite convincing. Further evidence supporting his view that intonational contours tend to be language-specific and conventional comes from recent work by Larry Hyman and Jean-Marie Hombart on Cameroonian languages where, in some cases, they have found downdrift to be eliminated or constrained due to certain language-specific tonal traits (Hyman and Hombert, personal communication).

Nevertheless, it is still tempting to think that there is some kind of universal substrate on which language-specific uses of fundamental frequency are superimposed. Although it may not be an obsolute universal, it seems very common that uncertainty and lack of self-assurance is signalled with a generally high Fo whereas certainty, self-assurance, even agression, is signalled with low Fo. It is interesting to note (as I did in Ohala, 1970) that much the same use of Fo is found in the animal kingdom as well, e.g., among dogs, raccoons, etc. That is, in an encounter between individuals, a low-pitched growl signals self-assurance and aggression whereas a high-pitched squeal is used to signal appeasement, surrender, lack of assuredness. Being an amateur ethologist, I would speculate that in emitting a high-pitched sound, the animal is trying to imitate the necessarily high-pitched sound of the young of the species in order, perhaps, to elicit some kind of maternal or paternal response from his antagonist. I say 'necessarily high-pitched' because younger individuals, being physically smaller, will have less massive vocal cords (or syringeal flaps in the case of birds) and consequently higher Fo. If we are carrying this type of innate programming around inside of us, it would not be surprising if some of it manifested itself in our linguistic use of Fo. Unfortunately, I can't think of any simple way to test these speculations.