Title
Phonetic Universals and Phonological Change

Author
Javkin, Hector

Publication Date
1977

Phonetic Universals and Phonological Change

By

Hector Raul Javkin

A.B. (University of California) 1972
M.A. (University of California) 1974
C.Phil. (University of California) 1977

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Linguistics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

John J. Ohala

Johanna Nichols

Karl Zimmer

Committee in Charge

DEGREE CONFERRED JUNE 16 1978

To my parents

ACKNOWLEDGEMENTS

Many people have made this work possible, worthwhile, and, at times, exciting. I have learned a great deal from their work, their probing questions, their suggested answers to the questions I had. They include: Arthur Abramson, Stephen Baron, Madison Beeler, Peter Benson, Susan Bowman, Wallace Chafe, Mathew Chen, Joy Chuck, John Crothers, C.K. Chuang, Pierre Divenyi, William Ewan, Osamu Fujimura, Louis Goldstein, Mel Greenlee, Eduardo Gutentag, Mary Haas, Ervin Hafter, Gary Holland, Haruko Kawasaki, John Kingston, Peter Ladefoged, Leigh Lisker, Cheryl Longinotti, James Lorentz, Ian Maddieson, Kuniko Miyawaki, Yakov Malkiel, Lloyd Rice, Carol Riordan, Jesse Sawyer, Alice Schlicter, Janice Shafer, Donald Sherman, Carol Simoson, Timothy Smith, Mary Streeter, Graham Thurgood, Ralph Vanderslice, Jeanne van Oosten, Robert Van Valin, Kenneth Whistler, and Malcah Yaeger. They have been my teachers and colleagues. If you don't find your name and expected to, my apologies. If you're surprised to find it, it is here because of the inspiration of your work or your example, or because of help given so selflessly that you don't remember. Thank you.

Thanks go to the people who taught me the needed skills from the fields of dentistry and computers: Walter Allen, Dr. Green, Dr. Reoli, Michael Von Tiesenhausen, and to the technicians who built the instruments a kept them working: Chris Netterson and Steve Pearson.

From Jean-Marie Hombert and Robert Krones I learned most of what I know about experimentation. Steven Greenberg and James Wright provided numerous insights and explained to me many things which I had not understood. Jeri Jaeger provided invaluable comments and advice. I have been lucky to have had such colleagues.

Special thanks must go to the members of my committee. Johanna Nichols was asked to join the committee late in this effort and provided extraordinary insights and advice. Karl Zimmer has been my teacher in every way almost from the beginning of my work in linguistics, and has set an example I hope to be able to follow. Larry Hyman taught me how to do linguistics, how to see language as a human endeavor, and how to love it. Larry will not have his signature here because the mail between Berkeley and Bamenda was a little to slow to allow it, but he will be one of my teachers always. I hope he would have approved.

My deepest thanks go to John Ohala. Georg von Békésy once said that what a scientist needs is a good enemy. Békésy was wrong. What a scientist needs is someone to approach his work with the critical rigor of an enemy, while giving it the time, and giving one the attention, that only a dedicated teacher, advisor, and friend can give. John Ohala has done that.

Special thanks go to Ronya Dell Robinson for the

support and encouragement that made much of my schooling and research possible.

I would like to thank Linda Coleman for helping turn the manuscript into finished typescript.

I am grateful to the organizations that supported my research and schooling. The President's Undergraduate Fellowship of the University of California provided the funding for the development of the dynamic palatograph. I was supported during my graduate career by National Science Foundation grants to the Phonology Laboratory, and by teaching assistantships and an associateship in the Linguistics department at the University of California, Berkeley. The UCLA Phonetics Laboratory, the UCSD Phonetics Laboratory, and the Stanford Phonology Archive made their resources available at a number of stages in my work.

TABLE OF CONTENTS

CHAPTER 1

A Taxonomy of Phonetically-Based Sound Changes

ABSTRACT

The development of phonetically-based sound changes has been viewed by a number of researchers to be a result of the speech process acting as a "noisy" transmission line between speaker and hearer. The listener has no way of separating the "noise" from the "signal" and may come to interpret the noise as part of the language. Articulatory factors create certain types of "noise"; perceptual factors create others. The beginnings of a taxonomy of different types of causes of sound change within this paradigm is suggested. Four types of phonetic causes of sound change are identified. Type (I) changes are caused chiefly by articulatory factors. Type (II) changes are also caused chiefly by articulatory factors, but perception is more heavily implicated: the perceptual "loss" of a distinction is necessary for the changes to occur. Type (III) changes involve both articulatory and perceptual factors, and in addition, are dependent on acoustic factors. Type (IV) are chiefly perceptually based. The idea that sound changes can occur primarily as a result of perceptual factors has not received much support outside of this thesis (although it has been acknowledged to be part of the paradigm). That idea is defended with some simple statistical models.

Introduction

A paradigm for the origin of certain sound changes has been advanced in recent years in a number of papers (Jonasson 1971, Andersen 1973, Ohala 1973, 1974a, 1974b, 1975, Michailovsky 1975, Hombert 1975). The paradigm (in the form presented by Ohala 1975), holds that speech may be viewed as a transmission line between a speaker (or transmitter) who knows the language, and a listener (or receiver) who is learning the language. Constraints of the speaker's articulatory mechanism and of the listener's perceptual mechanism cause a certain amount of perturbation (or noise) in the signal actually reaching the listener, so that the signal contains not only characteristics intended by the speaker, but also extraneous, fortuitous elements. Fig. 1 illustrates this point.

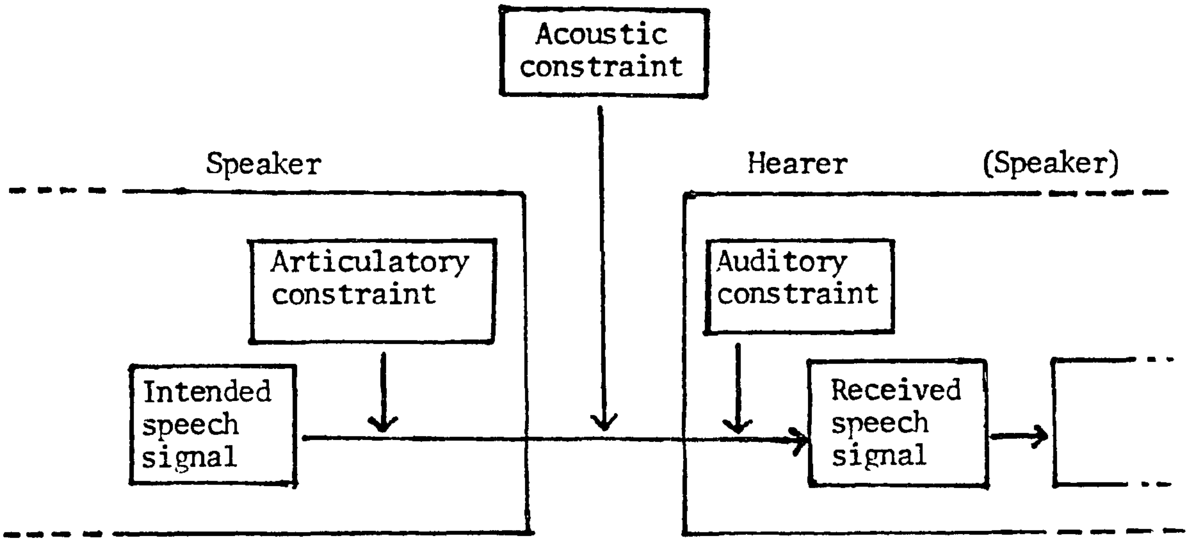


Figure 1. Speech communication viewed as a transmission system (version modified from Ohala 1975a)

The listener has no way of distinguishing the noise from the signal intended by the speaker. The listener may therefore, in certain circumstances, incorporate these elements into his own knowledge of the language, and begin intentionally producing them.

The paradigm suggests that the study of these sound changes involves the study of the characteristics of articulation and perception which cause noise, because certain types of noise are likely to occur and others are not. Determining what types of noise occur, and what their magnitude and likelihood are, will allow the explanation of these sound changes. Although the paradigm provides a framework within which the origin of certain types of sound change can be examined, the framework requires extension. A number of taxonomies of sound change have been proposed (e.g. Sweet 1889, Sturtevant 1917, Bloomfield 1933, Hoeningswald 1960) but none have been developed within the present paradigm, and none have taken into account recent developments in phonetic instrumentation and methods. Isolating the different types of phonetically-based changes should make it easier to determine what kinds of explanations can be sought for specific sound changes. The following pages discuss and attempt to do this.

The different types of sound changes break down roughly into four types. Type (I) changes are caused chiefly by articulatory factors, although the changes must be sufficiently perceptible in order to occur. Type (II) changes are similar to Type (I), although the perceptual mechanism is somewhat more involved -- it must lose a distinction in order for the change to occur. Type (III) changes are similar to Type (II), although (unlike Type (II)) they would not occur if speech were not transmitted via an acoustic medium. Type (IV) changes are chiefly perceptual in origin. Before elaborating on this classification, one point must be made. All areas of the speech transmission process are inextricably interrelated, so that in discussing sound changes it would be misleading to categorize a particular phenomenon as purely articulatory, acoustic, or perceptually-based.

The categories I will suggest are not all equally distinct. Type (I) is quite similar to Type (II). Type (III) is relatively similar to Types (I) and (II) (especially to Type (II)). Type (IV) is quite distinct from the other three. The immediate constituent structure of the categories is something like that in Figure 2.

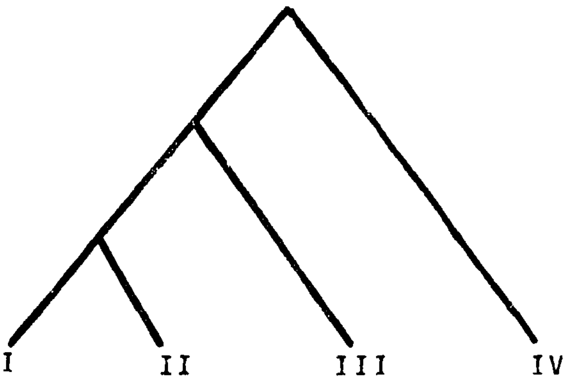


Figure 2

Type (I): Articulatorily-based changes

In Type (I) changes, perceptual factors are involved only in that the articulatory perturbation is perceptible, otherwise they are not implicated. In such a change, the role of the articulatory mechanism must be considered primary. Such a change is the development of epenthetic stops between nasals and following fricatives. Ohala (1974a, 1975) found that changes such as "something" /s mθIn/ to "something" /sʌmpθIŋ/ were due to an early closure of the velum, or denasalization, prior to the release of the labial closure for the /m/. This epenthesis occurs sporadically in the speech of many English speakers. When a stop does not occur, the opening at the velum is closed at the same time that the labial closure at the lips is released. The stop occurs when the velic port is closed prior to the release of labial closure. This results in a labial stop. The articulatory gestures for this stop are the same as the gestures for a stop made intentionally by the speaker. The stop may be somewhat different in closure duration from one occurring intentionally, but its other characteristics are the same. This contrasts with some of the changes discussed below, in which the sound resulting from the change requires a different articulation. In the case discussed here, listeners only re-interpret the unintended stop as intended. In the changes below, other speakers produce a new articulation as a result of misinterpreting the actual gestures of the original speaker.

To amplify this point, a Type (I) change would occur even if "listeners" used a three-dimensional X-ray machine with perfect resolution instead of relying on the acoustic signal. It would occur even if "listeners" used an accurate phonetic transcription as the basis for learning their speech. Of course, the articulatory perturbation must be sufficiently perceptible if it is to have an effect. This means that the perceptual mechanism is not totally irrelevant even in these changes. One Type (I) change is discussed in Chapter 2. The addition of creaky voice to contour tones or low tones, a development seen in a number of Chinese dialects, is, in one sense, similar to epenthesis. Neither of these changes involves a loss of a distinction. The perceptual mechanism is thus not "losing" or "failing to receive" any part of the signal. Therefore, our perfect x-ray machine would permit such a change to take place.

Type (II): Articulatorily-based changes

Type (II) comprises changes in which the articulatory mechanism is still primary, but in which some earlier distinction is in most cases lost by the perceptual system. The perceptual mechanism, or internal processes, therefore become responsible for the completion of the change. The development of tonal distinctions from prior initial voicing distinctions as described by Hombert (1975) and others is one important example of this kind of change. Articulatory factors cause vowels preceded by voiceless consonants to have a higher initial pitch than vowels preceded by voiced consonants. This has led to the development of tonal distinctions in a large number of languages, in which words with former voiceless initials developed high or falling pitch (retaining the high initial pitch of the vowel) and words with former voiced consonants developed low or rising pitch (retaining the low initial pitch of the vowel). In most of these cases, the voicing distinctions disappeared, although there is one example, described by Cappell (1949), in which the New Guinean language of Yabem maintained the voicing distinction, which co-occurs with the pitch distinction. Our X-ray machine would prevent the completion of this change. It would never fail to detect the presence or absence of voicing. It would, of course, detect the rate of vocal fold vibration, so that tonal distinctions might be added to distinctions in voicing, but would never replace them.

Another change of this type is the development of distinctive nasalized vowels in the environment of nasal consonants, with the nasal consonant frequently dropping out. This has occurred in French (Pope 1952), in many dialects of Chinese (Chen 1972), and in a number of other languages (see Ohala 1975, Bhat 1975, for references). The change occurs as a result of an articulatory effect (velic position for the nasal being maintained or anticipated in the vowel), causing a perturbation (nasalization of the vowel) which is interpreted by listeners as intended and part of the linguistic code.

Type (III): Acoustically-based changes

Type (III), similar to Type (II), occurs when articulatory factors cause a perturbation which results in an acoustic similarity to another sound. This kind of change is at least partly caused by the fact that speakers and listeners communicate through an acoustic medium. Perceptual factors then permit the reinterpretation of the perturbed older form as the new form. An example of this kind of change is the development of /pj/ into /t/ in the "Tet k" dialects of Czech, reported by Andersen (1973). The similarity between /pj/ and /t/ is acoustic, not articulatory. In some ways, this kind of change would seem to be an identical process to the one involving the development of pitch distinctions from voicing distinctions. The difference is that the effect of voicing distinctions on pitch differences is manifested in a consistent way in all facets of the signal: articulatory, acoustic, and perceptual, while this does not occur in the /pj/ to /t/ change. Specifically, the pitch perturbation is manifested as a changed rate of vocal fold vibration (an articulatory effect), as a changed fundamental frequency (an acoustic effect), and as a changed pitch sensation (a perceptual effect). The similarities causing the change from /pj/ to /t/ are not manifested articulatorily, only acoustically and perceptually. The reason the similarity appears in the acoustic signal is, in this case, a result of the fact that there is a many-to-one correspondence between some articulatory events and the resulting acoustic results. Specifically, the complex transformation from articulatory shape to acoustic output can cause two different articulations to "converge" in the acoustic domain. This convergence is the most important factor in the development of Type (III) changes.

Such a convergence has caused a set of changes in Lisu, which are discussed in Chapter 4, where /ɑp/ became /o/,

/ɑt/ became /e/, and /ɑk/ became /ae/. It is shown that the old and new forms are similar acoustically, although they are very different in articulatory terms. The comparison of what happens with an X-ray machine with what happens in acoustically transmitted speech is crucial. A perfect X-ray device would not show the similarities that show up in an acoustic image. If only an acoustic medium permits a change to occur, the change is acoustic-based. Accordingly the changes have been called acoustic-based, by Sweet (1899:23), Von Essen (1964), Ohala (1974b), Thurgood and Javkin (1975), and Michailovsky (1975). Only the fact that speech perception occurs through an acoustic medium permits these kinds of similarities to be detected and thus to influence phonological changes. In Thurgood and Javkin (1975), we argued that such explanations of changes should be accepted because of their similarity to the development of nasalized vowels in the environment of nasal consonants. The same argument is made in Chapter 4. However, the fact that there are similarities between the two explanations should not obscure an important difference. The development of nasalized vowels in the environment of nasal consonants is not due to two different articulations resulting in similar acoustic outputs. Rather, these cases of nasalization of vowels are Type (II) changes, in which an articulatory perturbation is the primary cause of the change. Another case of nasalization, however, the nasalization of vowels in the environment of /h/, may be a Type (III) change. Ohala (1975:303) has suggested a possible explanation for this change:

> Because of the open glottis during phonation accompanying an [h] (or breathy voice), the spectrum of the vowel will be changed in the following ways: there will be an upward shifting of the formants, especially F1 (Ohala 1974b), increased bandwidth of the formants, presence of anti-resonances in the spectrum and an overall lowering of the amplitude of the vowel (cf. also Fant 1973:8, Fujimura and Lindqvist 1971). This is identical to the effect of nasalization on vowels. Articulatory re-interpretation of the signal may occur, i.e., actual nasalization may be produced on the vowel.

Note that in the changes discussed above, the only relevant characteristic of the perceptual mechanism has been its resolution. In Type (I) and Type (II) changes, it was important to determine whether the perceptual mechanism was able to perceive an articulatorily-based perturbation. If the perceptual system did not have that resolution, or was not of a type that could perceive the changes, we would not expect a change to occur. In Type (III) changes, a change could not occur if the perceptual mechanism could too easily distinguish between the new and older forms. Although, as noted above, such lack of resolution does not appear amenable to quantification, resolution is still the issue. In other words, for Type (I) and Type (II) changes to occur, the magnitude of the perturbation must be _above_ some threshold determined by the perceptual system. For Type (III) changes to occur, the magnitude must be below some threshold.

Type (IV): Perceptually-based changes

The perceptual mechanism can, in addition, produce perturbations of its own other than the simple success or failure to note an articulatory based perturbation. The changes based on these perturbations I will put into Type (IV). One Type (IV) change involves a relatively weak tendency of the perceptual apparatus to distort the incoming signal. In Chapter 6, I will show that a tonal contour is perceived as occurring after a spectral contour when the two, in fact, occur simultaneously. There is, in other words, a "bias" on the part of listeners, so that they hear tone changes as occurring somewhat later than they actually oc in relation to changes in the spectrum. This appears to have led to a number of sound changes in which tones have spread into following syllables, and have prevented (for the most part) the appearance of tone spreading in which tones spread into preceding syllables. In a number of cases, the sound change now appears as a synchronic rule.

A defense of small differences

The results of the experiment reported in Chapter 6 show that the perturbation is relatively small, of the order of 10 ms. This is less than most accounts of the minimum difference required for the perception of temporal order (Hirsh 1959, Stevens and Klatt 1974, Pisoni 1976), which put the minimum difference at about 20 ms. However, Efron (1963a, b, 1970) found the minimum difference for judgments of simultaneity to be about 15 ms and reports that attentive listeners can better this considerably, to about 5 ms. It is not clear that minimum differences are the proper criterion for determining whether an effect is great enough to influence phonological developments. The concept of a threshold above which psychophysical events are perceived has come under serious attack in recent years. A number of investigators have suggested that such a concept of threshold does not take into account the distribution of subjects' responses. This has led to the development of the theory of signal detectability (TSD). I will briefly describe the theory both to defend the notion of ignoring jnd's and to lay the groundwork for the way in which the results described in Chapters 5 and 6 should be viewed. (For more complete treatments, cf. Swets 1964, 1973; Egan and Clark 1966; Tanner 1960, 1961.)

The proponents of TSD have noted that threshold is simply "a statistically defined point, meaning that it is an arbitrarily defined point and that some sensing and recognizing occurs below it" (Hake & Rodwan 1966:344).

These authors say, in addition (1966:344):

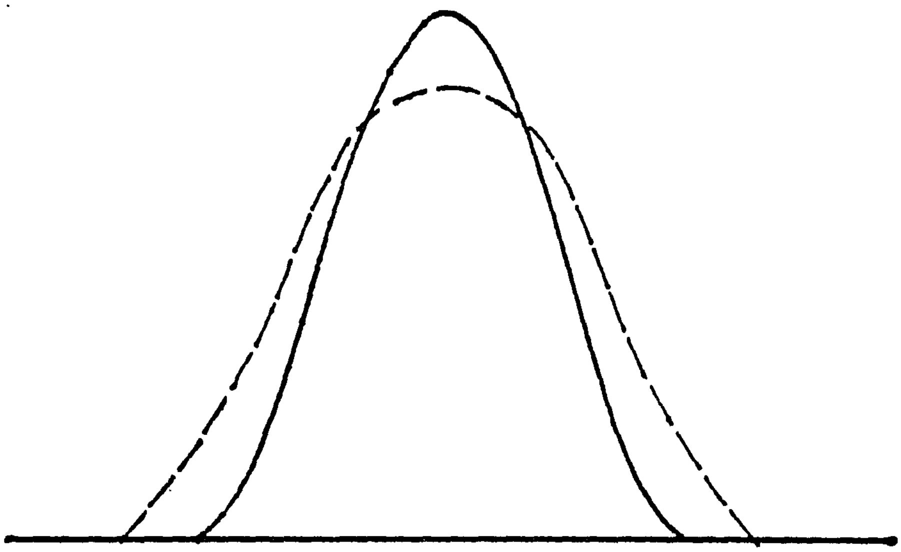
> Perhaps the major implications of TSD for perceptual research involve its insistence that the threshold concept itself may have dubious validity. In fact, the empirical shape of the ROC (relative operating characteristics) denies the existence of the threshold, i.e., denies the existence of occasions (trials) in the usual perceptual research situations in which the subject purely and simply guesses at the identity of the stimulus.

The theory holds that the quality which defines a subject's ability to detect a stimulus in noise is a ratio of two distributions rather than a threshold. (All input to the subject is held to contain noise. For auditory stimuli, even the quietest listening conditions are contaminated by the subject's own heartbeat, the spontaneous firing of neurons, etc.) The ratio defining the subject's ability to hear a sound in noise consists of the probability of his giving an "I hear it" response when a signal occurs along with the noise, divided by the probability of giving an "I hear it" response when noise alone is present. This is defined as the likelihood ratio. Listeners are assumed to have a certain expectation that a signal + noise will occur, or that noise alone will occur. The likelihood ratio is the ratio between the two probabilities, and thus itself a probability.

It should be noted here that, while the likelihood ratio as described above refers to the presence or absence of a signal and describes what has been termed "detection", the same ratio can describe the ability to recognize the difference between two signals. It is then termed "recognition". Once again, the question that detection theory seeks to answer is not whether the subject can or cannot "recognize" the difference between two stimuli, but rather, what the likelihood ratio of the subject's recognition is. An unbiased subject (i.e., one without a tendency in either direction) will have a likelihood ratio equal to 1.

Let us now return to the question of whether a small perceptual effect can cause a sound change. The distribution of the listener's input and his responses overlap. The means of the two distributions are different, but the distributions overlap to a large degree. Let us assume that the duration of the tone produced by the speaker has a normal distribution clustered around the end of a syllable. In addition, let us assume that the listener's perception of when the tone ends also has a normal distribution, but that the variance is somewhat greater. (The listener adds the effects of his errors to the effects of the speaker's errors. The result is a convolution of the distributions of the speaker's production and the listener's perception. The result is a larger variance than the variance of the speaker's production.) If both distributions have the same mean, the two distributions can be graphed as follows:

Figure 3

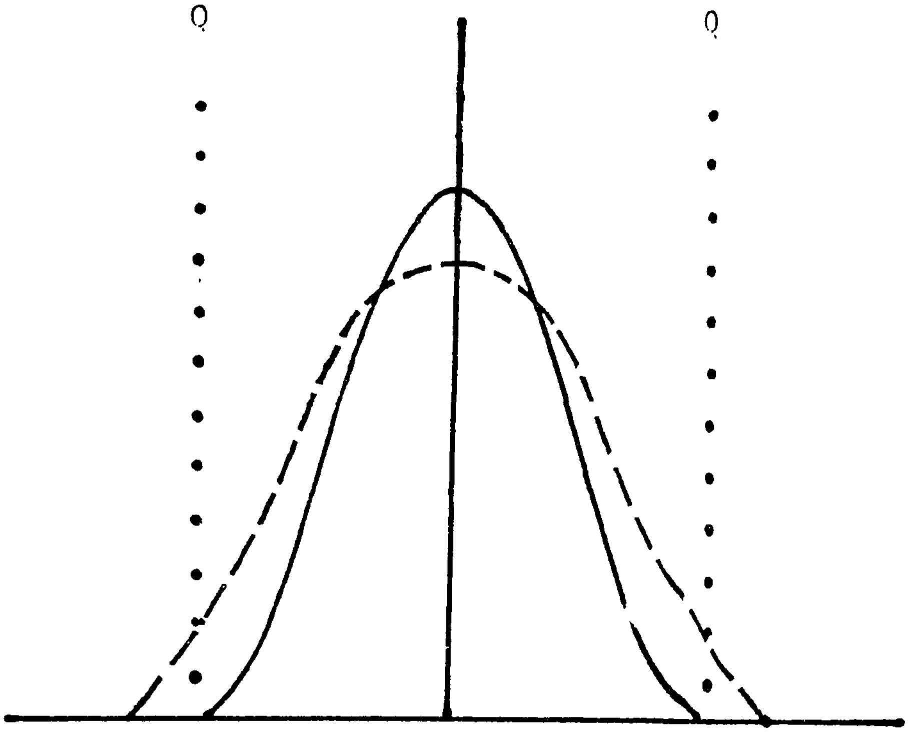


Distribution of Speakers' Production (solid lines) and
Distribution of Listeners' Perception (dashed lines)

(The solid line represents the variance of the speaker's production; the dashed line represents the variance of the listener's perception.)

Let us look at what happens toward the extreme points within the distribution, the so-called "skirts" of the bell-shaped curve. The question of relevance here, in terms of the theory, is what is the criterion for initiating a change. The criterion value relative to a perturbation, or relative to listeners' perception, has not been established. Let us assume a relatively large criterion, one greater than the difference between the means of the two distributions. In other words, only percepts of the end of the tone that differ by a relatively large amount can cause a sound change. Only effects greater than the criterion value can cause a sound change. The criterion values (Q) are given in Figure 4. These are effects to the left of the left dotted line or to the right of the right dotted line.

Figure 4

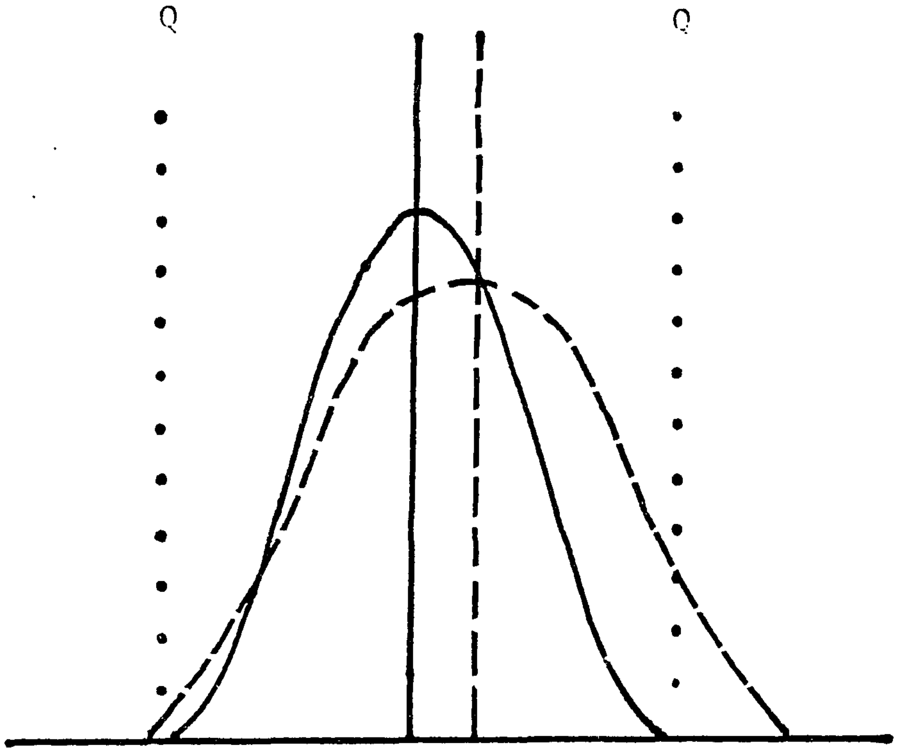


Distribution of Speakers' Production (solid lines) and
Distribution of Listeners' Perception (dashed lines)

Tone spreading could still occur, but in either direction, if listeners learn to produce their own output from values outside the criteria. The fact that the perceived change in tone sometimes falls outside the criteria can have a similar effect to a speaker sometimes producing an epenthetic stop. The spreading would be anticipatory if listeners perceived the tone as changing before the end of the syllable. The spreading would be perseveratory if listeners perceived the tone as changing after the end of the syllable.

Evidence will be given in Chapter 6 that tones do "linger" somewhat. The effect can be illustrated as in Figure 5.

Figure 5



Distribution of Speakers' Production (solid lines) and
Distribution of Listeners' Perception (dashed lines)

The solid line represents the end of the syllable for the speaker. The dashed line represents the mean of the distribution of the perceived values. The distance between these two lines is still less than the criterion value. However, observe the effects on the number of perceived values occurring earlier than the criterion before the speaker's production and the number of perceived values occurring later than the criterion following the speaker's production. It will be a very rare case when a listener will perceive a duration that falls to the left of the criterion, while the number of cases in which the listener perceives a duration falling to the right of the criterion will increase considerably. As a result, there will be more cases of perseverative tone spreading than anticipatory tone spreading occurs. This fits well with the linguistic data, as will be seen in Chapter 6.

The listener's perception of his own speech.

There is another problem that must be faced here. It can be argued that the listener's perception of his own speech is distorted in the same way that the speech of other speakers is distorted. This, so the argument goes, would prevent perceptual distortions from having any phonological effect. In other words, the speaker produces tones which the listener hears as "lingering". If the listener produces his own speech without adding the "lingering" effect, he will still hear that effect in his own speech. As a result, the listener would presumably learn to produce speech that is the same as that of the speaker, so that the distortion brought about by his perceptual mechanism should have no effect.

This is a cogent argument, and it would militate against the explanation offered here for perseverative tone spreading, as well as against the explanation offered for the longer duration of vowels before voiced consonants in Chapter 5, if it stood unchallenged. In response, it must be noted that listeners do not perceive their own speech in the same way that they perceive that of other speakers. The listener's perception of his own speech is not limited to the auditory channel; it includes kinesthetic information from his own vocal gestures. Such kinesthetic feedback provides an input which is not subject to the same bias which affects auditory perception. Moreover, the auditory information he does receive as to his own speech is of a very different sort than the auditory information he receives as to the speech of others. This information is different in at least two ways. First, the information is distorted by the very powerful transmission of sound from the vocal tract to the ears that occurs through the skull. Since skull transmission characteristically passes low frequencies much more than high frequencies (because of the skull's mass) a distortion is created. Second, there are two muscles, the tensor tympani and the stapedius, which protect the ear from very loud sounds such as those that reach the ear during speech by changing its mode of vibration (Dallos 1972:475). This also distorts the listener's perception of his own speech. In particular, these two effects, which lower the frequency response, reduce the temporal resolution available to the listener through his auditory system. These differences between a listener's perception of his own speech and his perception of the speech of other speakers shows that the two processes are not comparable.

Vowel duration differences associated with the voiced/voiceless distinction

Another example of a Type (IV) change is the result of a perceptual error made by listeners which is so pervasive that it causes a phonological universal. In Chapter 5, experimental evidence is provided which shows that vowels before voiced consonants are perceived as longer than vowels before voiceless consonants. This occurs even when artificially produced vowels having the same duration are presented to listeners. This explains the fact that vowels are longer before voiced consonants in every language that has been studied to the present time. It also explains the origin of a development in English and Latin, where, in many environments, the vowel duration difference replaces the voicing distinction of the following obstruent in many environments. It would seem that vowels before voiced consonants are always heard to be longer than they actually are.

Conclusion

In conclusion, four types of phonetic causes of sound changes have been identified. The taxonomy is described in Table 1.

Table 1

Type	Defining Characteristics	Features Necessary for Change	Loss of a Distinction	Examples
I	occasional articulatory perturbation -- when change occurs, perturbation simply comes to be used consistently	perturbation perceptible	no	[ls] --> [lts] in English [low tone --> [low creaky] in Cantonese
II	articulatory perturbation always occurs -- change occurs when speakers hear perturbation as primary cue	perturbation perceptible	yes	voicing contrast --> tone contrast, in many Chinese dialects
III	acoustic convergence of two sounds with different articulations	acoustic similarity between old and new sounds	yes	[ɑp] --> [o] [ɑt] --> [e] in Lisu [ɑk] --> [[illegible]] [pj] --> [t] in Czech dialects
IV	perceptual perturbation	sufficient perturbation	sometimes	vowel lengthening in English and Latin perseverative tone spreading in West Africa

Type (I) includes clear cases of articulatorily-based changes. Although a vast number of examples are not available, the explanations are very straightforward. The perturbation must be simply perceptible, and occur with sufficient frequency, in order to allow the change to occur. Type (II) changes are also articulatory-based, although the perceptual mechanism is somewhat more strongly implicated. Not only must the perturbation be perceived but, in addition, the distinction causing the perturbation leading to the change must be lost. Type (III) changes have been called acoustic-based by a number of researchers. Articulatory factors cause an acoustic perturbation which causes a sufficient acoustic similarity to another sound; this other sound comes to replace the original sound in the language. The "acoustic-based" label is justified, since such changes would not occur if an acoustic medium were not employed in the transmission of speech. Type (IV) changes are the result of a perceptual perturbation which comes to be produced by speakers.

CHAPTER 2

Epenthetic Stops Between /l/ and Sibilants [1]

ABSTRACT

A sound change appears to be developing in English in which a /t/ is inserted between /l/ and a following /s/, so that words such as "pulse" /pʌls/ are coming to be pronounced as /pʌlts/ by a number of speakers. In addition, a small number of words in some Indo-European languages have developed /stl/ for Indo-European /sl/. The epenthesis of /t/ is explained in terms of the failure of speakers to release the contact for /l/ (or /s/) before assuming the contact for /s/ (or /l/). The fact that English is not developing an epenthetic /d/ between /l/ and /z/, which have similar articulatory patterns is shown to have two possible explanations: either the weaker burst of /d/ makes it less perceptible in this environment, or the fact that /l/ and /z/ never come together in English except at word or morpheme boundaries causes a different temporal organization of the two gestures. Articulatory factors are claimed to be of primary importance in this type of change.

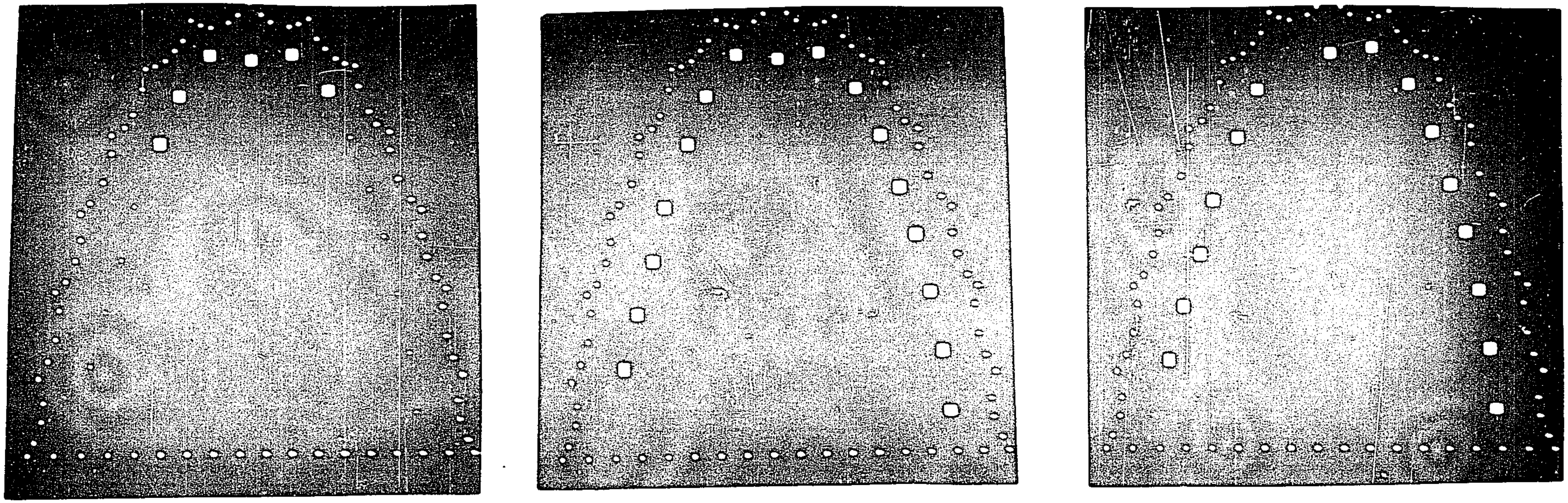
Introduction

A sound change appears to be occurring in English which, although it involves only a small number of words, is nevertheless significant as an example of a change whose origins are largely articulatory. A number of words with /l/ + /s/ sequences are developing epenthetic /t/, so that a word such as 'false' /fɔls/ is becoming /fɔlts/ in the production of a number of speakers of American English. A partial list of the words fulfilling the conditions for the change appears below:

else	ɛls	ɛlts
pulse	pʌls	pʌlts
false	fɔls	fɔlts
convulse	kənvʌls	kənvʌlts
calcify	ka͡elsIfaj	ka͡eltsIfaj

The epenthetic /t/ occurs rather consistently for certain speakers, who do not appear to form a coherent dialect. Other speakers have this form some of the time. This epenthesis is very similar to an epenthesis found by Phelps (1937) to occur in a number of Indo-European languages. [2] An /sl/ sequence is reconstructed for Indo-European, which has become /stl/ in a number of daughter languages, before undergoing further change in a number of them. Phelps reports the change to have occurred in Latin, Greek, Breton, and Czech. The explanation given by Phelps also applies (in reverse order) to the on-going change in English. Phelps noted that the areas of contact between the tongue and palate for /s/ and /l/ are complementary in that the gesture for /s/ closes off all but a small area at the front of the alveolar ridge, while /l/ closes off a small area only at the front, leaving an opening at the sides. If the /s/ contact is not released before the /l/ contact occurs, the result will be closure all around the alveolar ridge. Such contact constitutes a /t/ or /d/, depending on whether voicing began with the onset of alveolar contact or with the offset of lateral contact. The areas of contact on the palate are shown by the larger dots on the outline of the palate in Figure 1.

Figure 1



/l/ /t/ /s/

Phelps suggested that this kind of error in production would occur as a result of the fact that the tongue tip is a very fast articulator, which moves up to the /l/ position before the sides of the tongue can release their contact. This explanation, however, cannot apply to the case of /ls/ to /lts/ in English. For this to occur, the sides of the tongue must make contact too quickly, before the tongue tip releases its contact. This suggests that it is not the inherent rate of movement of the different parts of the tongue which are responsible for the insertion of /t/. Rather, speakers simply make an error in timing, and fail to release one area of contact before completing another. Presumably, listeners would learn words which originally had /sl/ (or in the case of English, /ls/) as having the epenthetic /t/, thus producing the change.

The change in English seems somewhat curious given the fact that a similar change has not occurred with /lz/ sequences, that words such as 'falls' /fɔlz/ have not developed into /fɔldz/ even in those speakers who regularly use the /fɔlts/ form for 'false'. In order to study this phenomenon, and to confirm the observations regarding the epenthesis of /t/, a method for determining contact between the tongue and palate was used.

Two palatographic studies

There were several questions that palatographic data might answer involving epenthetic stops between /l/ and alveolar sibilants. 1) Would stops occur accidentally in /sl/ sequences, thus confirming Phelps' observation? 2) Would stops occur accidentally in /ls/ sequences, thus confirming the observation here? 3) Would differences between /ls/ and /lz/ emerge that would show why epenthetic stops are developing between /l/ and /s/ but not between /l/ and /z/?

Methods

The method for determining tongue-palate contact used for these studies is an apparatus for dynamic palatography similar to systems developed by a few other investigators: Kuzmin (1962), Kozhevnikov and Chistovich (1965), Hardcastle (1968,1969), Shibata (1968), Fujimura, Tatsumi and Kagaya (1973), Fletcher (1975).

A dynamic palatograph consists of a sta-plate (or orthodontic retainer) with a series of electrodes embedded in the sta-plate so that they are exposed on the inferior (tongue-facing) side. The electrodes are attached to narrow-gauge wires which are joined onto a cable that leads out of one or both corners of the mouth. Figure 2 shows one such device.

Figure 2

Each wire is connected to a separate amplifier. A 20 KHz AC signal is applied to the body at 0.1 volt rms. At the time of this experiment, the current was applied at the wrist, using a copper electrode and standard electrode paste. More modern practice calls for the current to be applied at the back of the neck, in order to minimize the distance the current must travel.

When the tongue touches an electrode on the artificial palate, this completes a circuit which registers as an increase in voltage across the two leads coming from the subject: the lead from the electrode and the lead from the electrode at the wrist or neck. When the voltage is above a certain level (normally 60% of the voltage applied to the body) contact has taken place. When it is less than this, contact has not taken place. The change in voltage can be determined either by adjustable analog circuits or by software. Analog circuits were used in these studies.

The Linc-8 computer was used to record the palatographic record on digital tape together with a relatively coarse-grained form of the amplitude wave of the speech signal. The Linc-8 has a CRT display. Each sample of tongue-palate contact is analogous to a single movie frame. The amplitude waveform of the entire utterance is shown at the bottom of the display. A cursor on the amplitude waveform co-ordinates the tongue-palate sample to the proper place on the waveform. Figures 3 and 4 provide examples of the contact pattern for the words "false" and "falls".

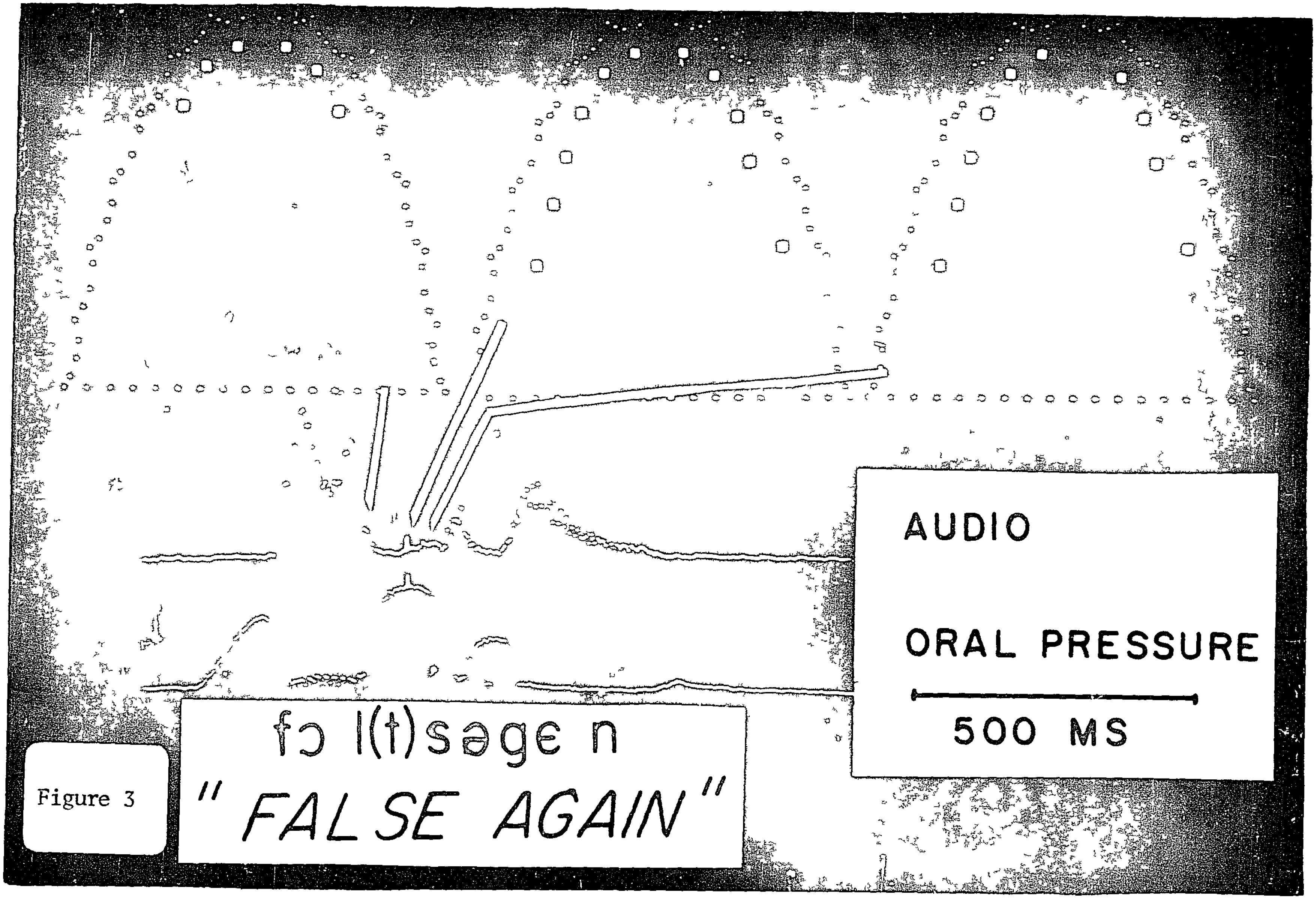


Figure 3

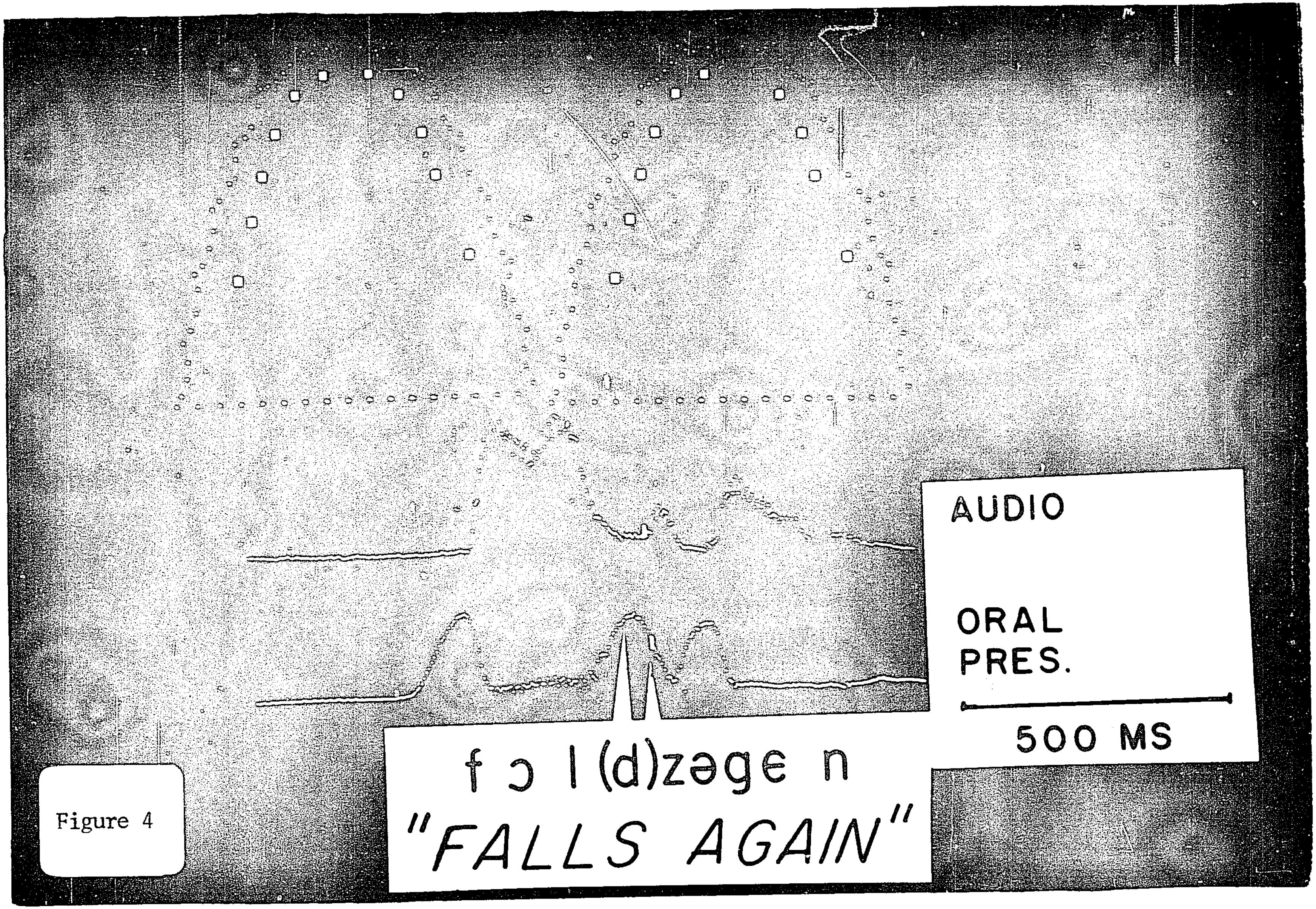


Figure 4

The sampling rate of the computer record was 2 ms. The amplitude waveform was also recorded on an FM tape recorder. The audio signal was recorded on a direct track of the FM tape recorder. The recordings were made in the computer room. Noise from the teletype attached to the computer and from the computer's cooling fan was evident in the amplitude signal. This was somewhat minimized by keeping the microphone as close as possible to the subject. All the relevant speech sounds were identifiable from the amplitude waveform.

First palatographic study

A number of words were selected and put into a randomized list. Each word occurred a minimum of 4 times and a maximum of 9. Two words were included which actually included the stops, in order to compare them to words that might have the eperthetic stops. The words were the following:

false	/fɔls/
falls	/fɔlz/
pulse	/phʌls/
pulls	/phUlz/
colts	/khowlts/
colds	/khowldz/
pills	/philz/
Ilse	/Ilsə/
utensils	/juthensəls/
slop	/slap/

Although it had criginally been planned that 3 subjects would participate in this test, two had to be eliminated. One subject, who appeared to have alveolar /l/ in a pretest, was found to substitute /w/ in most of the test words. (Such substitutions are not uncommon in American English, especially for /l/ preceding stops, in words such as 'milk'. See Ohala 1974a for discussion.) The palatograph of a second subject did not function reliably enough in pretesting to permit its use in this study. A shorter study was subsequently done with this subject, however, which is reported later in this chapter. The results of the one subject, for whom palatographic and amplitude data (but not air pressure data) was recorded, are given in Table 1. The column headed "Complete Closure" provides the number of cases in which the closure characteristic of an alveolar stop occurred. A drop in amplitude is also characteristic of a stop since closure results in a silent or near-silent interval. However, a similar drop can also occur at the onset of a fricative. The column headed "Amp Drop" provides the number of times that amplitude dropped to the level of the noise in the room.

Table 1

Results of first palatographic study

	Complete Closure	Amp Drop	Number of Samples
false	1	4	4
falls	0	3	5
pulse	0	8	8
pulls	0	2	4
colts	7	7	7
colds	9	8	9
pills	0	4	8
Ilse	0	4	5
utensils	0	5	6
slop	5	8	8

Of the words with orthographic /t/ and /d/ between /l/ and the sibilant, complete closure was recorded 7 out of 9 times in the case of 'colds' and 7 out of 7 times in the case of 'colts'. The stop occurred epenthetically only once between an /l/ and a following sibilant, in the word 'false' on one of the four times it was spoken. However, nothing in the contact pattern found here, nor in the timing of the gestures of /l/ and the two sibilants (with the exception of the one occurrence of the stop in 'false') suggest that it could not have as easily occurred with a voiced sibilant. In words both containing the voiced and the voiceless sibilant, the /l/ contact completely enclosed the small area which was released to form the narrow groove for /s/ or /z/. The change from /l/ to the following sibilant was always completed within 12 ms, in both the voiced and voiceless versions. This suggests that epenthetic stops are as likely to occur as an accident, in this environment, with either voiced or voiceless sibilants.

A further observation to be made from the data on /l/ and following sibilants is that the amplitude signal dropped to the noise level in most of the samples (52 of 64). Although the amplitude drop occurred in nearly all the samples in which closure occurred (20 of 21), this still left 32 occurrences of an amplitude drop developed without closure along the alveolar ridge, as determined by the palatograph. The effects of this will be discussed later in this chapter. The most interesting fact involved the surprising number of times that complete closure occurred in the word 'slop': 5 out of 8 times. Given the fact that [stl] and [sdl] are impermissible sequences for English phonology, the number of occurrences of the stop in this environment seems extremely high. Of further interest is the fact that these stops are not perceived by listeners -- several trained phoneticians did not note the appearance of these stops. There are several reasons for this. First, the transitional cues for an alveolar stop will not show up between two other alveolar consonants. The stops for this speaker were voiced, so that the release burst of the stop into the /l/ was considerably less than it might have been if the stop were voiceless. In fact, the question of the perceptibility of /d/ between /l/ and following a sibilant is also relevant to the failure of /ldz/ sequences to develop in English. Before proceeding with a discussion of this point, however, I will discuss a brief test which was run on the one subject for whom oral air pressure data was available, but whose palatograph did not behave with sufficient consistency to allow the use of the full word list.

Second palatographic study

This shorter study consisted of 12 words or phrases, within the frame "___________ again". The procedure was otherwise as described above. The words were:

false	/fɔls/
falls	/fɔlz/
pulse	/phʌls/
pulls	/phUlz/
halls	/halz/
sills	/sIlz/
slop	/slɑp/
calls	/khɔlz/
call Sue	/khɔlsu/
call zoo	/khɔlzu/

The results are given in Table 2.

Table 2

Results of second palatographic study

	Complete Closure	Amp Drop	Number of Samples
false	1	1	1
falls	1	0	1
pulse	1	1	1
pulls	1	0	1
halls	0	0	1
sills	1	0	1
slop	1	1*	1
call Sue	0	1	2
call zoo	0	0	1

* amplitude dropped into the noise _after_ closure ceased.

There was a bad recording of palatographic contact on the second occurrence of "call zoo". Complete closure occurred in all the utterances with the exception of the word "halls" and in the two-word utterances "call Sue" and "call zoo". Once again, the patterns of contact were essentially the same for the utterances containing the voiced and voiceless sibilants. There were differences, however, in the patterns of oral air pressure. The pressure peak was greater in the voiceless cases. In addition, while the air pressure in the voiceless stop-ard-sibilant combination quickly reached a near-maximum and then slowly increased, pressure gradually built up to a maximum in the voiced stop-and-sibilants and then gradually decayed.

Similar differences between the pressure patterns for voiced versus voiceless consonants have been presented by a number of investigators (Malécot 1955, 1966, Ladefoged 1963, Arkebauer, Hixon & Hardy 1967, Brown & McGlone 1969, Kent & Moll 1969, Lisker 1970, Malécot 1970, Warren & Hall 1973). There, although the data from the current experiment represents a very small sample, there is no doubt that the differences between the voiced and voiceless cases are to be trusted. Finally, before proceeding to the discussion of the two experiments, one further observation should be made. In the second experiment, the amplitude drop between the /l/ and the sibilant occurred for both of the voiceless sibilants and none of the voiceless.

Discussion

The palatographic data gathered in these studies confirms Phelps' observation concerning the origin of epenthetic stops between /l/ and sibilants. If the contact necessary for /l/ "overlaps" with the contact for /s/ or /z/, complete closure will occur, so that a stop closure will be formed. The palatographic patterns of contact do not reveal any reason why epenthetic stops should develop in the case of /l/ followed by a voiceless sibilant but not in the case of /l/ followed by a voiced sibilant. The voiced and voiceless sibilants had the same contact patterns. The articulations of both would lead to complete closure of the

/l/ contact were not released in time. Furthermore, the timing of the gestures were essentially the same, so that this also fails to explain the failure of an epenthetic /d/ to develop.

Possible reasons for the lack of /lz/ --> /ldz/

There are two possible explanations for the fact that epenthetic stops are developing between /l/ and /s/ but not between /l/ and /z/. One is based on the lesser perceptibility of /d/ versus /t/ between an /l/ and a following sibilant. The other is based on the fact that English permits /lz/ clusters only at morpheme boundaries, while /ls/ clusters can occur within a morpheme.

Is /ldz/ difficult to perceive?

The differences in pressure suggest an explanation. The cues for the presence of an alveolar stop in an environment between two alveolar consonants are relatively scarce. There are no transitional cues. The stops are of relatively short duration. One cue for the presence of a stop, of course, would be the presence of a period of silence between the /l/ and the following sibilant. However, the stops are of relatively short duration. Furthermore, the drop in amplitude (in other words, the silence) characteristic of a stop also occurs when a stop is not present. The silence in the cases in which stops do not occur is probably due to the fact that it takes a short time for sufficient pressure to build up in the oral cavity to provide fricative noise. A listener, therefore, hears almost exactly the same thing regardless of the presence or absence of a stop in this environment, with one exception. That exception is the burst characteristic of the release of the stop into the sibilant. If a stop is present, there will be a momentarily high fricative noise level, which will decay rapidly. This burst is smaller in the case of a voiced consonant, simply because the pressure build-up is smaller. The difference can be seen in the overall amplitude of the stop bursts measured by Halle, Hughes and Radley (1957). The result is that the one cue available to the listener that a stop has occurred is less strong for a voiced than for a voiceless epenthetic stop. Therefore, although both types of stops probably occur with similar frequencies accidentally, only the voiceless is likely to be noticed by listeners, so that such stops are less likely to become part of the listener's code.

The occurrence of /dz/ in other languages

Tests to determine the relative perceptibility of /lts/ clusters versus /ldz/ clusters have not been performed at this time. However, there is a way of assessing the relative perceptibility of the two sequences. The effect of the /l/ on the perceptibility of the /dz/ and /ts/ clusters should be similar. The /l/ will obscure the consonantal transitions of both sounds. If the smaller amplitude burst of /d/ relative to /t/ causes the /d/ to be less perceptible when the consonants are followed by a sibilant, one would expect /dz/ sequences to occur with less relative frequency than /ts/ sequences in the world's languages. Furthermore, if /d/ were relatively imperceptible in this environment, one would expect that phones consisting of [dz] sequences would tend to alternate with [z] phones and be members of the same phoneme more frequently than would be the case with the phones [ts] and [s]. In other words, if the contrast between [dz] and [z] is relatively imperceptible, that contrast would not be used in many languages.

In making this kind of assessment, it must be borne in mind that voiced consonants occur less frequently than voiceless consonants (Trubetzkoy 1939). This factor can be taken into account, however, by comparing the number of times /ts/ and /dz/ occur with the number of times that /s/ and /z/ occur. Even if [dz] sequences were perceptually salient for speakers, we would still not expect them to occur as frequently (relative to [ts]) as [d] and [z] (relative to [t] and [s]).

A count was obtained of the number of times that each of the sounds above occurred in the languages whose phonologies are described in the Stanford Phonology Archive.

The files of the Archive contain, at the time of this writing, rather reliable synchronic phonological descriptions of 221 of the world's languages. A computer printout prepared by the Archive staff provided the languages in which each of the sounds appeared as non-alternating phones (the only phone in a phoneme), main allophones, or secondary allophones. Some of the tabulations described below had to be done by hand. [3] The results are given in Table 3.

Table 3

The occurrence of dental/alveolar stops, affricates and fricatives

	t or d	ts or dz	s or z
Voiceless	198	84	198
Voiced	169	42	89

In assessing the numbers given in table 3, I will use a statistical approach used in Ohala and Lorentz (1977). One might suppose that the probability that /ts/ and /dz/ affricates will occur in the world's languages might be related to the probability that /t/ and /s/ will co-occur; the probability that /dz/ will occur would be related to the probability that /d/ and /z/ will co-occur. The probability that /t/, /s/, /d/ and /z/ occur can be taken from the table above. Dividing the number of languages in which each of these sounds occur by the total number of languages in the sample (221) provides the probability of occurrence. The probability of occurrence of /t/ and /s/ is calculated by multiplying the independent probabilities that /t/ and /s/ occur; the probability that /d/ and /z/ will co-occur is calculated by multiplying the independent probabilities that /d/ and /z/ occur. The probability of co-occurrence can then be translated into the number of languages expected to have a co-occurrence by multiplying by the sample size. The results are given in Table 4, rounded off to two significant digits.

Table 4

Statistical analysis

(p)t = 198/221 = .90

(p)s = 198/221 = .90

(p)t and s = (p)t x (p)s = .81 = 179.0 languages

(p)d = 169/221 = .76

(p)z = 89/221 = .40

(p)d and z = (p)d x (p)z = .31 = 68.5 languages

Thus, /ts/ occurs less frequently (84 languages), in relation to the number of languages in which its individual sounds can be expected to co-occur, than /dz/ does (44 languages). This suggests that /dz/ is not a disfavored cluster, but does not answer the question of how distinct it is from /z/.

To determine how distinct [dz] was from [z], I counted the number of times that [dz] was an alternant of [z], with both phones being part of the same phoneme. A computer program provided the segment inventories and a guide as to which languages should be examined, but the inventories were searched by hand. All languages in which [ts] and [dz] occurred were examined. The data leading to this count was checked twice. The results were the following: there were a total of 4 alternations between [dz] and [z], and 3 alternations between [ts] and [s]. This is very weak support for the hypothesis that [dz] and [z] are relatively indistinct.

I searched for all the languages which had all six phones: [t], [ts], [s], [d], [dz], and [z]. There were 24 such languages. In this group, there were no alternations between [ts] and [s], and two alternations between [dz] and [z]. This is very weak support for the hypothesis that [dz] and [z] are indistinct.

I searched for all the languages which contrasted [dz] and [z], and those which contrasted [ts] and [s]. This included all cases in which the contrasting pairs were members of different phonemes. This count should be examined with the count of the languages in which [ts], [dz], [s] and [z] occurred. The numbers obtained are given in Table 5.

Table 5

Constrasts between affricates and fricatives

	Voiceless	Voiced
All languages with [ts] or [dz]	84	42
All languages with [s] or [z]	198	89
All languages contrasting [ts] and [s] or [dz] and [z]	80	23

We can compare the number of languages which contrast the pairs /ts/ and /s/, and the pair /dz/ and /z/, to the probability of co-occurrence of each of the pairs. I have calculated the probabilities, which are given in Table 6, according to the method used in Ohala and Lorentz (1977).

Table 6

Statistical analysis

$(p)ts = 84/221 = .38$

$(p)s = 198/221 = .90$

$(p)ts$ *and* $s = (p)ts \times (p)s = .34 = 75.1$ languages

$(p)dz = 42/221 = .19$

$(p)z = 89/221 = .40$

$(p)dz$ *and* $z = (p)dz \times (p)z = .076 = 16.8$ languages

75.1 languages can be expected to have a co-occurrence of /ts/ and /s/; 80 languages actually have a contrast between these two sounds. 16.8 languages can be expected to have a co-occurrence of /dz/ and /z/; 23 languages have a contrast between them. Thus both pairs contrast more frequently than one would expect them to co-occur. Furthermore, co-occurrence is not the same thing as contrast. The number of languages in which two sounds contrast can never exceed the number of languages in which they co-occur, but the opposite can be true. The fact that the number of languages have a contrast between /dz/ and /z/ exceeds the number of languages which would be expected to have a mere co-occurrence shows that we cannot conclude that contrasts between /dz/ and /z/ are unlikely. Taking into account the fact that [dz] alternates with [z] in only four languages, but contrasts in 23, it is obvious that contrast, rather than alternation, is the general tendency. This does not support the hypothesis that [dz] and [z] are very similar perceptually.

The contrast of /dz/ and /z/

The preponderance of the evidence suggests that /dz/ is a cluster which can contrast with /z/. Although /dz/ occurs less frequently than /ts/, and participates in fewer contrasts, this is part of a general tendency favoring voiceless obstruents. Although part of the reason for that tendency is perceptually-based (i.e. the weaker burst of voiced stops, even if a sibilant does not follow), the tendency is not sufficient to prevent the occurrence of voiced obstruents or clusters of voiced obstruents such as /ldz/. Since that tendency is partly based on the lesser perceptibility of voiced obstruents, it could be argued that the perceptual factors helping to form that tendency make the development of epenthetic voiced stops less likely. This is partly true. However, while the tendency makes the development of such stops less likely, it does not prevent them. Voiced clusters such as /ldz/ do occur. They are heard by listeners, who learn to produce them. English listeners are accustomed to hearing such clusters. Moreover, the tendency is only partly based on perception. The fact that obstruents reduce the trans-glottal pressure drop needed for voicing may account for much of that tendency. Given the evidence just gathered that /dz/ clusters are not strongly "disfavored" by the world's languages, it seems unlikely that the reason for the failure of epenthetic /d/ to develop in English between /l/ and a following sibilant is due to the lesser perceptibility of this sound (vis-a-vis /t/) in this environment. [4]

Morphological constraints of English for /ls/ and /lz/

In English, /lz/ and /ls/ occur under different conditions. All the words which potentially could develop an epenthetic /d/ between an /l/ and a following /z/ contain a morpheme boundary between these two sounds. The sound /z/ follows /l/ only as the plural morpheme, as a possessive suffix, or as a third person singular marker. All the cases in which /s/ follows /l/ occur within a morpheme. This difference changes what is involved in the epenthesis of a stop in the two cases. For a stop to be inserted between an /l/ and an /s/, a speaker merely has to store the word in a new form in his lexicon. For a stop to be inserted between an /l/ and a /z/, the speaker must create a minor rule, since the /d/ cannot be assigned either to the stem or the suffix, as it does not occur in other environments. If the speaker were to store the /d/ as part of the stem or suffix, he would have to introduce a rule deleting the /d/ everywhere it did not occur. With either rule, the process would be quite different than the storing of words in a new form.

The fact that epenthesis does not occur when a word boundary separates /l/ and /s/ provides support for the hypothesis that the boundary between /l/ and /z/ also prevents epenthesis. Phrases such as "tall Sam", "tell Sandy", "yell softly" do not contain the epenthetic /t/, even for speakers who have the epenthetic /t/ within words. The non-occurrence of accidental epenthesis in "call Sue" in this experiment is suggestive. The organization of the motor movements by the speaker may be different in cases where a morpheme or word boundary intervenes. The result could be a delay in the tongue gesture for the sibilant until the /l/ contact is released. [5]

/sl/ and /stl/

The difficulty of perceiving alveolar stops between alveolar consonants, as well as the English constraint against /stl/ clusters, are relevant to the data regarding epenthetic stops between /s/ and following /l/. The appearance of complete closure in this environment in 5 out of 8 samples in English for one subject, and in the one sample for another subject, is surprising. These stops were not perceived during this study and are not even during careful listening to the speech of the subjects. The fact that /stl/ and /sdl/ are not permissible sequences in

English undoubtedly plays a role. However, Phelps has noted that the languages that may have developed /stl/ sequences from Indo-European /sl/ have frequently undergone further changes, in which the /t/ became another consonant in a number of languages in which this change originally occurred. In some cases, the /l/ changed or the /s/ dropped out. One of the most interesting cases involves the development of Vulgar Latin 'scloppus'. This is reflected in Italian 'schioppo' (a gun). The /t/ has become a /k/ possibly because of the difficulty in the perception of an /stl/ sequence. Incidentally, the /j/ is a regular reflex of /l/ in Italian in post-occlusive position. The phonetic basis of this change is one of the areas discussed in the following chapter.

Conclusion

In conclusion, the the results suggest that the epenthesis of /t/ between /l/ and /s/ is due to the occasional occurrence of such stops as a result of the failure to release /l/ contact before /s/ contact begins. It also suggests that Phelps' observation concerning the epenthesis of /t/ between /s/ and /l/ in some Indo-European languages is correct. In the case described by Phelps, there is a failure to release /s/ contact before /l/ contact occurs. Phelps suggested that the change was due to the faster rate of tongue-tip movement compared to the rate of movement of the tongue body. The fact that the same change occurs in English, but in the opposite direction, shows that it is not the different rates of movement of the two parts of the tongue, but rather the lack of precision in timing the two events, that is responsible for both changes. Finally, the failure of epenthetic stops to develop between /l/ and /z/ in English appears to be due to the fact that a morpheme boundary always occurs between these sounds in English.

The change, where it occurs, is essentially articulatorily-based. That its potential occurrence in English is restricted by facts about morpheme structure does not lessen its articulatory nature. It is therefore classified as a Type (I) change.

FOOTNOTES

1. An abstract of an earlier version of this chapter appeared as Javkin and Ohala (1973).

2. It has been brought to my attention (Gary Holland and Carol Justus, personal communications) that none of the words which Phelps gives as evidence for the epenthesis of [t] have an established Indo-European etymology. The languages he cites as having undergone the change have lost Indo-European [sl], but may have borrowed their (relatively) rare [stl] forms.

3. In preparing these other tabulations, I noticed that one language, Yurak, had the phone [dz], despite the fact that the computer printout did not list this language as having [dz]. This seems to have occurred because the records of the Archive for Yurak were not complete at the time that the printout was prepared. The phone [dz] thus occurs in 42 languages, although a reader obtaining the same printout from the Archive would only find 41.

4. This is a different view from that in Javkin and Ohala (1973), where we tentatively concluded that the lesser probability of [d] was responsible for the failure of [d] to develop between [l] and [z] in English. Drachman (1977) has noted that [lz] becomes [ldz] in some dialects of Greek, which shows that such a development is not impossible. We never claimed it was.

5. The fact that grammatical conditions can influence phonological processes is well established (cf. K. Pike 1947). That influence has been thought to be the result of the way phonological units are organized (e.g. Chomsky and Halle 1968) or of the necessity for signaling boundaries (Trubetzkoy 1939, Coker 1974). Regardless of the cause (an interesting question, although I will not try to answer it here), there are many phonological processes dependent on grammatical boundaries. In Javkin 1975b, I found that 346 out of 754 phonological rules in the files of the Stanford Phonology Archive made reference to grammatical boundaries.

CHAPTER 3

Palatalization of Post-Occlusive and Geminate /l/ [1]

ABSTRACT

Two related sound changes, the palatalization of geminate /l/ in Spanish and of post-occlusive /l/ in Italian, were investigated to determine whether they had a similar phonetic basis. The hypothesis was formed that the two changes occurred because these two types of /l/ involve a higher tongue position (and therefore a higher second formant) than non-geminate, non-post-occlusive /l/, which does not undergo the change. Acoustic data confirmed the hypothesis for Spanish, but not for Italian. The acoustic similarity of post-occlusive /l/ to a palatal glide in Italian appears to be responsible for the development of /Cl --> Cj/ in that language.

Introduction

In Chapter 2 I presented an example of a sound change which was strictly the result of an articulatory process. The sound change discussed in this chapter is somewhat more complicated. Articulatory events can cause two relatively distinct sounds to become more similar. That similarity, or rather the increase in similarity can lead to a situation in which one sound changes into the other. Such a change occurred in the development of Latin geminate /l/ into Spanish /j/. A closely related change in Italian, in which post-occlusive /l/ became /j/, was hypothesized to have had a similar origin. As will be shown, however, this involves an acoustic effect, with articulatory features not playing a major role.

Other sound changes involving /l/

The alternation of light or palatal /l/ with /j/ and of dark or velarized /l/ with /u/ or /w/ has been noted in several languages by a few investigators: Fortuguese and Spanish (Zauner 1921, cited in Von Essen 1964) have changed a number of velar /l/ sounds to /u/, as have French and Polish (Von Essen 1964). There are some examples of this in English. For many speakers of English, words such as "Bill" and "milk" are produced as [bIU] and [mIUk] (cf. Kenyon 1950:155-6). [2]

Some dialects of Southern Germany have changed /l/ to /i/ in certain environments (Von Essen 1964). Von Essen also noted the cases in Italian in which post-occlusive /l/ has changed into /j/: plenu>pieno, clave>chiave, etc. Spanish has undergone changes of geminate /l/ to /j/, as can be seen in the data from Lapesa (1962). The changes of light /l/ to /j/ or /i/ and of dark /l/ to /w/ or /u/ were explained by Von Esser (1964) (and by Jonasson (1971) in the case of the change to /w/) in terms of the acoustical similarities between the alternating forms. The similarities are most striking in the lower end of the spectrum. Von Essen found that dark /l/ and /u/ show very similar peaks at 250 and 400 Hz. Jonasson showed the similarity between dark /l/ and /w/. In the case of light

/l/ and /i/, Von Essen found a first formant peak at 250. The second formant for these sounds is different, however, with light /l/ showing a value of 1600 Hz, while /i/ has an F2 of 3150. The second formant value for /i/ found by Von Essen is questionable. [3]

Spanish and Italian

It seemed curious that the changes in Spanish and Italian were restricted to specific conditions for /l/. In Italian, /l/ following stops became /j/ while /l/ in other environments (as well as geminate /l/) did not change. In Spanish, short /l/ did not change in the same way, either when following stops or between vowels, but geminate or long /l/ became /j/, merging with the /j/ already in the language. Examples are given below. Note that in examples 6 and 7, the stop + /l/ clusters became palatalized /l/. [4]

	Latin	Spanish	Italian	
1.	bellum	bejo	bel:o	"beautiful"
2.	pullum	pojo	pol:o	"chicken"
3.	caballum	kabajo	kabal:o	"horse"
4.	templum	templo	tempjo	"temple"
5.	blancum	blanko	bjanko	"white"
6.	planum	jano	pjano	"flat"
7.	clave	jaβe	kjaβe	"key"

The "pl" and "cl" change to "ll" is a complicated development, which probably involved both the palatalization and gemination of post-occlusive /l/ prior to the loss of the occlusive. For more extensive treatments, see Lapesa (1962) and Malkiel (1964); suffice it to say here that the development is quite general, and does not contradict the tendency for post-occlusive /l/ to palatalize. Some further examples are given below:

Latin	Spanish (present-day)		
	Orthographic	Phonetic	
planta	llanta	janta	"a plant"
plaga	llaga	jaɣa	"ulcer"
pluvia	lluvia	juβia	"rain"
clamare	llamar	jamar	"to call"
clave	llave	jaβe	"key"
flamma	llama	jama	"flame"

It should be clear, then, that only geminate and post-occlusive /l/ underwent palatalization. We can hypothesize that geminate and post-occlusive /l/ each contained something that made them more likely candidates for palatalization than short, non-post-occlusive /l/. The situation presents a rare opportunity for the construction and testing of hypotheses concerning these kinds of changes. We can look at present-day geminate /l/ in Italian for the phonetic motivation of the change which occurred in Spanish. We can look at present-day Spanish for the phonetic motivation of the change in Italian. In other words, /bel:o/ and /pol:o/ in present-day Italian should contain some basis for palatalization, as should present-day /templo/ and /blaŋko/ in Spanish.

I formulated the hypothesis that these two conditions for /l/, gemination and post-occlusive position, had something in common which led to palatalization: namely, they both had a higher tongue position, leading to a higher second formant, than short, non-post-occlusive /l/. Geminate /l/ might have this if the tongue blade continued moving upward after /l/ contact was achieved. Post-occlusive /l/ might have a higher tongue position due to the fact that the jaw has to be in a relatively high position for the occlusive. According to Fant's data for palatalized and non-palatalized /l/ in Russian, a high tongue position for /l/ is associated with a higher second formant, which is the major cue of palatalization. Such an acoustic effect would make these two kinds of /l/ different (closer acoustically to /j/ or /i/) from intervocalic, short /l/.

Geminate /l/

Methods

One speaker of the Roman dialect of Italian was recorded to determine whether geminate /l/ has a higher second formant than the non-geminate /l/. A set of nine minimal pairs was put into a randomized word list, with three repetitions of each word. The speaker read the words within the frame sentence: "Ora diro __________ al microfono" ("I now say __________ into the microphone."). The following words were used:

calo	"loss, waste"
callo	"corn, hard skin"
cola	"strainer"
colla	"glue"
mola	"grindstone"
molla	"spring, spring lock"
molo	"wharf"
mollo	"flabby"
pala	"shovel"
palla	"ball"
polo	"polo"
pollo	"chicken"
scolare	"to drain"
scollare	"to unglue"
velo	"veil"
vello	"sheepskin"

The speech was analyzed using a Voiceprint 700 spectrograph over the range 0-4000 Hz. The first, second, and third formants were measured at the mid-point of the /l/. Values were rounded off to the nearest 50 Hz.

Results

The results are shown below:

	Geminate	Non-geminate
F1	361	418
F2	1357	1183
F3	2503	2324

A t-test for difference in mean F2 values in the paired words shows the results to be significant at the .001 level.

The second formant is considerably higher for geminate /l/ than for non-geminate /l/. It appears to continue to rise after the /l/ is formed. The fact that the midpoint of /l/ reaches a different formant value in the geminate might be compared to the phenomenon of vowel reduction, in which vowels attain different formant values depending on their length. The greater duration of geminate /l/ leads to a more nearly palatalized /l/. Although even geminate /l/ does not reach the values for full palatalization, the direction and magnitude of the change suggests that this is the origin of the change in Spanish.

Post-occlusive /l/

Methods

In order to determine whether post-occlusive /l/ also involves a more palatalized spectrum, eight pairs of words in Spanish having initial /l/ and initial stop + /l/ clusters were used. Four repetitions of each word were produced from a randomized list by a speaker of Buenos Aires Spanish in the frame "__________ otra vez." ("__________ once again."). Utterance initial position was used in order to minimize differences in the co-articulatory effects of preceding sounds. This also made possible the use of a nearly automatic measurement procedure. Since minimal pairs were not available in all instances, near-minimal pairs were utilized. The words used are given below.

clavo	"nail"
lavo	"wash"
cloro	"chlorine"
loro	"parrot"
clueca	"a brooding hen"
luego	"later"
clima	"climate"
lima	"file"
plancha	"an iron"
lancha	"a launch"
plata	"silver"
lata	"tin"
pluma	"a feather"
lumbre	"light"
plomo	"lead (the metal)"
lomo	"the small of the back"

The first 25 ms of each /l/ were analyzed using an LPC program on the PDP-12 computer at the UCLA Phonetics Laboratory. A subsequent analysis used maxima in the second derivative of the LPC spectrum to find the peaks (or formants) in the spectrum. An LPC analysis of consonants such as /l/ may sometimes yield misleading results, but a comparison of the LPC spectra with those produced by a Fourier analysis (which is not subject to the same kind of error) showed that this did not occur here. [5] The results correlated well with an FFT analysis of the data, so the output of the peak-picker was used.

Results

The results are given below:

	Post-occlusive /l/	Initial /l/
F1	302	277
F2	1577	1599
F3	2488	2458

The differences between mean F2's here is not significant. Note, in addition, that the F1 and F3 values for the two groups were also not significantly different. This part of the hypothesis is not confirmed. It was hypothesized that the tongue would be higher for post-occlusive /l/ than for initial /l/ due to the fact that the jaw must be relatively high for the occlusive. I would appear, however, that the jaw is already high for an initial /l/ so that the preceding stop does not have an effect. What, then, is responsible for the change in Italian of "plano" to "piano"? The change involves going from stop + /l/ + vowel to stop + /j/ + vowel. Such sequences can be quite similar acoustically, as is shown in Figure 1.

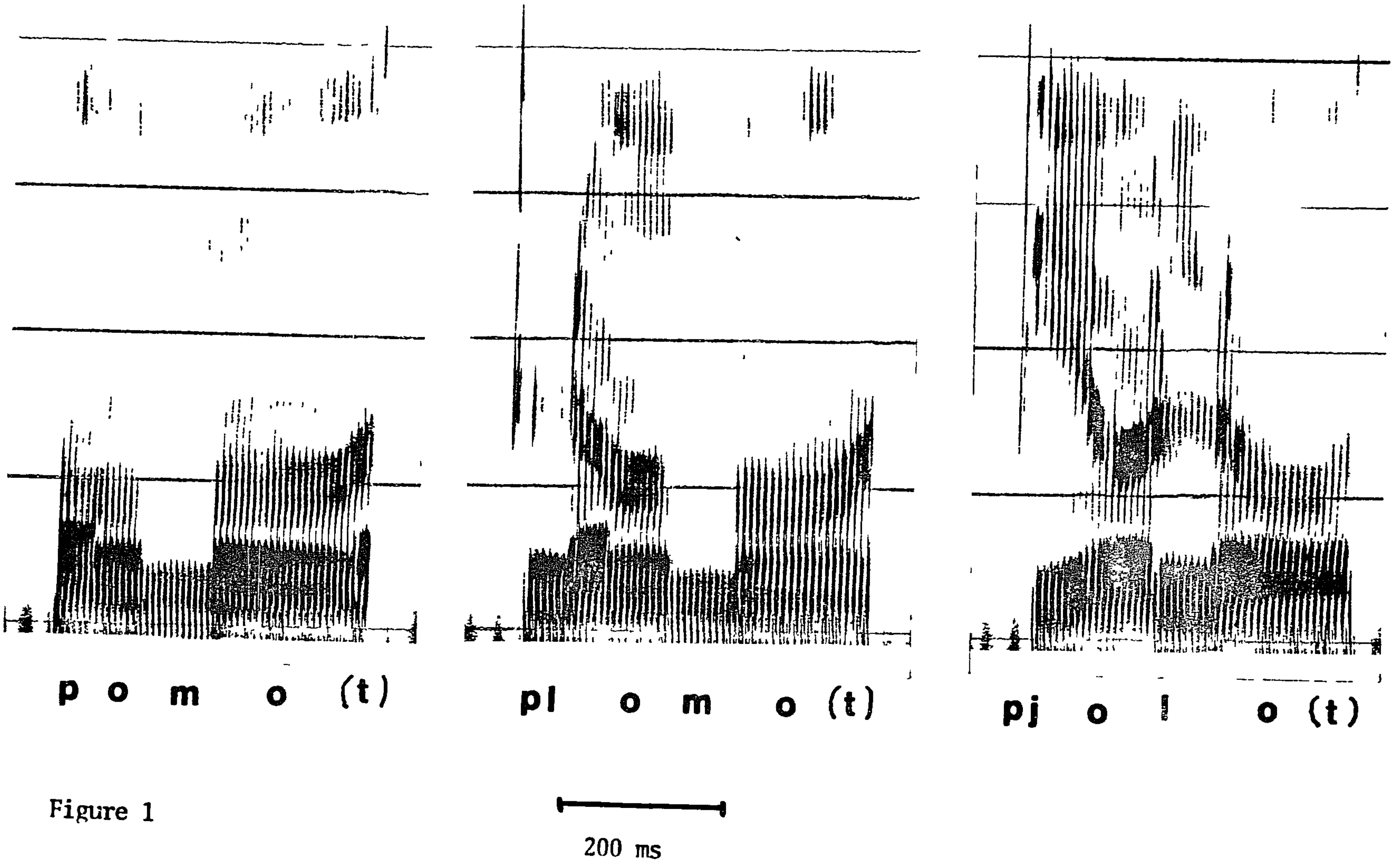


Figure 1

The /l/, and the transition from the /l/ to the vowel, can be quite similar to a palatal glide. The change occurred for /l/ following labial and velar consonants, and preceding all the Italian vowels /i,e,a,o,u/. Following velar consonants, and between velar consonants and back vowels, the presence of /l/ raises the second formant transition, in comparison with a stop + vowel sequence. At the time this change occurred, listeners may have reinterpreted the /l/ as a palatal glide. In other words, given the acoustic similarity of the sequences /plo/ and /pjo/, the two appear to have been confused by listeners, causing the sound change.

Discussion

Palatalization of geminate /n/

There has been a related change in Spanish, of geminate /n/ to palatal /n/. Examples are given below.

Latin	Spanish	English gloss
canna	cañja	"cane"
pannu	pañjo	"cloth"
annu	añjo	"year"

The development was concurrent with that of the palatalization of geminate /l/ and was almost certainly brought about in a similar way: gemination of the dental consonant caused a raising of the tongue body. Further work to determine actual tongue height values would not only help substantiate the explanation given here for the palatalization of /l/ but would (probably) explain the palatalization of /n/.

Conclusion

Note that the explanation for the Italian change is acoustic-based while the Spanish involves articulatory factors in addition. Note, also, that it has only been possible to explain the phonetic motivations behind these changes, not why each of the languages "chose" to make the changes it did. [6] The two changes, although closely related in terms of the sounds themselves, and occurring in two closely related languages of the same origin, show marked differences in their activating mechanisms. In Spanish, the fortuitous effect of gemination, the continuation of upward tongue movement during the extended duration of the /l/, caused the acoustic characteristics of the /l/ to change. This changed acoustical shape was interpreted by listeners as /j/, and came to be produced as /j/.

In Italian, the environment in which the /l/ occurred caused its acoustic shape to become more difficult to distinguish from the sound it eventually became. For a further illustration of this difference, assume that the reason for the continuation of movement occurring in the geminate /l/ resulted from the inertial properties of the tongue. [7] If speakers were endowed with a different articulatory apparatus, say, one with much less inertia, the change in Spanish would not have taken place, while the change in Italian would still have occurred.

Conversely, if the speech perception apparatus consisted of x-ray machines which allowed perceivers to accurately determine all articulatory activity, the change in Spanish might still have occurred, since the difference in gesture between /l/ and /lː/ would still be perceptible. The change in Italian, caused as it was by a difficulty in distinguishing /l/ transitions from /j/ offglides, would not have happened.

FOOTNOTES

1. This is a revised version of a paper presented at a meeting of the Acoustical Society of America, Spring 1977. The abstract appeared as Javkin (1977).

2. Some of the cases in which English appears to have changed /l/ into /u/ are spurious. The pronunciation of words such as "folk" /fouk/ and "yolk" /youk/ is due to the diphthongization of /o/, as can be seen in the word "yoke" /youk/. (Examples here are from Jones 1956).

3. The second formant value found by Von Essen for /i/ may be inaccurate. It is much higher than the value found by Lehiste & Peterson (1961) of 2200 Hz, and closer to their value for the third formant of 2800 Hz. Von Essen may have been measuring the third formant. This seems likely since he also finds what he terms an average spectral peak at 125 Hz for all his data. This is almost certainly just the fundamental frequency.

4. I am grateful to Eduardo Gutentag, who provided a number of examples used in this chapter, and Alessandro Durante, who helped construct the Italian word list for the experiment, and Professor Yakov Malkiel, who enlightened me as to the complexity of these changes in several discussions.I have used the following dictionaries in this chapter: Velasquez (1973), Corominas (1974), Lysle and Gualteri (1951), and Cassell's New Latin Dictionary: Simpson (1960).

5. LPC analysis is subject to two possible kinds of error if zeroes are present in the consonant being analyzed; this is the case with /l/. First, the zeroes are not detected: the output of the LPC produces something like a smoothed parabolic interpolation between the peaks of the spectrum. Second, the position of the peaks of the LPC output may differ somewhat from those in the actual spectrum if zeroes are present. Fourier analysis is not subject to these problems. Accordingly, the output of the LPC analysis was compared to a Fourier analysis of the same data. Although the results were not compared statistically, a careful visual inspection of the two outputs showed that the peaks found by the LPC analysis were within 50 Hz of the peaks found by the Fourier, with the exception of the two tokens of /l/. In these two cases, the zeroes were extremely close to the peaks representing the second formant. The exclusion

of these two tokens (and of two tokens of /i/ chosen arbitrarily) still did not yield significant results. One possible way in which the presence of zeroes could cause the change from /l/ to /i/ is if F2 were sufficiently reduced in amplitude, so that one of the higher formants might be misidentified as F2. The resulting percept would be one of a very high F2, which would be a cue for palatalization. Both analyses showed that F2 was not reduced relative to the higher formants. Therefore, listeners would be unlikely to re-interpret one of the higher formants as F2. Given that the LPC analysis provided similar results to a Fourier analysis, the LPC analysis was used, since this allowed the use of an automatic peak-picking procedure.

6. One possible explanation for the fact that Spanish palatalized geminate /l/ while Italian did not, is based on another sound change which occurred in Spanish. Spanish lost the former Latin palatalized /l/'s, which were retained in Italian. This is illustrated in the changes below.

Latin	Spanish		Italian		English gloss
	orthographic	phonetic	orthographic	phonetic	
meliore	mejor	mexor	meglio	meljo	"better"
folia	hoja	oxa	figlia	filja	"letter"
filiu	hijo	ixo	figlio	filjo	"son"
ampliu	ancho	antʃo	ampio	ampjo	"wide"
muliere	mujer	muxer	muliebre	muljebre	"woman"

It could be suggested that Spanish was able to palatalize geminate /l/ because it was losing its palatalized /l/. Since Italian did not lose palatalized /l/, it might resist the change, and thus keep the sounds apart. The problem with this explanation is that it requires another. How do we explain why Spanish lost its inherited palatalized /l/, while Italian did not?

7. This is, certainly, only one of several possible reasons; the upward movement could as easily be the result of active muscular effort.

CHAPTER 4

An Acoustic Explanation of a Sound Change in Lisu [1]

ABSTRACT

In Lisu -o is the modern reflex of *-ɑp, -e is the modern reflex of *-ɑt, and -ae is the modern reflex of *-ɑk. These changes lack a convincing articulatory explanation, but a simple and straightforward explanation exists: the steady-state values of the second-formant changed as a result of the acoustic influence of the following consonants on the vowel and vowel transitions.

Introduction

In this chapter I will discuss a set of related sound changes which occurred in Lisu which involve primarily the influence of a consonant on the acoustic characteristics of preceding vowels.

The change of *-ɑp to -o, *-ɑt to -e and *-ɑk to -æ.

Lisu is a member of the Lolo-Burmese subgrouping of Tibeto-Burman languages, which, in turn, are part of the Sino-Tibetan family. Lisu and the other Loloish languages have undergone a great deal of consonantal loss, but, because of the relative conservatism of related non-Loloish languages, Lisu can be reconstructed with a high degree of reliability. The changes which concern us are the following. In Lisu, -o is the modern reflex of *-ɑp, -e is the modern reflex of *-ɑt, and -ae is the modern reflex of *-ɑk. The basis for these reconstructions is illustrated in

Tables 1, 2 and 3.

Table 1

*-ɑp to -o

	Written Burmese	Lahu	Lisu	Akha	
*kjɑp	kjɑp			cɔ	"stick into"
*rɑp		ɣɔ̀ʔ	ɣɔ̀	ɣɔ	"needle"
*tsɑp		chɔ̂ʔ	tshɔ́	tsɔ	"stick into"
*C-kɑp	khɑp		khɔ̀	xɔ	"draw water"
*k-njɑp	hnjɑp	nɔ̂ʔ	njɔ́		"pinch"
*pjɑp			phɔ́	pjɔ	"turn over"

Table 2

*-ɑt to -e and -ɛ

	Written Burmese	Lahu	Lisu	Akha	
*k-r-wɑt	krwɑt	vɛ̂ʔ	wè	jɛ	"leech"
*cɑt	tshɑt	chɛ̂ʔ	chwé	tsɛ	"break in two"
*C-sɑt	sɑt		ʃè	sɛ	"kill"
*sə-wɑt	wɑt	ʃí-vɛ̂ʔ	sí-we	jɛ	"flower"
*tsɑt	tshɑt		tshɛ́	tsɛ	"deer"
*kjɑt			chɛ́	cɛ	"run"

Table 3

*-ɑk to -æ and -ɑ

	Written Burmese	Lahu	Lisu	Sani	
*zɑk	sɑk	jàʔ	jæ̀	ze	"descend"
*wɑk	wɑk	vɑ̀ʔ	á-væ̀	ve	"pig"
*ʃɑk		ʃɑ́	xɑ́	ʃe	"tear"
*nɑk	mənɑk	nàʔ	næ̀	ne	"early morning"
*mɑk	mɑk	màʔ	mæ̀	me	"army; war"
*s-nɑk	nɑk	nàʔ	ŋæ̀	ne	"black"
*ʔqɑk	ɑ-khɑk	qá	sí-kɑ́	qe	"branch"
*kɑk		qhâʔ	khɑ́	tʃ'e	"village"
*C-sɑk	sɑk		ʃæ̀	se	"breath"
lɑk	lɑk	làʔ	læ̀	le	"hand"

Of the languages given, Written Burmese is a Burmish member of the Lolo-Burmese subgrouping, while the others are Loloish. Note especially that Written Burmese has retained the endings that are reconstructed for Lisu. The correspondences for the *-ɑp reconstructions are very straightforward. Some variation exists in the reflexes. After velar initials the -o is sometimes replaced by -u, and in Fraser's dialect of Lisu (Fraser 1922) the reflex is graphically represented by -aw which, at that time, usually represented -ɔ. The Lahu reflex is -o, the Lisu reflex -o or -ɔ, and the Akha reflex -ɔ. Some variation exists for the *-ɑt ending, as can be seen in Table 2, but it is irrelevant to this discussion.

As can be seen in Table 3, the *-ɑk either becomes -ɑ in Lisu or becomes -ae. In Sani (another Loloish language) all *-ɑk rhymes to go -e. In modern Burmese, the reflex of *-ɑk is -ɛ followed by glottal stop. The vowel becomes a front vowel in all these cases in which it changes. Some of the variations will be discussed later in this paper. The changes from proto-Lolo-Burmese to Lisu are, therefore: *-ɑp to -o, *-ɑt to -e, and *-ɑk to -ae.

Historically, *-ɑp, *-ɑt and *-ɑk underwent similar changes. The final stops became glottal stops which later disappeared, leaving only laryngealization on the vowel. Each of these stages is attested in data from other Loloish languages. Phunoi, for example, still retains the final stops from the *-ɑp and *-ɑt rhymes, but the final -k from *-ɑk has dropped. The final glottal stops are found in the Lahu data above, and the laryngealization is found in a number of languages of the Loloish group. Accompanying these changes, there was frequently a change in vowel quality; it is this change in vowel quality which is the focus of this study.

Discussion

An examination of the articulatory facts shows that they do not account for these changes. One might think that /t/ has a tongue shape similar to /e/, but it is certainly no more like that vowel than any other non-low front vowel. In fact, though, x-ray studies show that the tongue configuration of dental stops is very different from those for vowels such as /i/, /e/, or /ɛ/ (see, e.g., Perkell 1969). For /t/, the tongue has a very flat and horizontal position against the rim of the alveolar ridge, while it is bunched toward the hard palate for the four vowels. The main constriction is palatal for the vowels, and dental for /t/. In the case of the change from *-p to -o, both sounds can be called labial, but the labiality is different in the two cases: -o involves rounding and -p does not. Furthermore, -p has no effect on tongue position. It is thus difficult to see what vowel might result from assimilation to the articulation of -p. Articulatory correlates of *-ɑk and -ae are simply nonexistent. The tongue position for /k/ varies with the preceding and following vowel, so that the influence of /k/ on vowels is difficult to define in articulatory terms. To the extent that one can make an articulatory determination, the different types of -k articulation that occur are most similar to the tongue configuration for -u and -o, and quite unlike the configuration for -ae (Ladefoged 1968, 1971).

The explanation of these changes apparently lies in the acoustic influence of the consonant on the vowel. In each case, the vowel has taken values determined by the acoustic characteristics of the stop. Subsequently, the final stop became a glottal stop and eventually disappeared. The explanation is tighter for the *-ɑp to -o and *-ɑt to -e changes than for the change from *-ɑk to -ae, but it fits all three.

A number of studies conducted at Haskins Laboratories some time ago (cf. Cooper, Delattre, Liberman, Borst and Gerstman 1952; Liberman, Delattre and Cooper 1952; Liberman, Delattre, Cooper and Gerstman 1954; Delattre, Liberman and Cooper 1955; Liberman 1957) showed that one of the characteristic cues for the place of articulation of stops and nasals consisted of the direction and rate of the different formant trajectories from the consonant to a following vowel or from the vowel to the following consonant. The first formant correlated less with place of articulation than with voicing distinctions. The transition of the second formant determined whether subjects listening to synthetic speech heard bilabial, alveolar, or velar consonants. Although the shape of the transitions varied within the same consonant, loci were found toward which the second formant was directed. These loci were: 720 Hz for bilabials; 1600-1800 Hz for alveolars; and 3000 Hz for velars in the environment of front vowels.

The steady-state value of the second formant is one of the most important cues for the identification of vowels. In an adult male speaker, the second formant for -o would be close to 700 Hz, while for -e or -ɛ it would be close to

1800 Hz. We can thus see an acoustic relation between the consonant transitions and the value for the vowels.

The acoustic influence of consonants on preceding vowels has been further documented and the original Haskins determinations refined by acoustic measurements by a number of investigators, including Lehiste and Peterson (1961), Lindblom (1963), Öhman (1966), Stevens and House (1963), and Stevens, House and Paul (1966). The first formant was found to be affected less than the second, and the place of articulation of the consonant had relatively little effect. The tendency of the second formant rising for -a followed by -t and lowering for -a followed by -p is generally supported in these studies, although the values given have differed. The second formant transition was generally higher for alveolars or dentals than for labials.

The *-ɑk to -ae change is similar although the data are more complicated. The Haskins perceptual data show that, after front vowels, there was a locus of about 3000 Hz. For back vowels, there was no single locus and the transitions were moderate. At the moment of consonant closure, in any case, the second formant value was slightly higher than the value for the vowel. Since the shift in the kind of formant transitions occurs somewhere between a front -a and a back -ɔ and since the original Lisu vowel is reconstructed as a low central vowel, the determination of a second formant value is problematical. Nonetheless, the historical and acoustic facts strongly suggest that a rising second formant transition (which is common to most varieties of -k) caused the raising in the second formant that is found in the *-ɑk to -ae change. Interestingly, the degree of variability for the transition of /k/ correlates nicely with the variability among the reflexes of *-ɑk; in the different languages in the Loloish branch *-ɑk becomes -e, ɛ, or -ae, depending on the language.

Locus theory has been criticized, chiefly because the formant transitions from real speakers differ from, and are not as consistent as, those of perceptual experiments in which the transitions were the only cues. The differences do not alter this analysis. The production data are very clear for labials and alveolars and dentals, with labials generally showing a lower value for the consonant transition than alveolars and dentals. Velars raise the second formant transitions sharply in the environment of front vowels, and have a variable effect in the environment of non-back vowels. In Lehiste & Peterson (1961) velars in the environment of back vowels showed the lowest formant transitions. Lindblom (1963) found more of an upward perturbation for /a/ in the environment of g_g than in the environment d_d. Öhman (1966) found a greater upward transition for /g/ than for /d/ in the environment a_a, but also found that /d/ had a higher transition than /g/ between back vowels. Stevens & House (1963) and Stevens, House & Paul (1966) found that for front vowels, velars showed higher perturbations than postdentals which in turn showed higher perturbations than labials. For back vowels, the effect of velars and dentals is reversed, but again both show a higher perturbation than labials.

It is clear from the early experiments of the 1950's that vowel transitions could effectively serve for the complex cues that occur in stops. Lisu speakers encoded the information present in the place of articulation of the consonant into the preceding vowel as the consonant became a glottal stop. The consonant transitions thus became interpreted as glides, with the vowels assimilating, in various degrees, to those glides.[2]

Cases of similar developments have been noted since this analysis was first presented (Thurgood & Javkin 1974). Michailovsky (1975) found that the analysis for alveolar and dental consonants also applied to a number of other languages. The phenomenon of a consonant leaving an imprint (acoustic or articulatory) on the preceding vowel is not at all uncommon. For example, final nasal consonants often drop out, leaving the preceding vowel nasalized. The changes described here are similar. In one case, the speakers of a language reinterpret nasalization as characteristic of the vowel rather than as the allophonic result of a final nasal; in the other, speakers reinterpret the second formant transition as an integral part of the vowel rather than as a transition to the final stop.

Conclusion

As an explanation of a sound change, the acoustic assimilation of the second formant of the vowel to the second formant of the consonant is not restricted to a handful of well-chosen examples from Lisu. The phenomenon is widespread in Sino-Tibetan languages and involves not just stop finals, but nasal finals as well. Thus, changes such as *-ɑm to -o and *-ɑn to -e paralleling the *-ɑp to -o and *-ɑt to -e changes are found in many Sino-Tibetan languages, among them Lahu and Lisu itself.

Three sound changes have been examined here: *-ɑp to -o, *-ɑt to -e, and *-ɑk to -o. The second formant values of the vowels took on the second formant values of the following consonants. Since the second formant transitions are determined, other things being equal, by the place of articulation of the consonant, these vowel quality changes are best explained in terms of the unit consisting of the vowel and final consonant.

These changes have arisen as a result of the fact that speakers of Lisu perceived the older forms as similar to the forms they came to pronounce. They could not have perceived articulatory similarities, only perceptual ones. Those similarities arose from some of the characteristics of the vocal tract as an acoustic resonator. These can cause a convergence of the output of more than one resonating cavity so that they become perceptually similar. A number of sound changes have therefore been explained as having been caused by acoustic factors. Nevertheless, a number of changes which are clearly based on this kind of convergence have been explained as being due to articulatory and/or perceptual factors. At times, researchers have presented exclusively acoustic data, and yet have failed to acknowledge the role of acoustics in the change.

Von Essen (1964) explained the changes of /l/ to /j/ and /l/ to /w/ which have occurred in various languages as being acoustic-based. He noted, among other cases, the change from /l/ to /w/ in Polish. Jonasson (1971) explained the same change in different terms. Jonasson presented spectrographic evidence showing that the two sounds are very similar acoustically. He presented no perceptual evidence. He suggested that the perceptual similarity between the two sounds implied by the acoustical similarity led to what he termed "articulatory reinterpretation" on the part of speaker-listeners. Although Jonasson's explanation provides an example of what I call an acoustic-based explanation, the terms "perceptual similarity" and "articulatory reinterpretation" emphasize two areas of the speech transmission process which are secondary in importance for this change compared to the effects of the acoustic medium.

This is unfortunate. Jonasson argues convincingly that these changes are very poorly described in articulatory terms. It is also clear that the perceptual similarity is not accidental. It is the result of the fact that two very different articulatory shapes can result in similar acoustic shapes. This is a fact of acoustics, which can be explained by examining the properties of resonators. The perceptual confusions between the two sounds are thus a secondary effect of the acoustic-based similarity.

As mentioned in Chapter 1, speech has been compared to other types of communication systems. Communication systems typically consist of a transmitter, a transmission line, and a receiver. In the sound changes discussed here, the characteristics of the "line", the acoustic medium, are primary in importance. The perceptual similarity of one articulatorily permissible sound to another are resultant factors and also important, but would not occur if it were not for the specific characteristics of the acoustic medium.

FOOTNOTES

1. This chapter is a revised version of a paper by Thurgood and Javkin (1975). The reconstructions exemplified here were prepared by Graham Thurgood.

2. The stage where the second formant transition becomes interpreted as a glide can be seen in some of the Burmish data. In Taungyo, a Burmish language, *-ɑt has gone to -ɑit.

*nɑt	nɑit	"spirit"
*sɑt	-ɑit	"kill"
*tɑt	tɑik	"be likely (be wont)"
*-wɑt	-wɑik	" clothes"

CHAPTER 5

The Perceptual Basis of Vowel Duration Differences Associated with the Voiced/Voiceless Distinction [1]

Abstract

Lisker (1974) has shown that there existed no viable explanation of vowel duration differences associated with the voicing state of the following consonant. Lisker's conclusion, however, that such differences are due to constraints specific to English, does not adequately explain such differences in other languages. Scully (1974) suggests that such differences are due to the different aerodynamic conditions that depend on the voicing state of the following consonant. Scully's aerodynamic model does not provide results entirely consistent with speech data, however, and is therefore open to question. A different kind of explanation may be needed. Evidence from perceptual experiments involving English speakers suggests that voiced vs. voiceless consonants make the preceding vowel appear to be longer vs. shorter. Although these results have yet to be tested with speakers of other languages, they support an interesting hypothesis: the voicing of the following consonant, by continuing the voicing present in the vowel, may create the illusion of greater length. Such an illusion, acting as an input to phonological acquisition, may explain the observed vowel duration differences.

Introduction

In this chapter I will be dealing with a phonological universal which has led to a sound change in at least two languages. The longer duration of vowels before voiced consonants, compared to their duration before voiceless consonants, has been reported in many languages. In certain environments, English and Latin have replaced the voicing distinction of the consonant following the vowel with a vowel duration distinction.

While in the case of the sound changes discussed earlier, listeners either heard a sound somewhat perturbed by articulatory factors or made certain sporadic errors in the discrimination of similar sounds, a different process occurred to cause the universals about to be discussed. In these cases, listeners made consistent and systematic errors in their perception. This is a strong claim, but I believe the following pages contain strong support for it. It should be noted especially that no alternative explanations have proved valid, despite the fact that a great deal of effort has gone into finding explanations of another kind.

The longer duration of vowels before voiced consonants, compared to their duration before voiceless consonants, has been observed in many languages by a large number of investigators, e.g. Sweet (1877) and Meyer (1903) for English, Meyer (1904) for German, Meyer and Gombocz (1909, cited in Elert, 1964) for Hungarian, Gregoire (1911) for French, Metz (1914) for Italian, Navarro Tomas (1916) for Spanish, Einarsson (1927, cited in Elert 1964) for Icelandic, Fintoft (1962) for Norwegian, Elert (1964) for Swedish, Fischer-Jórgensen (1964) for Danish, Slis and Cohen (1969) for Dutch, Maddieson and Gandour (1975) for Hindi. The observations of Meyer for German have been re-examined by Maack (1953). The observations of Sweet and Meyer for English have been re-examined by House and Fairbanks (1953), Peterson and Lehiste (1960), House (1961), Sharf (1962), Klatt (1973b). The more recent work generally confirmed the earlier, although the magnitude of the duration difference for English was found to be relatively small when the words were not in utterance-final position (Umeda 1975, Klatt 1975).

In these studies, English tended to show a greater vowel duration difference than other languages, and a similar trend appeared in comparative studies by Zimmerman and Sapon (1958), who examined Spanish and English, and by Chen (1970), who compared English to French, Russian and Korean. Chen's work added Russian and Korean to the languages in which greater vowel duration before voiced consonants was found to occur. Zimmerman and Sapon were criticized by Delattre (1962) for their failure to isolate the voiced/voiceless distinction as the only independent variable; and Chen, who compared, for example, Russian disyllables to English monosyllables, is open to similar criticism. Nevertheless, the greater difference for English seems rather well-established (Lehiste 1970). The relatively large duration difference associated with the voicing of the following consonant which occurs in English has been shown to function as a perceptual cue under certain conditions (Denes 1955, Raphael 1972), a function which, for the most part, has not been established for other languages. In Latin, vowels became distinctively long before clusters of the form voiced obstruent + voiced obstruent by Lachman's Law, while the voicing of the first obstruent was lost (Kent 1928).

Proposed explanations

A number of explanations have been offered for the greater vowel duration before voiced consonants. Lisker, in 1974, reviewed most of the explanations that had been proposed to that time and showed that all were seriously flawed.

Halle and Stevens (1967), for example, had suggested that "vowels are lengthered before voiced stops to allow time for laryngeal re-adjustment needed if voicing is to be maintained during closure."[2] Lisker pointed out that laryngoscopic and electromyographic data (Lisker, Sawashima, Abramson and Cooper 1970) showed no such adjustment for voiced stops. In fact, it showed the opposite, that the only adjustment noticed was the separation of the vocal folds necessary for voiceless stops. Lisker also discussed the various explanations that posit a shortening of vowels preceding the supposedly more fortis, or more strongly articulated, voiceless stops. Variations of this explanation have been suggested by a number of investigators (Meyer

1903, Delattre 1941, Belasco 1953, House 1961, Malecot 1970). Lisker showed that there was no agreed-upon measure of what constitutes articulatory effort, and that, when fortisness is defined, that definition is in terms of an earlier closure and later release. Since an earlier closure simply entails a shorter vowel, the putative explanation is no more than the fact for which an explanation is sought.

One explanation involving a faster closure was proposed by Chen (1970), who suggested that the closing gesture for voiceless stops is more rapid than that for voiced stops because the articulators have to withstand greater oral pressure in the case of voiceless stops. This argument fails on several grounds. If it were correct, we would expect that close vowels would be less affected than open vowels. Since a close vowel such as /i/ requires less articulator movement than an open vowel such as /a/ (for dental stops, for example), a difference in rate should have less effect. This prediction is not borne out by data presented by Peterson and Lehiste (1960) and by Sharf (1962). Both of these studies show approximately the same influence from voicing of dental consonants following /i/ and following /a/.

Furthermore, the rate of consonantal closure is not sufficiently different for voiced vs. voiceless stops to account for the vowel duration difference. Taking the combined rate of movement for both the jaw and lips for the closure of labials, Sussman, MacNeilage and Hanson (1973) found that, after /i/, voiced stops showed a faster rate of closure than did voiceless stops. Since close vowels show the same relationship as other vowels in their duration differences before the voiced and voiceless consonants, one would have to posit a different explanation for close vowels.

In any case, Chen may have confused cause and effect. Even if the final closing velocity of the jaw were faster for final voiceless consonants, this could be due to the requirement of shorter vowel duration for these consonants and not the cause of that shorter vowel duration.

A somewhat teleological explanation was proposed by Klatt (1973a) who suggested that:

> One possible reason for a universal tendency to encroach on the duration of a vowel if a voiceless segment follows concerns the relative timing of oral closure and glottal opening gestures at vowel termination. A slightly delayed glottal opening gesture is likely to produce a few cycles of vibration during closure, which is a cue to voicing. Since perfect synchrony of glottal and supraglottal activity is impossible, it is proposed that the glottis is normally opened somewhat early in order to avoid generation of a false voicing cue.

If this hypothesis were correct, we would not expect vowels to shorten before voiceless consonants in languages which use vowel length distinctively. German and Hungarian have such a contrast and have vowel duration differences based on the voicing of the following consonant. Although Meyer (1904) reports that the duration difference in German does not cause an overlap between short vowels preceding a voiceless consonant, the vowel duration difference associated with voicing surely weakens the perceptibility of the vowel length distinctions. To accept Klatt's explanation we would have to explain why German (and Hungarian) would sacrifice some of the perceptibility of one distinction in order to preserve another.

Lisker's own explanation for the phenomenon is that vowels become shorter before voiceless consonants in order to prevent such consonants from becoming pre-aspirated. It is based on the fact that the vocal folds begin to separate during the final portion of vowels preceding voiceless consonants. If the closure for the consonant were delayed, pre-aspiration would result. Since pre-aspiration is not permitted in English, the closing gesture occurs immediately after the vocal folds start to separate. It is maintained that this close coordination of vocal folds and vocal tract causes the shorter vowel duration.

Lisker's argument uses one fact of English to explain another. The constraint on pre-aspiration is said to explain the longer duration of vowels before voiced consonants. The constraint on pre-aspiration is then left unexplained (in fact, the phonetician's responsibility for explaining the pre-aspiration constraint is specifically denied). It thus puts the ultimate reason for the vowel duration difference outside the realm of phonetic explanation, an unsatisfying result at best. Not only does the explanation fail to explain why every language thus far examined shows the same kind of difference (albeit a smaller one), but also some important phenomena of vowel duration in English are left completely unexplained. Klatt (1975), using English data, found that the influence of the following consonant is highly dependent on the syntactic environment. Since the injunction against pre-aspiration in English does not vary according to these environments, it cannot be the major factor responsible for variable vowel duration differences.

The above explanation is limited to vowel duration differences in English. An explanation suggested by Scully (1974), which would apply to all languages, involves reference to an aerodynamic model that predicts vowel duration differences based on the different aerodynamic conditions that occur in voiced and voiceless stops. It predicts that vowel amplitude will decay sooner and build up later when a voiceless consonant occurs between vowels than when a voiced consonant does, even though the supraglottal consonantal closing gestures are timed identically.

There are some problems with the model. True, it predicts that vowels will be shortened when they precede voiceless fricatives. However, it also predicts that vowels will be shortened even more following voiceless fricatives. This last effect is not borne out by the measurement data of Peterson and Lehiste. The model also predicts that the duration of vowels is the same regardless of the voicing of the preceding stop, if aspiration is counted as part of the vowel. Peterson and Lehiste show that vowels following aspirated stops are shorter if the aspiration is not counted, but longer if it is counted. In neither case are the durations equal, which is the prediction of the model.

Scully's aerodynamic model deserves further investigation, in particular, to determine if the drop in vowel amplitude predicted by the model occurs in real speech, and to determine if simple modification would eliminate the model's discrepancies with speech data. Nevertheless, since the predictions of the model show discrepancies with duration data from real speakers, we must conclude that the model has not been successfully demonstrated.

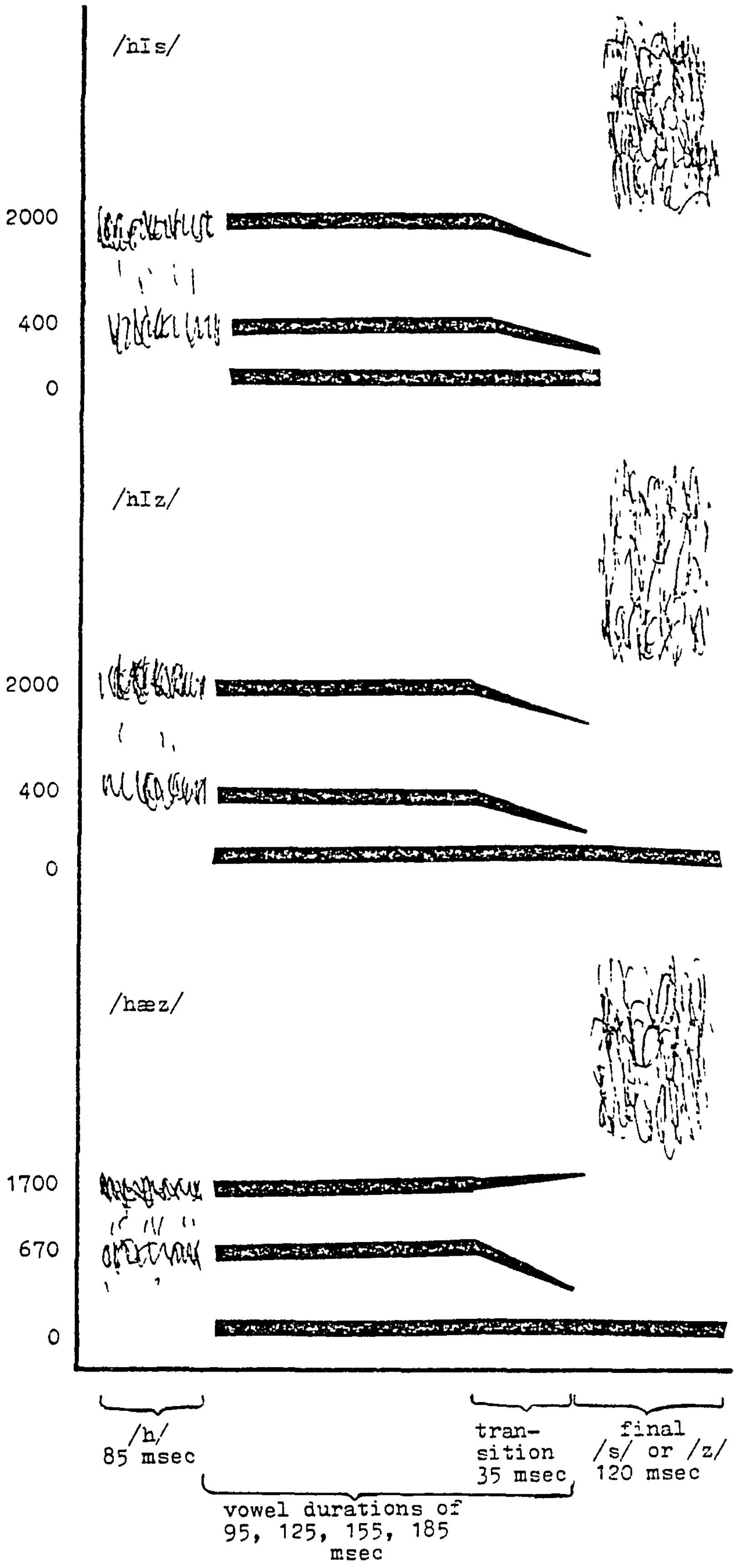
A hypothesis

It seems reasonable, then, to look elsewhere for the basis of vowel duration differences. It is possible that the explanation lies not in the constraints of speakers' articulatory mechanisms, but in the constraints of their auditory mechanisms. The fact that one of the characteristics of the vowel, voicing, continues into the consonant, might make the vowel appear to be longer than if voicing abruptly ended. [3]

Experiment 1

An experiment was constructed to test this hypothesis. The English words "hiss" (/hIs/) and "his" (/hIz/) were synthesized as shown in Figure 1.

Figure 1:
Synthesized stimuli



Both words had identical duration for the initial /h/ and the final sibilants, 85 and 120 ms, respectively; there were no formant transitions between the /h/ and the following vowel, the two segments differing only in voicing; there was a 35 ms transition between the vowel and the following sibilant. The words differed only in that voicing, in the case of /hIs/, ended at the end of the transition, whereas, in the case of /hIz/ voicing continued throughout the final /z/. The vowels were given durations of 95, 125, 155, 185 ms including the transition. This provided for two conditions and four durations. There were three repetitions of each stimulus, and the stimuli were presented in quasi-random order. The presentation of the stimuli and the recording of responses were controlled by a small laboratory computer. The subjects heard one of the two words, at one of the four durations, followed by a tone (250 Hz) a half second later. Their task was to match the duration of the tone to the duration of the vowel in the word. Subjects could control the duration of the tone by adjusting a knob. They could hear each word and following tone as many times as they wished by pressing a button. When they were satisfied that the vowel and the tone had the same duration, they pressed another button. This caused their last adjusted tone duration to be recorded and started the next trial.

Eight subjects participated in this experiment. In this and in the two experiments described below, the subjects were students, native speakers of American English, with no history of hearing pathology, who served with no compensation.

<u>Results and discussion of experiment 1</u>

Table 1 shows the mean response durations averaged over eight subjects and three repetitions for both the voiced and voiceless sibilants.

Table 1

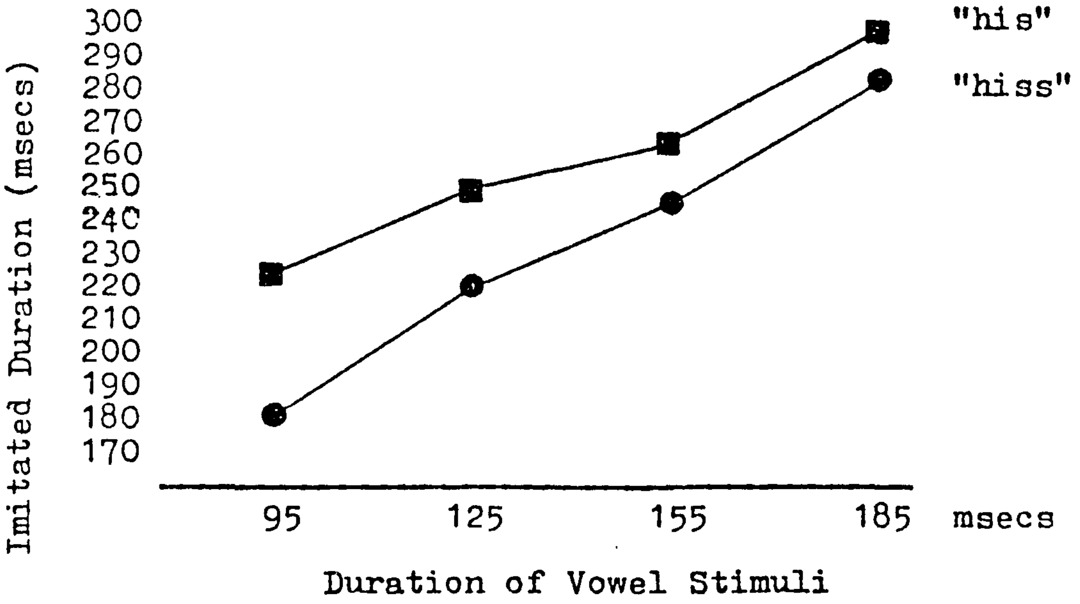
Imitation of Vowel Stimuli - Experiment 1

Duration of subject-produced tones in ms

	Stimulus duration			
	85	125	155	185
Stimuli				
"hiss" /hIs/	181	221	246	280
"his" /hIz/	222	250	264	298

Figure 2 is a graphical representation of response durations as a function of the different stimulus vowel durations.

Figure 2:
Averages of 8 subjects, 3 repetitions



The results were shown to be significant at the 0.02 level by a two-way analysis of variance. Subjects consistently made the imitating tone somewhat longer than the vowel they were imitating. This may have been due to their including the /h/ as part of the vowel, or their including part of the /h/ and part of the following sibilant. Note that if 85 ms of /h/ duration is added to the duration of the stimulus vowels, the subjects appear to be much more accurate, especially for the voiceless sibilant.

The results of the experiment support the hypothesis. Speakers perceive vowels preceding the voiced consonant as longer than vowels preceding the voiceless consonant. It is possible that listeners were influenced in their responses by their knowledge of English sound patterns, i.e., since English vowels are longer before voiced consonants than before voiceless consonants they may have expected the same pattern to be present in the stimuli and their judgments of vowel duration might have reflected their expectations rather than what they heard. This possibility is minimized by the fact that the stimuli were presented in isolation and in random order, so that the listeners really had no opportunity to compare a given stimulus to another. Nevertheless, a second experiment was conducted to determine how the sound patterns found in English may have affected these results.

Experiment 2

It is well known that low vowels, other things being equal, are longer in duration than high vowels (Peterson and Lehiste 1960). As Cordax and Krones showed (1973,1976), speakers are at least aware of the duration differences between /I/ and /æ/. The synthesized word "has" (/hæz/) was therefore added to the experiment. If listeners reported the low vowel to be longer than the high vowels in the experiment (even though the stimulus vowels were equal in duration), it would suggest that they were following their intuitions about English. If listeners did not report that vowel as longer, it would indicate that the judged duration differences for the voiced/voiceless difference were not influenced by knowledge of the phonetic pattern of the language. The second experiment followed the same format as the first, except that the words were: "hiss," "his" and "has." As in the first experiment, there were four durations and three repetitions.

Results and discussion of experiment 2

Twelve subjects participated in this experiment, including one who had participated in the first experiment. Table 2 shows the mean response durations for the three words averaged over all four stimulus vowel durations.

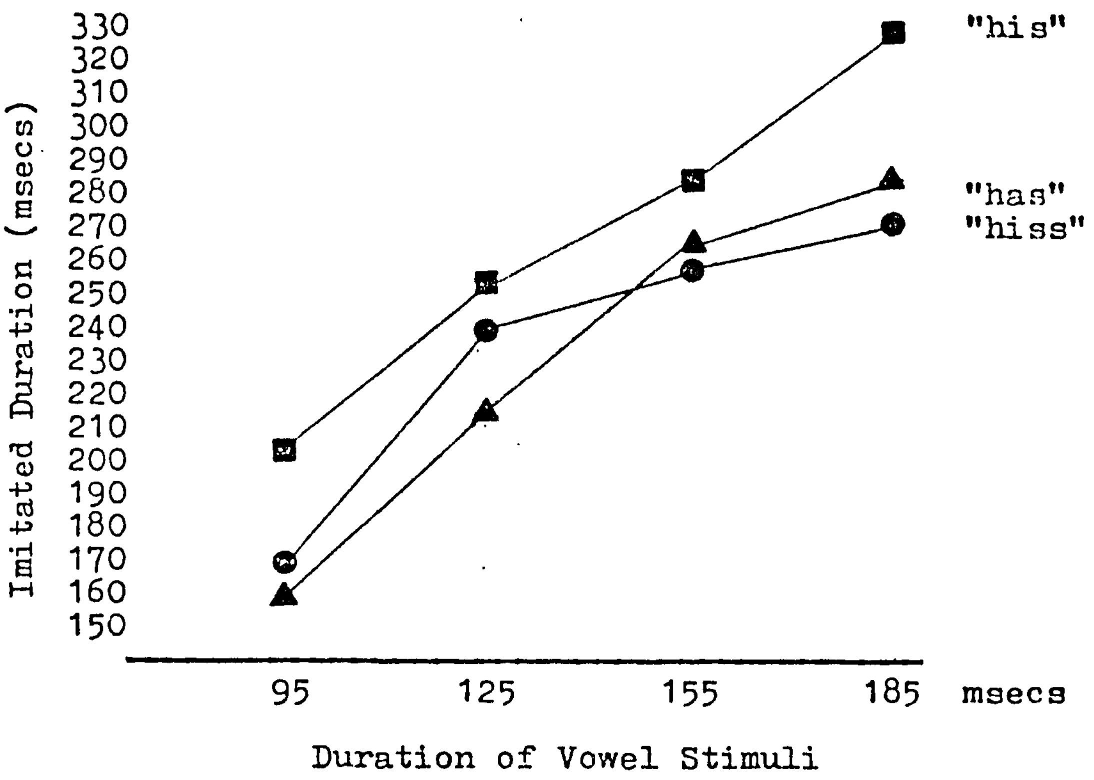
Table 2 Imitation of Vowel Stimuli - Experiment II

Duration of subject-produced tones in ms

Stimulus duration	95	125	155	185
Stimuli				
"hiss" /hIs/	164	238	261	277
"his" /hIz/	203	255	286	328
"has" /h z/	161	215	264	289

Figure 3 is a graphical representation of the response durations as a function of the different stimulus locations.

Figure 3:
Averages of 12 subjects, 3 repetitions



A two-way analysis of variance showed the results to be significant at the 0.01 level. Vowels were perceived to be shorter for "hiss" than for "his" by about 12 percent, which agrees closely with the results of Experiment 1. Surprisingly, "has" was perceived to have the shortest vowel duration of the three stimuli. This does show that listeners were not simply following their English intuitions, but it is somewhat puzzling.

It is possible that the sharp F1 transition for /æ/ (which, like the transition for /I/, took 35 ms) may have caused this. It is also possible that listeners thought the vowel in "has" to be short for that vowel, and thus judged it as very short. Although this suggests that the listeners were, to some extent, reacting to their knowledge of English, note that this judgment for "has" is opposite in direction from what was observed for the "his"/"hiss" difference. Listeners presumably knew that "has" contains a long vowel. As Peterson and Lehiste (1960) showed, /æ/ has an average duration of close to 300 ms. The /æ/ vowel in this experiment had an average duration of 140 ms. If the listeners were expecting a relatively long vowel, they may have judged the vowel they heard as very short. Now, they presumably knew that vowels are longer before voiced consonants. Application of speakers' knowledge to the voiced/voiceless distinction would have led to vowels before the voiced consonant being judged as shorter. In other words, given that the "his" and "hiss" stimuli were of the same duration, "his" might have been judged as having a short vowel for that word. If any such effect occurred, it was evidently overshadowed by the perceived lengthening due to the voicing of the following consonant. In any event, the second experiment suggests that listeners in this kind of experiment do not simply imitate the durations that occur naturally within English.

Experiment 3

For comparison with the results of Experiments 1 and 2, it would be of interest to see exactly what listeners' internalized knowledge of the duration patterns of English would lead them to do in judging the vowel durations of the words in Experiment 2, without the influence of the synthesized words. In order to do this, a third experiment was conducted. The subjects were asked to adjust the duration of the tone not to a vowel in a synthesized word, but to a vowel they said to themselves. In order to keep the experimental situation as comparable as possible to the first two experiments, a similar procedure was followed. Instead of hearing "hiss," "his" and "has," the subjects heard the cue words "a," "b," and "c," that is, /ej/, /bi/, and /si:/. If they heard "a," for example, they were to say "hiss" to themselves and copy the duration of their own vowel by, once again, adjusting the duration of the tone. If this experiment patterned like Experiment 2, it would suggest that the results of that experiment could be due to the listeners' intuitions about English. Twelve subjects, two of whom had also been subjects in Experiment 2, participated in this experiment. The results are given in Table 3.

Table 3

Imitation of Internal Production - Experiment III

Word	"hiss"	"his"	"has"
Duration of subject-produced tone (ms)	255	340	365

The word "hiss" showed the shortest duration, followed by "his," followed by "has." This pattern, and the magnitude of the differences, corresponds closely with the durations that these vowels have in spoken English. This suggests that the results in Experiment 2 were not merely due to listeners' intuitions about English, but rather to the inherently different auditory characteristics of the different stimuli. The perceived difference in the durations of the vowels in "hiss" vs. "his" therefore seem to be due to the continuation of voicing into the consonant in the case of "his."

General discussion and conclusion

There is some additional evidence which supports the two critical points of this study, that (1) listeners perceive vowels before voiced consonants to be longer than vowels before voiceless consonants, and that (2) this perceptual difference forms the basis for the vowel duration differences that are actually manifested in speech.

Raphael, Dorman and Liberman (1975) found that vowels were more likely to be judged as short when preceded by the voiceless consonant /s/ than when preceded by the voiced consonant /d/ (or when a vowel begins the syllable). This result, on the perceptual effect of the consonant, parallels the results of the first two experiments described in this paper. (It should be noted, however, that it is still necessary to determine why this perceived vowel duration difference does not lead to vowel duration differences based on the voicing of the initial consonant.)

For support of the second point, we may turn to the evidence in Sharf (1964) that whispered speech also manifests vowel duration differenced based on the "voicing" of the following consonant. Sharf correctly viewed this as evidence that the vowel duration difference is part of the linguistic code of English. If it were the case that articulatory factors contributed to the vowel duration difference, we would expect that the difference would diminish somewhat, even in English, when the supposed articulatory basis for it, a difference in voicing, was removed. As this did not occur, we can only conclude that articulatory factors are not involved. This evidence is much more consistent with the notion that vowel duration differences associated with the voicing of the following consonant stemming from perceptual rather than from articulatory factors.

It should be emphasized that the hypothesis put forth here states only that there is a universal auditory constraint which makes vowels before voiced consonants appear to be lorger than they are. It makes no claim that ALL languages will develop actual (not apparent) vowel duration differences due to this factor and thus come to use the vowel duration as a major distinctive feature in forming lexical contrast. English happens to have undergone these processes as the above-cited phonetic studies and the Denes perceptual tests have shown. However, I make no claims that all languages show the same results in such tests, only that all listeners will manifest the effects of the posited auditory constraint as revealed by the experiments reported here.

In English, of course, the tendency caused by that constraint has become exaggerated and integrated into the linguistic code, so that the vowel duration difference can be used to form lexical contrasts. For that development, there is at present no explanation. Rather, the explanation here is for the phonetic differences that formed the basis for the growth of that contrast.

FOOTNOTES

1. This is a revised version of a paper presented at the Fall meeting of the Acoustical Society of America, 1975. An abstract appeared as Javkin (1975a) and an earlier revised version appeared as Javkin (1976a).

2. Quoted from Lisker (1974).

3. Parker (1974) suggested a superficially similar proposal regarding the abruptness of termination of the vowel, vowel duration, and the voicing of the final consonant. Careful reading, however, shows that while the proposal presented here assumes the continuation of voicing into the consonant causes the perception of greater vowel length (eventually leading to the production of greater vowel length), Parker's proposal attempts to derive both greater vowel duration and the voicing of the final consonant from a loosely defined "gradual vowel offset". Any similarities between the two proposals are thus purely coincidental.

CHAPTER 6

The Perceptual Basis of Perseverative Tone Spreading [1]

ABSTRACT

Hyman and Schuh (1974) noted that tone spreading, in which a tone extends its domain in time, is always perseverative, never anticipatory. Some further evidence for this claim is presented. The phenomenon is shown to be a result of the slower processing of pitch vis-a-vis segments. Evidence for this hypothesis is drawn from the psychoacoustic literature, and the results of an original psychophysical experiment using speech stimuli.

Introduction

Hyman and Schuh (1974), in a study dealing with tonal processes in a large number of West African languages, note that tone spreading is always perseverative, never anticipatory. They examined 23 languages, members of four families. Perseverative tone spreading occurred in Bamileke and Kikuyu of the Bantu group; Hausa and Ngizim of the Chadic group; Gwari, Nuepe, and Yoruba of the Kwa group; and Mende of the Mande group. Thus, eight of the languages, representing all four families surveyed, underwent perseverative tone spreading. They view the process as the spreading in time of a tone's domain. Given the sequence of tones high-low-high, the low tone will spread to the following syllable to give high-low-low, but will not spread to the preceding syllable to give low-low-high. They claim that this tendency is not only true in the languages they have examined, but in the rest of the languages of the world as well. The claim is particularly interesting since segmental assimilations show a tendency to be anticipatory rather than perseverative. If true, this would suggest that tone behaves differently from other phonological units.

In this chapter, I will first discuss evidence from languages not examined by Hyman and Schuh which, on the whole, support their claim. I will then attempt to show the extent to which this contrasts with segmental assimilations and will propose a phonetic explanation for why tone should behave this way. Finally, I will report the results of an experiment which tends to support this explanation.

Further evidence from other languages

The supporting evidence

Tone spreading does not seem to be as common outside African languages as it is among them. However, there appears to be evidence that Hyman and Schuh's claim is generally, if not universally, true.

In addition to the languages with tone spreading found by Hyman and Schuh, a number of other African languages show perseverative spreading. Maddieson (1976a) noted a very general perseverative process of tone spreading in Ciluba, a Bantu language. Depending on the tone pattern, the spreading can occur over as few as one or as many as three syllables so that given the labels H (for high tone) and L (for low tone), the following tone processes can occur:

1) HLH --> HLL

2) HLL --> HHH

3) HLHL --> HLHH

4) HLLH --> HHLL

5) HLLL --> HHHH

6) HLHH --> HLLL

The perseverative spreading is thus quite extensive in Ciluba.

In another example from Bantu, Bennett (1976) found tone spreading occurring in Shona, although she reports the pattern is somewhat different than that found by Fivaz (1970, cited in Bennett).

Tone spreading is perseverative in Lango, a "Nilotic" language of Uganda (Maddieson, Shopen, and Okello 1973). A number of perseverative processes occur, for example, the sequence of tones HIGH-LOW becomes HIGH-FALLING so that the HIGH tone of the first syllable spreads into the second in disyllabic nouns.

Blood (1967) reports that the low tone on the first syllable of disyllabic words in Cham sometimes undergoes perseverative tone spreading. This case is particularly interesting for two reasons. First, there seems to be a general tendency for high tones to affect low tones more than vice versa (cf. Hyman and Schuh's low-raising, also Maddieson 1977b). The rule in Cham goes against that tendency. Second, Blood has examples in which the low tone spreads into final syllables ending in glottal stop, a segment which has a tendency to raise pitch (Matisoff 1970, Mohr 1971, Hombert 1975, Hombert, Ohala and Ewan 1976) with some exceptions (Maran 1971, 1973). The perseverative spreading seems to override these tendencies.

Lhasa Tibetan appears to have cases of perseverative tone spreading in disyllabic compounds.[2] Qu (1963) reports four tonal processes occurring in disyllabic compounds:

7) 35+15 --> 35+55

8) 55+15 --> 55+55

9) 55+35 --> 55+53

10) 53+35 --> 55+53

The first two of these look like spreading of the 5 tone, but are not. The first part of the tone in the second syllable simply gets raised to level 5 between two level 5 tones. The second two rules, however, are cases of perseverative spreading. The 53 pattern spreads to the end of the words.

Ballard (1975) studied the tone sandhi phenomena in the

Wu family of dialects of Chinese. Ballard found what he called tone spreading in three of the dialects, namely Shanghai, Tangsic and Ningpo, and found that this effect was always perseverative. Other types of sandhi occurred in both directions, both in some of the above three dialects and in some of the other Wu dialects.

In discussing Shanghai Ballard says (on the basis of data from Sherard 1972) that "within a phonological word, all syllables after the first ignore their citation tone and take their tone from the first (leftmost) syllable by its spreading out over the whole phonological word." In other words, a 42 tone on the first syllable of a word results in a falling tonal contour over the entire word. This is clearly a case of perseverative tone spreading.

In Tangsic, Sherard finds a similar pattern of spreading as occurs in Shanghai. In Ningpo, however, a case of spreading occurs which is exactly analogous to what Hyman and Schuh found in some of the African languages: the tone of a following syllable took on the tone of the last part of the preceding syllable. It should be noted that the leftward, or anticipatory, sandhi effects found by Ballard are completely different in kind from spreading; they simply involve one or more of the beginning syllables in phonological words being neutralized to some value, often simply a mid tone. No counterexamples, i.e., cases of true spreading which were anticipatory, occurred in any of the Wu dialects.

An extremely interesting case of tone spreading occurs in Telefol, a language of New Guinea, which was described by Healy (1964).[3] Some of the rules are straightforward cases of perseverative tone spreading:

11) 33 --> 55/55___

12) 55 --> 33/33___

These rules cause a perseverative spreading throughout an utterance, with each tone spreading into the following syllable, yielding processes such as the following.

13) 33 55 33 33## --> 33 33 55 32##

The reason the last tone does not become 33 is that the presence of a pause causes a lowering of the 33 tone:

14) 33 --> 32/___##

The sequence in (13) is thus the result of the application of two rules. First, each of the tones spreads into the following syllable:

15) 33 55 33 33 --> 33 33 55 33

Following the spreading, the presence of a pause causes the last 33 tone to become 32.

Ayulta Mixtec has a number of perseverative tone rules (Pankratz and E. Pike 1967), some of which are cases of tone spreading. For example, words not ending in glottal stop which end in a high tone cause a following mid tone to become high. Enclitics take a high tone when following high tone in the stem.

The languages whose tone spreading phenomena are discussed above (including the languages cited by Hyman and Schuh) total 18.

The ambiguous evidence

I examined a number of phonological descriptions of languages which use tone distinctively. Most of these (the number was not recorded) did not have tone spreading. Some descriptions provided evidence which I felt should not be counted, for several reasons. The possibility of the tone spreading in these languages is discussed below, together with my justification for not counting these languages. The languages with tone spreading in both directions are also included in this section.

Downer (1967) has noted a number of tone sandhi rules in White Miao, a language spoken in parts of Laos and which has some dialects spoken in China (Kratochvil 1968). Downer gives a number of examples of tone sandhi, which apparently typify a more general process in White Miao, in which nearly all of the effects are perseverative. One of these looks very much like tone spreading.

16) nton(33) --> nton(11) / nplon(51)_______

Even if we assume that this example represents a general process, however, there are so many tone sandhi rules in White Miao that do not look like any assimilatory process at all, that I would not want to count this example as relevant. Some of the effects of the 51 tone are:

17) 35 --> 33/51___
51 --> 31/51___
33 --> 11/51___
11 --> 31/51___

Tones other than 51 have other effects, none of which look like spreading. The system of alternations becomes so complicated that we would not want to single out one of those alternations and claim that it was due to some universal process. Similarly complicated tone sandhi systems, such as that occurring in Taiwanese (Robert Cheng 1973) were simply excluded here because they would be subject to the same problem.

Both anticipatory and perseverative tone spreading occur in Soyaltepec Mazatec, a language of central Mexico, reported by E. Pike (1956). Pike says: "Allomorphs with tone clusters 45, 34 and tone 4 are replaced by allomorphs with tone 5 when preceding a morpheme with tone 5." [p. 60] I have translated Pike's system into the Chao tone letters.

Her term "tone clusters" refers to what are called contour tones by others. Pike also found cases of perseverative spreading, however. The tone cluster 45 causes the tone of the following word to become tone 5, and tone clusters 34 and 24 cause a tone 3 in the following word to become tone 4. In both these cases, the tone level at the end of the first word spreads to the tone of the following word.

Both anticipatory and perseverative tone spreading occur in the San Esteban and San Miguel dialects of Mixtec (Mak 1953). The spreading rules are said to be optional.

A problem arises in determining the direction of spreading in Villa Alta Zapotec. E. Pike (1948) reports only anticipatory spreading in the dialect spoken in Oaxaca. A preliminary investigation by Jeri Jaeger and Robert Van Valin (personal communication) has thus far found only perseverative spreading in the dialect spoken in Yace, which is also in the Villa Alta group. The data of Jaeger and Van Valin is limited at this time, and that from Pike is somewhat difficult to analyze. Although it is likely that dialect differences are responsible for the different patterns, the possibility remains that both anticipation and perseveration occur in speakers of the same dialect.

There are apparent counterexamples in synchronic alternations in other Chinese dialects, although I have not found any that were actual cases of spreading. In

Cantonese, there are cases of tone sandhi such as the following (Steve Baron, personal communication):

18) 53 55 > 55 55
syl syl syl syl

If we view the tones of each syllable as units, i.e., that 53 is a unit, that 55 is a unit, the case in (1) looks like anticipatory tone spreading. In other words, if we call the tone with a 53 contour Tone A and the tone with contour 55 Tone B, the following rule results.

19) A-->B/___B

As Hyman (1975) points out, if we consider the contour tones as a sequence of level tones, level 3 is simply going up to level 5 when occurring between two level 5 tones in a process that has been labelled "vertical assimilation" by Hyman and Schuh. I will ignore the issue here of whether or not contour tones should be considered as sequences of level tones. The fact is that one cannot make a case for phonetically motivated tone spreading in either direction from the example in (1). In terms of phonetic values, the tone value described as 3 simply took on the value that both preceded and followed.

In Mandarin, in a similar way, the tone sandhi cannot be analyzed as tone spreading. However, in discussing tone sandhi, Chao (1968) says: "In Mandarin, as well as in other

Chinese dialects, the tone of the syllable is affected much more by the tone of the syllable following than by the preceding one." An exception to this occurs in the neutral tone, whose value is determined by the value of the preceding tone, but is not simply a continuation of the end-point of the following tone.

Counterexamples

Chichewa, a Bantu language, has anticipatory tone spreading. Trithart (1976) describes a rule which makes Low-High tone patterns on disyllabic nouns very rare, changing most of them to High-High.

20) Low --> high/___High#

Another Bantu language, Kinyarwanda, has a number of processes of anticipatory tone spreading (Kimenyi 1976). In some cases, High-Low sequences become Falling-Low. In others, level tones simply spread to the preceding syllables.

Perseveration versus anticipation--the tally

Although the cases of anticipatory tone spreading show that the tendency for perseveration is not universal, it is clear that perseverative tone spreading is encountered more frequently. It might be useful to provide a count of the examples above. I have not included White Miao since it is not clear that the "spreading" in that language is a general process. I have not counted Mandarin and Cantonese since, as I showed above, it is not correct to view the sandhi in these languages as spreading. I have counted languages which have both anticipatory and perseverative spreading twice: once in each group. I have counted the different dialects of Wu and Mixtec as different languages. The tally thus results in 22 cases of perseverative tone spreading versus 6 cases of anticipatory spreading. (If the different dialects of Wu and Mixtec are counted as single languages, the numbers are 20 and 4). In either case, the difference is significant at the .01 level.

A tendency does not have to be universal, of course, in order to be interesting and worthy of explanation. In this particular case, however, it is important to establish that the behavior of tones differs from the behavior of segments. If perseverative tone spreading were simply part of a general tendency toward perseverative assimilation, any explanation would have to encompass the related tendency for segments. It is thus important to see whether segmental assimilations are generally perseverative.

Assimilation in segments

Vowel harmony is comparable to tone spreading. As will be seen below, the fact that vowels and tones are communicated chiefly by frequencies in the lower part of the spectrum makes the comparison particularly important. Ultan (1973) notes that vowel harmony has been largely thought of as perseverative because it has been studied so extensively in Altaic and Uralic, which are primarily suffixing language families. He notes that a very similar process is called umlaut when it is anticipatory. Ultan notes that it is nearly always the stem vowel that determines the shape of the harmonic vowel, so that the directionality of vowel harmony within utterances largely depends on whether languages tend to use prefixes or suffixes. Lindau (1975) has a similar view concerning the vowel-harmonizing Niger-Congo and Nilo-Saharan languages, in which, she notes, harmony is largely anticipatory, with a greater tendency for stems to affect prefixes than suffixes.

There are cases, however, in which vowel harmony tends to be anticipatory without involving an affix harmonizing to a stem. One is German umlaut, in which the stem vowel takes on some of the properties of the suffix vowel. English, of course, had umlaut very similar to German in forming plurals, with /fot/:/foti/ becoming /fot/:/føti/, manifested in present day 'foot: feet.' According to Sapir (1921:172) the English umlaut developed independently of the German, so that these can be considered different examples. We can at least conclude that vowel harmony, unlike tone spreading, is not predominantly perseverative. Vowel harmony probably develops from phonetic co-articulatory effects between vowels. Öhman's (1966) spectrographic study of VCV utterances showed that such co-articulatory effects go in both directions. In other words, in an utterance of the type V1CV2, V1 affects V2 and V2 affects V1. Broad and Fertig (1970) show a similar effect. Given this evidence, together with the phonological evidence above, we can conclude that vowel harmony, unlike tone spreading, is not predominantly perseverative.

Another process quite similar to tone spreading occurs in the development of geminate consonants from two adjacent dissimilar consonants. If we have the sequence C1C2 becoming the sequence C1C1, we would have a case of complete perseverative assimilation, whereas if the same sequence became C2C2, we would have complete anticipatory assimilation. In these cases, it can be said that the consonant that remains (as a geminate) has spread its domain in time. There are obvious differences, of course, in that it is very unlikely that the "surviving" consonant gradually encroaches on the time available to the "conquered" consonant. In the cases of tone spreading reported in Hyman and Schuh, something very much like this occurs with tones, so that a high-low sequence on two syllables becomes high-falling, with the high tone continuing into a domain at some point held by the low tone.

Nevertheless, a comparison with consonant gemination might tell us something about the nature of similar assimilations in tone. Thurgood (1974) studied consonant assimilations in several language families: Italian and Sardinian within Romance, several languages of the Muskogean family, and Pali, an Indo-Aryan language. He found that "other things being equal, the first consonant assimilates to the second."

Although Thurgood found a few exceptions, his analysis reveals that these occurred in a limited number of consonant combinations. Thus in consonant clusters in which /s/ was present, the result was always a geminate /s/, regardless of whether the /s/ was the first or second consonant. Consonantal gemination thus appears to be generally anticipatory in sharp contrast to tone spreading.

The most conclusive way to determine whether segments have the same perseverative tendencies that tones do would be to count the number of anticipatory vs. perseverative assimilation rules in a large number of languages. Accordingly, I examined 365 segmental assimilatory rules culled from 60 languages whose phonologies are documented in the Stanford Phonology Archive. 195 of these rules involved anticipatory assimilation of a segment to a following segment. 89 of these involved the perseverative assimilation of a segment to a preceding segment. 81 of the rules fell into both categories, in that segments were affected both by preceding and following segments. The conclusion must be that segmental assimilation is generally anticipatory, in sharp contrast to the situation in tonal assimilation.

Toward an explanation

Why should tone spreading be perseverative? Hyman and Schuh (in a footnote obviously meant to provoke further research) suggest that a possible explanation might be that the physiological adjustments necessary to change pitch take longer than the physiological adjustments necessary to produce different segments. They did not present evidence for this suggestion, and indeed, it is difficult to see how this notion might be confirmed. Even if the vocal folds were slow relative to supra-glottal articulators, it is not clear that this would cause perseverative tone spreading. A comparison might be drawn with the relative speed of other types of speech gestures. The tongue tip is an extremely fast articulator, as evidenced by the very short duration of alveolar taps occurring in certain languages (see Lehiste 1970 on Estonian, for example). The velum is almost certainly slower, given the relative slowness of its movements (see, for example, the tracings in Ohala 1974a). Yet there are numerous cases of anticipatory nasalization before alveolar nasals, suggesting that the different speeds of articulators are in some way taken into account by speakers.

A perceptual asynchrony between tones and segments might explain perseverative tone spreading. If tones were processed more slowly than segments, a listener would perceive a tone as occurring later within a sequence of segments than when it was produced by the speaker. Such misperceptions, if they occurred constantly within a language community, would favor perseverative tone spreading, and disfavor anticipatory tone spreading.

There is evidence suggesting the possibility that tones are processed more slowly. There are at least two different ways in which a slower processing of tones could come about. One involves the fact that high and low frequencies are processed by different mechanisms, the other relates to the fact that high and low frequencies have different effects on the ear even within the same general mechanism. Before proceeding, I will note some general facts of hearing.

The middle ear amplifies the amount of motion caused at the eardrum by the pressure variations known as sound, and transmits these pressure variations to the cochlea. Different frequencies cause the basilar membrane, inside the cochlea, to be set into maximum motion at different locations. The displacement at different locations along the basilar membrane cause the excitation of different nerve endings. A complex sound, such as a vowel, causes excitation maxima at several different locations along the basilar membrane, corresponding to the different formants, or concentrations of energy, of the vowel.

It is generally agreed (Plomp 1966, Tobias 1970, Broadbent 1971) that there are different mechanisms used for processing of relatively high frequencies (comprising most of the frequencies involving speech segments) and for processing of relatively low frequencies (comprising the frequencies involved in speech tone). A place mechanism, which uses the areas of maximum excitation of the cochlea, appears to be the only way in which high frequencies are processed. Although the place mechanism theoretically can process low frequencies, a mechanism of time-and-frequency analysis appears to be used for low frequencies. This mechanism, working at the brain rather than at the cochlea, determines frequency by measuring how often the incoming pattern repeats itself. The reason two mechanisms are believed to be present is that neither is capable of doing alone what the auditory system is known to do. The place mechanism cannot account for the remarkable accuracy with which sounds are discriminated.[4] Time and frequency analysis cannot account for the processing of high frequencies, since the "recovery time" (refractory period) of nerve fibers does not allow them to respond accurately to rapidly varying signals. [5]

The characteristics of the two mechanisms suggest that there may be differences between the processing of fundamental frequency, or tone, and the processing of segments. Given that the fundamental frequency of the human voice hardly ever exceeds 800 Hz, even in small children, it is safe to assume that a time and frequency analysis is capable of determining tone. Miller & Nicely (1955) found that the frequencies above and below 2200 Hz made equal contributions to the perception of duration of segments. This means that most of the cues for duration are above the 1400 Hz limit suggested by Plomp (1966). Moreover, there is evidence that segments, regardless of frequency, are processed by a different, and less accurate, mechanism than the mechanism which processes tone. Flanagan (1972), in comparing the results of Flanagan (1955) and of Flanagan and Saslow (1958), notes that the accuracy of perception is about 10 times greater for fundamental frequency than for formant frequencies (a difference limen of approximately .3 to .5 percent for fundamental frequency but 3 to 5 percent for formants). [6] Since the mechanism involving time and frequency analysis is capable of far more accuracy than the place mechanism, it would seem reasonable to assume that fundamental frequency is probably processed by a time and and frequency analysis, while formants are processed by a place mechanism. It must be remembered that the place mechanism performs its frequency analysis at the cochlea, and that the time and frequency analysis must still use the input from the cochlea to determine frequency. It is likely, then, that the time and frequency analysis takes more time than the place mechanism, since the brain has to get the input from the cochlea to even begin a time and frequency analysis.

There is a complication in all this, in that the place mechanism apparently undergoes a "sharpening" or contour enhancement that increases its accuracy somewhat. This sharpening, since it also presumably takes time, might take as much time as a time and frequency analysis, thus equalizing the time taken by the two mechanisms. On the other hand, it is likely that a time and frequency analysis uses the "sharpened" output of the cochlea, so that it would occur after the sharpening process. However, this is not the only evidence. Although it is possible that the fact of two mechanisms in the ear would explain a slower processing of tone, this is not the only way in which a slower processing of tone could occur.

Dallos (1972:517, 211-217) suggests, on the basis of evidence from Kiano (1965) and from Møller (1970) that the cochlea's transient response is much slower for low frequencies than for high frequencies. This stems directly from the frequency-analyzing properties of the cochlea. The cochlea may be thought of as a set of resonators, each with a different resonating frequency. The higher the resonating frequency, the more quickly a resonator will begin and end its oscillation. Conversely the lower the resonating frequency, the more slowly a resonator will begin and end its oscillation.

Dallos goes on to suggest (1972:517) that an oscillation in the cochlea will continue for a given number of cycles after a sound input has ceased. He says:

> At any point along the membrane the vibration is in the form of a decaying oscillation. The frequency of this oscillation is precisely the same as the best frequency (as far as maximum amplitude with sinusoidal stimulation is concerned) of the given site. There are several (5-10) cycles of this oscillation, and thus any point along the partition is in actual vibration for a length of time that is determined by its characteristic frequency.

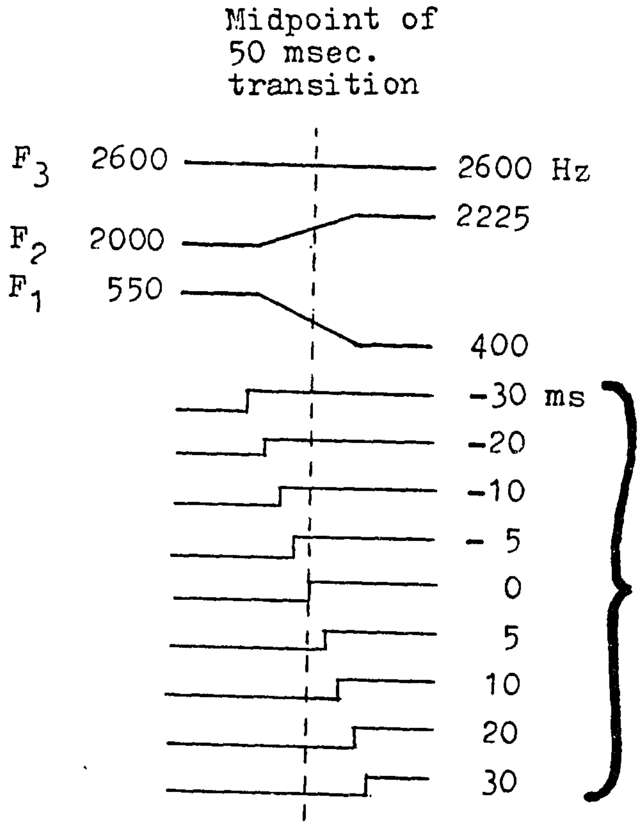
This could explain the tendency of tones to "linger" in the ear and thus undergo perseverative spreading. There is good reason for caution here, however. First, the experiments on which Dallos has based his conclusions were done with animals. Second, the rate of decay of the oscillations that follow an input to the ear is relatively high (20 dB per cycle). If the rate of decay is much higher than that in humans, the effect would be minimized. Worse, the rate of decay (measured in cycles) could be different for different frequencies, so that it is possible that all frequencies linger approximately the same amount of time.

These facts of hearing thus seem promising but not conclusive as an explanation for perseverative tone assimilation. Accordingly, an experiment was conducted to determine if tones are processed more slowly than segments, in this case vowels, although no attempt was made to determine which of the possible explanations above might cause this effect. The assumption was made that, if fundamental frequency were processed more slowly than the formants in vowels, a tone change and a formant change could be perceived as occurring simultaneously if the tone occurred somewhat before the formant change. The amount by which an F0 (fundamental frequency) change would have to lead in order for the changes to be perceived as simultaneous should reveal the difference in the time required for processing F0 vs. formants.

Experimental procedure

The vowel glide /ej/ was synthesized with a tone change occurring near the middle of the glide. The glide was actually composed of a steady-state /e/ portion, a transition portion of 50 ms. and a final /I/ portion. The formant values for /e/ were 550, 2000, and 2600 Hz. The formant values for /I/ were 400, 2200, and 2600 Hz. The transition consisted of a linear interpolation between the /e/ and /I/. This transition was the shortest that could be produced without giving the subjective impression of a consonant. F0 changed abruptly from 120 to 140 Hz at different times in relation to the midpoint of the formant transition: -30, -20, -10, -5, 0, +5, +10, +20 and +30 ms. These are given in Figure 1.

Figure 1



Abrupt 120 to 140 Hz change in F_0. Time given relative to midpoint of formant transition.

A tape was prepared in which each of these stimuli appeared 4 times, with the entire set in random order. A training tape was prepared in which the following sequence was repeated 4 times.

a) The vowel /e/ with no change in quality or pitch.

b) The vowel /e/ with a pitch change toward the beginning.

c) The vowel /e/ with a pitch change toward the end.

d) The glide /ej/ without a pitch change.

e) The glide /ej/ with a pitch change from 120 to 140 Hz occurring 5 ms. before the beginning of the formant transition (and thus 30 ms. before the midpoint of the transition). Instructions told the listeners that the pitch change here occurred before the glide.

f) The glide /ej/ with a pitch change from 120 to 140 Hz occurring 25 ms. into the formant transition (and thus at the midpoint of the transition). The subjects were told that the pitch change here occurred at the same time as the vowel change.

g) The glide /ej/ with a pitch change from 120 to 140 Hz occurring 5 ms. after the end of the formant transition (and thus 30 ms. after the midpoint of the transition. The subjects were told that the pitch change here occurred after the vowel change.

27 subjects participated in the experiment. They were

students and faculty at the University of California, native speakers of American English, with no history of hearing pathology, who served with no compensation. Each listened to the stimulus tape three times. There was thus a total of 12 repetitions of each stimulus per subject. The training tape was played to subjects immediately before the first and second administrations of the stimulus tape. The subjects heard the tape through earphones in a sound-treated room and were asked to make a forced-choice decision as to whether or not the tone change and the vowel change were simultaneous. They were told to guess if they were not sure. A number of subjects reported the task to be extremely difficult but all completed the experiment. The responses to the first playing of the stimulus tape for one subject, however, were not counted due to failure to understand the instructions. The results are given in Table 1 and Table 2.

Table 1

Time of F0 change relative to midpoint of formant transition	-30	-20	-10	-5	0	5	10	20	30
No. of responses of simultaneity	116	167	180	165	172	159	168	143	113

[not significant]

Table 2

Simultaneity judgments when tone change preceded formant change	Simultaneity judgments when tone change followed formant change
635	583

[significant at 5%]

Results and discussion

Table 1 gives the number of responses, pooled for all subjects, to each of the different stimuli. In the -30 stimulus, the F0 change occurred 30 ms. prior to the midpoint of the formant transition. In the +30 stimulus, the F0 change occurred 30 ms. following the mid-point of the formant transition. The particular pattern indicated in Table 1 is not significant by a chi-square test. However, if we compare all the cases in which the F0 change occurred before the mid-point of the transition with all the cases in which the change occurred after that mid-point, as in Table 2, the results are significant at the 5% level. The tone and vowel change were judged to be simultaneous 635 times when the tone change preceded and 583 times when the tone change followed. The results thus show a tendency for tone to be processed more slowly than formants, but they do not reveal, in a statistically significant way, the magnitude of the effect. The experiment thus provides supporting, if incomplete, evidence for the hypothesis that perseverative tone spreading is due to a slower auditory processing of tone compared to the processing for segments. [7]

Two questions remain unanswered. First, the magnitude of the effect needs to be determined, since a relatively small difference in processing time for fundamental frequency would weaken the hypothesis. This could be done by having the subjects control the asynchrony between a tone change and some segmental event until they were satisfied that the two were simultaneous. Although this was not done (instead of the experiment reported) for lack of some of the necessary equipment, it is the obvious method for determining the magnitude of the effect.

Second, it is important to determine which of the two causes suggested here account for the longer processing time for fundamental frequency. One method to test for this is relatively straightforward. The mechanism suggested by Dallos's findings, that the transient response of the cochlea is slower for low frequencies, requires that the fundamental frequency be present in the cochlea. However, we can hear fundamental frequency, using a time and frequency analysis, even when that frequency is only represented in higher harmonics as occurs when we hear through a telephone. By repeating the present experiment with the fundamental frequency missing, we should be able to determine which of the two explanations is correct.

Conclusions

Several conclusions can be drawn concerning tone spreading. First, there is a definite tendency for tone spreading to be perseverative rather than anticipatory. The tendency is certainly not a universal one, as there are too many counterexamples. Second, this tendency contrasts with the tendency of segmental assimilations to be anticipatory. The experimental results show that the perception of when tonal contours occur with respect to when segments occur is not very accurate. The third conclusion is that such results suggest a mechanism for tone spreading in either direction. If listeners have a difficult time determining where a tonal contour occurs, they may hear it and learn it as occurring at a different time from that produced by speakers. Fourth, the experimental results, as well as certain characteristics of hearing, suggest that the tendency toward perseverative spreading is due to the slower processing of tones versus segments.

FOOTNOTES

1. This is a revised version of a paper presented at the Fall meeting of the Acoustical Society of America, 1976. An abstract appeared as Javkin (1976b).

2. I have used the Chao tone letters, in which 1 represents low tone and 5 represents high tones, throughout this chapter where appropriate.

3. I would like to thank John Crothers for bringing this to my attention.

4. Even with a "sharpening" mechanism, to be discussed below, which increases its accuracy.

5. Estimates as to the maximum frequency which could be processed by a time and frequency analysis have been as high as 5000 Hz (Wever 1950) but Plomp (1966) concluded, from a wide survey, that the limit is probably around 1400 Hz. This figure has convinced other investigators, including Broadbent (1971).

6. In fact, if there is a tendency for tones to be processed more slowly than segments, it may be due not to the frequency differences between these two types of sounds. Hirsh (1959) found no tendency for anticipation or perseveration for pure tones of different frequencies. Pisoni (1976) found effects that would probably lead to anticipatory spreading if they occurred for speech sounds. His results are difficult to reconcile with those of Hirsh. It is not clear, however, that the stimuli in these experiments would reflect differences in the processing of tones and segments.

7. However, see Maddieson (1976b, 1977a) for a report on different experimental results bearing on this question.

CHAPTER 7

Conclusion

ABSTRACT

This chapter concludes this thesis and summarizes the sound changes explained here.

In conclusion, it might be useful to note the purposes which the preceding chapters have served. First, the explanations of sound changes offered here have provided additional examples under a paradigm which holds that speech communication may be viewed as a transmission line. Examples have been given of sound changes that are caused by articulatory, acoustic and perceptual factors.

The change from [ls] to [lts] in English and the change of geminate [l] to [j] in Spanish add to the number of sound changes which have already been explained in terms of articulation; e.g. nasalization of vowels in the environment of nasal consonants, the development of tone contrasts from former voicing contrasts, the epenthesis of stops between nasals and following fricatives. I have chosen to separate the articulatory-based explanations into Type (I) and Type (II), depending on whether the loss of a distinction is part of the change (Type II), or whether a segment (as in cases of epenthesis) or a feature (as in the development of creakiness on low tones) is added to words fulfilling certain conditions. This is not the only possible classification. One might well want to put the two kinds of changes into the same category, since both kinds of changes share the major characteristic of being articulatory in origin. Classifying them separately suggests a way to look for more examples, however, and I have chosen to do that here.

The changes of *- p to -o, *- t to -e, and *- k to -ae add to the small number of sound changes which have been explained as being based on an acoustic similarity between the older and newer forms; e.g. the change of [l] to [w], of [sl] to [sl], of vowels nasalizing in the enviroment of [h]. I am somewhat more cautious in suggesting that the change of post-occlusive [l] in Italian also adds to this group. The original hypothesis for explaining this last change was that the basis was articulatory: that the tongue was higher for post-occlusive [l] than for non-post-occlusive [l] in other environments. While the study of this change did not confirm the hypothesis, it is possible that a more complete study would confirm it. I have classified these changes as Type (III).

The changes in which vowels lengthen before voiced consonants and in which tones undergo perseverative spreading opens an area for explanation which has not been very thoroughly explored. These changes (certainly, I think, the former and probably the latter) are caused by constraints of the listener's perception. I have classified them as Type (IV). By providing evidence of changes of a different type, these explanations serve to extend the paradigm into an area in which explanations might have been expected, but had not been developed.

References

Ainsworth, W.A. 1972. Duration as a cue in the recognition of synthetic vowels. Journal of the Acoustical Society of America 51.2:648-652.

Andersen, H. 1973. Abductive and deductive change. Language 49:765-793.

Arkebauer, H.J., T.J. Hixon, and J.C. Hardy. 1967. Peak intraoral air pressures during speech. Journal of Speech and Hearing Research 10:196-208.

Ballard, W.L. 1975. Wu tone sandhi. Paper presented at the 8th International Conference on Sino-Tibetan Languages and Linguistics, September 1975.

Barnwell, T.P. 1971. An algorithm for segment duration in a reading machine context. Technical Report 479, M.I.T. Research Laboratory of Electronics. Cambridge.

von Bekesy, G. 1960. Experiments in hearing. New York: McGraw-Hill.

Belasco, S. 1953. Influence of force of articulation of consonants on vowel duration. Journal of the Acoustical Society of America 25:1015-1016.

Bell-Berti, F. 1975. Control of pharyngeal cavity size for English voiced and voiceless stops. Journal of the Acoustical Society of America 57.2:456-461.

Bennett, T.L. 1976. Tonally irregualr verbs in Chishona. In: L. Hyman, ed. Studies in Bantu Tonology; Southern California Occasional Papers in Linguistics 3:287-320.

van Bergeijk, W.A., J.R. Pierce and E.E. David. 1958. Waves and the ear. Garden City, New York: Anchor Books, Doubleday and Company.

Bhat, D.N.S. 1975. Two studies of nasalization. In: C. Ferguson, L. Hyman, and J.J. Ohala, eds. Nasalfest: Papers from a symposium on nasals and nasalization. Language Universals Project, Committee on Linguistics, Stanford University, Stanford, California. Pp. 73-104.

Blood, D.L. 1967. Phonological units in Cham. Anthropological Linguistics 9.8:15-32.

Bloomfield, L. 1933. Language. New York: Holt, Rinehart and Winston.

Broad, D.J. and R.H. Fertig. 1970. Formant-frequency trajectories in selected CVC syllable nuclei. Journal of the Acoustical Society of America 47:1572-1582.

Broadbent, D.E. 1971. Decision and Stress. London: Academic Press.

Brown, W.S. and R.E. McGlone. 1969. Constancy of intraoral air pressure. Folia Phoniatrica 21:332-339.

Capell, A. 1949. Two tonal languages of New Guinea. Bulletin of the School of Oriental and African Studies 8:184-199.

Chao, Y.R. 1948. Mandarin Primer. Cambridge, Mass.: Harvard University Press.

__________. 1968. A grammar of spoken Chinese. Berkeley: University of California Press.

Chen, M. Vowel length variation as a function of the consonantal environment. Phonetica 22:129-159.

__________. 1972. Nasals and nasalization in Chinese: Explorations in phonological universals. University of California doctoral dissertation.

Cheng, C.C. 1973. A synchronic phonology of Mandarin Chinese. The Hague: Mouton and Company.

Cheng, R. 1973. Some notes on tone sandhi in Taiwanese. Linguistics 100:5-25.

Chiba, T. and M. Kajiyama. 1941. The vowel--its nature and structure. Tokyo: Phonetic Society of Japan.

Chistovich, L.A. 1968. Variation of the fundamental voice pitch as a discriminatory cue for consonants. Journal of Soviet Physics--Acoustics 14.3. Chomsky, N. and M. Halle. 1968. The sound pattern of English. New York: Harper and Row.

Chuang, C.K. and W. S.-Y. Wang. 1975. A distance-sensing device for tracking tongue configuration. Journal of the Acoustical Society of America 58:S11

Cohen, A., J.F. Schouten and J. Hart. 1962. Contribution of the time parameter to the perception of speech. Proceedings of the 4th International Congress of Phonetic Sciences, Helsinki 1961. Pp. 555-560.

Coker, C.H. 1974. Speech as an error-resistant code. Journal of the Acoustical Society of America 55.2:476.

Condax, I.D. and R.R. Krones. 1973. Central vs. peripheral control of vowel duration. Journal of the Acoustical Society of America 53.1:378.

___________. 1976. Duration of four vowels in manually produced synthetic speech. Journal of Phonetics 4:255-264.

Cooper, F.S., P.C. Delattre, A.M. Liberman, J.M. Borst and L.J. Gerstman. 1952. Some experiments on the perception of synthetic speech sounds. Journal of the Acoustical Society of America 24:597-606.

Corominas, J. 1954. Diccionario critico etymologico de la lengua castellana. Bern: Editorial Francke.

Dallos, P. 1972. The auditory periphery: biophysics and physiology. New York: Academic Press.

Delattre, Pierre. 1941. La force d'articulation consonantique en francais. French Review 14.3:220-232.

___________. 1962. Some factors of vowel duration and their cross-linguistic validity. Journal of the Acoustical Society of America 34:1141-1143.

___________, A.M. Liberman and F.S. Cooper. 1955. Acoustic loci and transitional cues for consonants. Journal of the Acoustical Society of America 27:769-773.

Denes, Peter. 1955. Effects of duration on perception of voicing. Journal of the Acoustical Society of

America 27.4:761-764.

Downer, G.B. 1967. Tone-change and tone-shift in White Miao. Bulletin of the School of Oriental and African Studies, London, 30.3:589-599.

Drachman, G. 1977. On the notion of "phonological hierarchy". In: Herok, T., ed. Phonologica 1976 [Akten der dritten Internationalen Phonologie-Tagung, Wien, 1-4 September 1976]

Efron, R. 1963a. The effect of handedness on the perception of simultaneity and temporal order. Brain. 86:261-84.

___________. 1963b. The effect of stimulus intensity on the perception of smultaneity in right- and left-handed subjects. Brain. 86:285-94.

___________. 1972. The measurement of perceptual durations. Studium Generale 23:550-561.

Egan, J.P. and F.R. Clarke. 1966. Psychophysics and signal detection. In: J. D. Sidowsky, ed. Experimental methods in psychology. New York: McGraw-Hill. Pp. 211-246.

Einarsson, S. 1927. Beitrage zur Phonetik der islandischen Sprache. Oslo.

Elert, C.-C. 1964. Phonologic studies of quantity in Swedish. Stockholm: Almquist and Wiksell.

von Essen, O. 1964. An acoustic explanation of a sound shift: [ɫ] > [u] and [l] > [i]. In: D. Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott and J.L.M. Trim. In Honour of Daniel Jones. London: Longmans, Green and Co. Ltd. Pp. 53-58.

Fant, G. 1960. Acoustic theory of speech production. The Hague: Mouton and Company.

________. 1973. Speech sounds and features. Cambridge: M.I.T. Press.

Fintoft, K. 1962. The duration of some Norwegian speech sounds. Phonetica 7:19-39.

Fischer-Jørgensen, E. 1964. Sound duration and place of articulation. Zeitschrift fur Phonetik, Sprachwissen-schaft und Kommunikationsforschung 17.2-4:175-207.

Fivaz, D. 1970. Shona morphotonemics and morphosyntax. Unpublished Doctoral Dissertation, Department of African Languages. University of Witwatersrand, Johannesburg.

Flanagan, J.L. 1955. A difference limen for vowel formant frequency. Journal of the Acoustical Society of America 17:613-617.

__________. 1972. Speech analysis, synthesis and perception. New York: Springer-Verlag.

__________, and M.G. Saslow. 1958. Pitch discrimination for synthetic vowels. Journal of the Acoustical Society of America 30:435-442.

Fletcher, S. 1975. Dynamic palatography. Journal of Speech and Hearing Research 18.4: 812-19.

Fraser, J.C. 1922. Handbook of the Lisu (Yawyim) language. Rangoon.

Fujimura, O. and J. Lindqvist. 1971. Sweep-tone measurements of vocal-tract characteristics. Journal of the Acoustical Society of America. 49: 541-48.

____________, I.F. Tatsumi, and R. Kagaya. 1973. Computational processing of palatographic patterns. Journal of Phonetics 1:47-54.

Gregoire, A. 1911. Influence des consonnes occlusives sur la duree des syllabes precedentes. Revue de phonetique 1:260-292.

Hake, H.W. and A.S. Rodwan. 1966. Perception and recognition. In: J.D. Sidowski, ed. Experimental methods in psychology. New York: McGraw-Hill.

Halle, M. and K. Stevens. 1967. On the mechanism of glottal vibration of vowels and consonants. M.I.T. Quarterly Progress Reports 85:267-271.

__________, G.W. Hughes and J. Radley. 1957. Acoustic properties of stop consonants. Journal of the Acoustical Society of America 29:107-116.

Hardcastle, W. 1968. Dynamic palatography. Work in Progress No. 2. Department of Phonetics and Linguistics, Edinburgh University. Pp. 53-59.

______________. 1969. A system of dynamic palatography. Work in Progress No. 3. Department of Phonetics and Linguistics, Edinburgh University. Pp. 47-52.

Healy, A. 1964. Telefol phonology. Pacific Linguistics, series B, No. 3.

Hirsh, I.J. 1959. Auditory perception of temporal order. Journal of the Acoustical Society of America 31:759-767.

Hoenigswald, H.M. 1960. Language change and linguistic reconstruction. Chicago: University of Chicago Press.

Hombert, J.M. 1975. Towards a theory of tonogenesis: an empirical, physiologically and perceptually-based account of the development of tonal contrasts in language. Ph.D. dissertation, U.C. Berkeley.

___________, J. Ohala and W. Ewan. 1976. Tonogenesis: theories and queries. Report of the Phonology Laboratory 1:48-77.

House, A.S. 1961. On vowel duration in English. Journal of the Acoustical Society of America 33:1174-1178.

___________, and G. Fairbanks. 1953. The influence of consonant environment on the secondary acoustical characteristics of vowels. Journal of the Acoustical Society of America 25:105-113.

Huttar, G. and M. Huttar. 1972. Notes on Djuka phonology. In: J. Grimes, edd Languages of the Guianas. S.I.L. University of Oklahoma, Norman. Pp. 1-12.

Hyman, L.M. 1975. Phonology: theory and analysis. New York: Holt, Rinehart and Winston.

____________, and R.G. Schuh. 1974. Universals of tone rules: evidence from West Africa. Linguistic Inquiry 5.1:81-115.

Jakobson, R. 1928. The concept of the sound law and the teleological criterion. First published in Czech in Casopis pro moderni filologii 14 (Prague, March 1928). Translation in: Jakobson, R. Selected Writings. 1962. The Hague: Mouton and Company.

Javkin, H.R. 1975a. The perceptual basis of vowel duration differences. Journal of the Acoustical Society of America 58:S58.

___________. 1975b. Phonological rules influenced by boundaries. Journal of the Acoustical Society of America 58:S119.

___________. 1976a. Perceptual basis of vowel duration differences associated with the voiced/voiceless distinction. Report of the Phonology Laboratory, U.C. Berkeley 1:78-92.

___________. 1976b. Auditory basis of progressive tone spreading. Journal of the Acoustical Society of America 60:S45.

___________. 1977. Palatalization of /l/ in Italian and Spanish. Journal of the Acoustical Society of America 61:S90.

___________ and J.J. Ohala. 1973. Epenthetic stops between [l] and following sibilants. Journal of the Acoustical Society of America 53.1:378.

Jonasson, J. 1971. Perceptual similarity and articulatory reinterpretation as a source of phonological innovation. Speech Transmission Laboratory, Stockholm. Q.P.S.R. 1:30-41.

Jones, D. 1956. Outline of English phonetics. Cambridge: Heffer and Sons.

Kent, R.D. and K.L. Moll. 1969. Vocal tract characteristics of stop cognates. Journal of the Acoustical Society of America 46:1549-1555.

Kent, R.G. 1928. Lachman's Law of vowel lengthening. Language 4:181-190.

Kenyon, J.S. 1950. American Pronounciation. Ann Arbor: George Wahr Publishing Co.

Kiang, N. Y.-S. 1965. Discharge patterns of single fibers in the cat's auditory nerve. Cambridge, Mass.: M.I.T. Press.

Kimenyi, A. 1976. Tone anticipation in Kinyarwanda. In: L. Hyman, ed. Studies in Bantu Tonology, Southern California Occasional Papers in Linguistics 3:167-82.

Klatt, D. 1973a. Voice-onset time, frication and aspiration in word-initial consonant clusters. M.I.T. Quarterly Progress Reports 109:124-136.

___________. 1973b. Interaction of two factors that influence vowel duration. Journal of the Acoustical Society of America 54:1102-1104.

___________. 1975. Vowel lengthening is syntactically determined in a connected discourse. Journal of Phonetics 3:129-140.

Kratochvil, P. 1968. The Chinese language today. London: Hutchinson University Library.

Kozhevnikov, V.A. and L.A. Chistovich. 1965. Speech: Articulation and perception. Moskow-Leningrad. Translated by the U.S. Department of Commerce, Joint Publication Research Service, Washington, D.C.

Kuzmin, Y.I. 1962. Mobile palatography as a tool for the acoustic study of speech sounds. Report of the Fourth International Congress on Acoustics, Copenhagen. Paper 635.

Ladefoged, P. 1963. Some physiological parameters in speech. Language and Speech 6:109-119.

___________. 1968. A phonetic study of West African languages. Cambridge: Cambridge University Press.

___________. 1971. Preliminaries to linguistic phonetics. Chicago: University of Chicago Press.

Lapesa, R. 1962. Historia de la lengua española. Madrid: Escelicer, S.A.

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

__________, and G.E. Peterson. 1961. Transitions, glides and diphthongs. Journal of the Acoustical Society of America 33:268-277.

Liberman, A.M. 1957. Some results of research on speech perception. Journal of the Acoustical Society of America 29:117-123.

__________, P. Delattre and F.S. Cooper. 1952. The role selected stimulus-variables in the perception of the unvoiced stop consonants. American Journal of Psychology 65:497-516.

__________, P. Delattre, F.S. Cooper and L. Gerstman. 1954. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychological Monographs 68.8:1-13.

Lindau, M. 1975. Features for vowels. U.C.L.A. working Papers in Phonetics 30. Los Angeles.

Lindblom, B. 1963. Spectrographic study of vowel reduction. Journal of the Acoustical Society of America 35:1773-1781.

Lisker, L. 1970. Supraglottal air pressure in the production of English stops. Language and Speech 13:215-230.

__________, 1974. On "explaining" vowel duration variation. Glossa 8.2:223-246.

__________, M. Sawashima, A.S. Abramson and F.S. Cooper. 1970. Cinegraphic observations of the larynx during voiced and voiceless stops. Status Report on Speech Research, Haskins Laboratories. SR 21/22, Pp. 201-210.

Lysle, A. and Gualteri, F.M. 1940. Nuovo dizionario moderno delle lingue Italiana e Inglese. Torino: F. Casanova and C.

Maack, A. 1953. Die Beeinflussung der Sonantendauer durch die Nachbarkonsonanten. Zeitschrift fur Phonetik, Sprachwissenschaft und Kommunikationsforschung 7:104-128.

Mak, C. 1953. A comparison of two Mixtec tonemic systems. International Journal of American Linguistics 19:85-100.

Maddieson, I. 1976a. Tone reversal in Ciluba - a new theory. In: L. Hyman, ed. Studies in Bantu Tonology; Southern California Occasional Papers in Linguistics 3:141-166.

__________. 1976b. Tone spreading and perception. Journal of the Acoustical Society of America 60:S45.

__________. 1977a. Tone spreading and perception. U.C.L.A. Working Papers in Phonetics 36:84-90.

__________. 1977b. Universals of tone. Los Angeles: U.C.L.A. Doctoral Dissertation.

__________, and J. Gandour. 1975. Vowel length before stops of contrasting

series. Journal of the Acoustical Society of America 58:S61. __________, T. Shopen and J. Okello. 1973. Suprasegmentality, Lango tonology and paradigms. Paper presented at the Fourth Africa Linguistics Conference, New York.

Malecot, A. 1955. An experimental study of force of articulation. Studia Linguistica. 9:35-44.

__________. 1966. The effectiveness of intra-oral air-pressure-pulse parameters in distinguishing between stop cognates. Phonetica 14:65-81.

__________. 1970. The lenis-fortis opposition: its physiological parameters. Journal of the Acoustical Society of America 47.6:1588-1592.

Malkiel, Y. 1964. The interlocking of narrow sound change, broad phonological pattern, level of transmission, areal configuration, and sound symbolism. Archivum Linguisticum 15.2:1-30 and 16.1:31-63.

Maran, L.R. 1971. Burmese and Jinghpaw: A study of tonal linguistic processes. Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics.

__________. 1973. On becoming a tone language: A Tibeto-Burman model of tonogenesis. In: L. Hyman, ed. Consonant types and tone, Southern California Occasional Papers in Linguistics 1:97-

114.

Matisoff, J. 1970. Glottal dissimilation and the Lahu high-rising tone. Journal of the American Oriental Society 90:13-44.

Massaro, D.W. 1972. Perceptual images, processing time, and perceptual units in auditory perception. Psychological Review 79:124-145.

Metz, C. 1914. Ein experimentell-phonetischer Beitrag zue Untersuchung der italienischen Konsonantengemination. Vox 23:201-270.

Meyer, E.A. 1903. Englische Lautdauer. Uppsala Humanistika Vetenskaps Samfundet Schrifter. Uppsala.

__________. 1904. Zur vokaldauer im Deutschen. Nordiska studier tillegnade A. Noreen. Uppsala. Pp. 347-356.

__________. and Gombocz. 1909. Zur Phonetik der Ungarischen Sprache. Le Monde Oriental 1907-8, Pp. 122-187.

Michailovsky, B. 1975. Notes on some Tibeto-Burman sound changes. Proceedings of the First Annual Meeting of the Berkeley Linguistics Society. Pp. 322-332.

Miller, G.A. and P.E. Nicely. 1955. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America 27:338-352.

Minifie, F.D., T.J. Hixon, C.E. Kelsey and R.J. Woodhouse. 1970. Lateral pharyngeal wall movement during speech production. Journal of Speech and Hearing Research 13:584-594.

Mohr, B. 1971. Intrinsic variations in the speech signal. Phonetica 23:65-93.

Møller, A.G. 1970. Unit responses in the cochlear nucleus of the rat to noise and tones. Acta Physiologica Scand. 78:299-314.

Navarro Tomas, T. 1916. Cantidad de las vocales acentuadas. Revista de Filologia Espanola 3:387-

408.

Nordmark, J.O. 1970. Time and frequency analysis. In: J. Tobias, ed. Foundations of modern auditory theory. New York: Academic Press. Pp. 55-83.

Ohala, J.J. 1973. The physiology of tone. In: L. Hyman, ed. Consonant Types and Tone; Southern California Occasional Papers in Linguistics 1:1-14.

__________. 1974a. Experimental historical phonology. In: J.M. Anderson and C. Jones, eds. Historical Linguistics II: Theory and Description in Phonology (Proceedings of the First International Conference on Historical Linguistics, Edinburgh). Amsterdam: North Holland Publishing Co. Pp. 352-389.

__________. 1974b. Phonetic explanation in phonology. Papers from the Parasession on Natural Phonology. Chicago Linguistic Society, Chicago. Pp. 251-274.

__________. 1975a. Phonetic explanations for nasal sound patterns. In: C. Ferguson, L. Hyman and J.J. Ohala, eds. Nasalfest: Papers from a symposium on nasals and nasalization. Language Universals Project, Stanford University. Stanford, California. Pp. 289-316.

__________, and W. Ewan. 1973. Speed of pitch change. Journal of the Acoustical Society of America 53:345.

__________, and J. Lorentz. 1977. The story of [w]. Proceedings of the Third Anualnual Meeting of the Berkeley Linguistics Society. Pp. 577-599.

Ohman, S.E.G. 1966. Coarticulation in VCV utterances: spectrographic measurements. Journal of the Acoustical Society of America 39:151-158.

Pankratz, L. and Pike, E.V. 1967. Phonology and morphotonemics of Ayulta Mixtec. International Journal of American Linguistics. 23:287-99.

Parker, F. 1974. The coarticulation of vowels and stop consonants. Journal of Phonetics 2.3:211-221.

Perkell, J.S. 1969. Physiology of speech production: results and implications of a quantitative cineradiographic study. Cambridge, Mass.: M.I.T. Press.

Peterson, G.E. and I. Lehiste. 1960. Duration of syllable nuclei in English. Journal of the Acoustical Society of America 32:693-703.

Phelps, J. 1937. Indo-European si. Language 13:279-284.

Pike, E.V. 1948. Problems in Zapotec tone analysis. International Journal of American Linguistics. 14.3:161-70.

Pike, K.L. 1947. Grammatical prerequisites to phonemic analysis. Word 3:155-172.

__________. 1956. Tonally differentiated allomorphs in Soyaltapec Mazatec. International Journal of American Linguistics 22.1:57-71.

Pisoni, D.B. 1976. Identification and discrimination of the relative onset time of two-component tones: Implications for voicing perception in stops. M.I.T. Research Laboratory of Electronics, Quarterly Progress Report 118:212-31.

Plomp, R. 1966. Experiments in tone perception. Ph.D. dissertation, University of Utrecht; and Institute for Perception, R.V.O.-T.N.O., Soesterberg.

Pope, M.K. 1952. From Latin to Modern French. Manchester: Manchester University Press.

Qu Aitang. 1963. An Outline of Tibetan. Zhongguo Yuwen 127:511-28.

Raphael, L.F. 1972. Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in English. Journal of the Acoustical Society of America 51.4:1296-1303.

__________, M.F. Dorman and A.M. Liberman. 1975. Perception of vowel and syllable duration in VC and CVC syllables. Journal of the Acoustical Society of America 58:S57.

Sapir, E. 1921. Language. New York: Harcourt, Brace and World.

Scully, Celia. 1974. A synthesizer study of aerodynamic factors in speech segment durations. In: G. Fant, ed. Speech Communication. Proceedings of the Speech Communication Seminar, Stockholm, 1-3 August 1974. Volume 2: Speech Production and Synthesis by Rule. Stockholm: Almqvist and Wiksell. Pp. 227-234.

Sharf, D. 1962. Duration of post-stress intervocalic stops and preceding vowels. Language and Speech 5:26-30.

-----------. 1964. Vowel duration in normal and whispered speech. Language and Speech 7:89-97.

Sherard, M. 1972. Shanghai Phonology. Ph.D. dissertation, Cornell University, Ithaca, N.Y.

Shibata, S. 1968. A study of dynamic palatography. Annual Bulletin No. 2, Research Institute of Logopedics and Phoniatrics, University of Tokyo. Pp. 28-36.

Simpson, D.P. 1960. Cassell's new Latin dictionary. New York: Funk and Wagnalls.

Slis, I.H. 1970. Articulatory measurements on voiced, voiceless and nasal consonants. Phonetica 21:193-210.

-----------, and A. Cohen. 1969. On the complex regulating the voiced-voiceless distinction. Language and Speech 12:137-155.

Small, A.M. 1970. Periodicity pitch. In: J. Tobias, ed. Foundations of Modern Auditory Theory. New York and London: Academic Press, Pp. 1-54.

Stevens, K.N. and A.S. House. 1955. Development of a quantitative description of vowel articulation. Journal of the Acoustical Society of America 27:484-493.

-----------, and A.S. House. 1963. Perturbation of vowel articulations by consonant context: an acoustical study. Journal of Speech and Hearing Research 6:111-128.

__________, and D.H. Klatt. 1974. The role of formant transitions in the voiced-voiceless distinction for stops. Journal of the Acoustical Society of America 55:653-659.

__________, and A.P. Paul. 1966. Acoustic description of syllable nuclei: an interpretation in terms of a dynamic model of articulation. Journal of the Acoustical Society of America 40:123-132.

Sussman, H.M., P.F. MacNeilage and R. Hanson. 1973. Labial and mandibular movement dynamics during the production of bilabial stop consonants: preliminary observations. Journal of Speech and Hearing Research 16:397-420.

Sturtevant, E.H. 1917. Linguistic change. Chicago: University of Chicago Press.

Sweet, H. 1877. A Handbook of Phonetics. Oxford: Clarendon Press.

__________. 1899. The history of language. London: Aldine House; J.M. Dent.

Swets, J.A. 1964. Signal detection and recognition by human observers. New York: Wiley.

__________. 1973. The relative operating characteristic in psychology. Science 189:990-999.

Tanner, W.P. 1960. Theory of signal detectability as an interpretive tool for psychophysical data. Journal of the Acoustical Society of America 32:1140:1147.

__________. 1961. Application of signal detectability to amplitude discrimination. Journal of the Acoustical Society of America 33:1233-1234.

Thurgood, G. 1974. The origins of geminates: some speculations. Unpublished manuscript.

__________, and H. Javkin. 1974. An acoustic explanation of a sound change. Paper pesented at Linguistic Society of America meeting, December, 1974.

__________. 1975. An acoustic explanation of a sound change - *-ɑp to -o, *-ɑt to -e, and *-ɑk to -æ. Journal of Phonetics. 3:161-165.

Tobias, J.V. 1970. Foundations of Auditory Theory, Vol. 1. New York and London: Academic Press.

Trithart, L. 1976. Desyllabified noun class prefixes and depressed consonants in Chichewa. In: L. Hyman, ed. Studies in Bantu Tonology, Southern California Occasional Papers in Linguistics. 3:259-286.

Trubetzkoy, N.S. 1939. Principles of phonology. Originally published in German. Translated (1969) by Christiane Baltaxe. Berkeley: University of California Press.

Ultan, R. 1973. Some reflections on vowel harmony. Working Papers in Language Universals. Stanford University: Language Universals Project. Pp. 37-68.

Umeda, Noriko. 1975. Vowel duration in American English. Journal of the Acoustical Society of America 58:434-445.

Velasquez de la Cadena, M. 1973. A new pronouncing dictionary of the Spanish and English languages. Englewood Cliffs, New Jersey: Prentice-Hall.

Warren, D.W. and D.J. Hall. 1973. Glottal activity and intraoral pressure during stop consonant productions. Folia Phoniatrica 25:121-129.

Wever, E.G. 1950. Theory of Hearing. New York: John Wiley and Sons.

Zauner, 1921. Romanische Sprachwissenschaft. Leipzig.

Zimmerman, E. and S.M. Sapon. 1958. Note on vowel duration seen cross-linguistically. Journal of the Acoustical Society of America 30:152-153.