

CHAPTER 5: DISCUSSION AND CONCLUSION

5.1. Introduction

The discussion and experimental results presented in the previous chapters support the hypothesis that velar palatalization arises out of a perceptually conditioned reanalysis of faster speech. First, accounts which rely solely on gradual articulatory steps fail on their own terms to account for this change: not each discrete step needed by a purely articulatory account can be well motivated by articulatory means (§2.3). Second, velar stops before front vowels and palatoalveolar fricatives were shown to have an acoustic affinity (§3.4). The similarity is greater for the voiceless consonants than the voiced which echos the most common target of the sound change, the voiceless [k]. Faster speech tokens of fronted velars were also shown to be more acoustically similar to palatoalveolars than citation speech tokens. Thirdly, in perception studies, velar stops before high front vowels were shown to be highly confusable with palatoalveolar affricates when the signal was degraded by noise masking or gating (§4.2 and §4.3). The voiceless velars were more susceptible to mishearing than the voiced velars.

In this chapter, I consider several models of sound change to see how well they predict a perceptual conditioning of the sound change $k > tʃ / _ [front\ vowels]$. I limit the models of sound change discussed along two lines. First, I only consider models which allow for phonetic conditioning. Many models in the generative tradition which characterize sound change in purely phonological terms such as rule addition or feature spreading (in a feature geometry model) propose accounts of the sound change

velar palatalization (Calabrese 1992, Chomsky and Halle 1968, Clements 1991, Lahiri and Evers 1991). These accounts are not considered here since findings from phonetic investigation do not speak to such models. The models do not make predictions that can be evaluated in terms of phonetic investigation; instead they are evaluated in terms of the formalism used and synchronic patterning. Second, I begin by focusing the discussion on models of sound change within an individual speaker. The chapter begins with the individual speaker in order to narrow the scope of the discussion as well as to make a theoretical stance which distinguishes between sound change in the individual and the spread of sound change through a speech community. Other approaches to sound change (e.g., Weinreich, Labov, and Herzog 1968, Labov 1994) consider the sound change to be the property of the speech community. The relationship of sound change to the speech community raises important questions that I will return to at the end of the chapter.

5.2. Evaluation of Some Current Models

In this section, I outline the models of sound change proposed by Ohala (1971, 1974, 1986, 1989, 1993 and Hombert *et al.* 1979), Kiparsky (1988, 1995), and Lindblom *et al.* (1995) and assess how their models deal with the sound change under investigation here, namely the perceptually motivated change of a velar stop to a palatoalveolar affricate before front vowels. This sound change does not fit nicely into the models of sound change proposed by Kiparsky and Ohala which rely on listener misperception. The change is not necessarily “structure-preserving” as Kiparsky would predict. And the conditioning environment is not usually lost as Ohala would predict.

Contrary to their predictions, the result of the change ([tʃ]) is often a newly created segment in the language and the following vowel is not usually lost (§2.2). Lindblom's model, on the other hand, can easily handle this kind of sound change since it does not rely on misperception, but on collection of alternate forms gleaned from moments of veridical perception.

5.2.1 Ohala's Model

Ohala proposes a theory of sound change which focuses on listeners' misperceptions or misparsings of auditory input. The scope of his investigation is limited to the "mini" sound change in which "one listener hears one word spoken by a single speaker incorrectly" (Ohala 1993:242). Thus, his primary concern is with the preconditions for sound change. These preconditions are, by hypothesis, language universal. In other words he is investigating the physiological, acoustic, and psychological factors common to all human speakers at any time. These factors provide the variation in speech which can give rise to misperceptions or mini sound changes.

Ohala proposes that variation in the speech signal due to coarticulation, speaker specific effects, etc. normally gets factored out. He writes that "listeners *normalize* or *correct* the speech signal in order to arrive at the pronunciation intended by the speaker minus any added contextual perturbation" (Ohala 1993:245, italics in original). There are other possibilities, however. One possibility is that the listener fails to normalize the input and interprets the actual signal as intended by the speaker. Ohala calls this "hypo-correction" and proposes it as one of the basic mechanisms of sound change. First, the

listener misinterprets the speaker's intended pronunciation of a word. Then, the listener adopts the misperception as their own new intended pronunciation.

[I]f the listener fails to correct the perturbations in the speech signal, then they will be taken at face value and will form part of his [sic] conception of its pronunciation. Via such 'hypo-correction', as I call it, the phonetic perturbation, originally just fortuitous results of the speech production process, become part of the pronunciation norm. (Ohala 1993:246)

Ohala offers two main reasons why hypo-correction would happen. The first possibility is that the listener simply does not have the experience to factor out details of speech due to the influence of physical phonetic properties of the speech production system. This would be the case of children acquiring the phonology of the language, as well as adult second language learners. The second possibility is that listeners fail to perceive the conditioning environment of the variation. In the absence of the conditioning environment, the variation can not be factored out. In this case, the mini sound change would involve the loss of the conditioning environment.

“A second reason for hypo-correction is that a listener may, for various reasons, fail to perceive or to attend to the phonological environment which causes, or as phonologists usually put it, ‘conditions’ the variation. **Thus it is frequently the case that sound changes which are the result of assimilations and other effects assignable to the domain of speech production take place with the simultaneous loss of the conditioning environment...** Many (but certainly not all) of the most common sound changes have this feature, e.g. the development of distinctive nasal vowels is generally correlated with the loss of the nasal consonant which conditioned the change from oral to nasal in the vowel.” (Ohala 1993: 247, bolding mine)

Another possible scenario for sound change is what Ohala terms “hyper-correction”. This is the case of “correction” erroneously implemented. The canonical example of hyper-correction is dissimilation. In dissimilation, the listener misapplies the corrective process normally used in perception and attributes a pronunciation intended by the speaker to be an effect of coarticulation or the like. In hyper-correction, the conditioning environment is not lost.

Figures 5.1-5.3 illustrate Ohala’s model of normal perception, hypo-correction and hyper-correction. The figures are slightly modified redrawings of those in Ohala (1981, 1986). In the normal case, the fronting of /u/ in the context of a coronal consonant is factored out or corrected by the listener. This is the case in Figure 5.1. In the case of hypo-correction in Figure 5.2, the conditioning environment is not heard or not attended to and the listener interprets the intended utterance as /y/. Then, Ohala proposes that the listener adopts this new production as the “listener-turned-speaker”. The illustration in Figure 5.3 models the process of hyper-correction. In this case, the listener over applies the corrective process. The listener attributes the fronted vowel to the effects of the following consonant and determines that the speaker really meant [ut]. Then, this misperception is adopted as the new lexical item and the listener uses it when they turn to speak.

Normal Perception

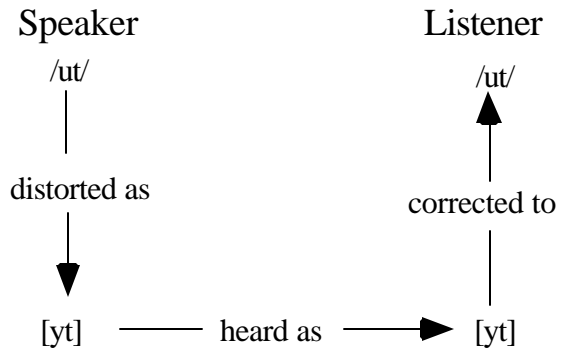


Figure 5.1

Hypo-correction

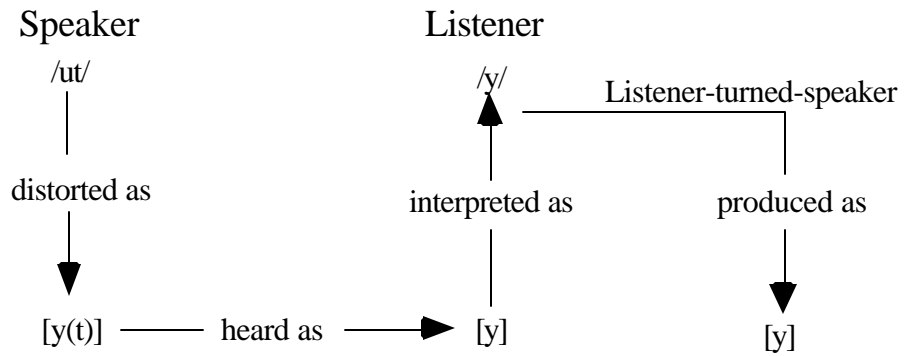


Figure 5.2

Hyper-correction

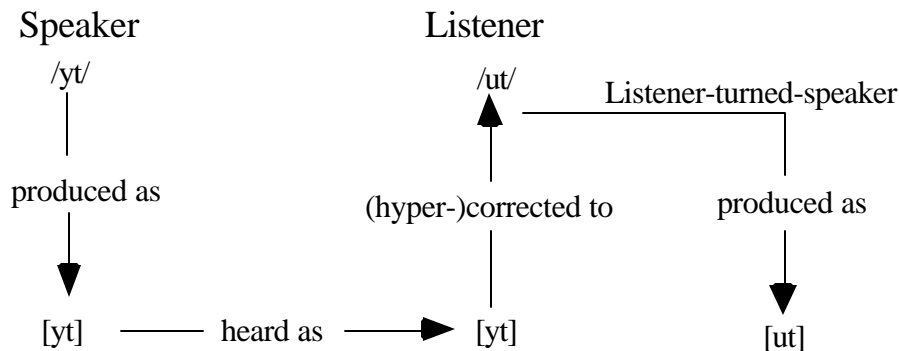


Figure 5.3

How does velar palatalization fit into this model? The facts would seem to be the most consistent with the case of hypo-correction since the effects of the following front vowel on the velar are not factored out, but misinterpreted as intended by the speaker and, hence, the [k] misperceived as a [tʃ]. The acoustic effects of the velar fronting produce bursts and F2 transitions similar to palatoalveolars. These acoustic effects have an impact on perception: the velars can be misinterpreted as palatoalveolars. The perceptual and acoustic experiments discussed in the last chapters abundantly attest to the acoustic similarity and perceptual confusability of the velar and palatoalveolar. For this to be a case of hypo-correction, the listener needs to fail to factor out the effects of the following front vowel on the velar and in some sense decide that the speaker really meant [tʃ]. Ohala's model predicts that this will most often occur when the conditioning environment (in this case the following front vowel) is not heard or not attended to. Thus, the model does not predict the observed cases of

palatalization in which the conditioning environment is **not** lost.¹ The scenario is illustrated in Figure 5.4. . Granted, Ohala does not make the strong claim that all cases of hypo-correction will involve loss of the conditioning environment, but he states that most will. Since velar palatalization is such a common sound change, it presents a problem for Ohala's model if he would predict the most common diachronic processes. Ohala does offer another way in which hypo-correction happens: language learners who do not have a fully developed phonological system are not able to correct for coarticulation. Thus, the model would predict that most cases of velar palatalization are the result of the immature processes during child language acquisition or result from effects of second language acquisition. On this analysis, velar palatalization is really not an example of hypo-correction since there is no inappropriate application of the corrective rules of perception. In fact, there is no correction at all since the hypothesized initiators of the sound change do not possess the corrective rules.

¹While the loss of the conditioning environment in cases of palatalization before front vowels is rare, it is rather more common in cases of palatalization before a palatal glide.

Velar Palatalization as Hypo-Correction?

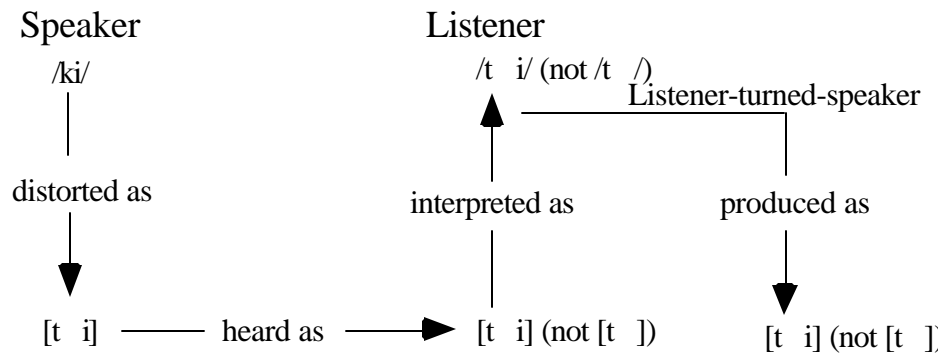


Figure 5.4

Another observation about velar palatalization to be weighed against proposed models of sound change is the fact that velar palatalization often introduces a new phone into the phonological inventory of a language. An analysis of velar palatalization as a case of hypo-correction does not run into trouble with this observation. Ohala writes that “although sound changes due to hypo-correction may often result in the introduction of new segments, for example nasalized vowels from loss of a post -vocalic nasal consonant, sound changes caused by hyper-correction apparently do not.” (Ohala 1993:255).

5.2.2. Kiparsky’s Model

Now let us move on to a discussion of Kiparsky’s model of sound change (Kiparsky 1988, 1995). He proposes a typology of sound change with two main groups. On the one hand, there are major sound changes which are articulatorily conditioned and on the other hand, there are minor sound changes which are

perceptually conditioned or are the product of language acquisition. Table 5.1 is adapted from Kiparsky (1995:659).

Kiparsky's Typology of Sound Changes

	Major changes	Minor changes
Source in speech:	Production	Perception and acquisition
Parameter of Change:	Articulatory similarity	Acoustic similarity
Gradiency:	Gradient	Discrete
Effect on system:	New Segments and Combinations	Structure-preserving
Regularity:	Exceptionless	Can be sporadic

Table 5.1

Kiparsky proposes that major sound changes are articulatorily conditioned, gradual and exceptionless, while minor sound changes are conditioned by perception or language acquisition and can be abrupt and sporadic. In addition, articulatorily conditioned sound changes can introduce new segments into the language whereas perceptually conditioned sound changes can not introduce new segments into the language. Kiparsky proposes that perceptually conditioned sound changes are essentially the result of misparsing or misperception on the listener's part. Kiparsky attributes this characterization to Ohala, but differs from him in at least one regard. Kiparsky proposes that the misparsings are subject to top down processing and therefore can not create segments which are not already in the language; i.e. they must be structure-preserving.

In addition to articulatory variation, speech is subject to variation that originates in perception and acquisition, driven by the possibility of alternative parsing of the speech output (Ohala 1986, 1989). Sound changes that originate in this fashion clearly need not be gradient, but can proceed in abrupt discrete steps. Moreover, like all reinterpretation processes, they should be subject to inherent top-down constraints defined by the linguistic system: the “wrong” parses that generate them should spring from a plausible phonological analysis. Therefore, **context-sensitive reinterpretations would be expected not to introduce new segments into the system**, and context free reinterpretations (such as British Celtic $k^w \rightarrow p$) would be expected not to introduce new features into the system; and neither should introduce exceptional phonotactic combinations.” (Kiparsky 1995:658, my bolding)

As mentioned above, there is sufficient evidence to consider velar palatalization a perceptually conditioned sound change. Velars before front vowels and palatoalveolars are similar in both in the spectral characteristics of the consonantal release and in terms of the following formant transitions. Fronted velars are also highly confusable with palatoalveolars under conditions of noise-masking and truncation. Thus, velar palatalization would fit under the heading of a “minor sound change” in Kiparsky’s typology and as such should not introduce new phones into the phonological system. Kiparsky’s predictions in this case are falsified: velar palatalization often introduces new segments into a language.

An obvious solution to the problem velar palatalization poses to Kiparsky’s model is to admit perceptual conditioning of major sound changes. Since major changes can create new segments, velar palatalization would be consistent with the typology. Indeed, given the wide-spread attestation of velar palatalization, it would be desirable to consider it a major change.

To sum up thus far: both Kiparsky's and Ohala's models have difficulty accounting for the sound change of velar palatalization. Kiparsky's model does not predict the formation of a new segment by perceptual mechanisms and Ohala's model does not predict that the conditioning environment (the front vowel) will be kept. The reliance that both models place on "misperception" also poses a problem. In general, both models have a problem with fully competent listeners who have fully acquired the language. Why would the "listeners-turned-speakers" change their intended pronunciation of a word that they heard pronounced differently (due to coarticulation) when it did not match their own articulatory/acoustic target? If the listener can understand the word, then they must have performed lexical access. Thus they will be able to compare (in some sense) the misperceived word with their own stored form. Ohala does not offer an explanation for the replacement of a lexical item with a misheard form. This falls outside the scope of his work since he defines his goal as studying the preconditions of sound change and not the change itself. Nevertheless, his proposal is problematic when carried to the next logical step.

5.2.3. Lindblom's Model

Lindblom *et al.* (1995) address the problems posed by a misperception analysis. We note that listeners could not decide to change the pronunciation of a word that they had misperceived unless they knew what that word was. If the utterance is understood, lexical access must successfully occur despite a phonetic error in perception. If the utterance was not understood, the listener-turned-speaker would not know which word to pronounce (i.e. they would not know which old form to replace

with the new) and no mini-sound change would happen. Lindblom's model, however, still keeps Ohala's basic insight that sound changes can arise out of listeners' perceptions of the inherently variant speech signal.

Lindblom's model is based on the premise that speech percepts tend to be bimodal: there is a context- and knowledge-dependent mode which is content-oriented and focused on **what** is being said, and there is a signal-oriented mode which is focused on **how** something is said. The "how"-mode of perception provides a breeding-ground for new pronunciations (or mini sound changes in Ohala's terminology). Listeners hear a large range of phonetic variation which results from listener-dependent adaptations which distribute phonetic shapes along the hypo and hyperspeech continuum (Lindblom 1990) and other dimensions. Normally, listeners are unaware of the phonetic variability because the signal interacts with the listener's stored knowledge which usually has the effect of reversing the contextual transforms applied by the speaker. This is the "what"-mode. Occasionally, however, modulation by signal-independent information is somehow inhibited, or becomes superfluous (because intelligibility demands have already been redundantly satisfied, or are of secondary importance for social or speaker-related physiological or cognitive reasons). In this case, phonetic variants have the opportunity of presenting themselves to listeners in close to a "raw" acoustic/auditory form. This is the "how"-mode. Lindblom's model proposes that speakers store in their phonetic memories, not only motor-perceptual information on the "canonical" pronunciation of each lexical item, but also unprocessed phonetic patterns captured in the sporadic moments of acoustic/auditory "truth". In conformity with Ohala's thinking, Lindblom proposes that it is from this pool that "new pronunciations",

or mini sound changes, are selected. The selection is performed as the level of the speech community by evaluating the “new pronunciations” in terms of behaviorally based criteria such as intelligibility and articulatory energetics.

Figure 5.5 and 5.6 are adapted versions of Figures 3A and 3B in Lindblom *et al.* (1995). Figure 5.5 models the origin of phonetic variation which provides the pool of variation out of which sound changes are selected. The range of variant productions can be perceived by the listener in a “what” or a “how”-mode. The veridical perceptions (from the “how”-mode) are then subject to reproduction by the speaker. These new productions are evaluated by the speech community as Figure 5.6 models.

Origins of Phonetic Variation

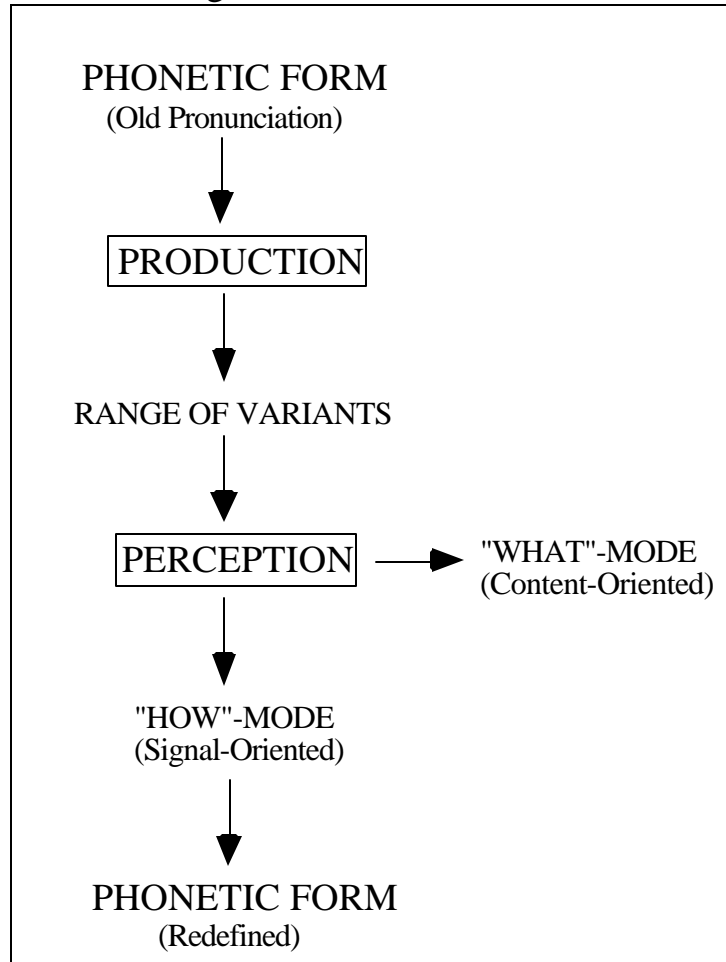


Figure 5.5

Selection

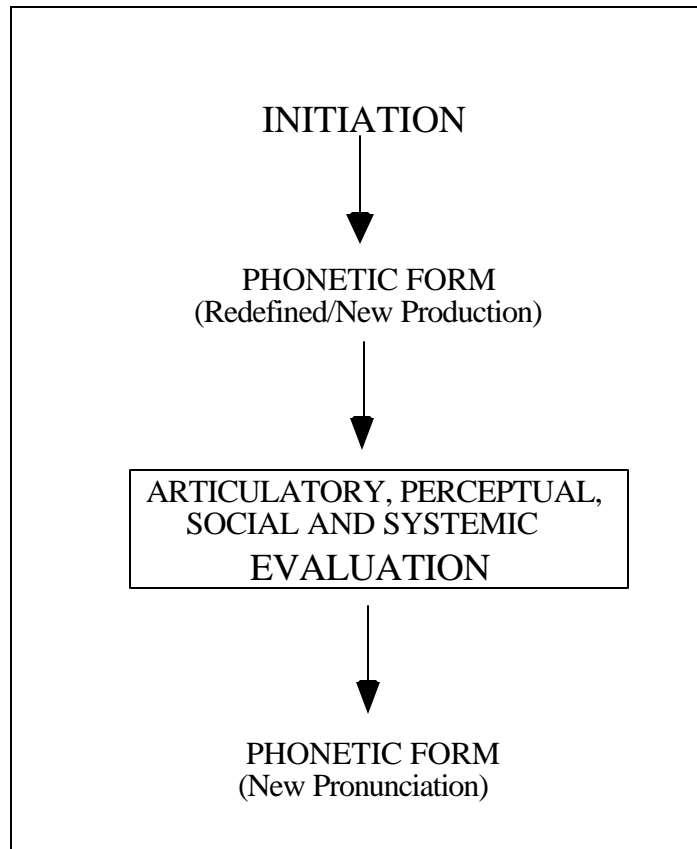


Figure 5.6

Using the veridical “how”-mode does not necessarily lead to sound change, however. For listeners to gain access to the surface value of a particular pronunciation variant, they do not necessarily have to make a perceptual error. For example, a phonetician doing narrow transcription is tuning in to the “how”-mode without replacing an old canonical form with the new veridical perception. Automatic speaking styles (Bates 1979) such as nursery rhymes, poems, memorized lines and lists, overlearned phrases, idioms, greetings, formulas and clichés, exclamations, expletives and curses

also provide listeners with an opportunity to tune into the phonetic shape of the utterances. Mimicry is another example of veridical perception. In mimicking the speech of another, the mimicker has listened to, stored and then reproduced an alternate production of a word or utterance. It does not follow, however, that the mimicker then adopts the alternate production in their everyday speech.

Evidence that listeners are able not only to perceive what is being said, but how it is said is provided by Williams (1986), expanding on a previous study by Lindblom and Studdert-Kennedy (1967/1991). He shows that listeners are capable of hearing speech-like input in two ways, either as speech (contextualized what-mode) or as purely acoustic phenomena (veridical how-mode). Synthesized vowels with formant values ranging from /ɪ / to /ʊ/ were played for subjects both in isolation and in the frames /w_w/ and /j_j/. The subjects were then asked to identify each token as an /ɪ / or /ʊ/. The boundary line was different depending on the context; they allowed lower frequencies for an /ɪ/ in the /w_w/ context, and higher frequencies for an /ʊ/ in the /j_j/ context than were allowed for the vowels in isolation. Williams then replaced the formants with sine waves and first asked the subjects to designate the midpoint of the signal as 'high' or 'low' pitch. The subjects' responses were the same for both vowels in isolation and vowels in /w_w/ and /j_j/ contexts. Then he asked subjects to listen to the same signals as speech by asking them to identify vowel type. This time the results were the same as those with synthesized speech, i.e., the subjects had contextualized the signal. Thus, when the listeners listened to the sine waves as non-speech they were using the how-mode and when they listened to the sine waves as speech, they were using the what-mode.

Now let us turn to the predictions made by Lindblom's model of sound change. The model proposes that listeners store variant productions of words which are gleaned from moments of veridical perception. These stored words are sampled from productions along the hypo and hyperspeech continuum. This continuum is part of the Theory of Hyper and Hypospeech set forth by Lindblom (1990) which states that speakers manipulate the energy expended in production based on considerations of perception as well as production. Speakers expend only enough energy to produce a contrast sufficient for the listener to perform successful lexical access. Thus, phonetic variation is determined by the speaker's dynamic tuning of production to more output-oriented (hyperspeech) or production-oriented speech (hypospeech) based on an estimate of the running contribution of signal complementary processes.

[T]he speaker estimates the running contribution that signal-complementary processes will make during the course of an utterance, and dynamically tunes the production of its elements to the short-term demands for either output-oriented control (hyperspeech) or system-oriented control (hypospeech). What he/she needs to control is — not that linguistic units are actualized in terms of physical invariants ...—but that their signal attributes possess sufficient contrast, that is discriminative power that is sufficient for lexical access (Lindblom 1990:405).

The prediction made by the model is that sound changes originating in hypospeech should result in lower cost articulation without necessarily increasing perceptual saliency. Sound changes originating in hyperspeech, on the other hand, should result in greater perceptual saliency without necessarily being easier to produce.

How does velar palatalization fit into this picture? As the experimental results discussed in the previous chapters show, velar palatalization has its origins in hypospeech. The velars before front vowels were more acoustically similar to palatoalveolars in the faster speech tokens than the citation speech tokens. This acoustic similarity is the result of coarticulation in which the velar is produced in more anterior location due to the influence of the following front vowel. Thus we would expect a sound change that minimizes articulatory effort without necessarily increasing perceptual saliency. It is not clear that the result of the sound change, a palatoalveolar affricate, meets these criteria. While perhaps an anterior consonant and front vowel are less demanding articulatorily than a dorsal stop and a front vowel, a palatoalveolar affricate [tʃ] takes more articulatory effort than a velar stop [k] (Willerman 1994). A sibilant affricate, such as [tʃ] is also highly perceptually salient. The sibilant fricative release is perceptually robust due to its high intensity, high frequency noise created by air turbulence at both the place of constriction and at the teeth. The greater perceptual salience of sibilants was shown by Miller and Nicely (1955), who found sibilants were more easily perceived under greater signal-to-noise ratios than non-sibilant fricatives.

Thus, velar palatalization does not seem to be a nice, neat case of the replacement of an old canonical form with a new hypoform. I suggest that there is a step missing from this scenario. The result of velar palatalization is not simply the coarticulated version of a velar stop. If that were the case, we would expect a palatal stop as the result of the change. As discussed previously, a palatal stop is a rare result of velar palatalization (Bhat 1978). The more common result of velar palatalization, namely a palatoalveolar affricate, seems to be an amplification of certain elements of a

fronted velar: a noisy release, a high peak spectral frequency of the burst, and a high F2 onset.

5.3 Proposal for a New Model of Sound Change

In this section, I propose a model of sound change that draws on the models of Ohala, Kiparsky and Lindblom but address the problems discussed in §5.2. I first outline the goals of the model, then go over some background assumptions that will be used in the new model. Finally, I give an outline of the model. The proposed model basically takes Lindblom et al. (1995)'s model and adds a role for phonetic categories. This move is an improvement since it will predict the regularity of sound change and allow investigation into adaptive forces which act on systems as a whole (not individual words) such as auditory enhancement and adaptive dispersion.

5.3.1. Goals for the Model

The new model needs to address the problems of velar palatalization encountered by models of misperception. The fact the velar palatalization does not usually result in the loss of the conditioning environment and is not structure preserving is a problem for the misperception models of Ohala and Kiparsky. The new model should also be able to explain why the change from [k] to [tʃ] is more common than the change from [k] to [ç].

The new model also needs to address one of the most notable characteristics of sound change. Namely, changes are not generally restricted to single lexical items but tend to be generalized to every item in the lexicon that contains a contextually relevant

instance of the phone in question. The observation that sounds in related languages stand in a regular relation to one another is usually first attributed to Rask (1818) and Grimm (1822) who noted the regular correspondences between the consonants of Germanic and other Indo-European languages. The notion of regular sound change was developed by the neogrammarians (e.g., Ostoff and Brugmann 1878) who are most noted for their belief that sound changes have no exceptions. Whether or not we accept the proposal that sound changes are exceptionless, most historical linguists would support the proposal that sound changes are, in general, regular. (See Kiparsky 1988 for discussion). Bloomfield aptly sums it up by saying that “*the phonemes of a language are subject to historical change [and not individual words]*” (Bloomfield 1933 [1984]:351, italics in original). A working hypothesis of regular sound change is the cornerstone of comparative reconstruction, a practice which is generally agreed to be one of the greatest advancements in the linguistic sciences. Clearly, therefore, a more fully developed model of sound change needs to include some proposal to account for the regularity of sound change. The models of Ohala and Lindblom, which are based on individual words, do not predict the regularity of sound change.

5.3.2. Background

Before moving to the actual proposal for a new model of sound change, I give a brief review of some concepts that are used in the formation of the model proposed here. In this section, I outline two theories of phonetic categories and the theories of auditory enhancement and adaptive dispersion.

Research on developmental speech perception provides evidence for the formation of phonetic categories based on language specific input. It is hypothesized individual that stored perceptions, instances or “exemplars”, organize themselves into categories specific to the ambient language. It has been shown that a given linguistic input shapes the perceptual system of an infant by 6 months of age (e.g., Kuhl 1991, Kuhl *et al.* 1992). By 10 to 12 months of age, infants show a marked reduction in their ability to discriminate some non-native contrasts (Werker and Tees 1984, Werker and Lalonde 1988). This change probably involves a reorganization of perceptual biases rather than a loss of auditory capabilities since, under certain conditions, listeners can discriminate non-native contrasts (Pisoni *et al.* 1994) as well as acquire distinctions not present in their native language (Flege *et al.* 1995, Logan *et al.* 1989).

One school of thought holds that a “prototype” emerges in category formation (Kuhl 1993, 1995). Kuhl defines the prototype as “exceptionally “good” instances of members of categories, ideal exemplars” (Kuhl 1993:125). The prototype is thought to act like a perceptual magnet which draws other exemplars of the category toward it. Kuhl proposes that a “perceptual magnet effect” helps explain how language experience affects speech perception and production. She suggests that linguistic experience alters the perceived distances between speech stimuli. Kuhl (1995) writes:

In effect, our results suggest that linguistic experience “warps” the perceptual space underlying speech. The result is that perceptual categories are formed, ones that begin to mirror the phonological categories of the ambient language. The experimental data that support these claims derive from a phenomenon I have called the *perceptual magnet effect*. It shows that phonetic “prototypes” (the best of most representative instances of phonetic categories) play a unique role in speech perception. They function like “perceptual magnets” for other sounds in the category...When listeners hear a phonetic prototype and attempt to discriminate it from sounds that surround it in acoustic space, the prototype displays an attractor effect on the surrounding sounds. It perceptually pulls other members of the category toward it, making it difficult to hear differences between the prototype and surrounding stimuli. Poor instances from the category (nonprototypes) do not function this way.” (Kuhl 1995:133)

Another school of thought argues that categories are constructed by exemplars alone. Lacerda (1995) argues that prototypes are implicit in an exemplar-based categorization process and, as such, do not need to be stipulated as separate entities. He outlines a model in which the “magnet effect” results from a simple similarity (or distance) metric operating on collections of exemplars stored in memory. In his model, the movement of a category in perceptual space is seen as an effect of the amount of experience with exemplars. In the prototype model, the movement of a category in perceptual space has to be described as a dissolution and reformation of a prototype. The exemplar-based model allows continual refinement and updating of a category through the addition of new exemplars. A model which easily accounts for changes in the perception as well as the production of phonetic categories is desirable since there is evidence of category shift during acquisition of L1 in infants (Kuhl and Meltzoff 1995) and as the consequence of L2 acquisition (Flege 1987), as well as in sound change.

Jusczyk (1994) also suggests that categories are modified by adding new exemplars. Support for the proposal that categories are formed by specific exemplars is constituted by the finding that listeners do retain very specific information about acoustic input they hear, suggesting that they may indeed store alternate productions. For example Jusczyk *et al.* (1993), Nygaard *et al.* (1992), and Pisoni (1995) demonstrate that listeners retain fine detailed information concerning the speaker of utterances that they hear.

It is beyond the scope of this work to evaluate whether a prototype is central to a category or not. But there is substantial evidence arising from the body of work on category formation that phonetic categories are created by storing perceptions (exemplars) of a given phone, even if there is disagreement on the internal structure of the category.

The parallel between the nature of category formation and the model of sound change proposed by Lindblom is quite striking. Both propose that stored perceptions of variant productions constitute the variation out of which new linguistic entities arise: phonetic categories or new words respectively. Jusczyk (1994) brings out another similarity between category formation and Lindblom's model. He outlines the addition of new exemplars as a process requiring extra effort on the listeners part.

[N]ot every utterance that an infant hears will be recorded as an episodic trace [i.e. exemplar]. Although some random storage of the sound structure of the processed input may occur, in general storage of sound patterns requires that some extra effort be given to processing the sound. (Jusczyk 1994:252)

Thus, the number of stored perceptions is limited in much the same way as the “what”-mode of the Lindblom model. In that model, only instances of words collected in moments of veridical perception are stored.

The adaptive forces which act on systems as a whole are also investigated in the model of sound change to be proposed. Here I briefly introduce the theories of auditory enhancement and adaptive dispersion. Diehl and his colleagues (Diehl and Kluender 1989, Diehl *et al.* 1990) have proposed the auditory enhancement hypothesis which states that properties of vowels and consonants covary as they do largely because speakers/listeners tend to select properties that have mutually enhancing auditory effects. For example, across languages, back vowels are usually produced with lip rounding, while front vowels usually are not. The rounding of back vowels enhances its 'backness' by creating a larger front cavity in the vocal tract thereby lowering the first two spectral peaks of the vowel. It is these low spectral peaks that encode 'backness'.

The theory of adaptive dispersion (Lindblom 1986) proposes that sound systems evolve through adaptations to the developmental and adult mechanisms of speech perception and production. Lindblom outlines the process as it applies to vowels.

We hypothesize that vowel systems tend to evolve so as to make the process of speech understanding efficient and to ensure speech intelligibility under a variety of conditions and disturbances. Such efficiency depends in part on vowel identification, which can be assumed to be facilitated by the ontogenetic and diachronic development of perceptual difference among the targets of a vowel system that are maximally, or, perhaps, sufficiently large. (Lindblom 1986:20)

Crucially, the adaptations are systemic. The sound system itself, and not individual words, is subject to evolution.

5.3.3. The Proposal

The model proposed here is an adaptation of the Lindblom *et al.* (1995) model. This model is used since it avoids the problems associated with a misperception model by locating the origin of sound changes in veridical perceptions. Since sound changes are not seen as “mistakes”, they need not be attributed to the misapplication of corrective/normative perceptual devices, nor need they be bound by the constraint of structure preservation as in a misparsing account. Thus, the problems encountered by Ohala’s and Kiparsky’s accounts are solved. The fact that the conditioning environment is not lost in velar palatalization is a problem for Ohala’s misperception (hypo-correction) account and the fact that velar palatalization results in a new segment is a problem for Kiparsky’s misparsing account. However, the basic insight shared by all of the accounts is preserved. The origin of sound change is thought to be located in the perceptions of on-line phonetic variation.

The model proposed by Lindblom *et al.* (1995) needs to be extended beyond the word to account for the observation that sound change is regular. It is not the usual case that a sound changes only in one word, but that the sound changes in every word in the lexicon which provides the appropriate phonetic context. Extending the model beyond the word will have the additional effect of addressing diachronic systemic effects such as auditory enhancement and adaptive dispersion. Auditory enhancement theory

offers an explanation for regularities in language systems. Thus, we would expect processes of auditory enhancement to apply to phonetic/phonological categories which are part of the linguistic system as a whole and not to individual lexical items. Models of sound change which are based on individual words (Lindblom *et al.* 1995, Ohala 1993) cannot explain how systemic effects such as auditory enhancement shape languages. The theory of adaptive dispersion (Lindblom 1986) also poses a problem to models of sound change based on the word since it affects systems as a whole.

I suggest that the insights on the formation of phonetic categories be used to extend Lindblom's model in a way that views sound change as a type of phonetic category reorganization. Perhaps the exemplars which form phonetic categories during language acquisition can also reshape categories throughout the lifetime of a speaker: a preponderance of stored perceptions in one direction could shift a category in that direction. This could be accomplished through the formation of a new prototype or through continual refinement and updating of a category through the addition of new exemplars.

Once we admit a role for phonetic categories in sound change, the regularity problem goes a long way towards being solved. If a category shifts, then, by hypothesis, all the words containing that phonetic category will shift as well. Let me be clear that I am not necessarily equating the phonetic category with the "phoneme". The phonetic category could be positional variants of what a phonemic analysis would call an allophone. For example, there might be separate syllable initial and syllable final categories for stop consonants. Or perhaps the phonetic category is a "syllable-like unit" as Jusczyk (1986, 1994) suggests, or perhaps a demisyllable (Fujimura 1976).

Origins of Phonetic Variation

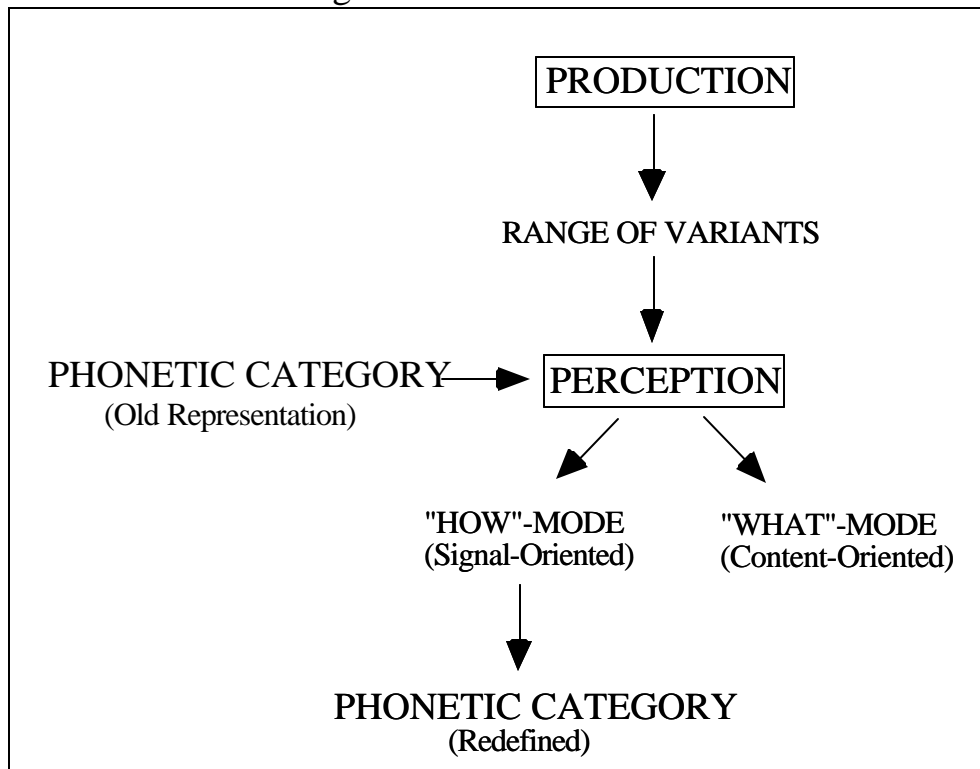


Figure 5.7

Figure 5.7 is a revised version of the Figure 5.5 from Lindblom's model. It represents the ongoing process of phonetic category refinement in an individual speaker. The flow chart begins with production. This box represents the various productions of utterances from speakers of the speech community. Production is subject to a large range of phonetic variation. In the Theory of Hyper- and Hypospeech (Lindblom 1990) the variation results from listener-dependent adaptations with the goal of making the utterance sufficiently distinct while obeying principles of economy. Of course, other

factors such as the speakers emotional state and the socio-linguistic context will also induce variation in production.

The next step in the flow chart is perception. By hypothesis, the phonetic categories involved in a particular utterance will be activated during perception and processing of the signal. As outlined in §5.2.3, listeners can engage in two different modes of perception, one that is content-oriented (the default case) and one that is signal-oriented. Following, Lindblom *et al.* (1995), variant forms perceived through the signal oriented “how”-mode of perception are stored in the lexicon. The variant forms gleaned through veridical perception (i.e. exemplars or instances) are then incorporated into the representation of the phonetic category. Stored perceptual variants or exemplars redefine and shift phonetic categories based on a type of weighting. For example, if the listener happens to store more hypofoms in a certain category, that category will shift in the hypodirection and *vice-versa*.

The exact nature of the interaction between lexically stored items and phonetic categories is uncertain. Lindblom (1992, 1996) proposes that phonetic categories arise through the unique storage of gestures shared by minimal pairs. This shared phonetic category is stored only once and a mapping function combines groups of categories to form lexical items. Since the gesture is stored only once, a change in the motor score would change all the lexical items which use that gesture. If we grant a perceptual component to these categories, we are able to model the regularity of sound change. As the perceptual component of a phonetic category shifts due to the incorporation of new exemplars, the connected motor score follows suit. As the motor score is changed, so are all the words to which the gestures are mapped.

Category shift in an individual speaker is the first step in this model of sound change and provides the variation out of which new phonetic categories can be selected. I propose that different listeners/speakers will have different phonetic categories based on their own unique linguistic experience. Thus, the set of speakers in a given speech community will have different phonetic categories at any given time. Following the proposal of Lindblom *et al.* (1995), we hypothesize that the speech community serves to select the most viable variants in terms of how well they fit the social, communicative, articulatory and linguistic-systemic criteria that speech community members tacitly apply to the variants they encounter as listeners, and when, and if, they try them out as speakers. We hypothesize that an implicit evaluation takes place between speech community members, and that it is performed with respect to, among the other things listed, the phonetic shape of an utterance. Thus we have a model in which speakers' category redefinitions (or "mutations" in terms of Darwinian variation and selection) are selected by the speech community for their adaptive properties. It is at this level of evaluation that the adaptive processes such as auditory enhancement and adaptive dispersion come into play. The most viable categories in terms of their perceptual distinctiveness and saliency will tend to be selected. As Lindblom and Maddieson (1988) remind us, however, perceptual distinctiveness is not the only adaptive concern in the evolution of languages: the principles of economy also come into play. They propose that "consonant inventories tend to evolve so as to achieve maximal perceptual distinctiveness at minimum articulatory cost." (p.72).

Kiparsky's model also admits a role for systemic evaluation in sound change. He proposes that language learners select variants which best meet the linguistic-systemic constraints:

[V]ariation at the level of speech production is conditioned purely by phonetic factors, independently of the language's phonological structure...the learner in addition selectively intervenes in the data, favoring those variants which best conform to the language's system. Variants which contravene language-specific structural principles will be hard to learn, and so will have less of a chance of being incorporated into the system. (Kiparsky 1995:655)

Lindblom's model, however, also allows for selection driven by social factors within the speech community beyond language acquisition. It is proposed that the socially motivated spread of sound change is itself adaptive.

There is a large literature dealing with the "socio-genesis" of sound change within a given groups and its subsequent spread to other social groups (Labov 1972, 1981). This work shows that, by adopting a new pronunciation, speakers signal their "solidarity" with a peer group. They benefit by from so doing in that they thereby increase their social "fitness" and signal their status and identity relative to members of other groups. This is evidence suggesting that speech communities do indeed judge phonetic forms with respect to their social value and that such evaluations do in fact result in sound changes. With respect to social variables sound changes can thus be said to be "adaptive". (Lindblom *et al.* 1995:28)

Figure 5.8 gives a synopsis of the second stage in sound change, the selection of variants by the speech community. The variations in the phonetic categories are

evaluated on the basis of their articulatory, perceptual, social and systemic adaptiveness.

If a new form is selected, it will be adopted by the speech community as a whole.

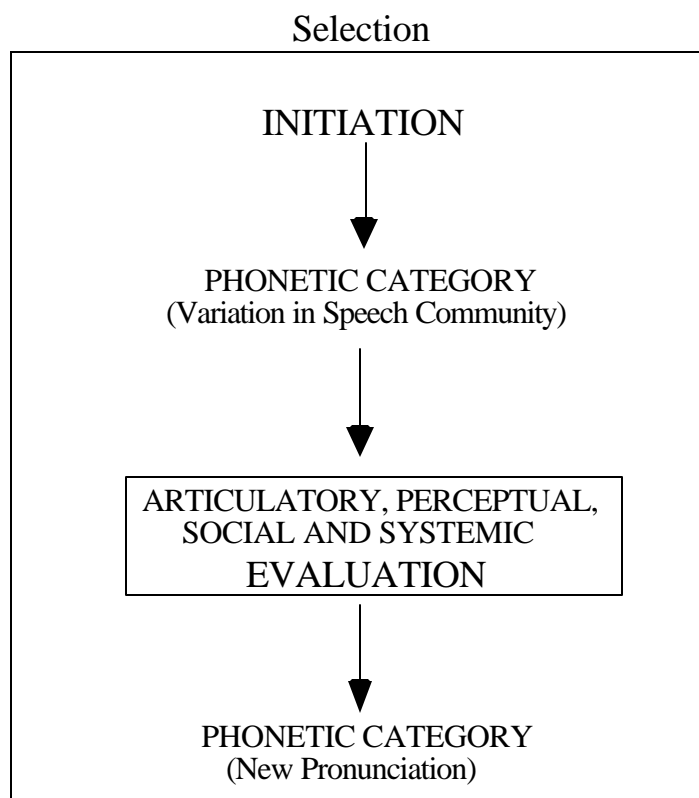


Figure 5.8

Let us consider what velar palatalization would look like in a model of sound change based on phonetic categories. Suppose that there is a phonetic category consisting of exemplars (or veridical perceptions) of syllable initial /k/. Perhaps there is one category for /k/'s before front vowels and one category for /k/'s before back vowels, since, as we have seen, the two groups are statistically different in terms of peak frequency of the burst and in terms of F2 transition. Or perhaps, the syllable initial

(and perhaps syllable final) /k/'s constitute a single category.² The dissimilar production (and the acoustic effects) of /k/'s before front vowels and /k/'s before back vowels, however, would serve as an added motivation for a categorical split. As the listener amasses exemplars of /k/'s before front vowels the category is continually redefined. Under certain circumstances the fronted /k/'s will split away from the /k/ category and shift to a separate category, the most typologically common being a palatoalveolar affricate. I suggest that a preponderance of hypoforms will move the category toward a higher peak spectral frequency of the burst and higher F2 onset. Figure 5.9 offers a schematic representation of the proposed shift. As more hypoforms (represented by little *k*'s) are added to the category, the category moves in the direction of the acoustically similar and perceptually confusable /tʃ/.

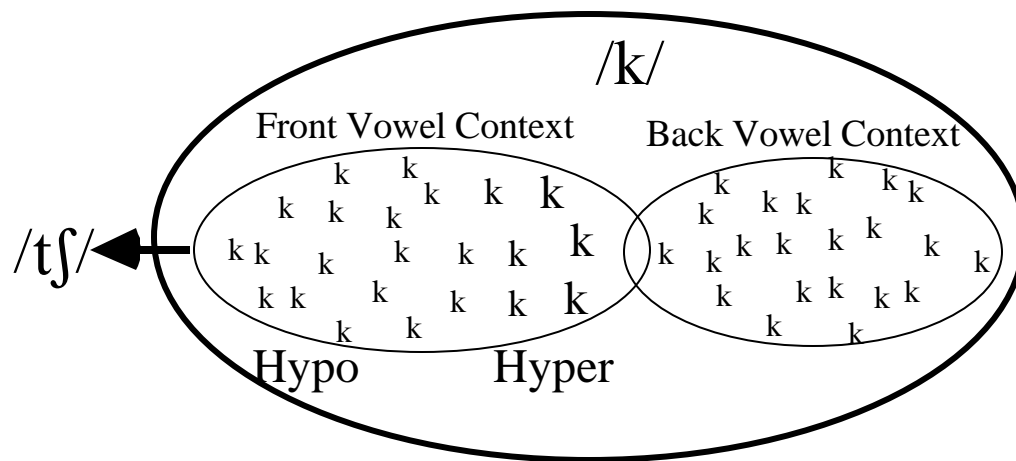


Figure 5.9

²See Sussman et al. (1996) who found the locus equations for syllable initial and syllable final velars to be different.

All the issues raised by velar palatalization have not yet been fully addressed. As mentioned above, the new category is not articulatorily identical to the coarticulated forms found in hypospeech. In the case of the voiceless velar, the coarticulated form is a palatal stop [c], and in the case of the voiced velar, we find a palatal stop [ɟ] or palatal glide [j] (see discussion in last chapter). The most common outcome of the palatalization however is a coronal affricate [tʃ] and sometimes [ts], [ʃ], or [s] in the case of the voiceless velar and a [dʒ] or [j] in the case of the voiced velar (Bhat 1978). Why is it rare, in the case of the voiceless stops, to get a palatal as a result of the sound change? As expected, palatal stops and [k]'s before front vowels are acoustically similar in terms of peak spectral frequency and F2 onset as Blumstein (1986) has shown for Hungarian [ki] and [ci]. As we have already seen, fronted [k]'s and the palatoalveolar affricate [tʃ] are also acoustically similar along certain dimensions (peak spectral frequency and formant transitions). The two segments do, however, differ along other dimensions such as duration, amount of frication, and overall intensity.

Ohala also comments on the exaggeration of the coarticulated form that can occur in sound change. "It must be allowed, however, that after a sound change has occurred—that is, that listener misinterprets the function of these phonetic details—the shape of these phonetic events and features may be different, perhaps exaggerated *vis-à-vis* their original state." (Ohala 1993:259-60) However, an account of sound change which relies on misperceptions cannot explain the exaggeration. If the listener-turned-speaker merely adopts the misheard word, there is no motivation for the observed exaggeration.

Now that the model of sound change proposed by Lindblom *et al.* (1995) has been extended to include phonetic categories, the outcome of velar palatalization can be addressed in terms of systemic processes such as auditory enhancement and adaptive dispersion. In the model proposed here, different listeners/speakers are predicted to have different phonetic categories based on their own unique linguistic experience. So, in our particular case (that of velar palatalization), at a given time, some individuals will have a phonetic category of a velar stop, some will have a palatal stop (or affricate) and some will have a palatoalveolar affricate. This is plausible, since, as we have seen, the fronted velar, palatal stop, and palatoalveolar affricate are all acoustically quite similar.

Out of this variation, that variant which offers the greatest perceptual saliency (among the other things discussed above) will be selected. In this case, given the choice between the [tʃ] and [c] (or [cç]), the palatoalveolar affricate is more robust perceptually due to its sibilant release. As noted in Chapter 4, the palatoalveolar affricate is found to be one of the most perceptually salient sounds in perception experiments. Thus, the theory of sound change being proposed here offers an explanation for the typological observation that [tʃ] is the most common outcome of velar palatalization, even though [c] (or [cç]) in some sense would be expected due to its greater articulatory similarity to fronted [k].

In this way, auditory enhancement (as part of larger process of adaptive dispersion) plays a role in this sound change. I suggest that, in the process of category selection speakers/listeners tend to select properties that have mutually enhancing auditory effects. In the case of velar palatalization, the mutually enhancing properties of sibilancy and frication are selected. As discussed in Chapter 2, velars before high front

vowels show more frication and are more likely to have multiple transients than velars before other vowels (Fischer-Jorgensen 1954, Fant 1973, Klatt 1975, Ohala 1983, Tekieli and Cullinan 1979). The frication of the fronted [k] is enhanced by sibilancy in a shift to [t̪]. Typologically we find that sibilancy covaries with affrication. In Maddieson's (1984) survey of 317 languages, he reports that 92.5% of all affricates have a sibilant release. This covariation serves to enhance the fricative since sibilant fricatives are more salient than non-sibilant fricatives (Miller and Nicely 1955). The greater perceptual saliency of the sibilant fricative release serves an adaptive function and is therefore selected from the other competing forms.

5.4. Conclusion

The model proposed here keeps Ohala's basic insight that sound changes happen due to listeners (mis-)perceptions of variant forms. It also incorporates Lindblom's proposal that variant productions are stored by listeners in moments of veridical perception. The proposed model then adds a role for phonetic categories so that variant perceptions have the function of redefining phonetic categories. This leads to variation in terms of phonetic categories among speakers who have all had different linguistic experiences. Note that this variation is arrived at through purely automatic, self-organizing processes and is not the result of listeners imposing their "will" on the linguistic system. The addition of phonetic categories also predicts regular sound change since as a phonetic category changes, so will all the words which are formed from that category.

The proposed model also handles the problems encountered by misperception models in which the sound change arises from misparsing of the input. A model which allows for the redefinition of phonetic categories does not predict that the conditioning environment will be lost (as Ohala's model) or that the result of perceptually driven sound change must be structure-preserving, i.e. not introduce new sounds into the system (as Kiparsky's model). Since the model of sound change proposed here relies on the adjustment of categories with each new stored perception, the result of a veridical perception (in which contextual variation is not factored out) does not result in the adoption of the veridical percept as the new canonical form. Since a category will shift over time as new perceptions are added to the category, there is no reason to predict that the conditioning environment will be lost or that the category can not be redefined as a "new" sound in the language.

The experimental work presented in the previous chapters has been instrumental in providing evidence with which to evaluate the models of sound change discussed here. The acoustic and perceptual similarity of velars and palatoalveolars support the hypothesis that velar palatalization has a perceptual component, thereby furthering our understanding of the nature of perceptual conditioning in sound change. Without the phonetic investigation we would not have been able to assess the implications that velar palatalization has on models of perceptually conditioned sound change.