On the phonetic and cognitive nature of alveolar stop allophony in American English

DANIEL SILVERMAN

Abstract

I offer an evolutionary approach to the English alveolar stop allophony pattern—with one value evolving towards context-dependent naturalness, and the other value evolving towards context-dependent contrastiveness (provided cue expression is sufficient to avoid neutralization)—arguing that patterns of allophony may be the product of an interaction among phonetic, cognitive, and evolutionary forces.

Keywords: evolutionary phonology; allophony; probability matching; sound change; language acquisition.

> *Language ... is a combination of physiological and acoustic phenomena governed by phonetic laws, and of unconscious and psychical phenomena governed by laws of an entirely different kind.*
>
> Mikolaj Kruszewski,
> *On Sound Alternation* (1881)

1. Introduction

In this study I investigate the origins of alveolar stop allophony in American English, especially focusing on the word-initial realization of the lenis/fortis distinction. I propose that particular allophonic values may diachronically arise as a consequence of their success at maintaining a sufficient phonetic distance from other values in the system. As phonetic forces may, over time, drag one value towards a phonetically natural state (say, from voicing to voicelessness in word-initial stops: *#b > #p), an opposing value may wend its way toward a less natural state (say,

Cognitive Linguistics 15–1 (2004), 69–93 0936–5907/04/0015/–0069

from voicelessness to aspiration in word-initial stops: *#p > #p^h); as *b begins to naturally devoice, those tokens of *p-initial words which contain more aspiration are communicated more reliably to listeners, who are in turn more likely to reproduce this aspiration in their own speech.

I thus argue that American English stop allophony has arisen through Darwin-like evolutionary means, whereby speakers largely recapitulate perceived allophonic values that may nonetheless evolve over time towards a better-adapted state. Through a combination of factors—(1) synchronic phonetic variation, (2) the consequences of ambiguity and misunderstanding, and (3) *probability matching* (Gallistel 1990; Labov 1994)—sounds may change their context-specific properties. The small phonetic variations in which allophones naturally engage are a means by which they take on new properties. Those variants which are more effective in keeping contrastive forms distinct are more likely to survive and flourish. The words associated with these functionally advantageous variants are more likely perceived unambiguously by listeners, and so these variants are more likely to be produced as these listeners become speakers. Slowly then, over the generations, the value may evolve toward its new, better-adapted state. By contrast, there are also contexts with a diminished capacity to accommodate a wide array of acoustic distinctions, for example, in stressless domains, and pre-consonantally. In these cue-deprived contexts, similar evolutionary forces may result in neutralization or merger.

In section 2 I present the phonological patterns of American English alveolar stop allophony, while in section 3 I discuss the phonetic and functional forces (and their diachronic interaction) that may have given rise to these patterns.

2. The pattern of American English alveolar stop allophony

In this section I show that the lenis value of the alveolar stop is always the aerodynamically and/or articulatorily *more natural*, or *easier* value, while the fortis form is always an aerodynamically/articulatorily *less natural, more difficult* value. In contexts where the two values do not contrast, it is always the more natural value which is present.

American English alveolar stop contrasts possess several context-dependent manifestations, the majority of which are shown in Table 1.

Word-initially (as in Table 1, row a), we see a primary aspiration contrast. The lenis stop is typically a plain voiceless stop or only slightly prevoiced (Lisker and Abramson 1964; Flege 1982), while the fortis stop is aspirated. Syllable and word-finally (row b) the lenis stop is also typically realized as a plain voiceless or partially devoiced stop (Ladefoged 1975),

Table 1. *American English alveolar stops*

	lenis			*fortis*		
	form	example		form	example	
(a) word-initial:	t	ˈtɑk	*dock*	tʰ	ˈtʰɑ̆p	*top*
	d̥	ˈd̥ɑk				
(b) syllable-final (including word-final):	t	ˈnɑt	*nod*	V̆t⌝	ˈnɑ̆t(⌝)	*knot*
	d̥	ˈnɑd̥		V̆ˠ̰t⌝	ˈnɑ̆ᶢ̰t(⌝)	
				V̆ʔ	ˈnɑ̆ʔ	
(c) word-internal stressed syllable–initial:	d	əˈdɑpt	*adopt*	tʰ	əˈtʰɑ̆p	*atop*
(d) word-internal unstressed syllable–initial:	Vɾ	ˈɑɾɹ̩	*odder*	ɾ	ˈɑɾɹ̩	*otter*
	V̆ɾ	ˈɑ̆ɾɹ̩	(neutralized)			
(e) following s:	form: t		example: stɑp(⌝)	*stop* (noncontrastive)		

while the fortis stop is normally realized with glottalization and/or unrelease and a shorter vowel (Yoshioka, Löfqvist, and Hirose 1981 discuss the laryngeal posture of English word-final [k] in these terms). Stressed syllable-initially (row c), the lenis stop is usually fully voiced, while the fortis stop is voiceless and aspirated. Word-internal unstressed syllable-initially (row d), there is complete or near-neutralization: both the lenis and fortis stops are tapped, with or without a length distinction on the preceding vowel. Finally, following [s] (row e), there is no contrast between the fortis and lenis stop. There are various additional aspects of the fortis/lenis distinction, including differences in F_1 cutback durations, differences in F_0 perturbations, and differences in closure durations. Kingston and Diehl (1994) provide a very useful overview of these phonetic characteristics. In the remainder of this section I consider in some detail the phonetic properties of both the lenis and fortis categories in all relevant contexts. I show for each context how the lenis value is always realized in a more natural way than its fortis counterpart, and how neutralization is always towards the more natural realization. This discussion of phonetic "naturalness" is necessary in order to motivate a major claim made in section 3, that phonetically natural sound changes may induce less natural ones.

2.1. *Word-initial position*

According to Westbury and Keating's (1986) computer version of Rothenberg's (1968) circuit model of the aerodynamics of plosive voicing,

voicing in initial position is aerodynamically less natural than voicelessness. Even if the vocal folds are appropriately postured for the voicing state, Westbury and Keating calculate that the subglottal-supraglottal pressure drop is insufficient to generate voicing until closure release. During oral closure in initial position, sub and supra–glottal pressure rise in tandem, resulting in a pressure drop which is insufficient until oral release finally induces voicing. These researchers estimate that, even with the vocal folds fully adducted well before stop release, perhaps only 30 to 40 milliseconds of voicing may be achieved during the middle portion of initial stop closure. In English, such natural aerodynamic constraints have been conventionalized as voicelessness throughout the closure, with voicing ensuing after, or only just before, stop release. The result is a voiceless, unaspirated stop or a largely devoiced stop, which is the cross-linguistic norm for utterance-initial stops. As Lisker and Abramson (1964: 384) write, "Although in medial position English [b d g] are voiced and [p t k] are voiceless, in initial position both sets are commonly produced with silent closure intervals and should therefore be classed as voiceless".

Regarding the fortis value in initial position, the glottis is wide open, and remains open beyond the stop release (Kim 1970). The result is aspiration. Then the vocal folds are approximated and voicing begins for the vowel.

2.2. *Syllable-final (including word-final) position*

Westbury and Keating calculate that, in final position (see Table 1, row b), and regardless of the positioning of the vocal folds, transglottal flow may naturally persist for the first portion of the closure (about 35 milliseconds after closure), until the supralaryngeal cavity is near filled. In English, these natural forces have been conventionalized such that the lenis value consists of a plain or devoiced stop. For the fortis stop, vowel shortening (glottal constriction) may be also present ([ɑ̆t⁽˺⁾], [ɑ̆ᵍ̰t⁽˺⁾]), which renders this value more distinct from its lenis counterpart. Jones (1950: 121) states that "words like heed and heat … are distinguished solely by the length of the vowel". Subsequent research has shown that Jones's characterization is only slightly overstated. Thus Raphael (1971), in a perceptual study on the topic, reports that vowel duration was both necessary and sufficient for listeners to perceive the so-called voiced/voiceless distinction in the synthesized stimuli he presented, although other cues may be utilized as well. House and Fairbanks (1953), Denes (1954), Malécot (1970), Cole and Cooper (1975), and Klatt (1976) come to similar conclusions with respect to obstruent "voicing" contrasts here.

Alternatively, in this context the glottal constriction can take the place of the oral occlusion, [aʔ]. Here, the tongue tip is down throughout. This configuration mimics the acoustic properties of a voiceless alveolar stop—silence with less pronounced formant transitions than either labials or velars—while running a minimal (though existing) risk of neutralizing with the labials and velars. As the tongue tip may be raised toward the alveolar ridge without effecting a significant acoustic modification of the surrounding vocalic context, it is quite free to coarticulate here, thus minimizing the time and distance needed for tongue-to-roof contact (Lindblom et al. 1975). This, in turn, results in shorter and less pronounced formant transitions, comparatively similar to the formant structure into and out of a glottal stop, which consist of virtually no closure-induced excursions at all. So, in final position the fortis value is again less natural in laryngeal configuration than is the lenis value.

2.3. *Word-internal, stressed syllable–initial position*

Stressed-syllable initially (Table 1, row c), we have what might be regarded as an "embarrassment of riches" scenario: stressed syllables, with their increased energy, duration, and articulatory force (e.g., De Jong 1991), allow for maximally distinct values to be readily implemented. Cueing potential is maximal here, and so English has a voicing versus aspiration contrast, which are at opposite ends of the laryngeal continuum, [da] versus [tʰa]. The lenis value is the natural value, as voicing is natural for intervocalic stops, where short closure duration allows vocal fold vibration to continue from one vowel to the next.

2.4. *Word-internal, unstressed syllable–initial position*

Medially (intervocalically) before a stressless vowel (Table 1, row d), Westbury and Keating (1986) claim that voicing is aerodynamically natural for the first portion of the closure; there may be more voicing for oral closures towards the front of the cavity, less voicing for oral closures towards the back of the cavity. Stresslessness often induces neutralization or merger of laryngeal contrasts, due to its decreased aerodynamic and articulatory force, and its shorter duration compared to stressed domains. Here, the oral closure is typically short enough that transglottal flow does not markedly dissipate, which establishes a natural environment for (obstruent) alveolar stops to turn into (sonorant) taps. Effective cueing of the distinct values is more difficult under these circumstances. Not surprisingly, in such cue-deprived contexts the contrast barely survives (often in the vowel length of the *preceding* [*stressed*] vowel); in some dialects the contrast is lost and only the natural tap survives.

2.5. *Following [s]*

After [s] (Table 1, row e), the fortis/lenis distinction does not exist. Kim (1970), and Yoshioka, Löfqvist, and Hirose (1981) report that in contexts such as [sp] and [ps], only a single laryngeal spreading gesture is present. Kingston (1990) suggests that, due to the laryngeal articulatory and aerodynamic demands of the voiceless sibilant (sustained laryngeal spreading to increase airflow), lack of cueing potential becomes the overriding factor here, and laryngeal spreading is thus maintained right up to stop release, being difficult to manipulate in any further sort of linguistically significant way. Browman and Goldstein (1986, 1990) suggest that the laryngeal spreading gesture observed in aspirated stops is shifted back onto the fricative in these clusters. Indeed, contrastively aspirated fricatives are quite rare cross-linguistically (Korean and Burmese are two of the few languages which possess them). It might be the presence of the fricative and its concomitant open glottis which account for the rarity of both aspirated fricatives and aspirated fricative–stop sequences.[1]

Table 2 offers a summary of the distribution of alveolar stop allophones in American English.

The upshot here is that stop consonants may be characterized as naturally susceptible to voicing in some contexts, but naturally resistant to it in other contexts. If there is a laryngeal contrast in stop consonants, one value is always the natural value while the other value is always less natural.

Given the potential for cueing in the cases shown in rows (a), (b), and (c) of Table 2, the fortis/lenis contrast is readily maintained, accommodating to context-specific natural constraints on stop production. In these contexts, the lenis stop is implemented aerodynamically "naturally", and cueing potential is sufficient to accommodate a push of the fortis stop to a somewhat less natural realization. In rows (d) and (e) of Table 2, how-

Table 2. *American English alveolar stop allophony*

Context	lenis	fortis
(a) word-initial	[d̥] or [t] (natural)	[tʰ] (less natural)
(b) syllable-final (including word-final)	[t] or [d̥] (natural)	[V̌t] or [V̌ˠ̰t] or [V̌ʔ] (less natural)
(c) word-internal, stressed syllable–initial	[d] (natural)	[tʰ] (less natural)
(d) word-internal, unstressed syllable–initial	(V)ɾ (natural)	(V̌)ɾ (natural)
(e) preceding [s]	[t] (natural)	[t] (natural)

ever, cueing potential becomes an overriding (or almost overriding) factor. In these contexts, the contrast is often lost, as in row (d), or the contrast simply does not exist, as in row (e). In both these contexts, neutralization is towards the natural, or easier realization.

To highlight the most important generalizations that this investigation of the English alveolar stop system has revealed:

- Lenis stop allophones are context-dependent natural values.
- Fortis stop allophones are context-dependent unnatural values.
- Neutralization is towards the context-dependent natural value.

An interesting result of this brief discussion is the observed parallel between Westbury and Keating's "naturalness" criteria, and cross-linguistic tendencies in patterns of stop allophony. The behavior of the American English lenis series, for example, is in fairly strict accordance with Westbury and Keating's computational model: as stated, this series is typically voiceless in initial position, voiced medially, and again voiceless finally (see also Keating et al. 1983; Gurevich, 2003). Based on his survey of over three hundred languages, Maddieson (1984: 31) states that "languages nearly always include a plain voiceless series of stops. If there is only one series it is of this kind". However, the plain "voiceless" series may be subject to underreported allophonic variation. For example, while Spanish is traditionally characterized as maintaining stop voicelessness in a VCV context, many phonetic studies have actually observed voicing here (see Lewis 2001 for a thoroughgoing review, as well as new experimental evidence in support of this characterization of Spanish VCV contexts). Final stops often must be voiceless, as in Basque, Bulgarian, Cantonese, Choctaw, Dutch, Efik, Ewondo, Finnish, Gaelic, German, Polish, Russian, Zoque, Korean, Nama, Thai, Tikar, and Vietnamese. In these senses then, English should be regarded as largely normal in terms of its stop allophony system.

3. The origin of American English alveolar stop allophony

It has often been proposed that phonetic variation is an engine of system-internal sound change. For example, Antilla (1972: 53) states that "variation is a prerequisite of change"; Ohala (1989) argues that "sound change drawn from a pool of synchronic variation"; Hock (1991: 648) writes that "the basis for linguistic change lies in the same ever-present low-level variability of ordinary speech"; Janda and Joseph (2001: 3) propose that "sound change originates in a very 'small', highly localized context ... purely phonetic conditions govern an innovation [and] partially [determine] its future trajectory"; Martinet said in 1975 (published

1988: 25), "Only those who know that linguistic identity does not imply physical sameness ... can accept the notion that discreteness does not rule out infinite variety and be thus prepared to perceive the gradualness of phonological shifts"; Hockett writes (1968: 83), "The distinction between system-conforming and system-changing events cannot, in principle, be made". I interpret this statement to mean that the variation inherent in speech production at once fulfills the synchronic requirements of effective transmission, and contributes to the process and direction of sound change. Paul (1886 [1970]: 43), writing extensively on the topic, offers the following about variability in speech, and its relevance to sound change:

> ... [H]owever much movement may be the result of training ... it still remains left to chance whether the pronunciation be uttered with absolute exactness, or whether slight deviation from the correct path towards one side or the other manifests itself .... This variability of pronunciation, which remains unnoticed because of the narrow limits in which it moves, gives the key to our comprehension of the otherwise incomprehensible fact that a change of usage in the sounds of a language sets in and comes to its fulfillment without the least suspicion on the part of those in whom this change is being carried out .... There thus gradually arises, by adding together all the displacements ... a notable difference .... The reason why the inclination to deviation is greater on one side than the other must be probably sought in the fact that the deviation towards the side to which it tends is in some respect more convenient .... It must not, however, be supposed that it is not at the same time conditioned by psychology.

A remarkable aspect of the inherent variability in speech is that it is largely recapitulated from generation to generation, in a form of *probability matching*. That is, learners come to largely reproduce the nuances of variation engaged in by their elders. So, for example, if elders produce seventy-five percent of their word-initial lenis stops without voicing, twenty percent with minimal voicing, and five percent with more robust voicing, learners are likely to largely recapitulate these percentages in their own speech. Consequently, the inherent variation in speech production may not be as free or as uncontrolled as it is often thought to be, but may instead be conventionalized to a significant degree. Labov (1994, *pace* Gallistel 1990, especially chapter 11) shows how such probability matching—which is also observed in the foraging behavior of scavenging animals—may be affected by ambiguities of meaning in morphologically complex contexts, and by sound changes in progress. But what exactly is probability matching?

Gallistel (1990: 352) reports on a study in which rats in a T-maze were rewarded with food 75 percent of the time at one end, 25 percent of the time at the other. When provided with feedback, rats matched the probability of reward—running to the one end 75 percent of the time, the other end 25 percent of the time—despite the fact that they would receive more rewards if they ran to the one end 100 percent of the time.[2] It turns out that similar statistical calculations underlie aspects of human linguistic behavior: even though certain variants are better at communicating the intended word to listeners, speakers largely match their own variability of production probabilities with that which they perceive, including both "better" (more distinct) and "worse" (less distinct) variants. Indeed, as Labov writes (1994: 583), "[i]t is not a hypothesis that children do probability matching [during language learning]. It is simply a description of the observed facts," and linguists, psychologists, and speech scientists have begun to see the descriptive and explanatory usefulness of employing stochastic or probabilistic approaches to knowledge of language, for example, Miller (1994) Kelly and Martin (1994), Pierrehumbert (1994, 1999) Steels (2000), Jurafsky, Bell, Gregory, and Williams (2001), Bybee (2001), and Munson (in press). Further issues in the mental representation of category-internal variability are explored in "episodic", "exemplar", or "multiple trace" theories, in which perceptual categories are defined as the set of all experienced instances of the category, such that variability across exemplars actually contributes to the categorical properties themselves (e.g., Gluck and Bower 1988, Kruschke 1992, Goldinger 1997, 1998, Johnson 1997, and Pierrhumbert 2001).

Probability matching in speech betrays an extremely sophisticated statistical analytic ability on the part of learners. Moreover, learners' actual productions betray evidence that they are able to implement their calculated probabilities in their own speech. It may be said that the exquisite articulatory control that speakers display in their productions is best evidenced by the fact that they are able to largely match the variability present in the ambient pattern. On this view, learners' articulatory talents are harnessed in service to *copying* or *imitating*, not *modifying* (improving upon or otherwise) the ambient speech pattern.

However, if learners were able to perfectly match probabilities present in speech that is produced around them, then sounds would never have the opportunity to change in the proposed fashion. Rather, perfect reproduction would yield perfect diachronic stability. So, either probability matching is imperfect, or learners do not match their productions to ambient productions, but to something else. Let us suppose for the moment that probability matching is indeed largely perfect, but instead of learners matching ambient *productions*, they match their own *perceptions* of these ambient productions. Since perception is demonstrably imperfect, then reproduction is imperfect as well. I am not proposing that these sorts of sound changes are triggered by *mishearing the speech signal.* Instead, they are rooted in *misunderstanding word meanings* intended by speakers. When two different words are acoustically similar, some tokens of the one word may be misperceived as the other word. So it's the imperfection of the *system* of communication—not the *medium*—that leads to the sorts of sound changes I am proposing. To illustrate how ambiguity of meaning may effect sound changes, in this section I explore in considerable detail the proposed interplay of the phonetic and functional forces that may have given rise to the English alveolar stop pattern in evidence today, focusing especially on the word-initial allophones.

3.1. *Word-initial allophones*

The aspiration contrast in word-initial position might derive from an earlier voicing contrast. That is, an earlier voicing contrast may have evolved into an aspiration contrast as word-initial voiced stops underwent natural diachronic devoicing. However, upon researching the issue, it becomes immediately apparent that the existing historical record is frustratingly scanty regarding phonetic descriptions of English stop allophony (for example, none of Emerson 1906, Robertson 1934, or Jespersen 1956 discuss the issue). Furthermore, most historians have even neglected to address the paucity of evidence—textual, comparative, or even internal reconstructive—and so the issue remains, perhaps permanently, unresolved. Nonetheless, internal reconstructive hypotheses might be able to illuminate the issue somewhat.

The hypothesis that the present-day aspiration contrast derives from a historic voicing contrast gains some support from two separate, though related, findings. First, Abramson and Lisker (1985) report that while voice onset time (henceforth VOT) is the primary determinant of the fortis/lenis distinction in initial position, category ambiguities (at around 20 milliseconds VOT) can be partially resolved by synthetically manipulating fundamental frequency at stop release. It is well established that F_0 is lower at the release of a voiced stop, and higher at the release of a voiceless stop. As already noted, during the word-initial lenis stop, the vocal folds may be adducted in the posture for voicing but, nonetheless, voicing may not begin until the stop is released. Lisker and Abramson reason that a lowered F_0 at release may induce the perception of the lenis stop, while a raised F_0 may induce the perception of the fortis stop, and indeed, Caisse (1982; reported in Kingston and Diehl 1994) finds such pitch perturbing effects in English stop production. These findings suggest that speakers might employ the appropriate articulatory posture to lower

F_0 and hearers might be sensitive to it, despite the fact that the lenis stop is typically voiceless in this context.

Second, Hombert (1978) reports that the pitch perturbations at stop releases extend over 100 milliseconds into the following vowel, well surpassing their perceptual limen, which Hombert determines to be around 40 milliseconds. As Hombert notes, the English pattern cannot be accounted for by proximate phonetic forces alone, since aerodynamic theory does not predict such extensive pitch effects. Hombert further reports that the expected pitch perturbations are significantly curtailed in tone languages such as Thai and Yoruba. He notes that if stop-induced pitch perturbations went uncurtailed in tone languages, tonal contrasts might be jeopardized. By hypothesis then, such perturbations are curtailed here for this reason. In non-tonal languages such as English, however, there is no tonal system to disrupt. Instead, extending these pitch perturbations well into the following vowel may actually serve to enhance the contrast between prevocalic voiced and voiceless stops with no antifunctional consequences.[3]

What I am suggesting is (1) that the present-day posture of the vocal folds may be seen as a relic of an era in which the lenis value was genuinely voiced in word-initial position, and (2) that the present-day pitch-perturbing effects on the following vowel may be a relic of the immediate post-voicing era, in which pitch perturbations became an increasingly important cue. As natural diachronic devoicing began, the lenis value was becoming increasingly similar to the (voiceless) fortis value. Consequently, those lenis tokens which retained vocal fold adduction and pitch lowering and extended these pitch-lowering effects provided a greater acoustic contrast with the fortis value, and so were more likely to be perceived unambiguously by listeners. In turn, such variants were more likely to be employed by listeners, and thus the system moved towards its present-day state. So what might superficially appear to be a synchronically controlled response on the part of speakers to implement vocal fold adduction, and extend the pitch lowering effects in order to enhance the cues of the lenis value (as has been proposed, for example, by Kingston and Diehl 1994), may instead be deeply rooted in paleophonetic history; the present state of the system contains vestigial remains of the past which were at some point functionally beneficial.

So far I have suggested that phonetically natural aerodynamic forces may have effected changes in the realization of the lenis value, and that probability matching may have effected the conventionalization of these changes. But how might these phonetic and cognitive forces have diachronically interacted to produce the sound changes proposed? An important point to recall is that, in probability matching, production

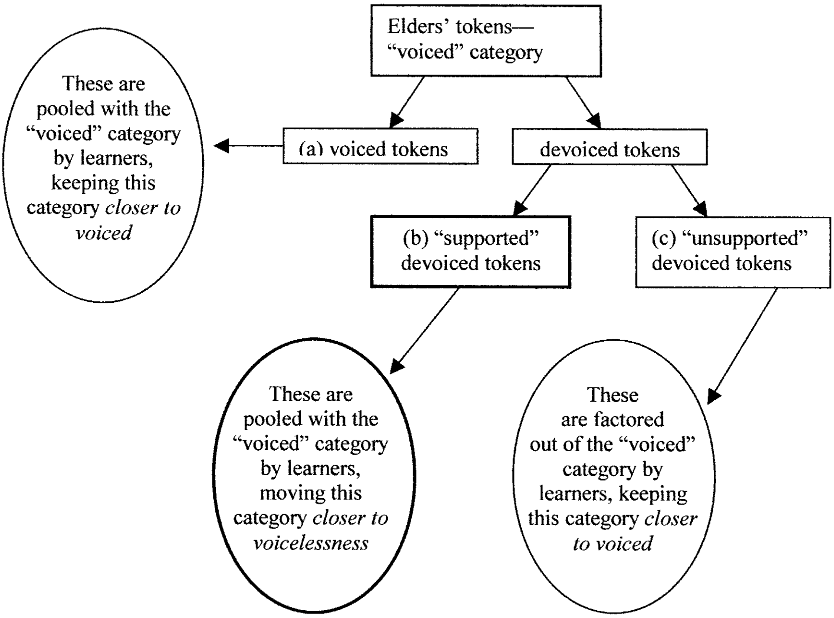

Figure 1. *Probability matching and devoicing*

matches listeners' *perceptions*, not speaker's *productions*. In the past, younger generations may have largely matched the "voiced" category variations present in preceding generations, although all the while there existed a natural tug toward an increased number of devoiced productions. So, probabilities may have been largely matched, but devoicing was still gaining ground. Specifically, among elders' "voiced" category tokens, those phonetically voiced tokens (see Figure 1a) are interpreted as such by learners, and thus learners add them to their pool of "voiced" tokens, reinforcing the voicing of this category. Devoiced tokens fall into two subcategories. One subcategory consists of so-called "supported" tokens (Labov 1994), which, despite consisting of phonetically devoiced tokens, are nonetheless disambiguated with grammatical or pragmatic information. These tokens are thus pooled with the "voiced" category (see Figure 1b). As learners' productions match their calculated probabilities, these devoiced "voiced" tokens serve to tug the "voiced" category towards an increasingly devoiced state. The other subcategory of voiceless tokens leaves learners in the dark: without "support", such tokens cannot be added to the pool of "voiced" tokens (cf. Figure 1c). However, the assumption that *all* such tokens are misinterpreted as "voiceless" is not well founded. Some of these tokens might simply remain uninterpreted. But either way, all these tokens may be factored out of the pool over which voicing probabilities are matched within the "voiced" category. This factoring-out procedure might render the voicing contrast *more* robust as generations proceed: the pool of reliably interpreted "voiced" tokens is smaller for learners than for elders, and so the percentage of phonetically voiced "voiced" tokens is greater for learners than for elders. Even though there is a naturally induced devoicing in progress, the effect of these factored-out tokens is to slow the rate of change toward the voiceless state. As Steels proposes (2000: 20),

> [t]here is a positive feedback between use and success. Sounds that are … successful propagate. The more a sound is used the more success it has and it will be used even more … The scores of [sounds] that can be successfully distinguished and reproduced given a specific sensorimotor apparatus have a tendency to increase and they hence survive in the population. Novel sounds or deviations of existing sounds (which automatically get produced due to the unavoidable stochasticity) create variation, and sensorimotor constraints select those that can be re-produced and recognised.

In Figure 1, heavily outlined cells highlight those forces which effect the *change* toward voicelessness, whereas the more lightly outlined cells contain those forces which maintain voicing.

Consider how such a gradual devoicing of the English "voiced" series affects the "voiceless" series. Since the "voiceless" series also engaged in variations that were largely matched from generation to generation, we might at first conjecture that the end result would be a sound merger: as voiced stops devoiced, they would ultimately merge with the other series. But this is not what we find, of course. Instead, just as phonetic forces may have dragged the VOT forward in the present-day lenis series, functional forces may have pushed the VOT forward in the present-day fortis series. Specifically, those "voiceless" variants which possessed a slightly later voice onset time came to be functionally beneficial as the "voiced" series began to creep towards voicelesness: exactly those tokens that were more distinct from the "voiced" series were successful in cueing lexical distinctions for learners. Learners, interpreting the signal unambiguously in such contexts, were more likely to reproduce these forms in their own speech. Consequently, such stray tokens served to promote the shift toward the aspirated state (see Figure 2). That is, when provided with a clear phonetic contrast between phonologically distinct forms, probabilities were matched accordingly, while phonetically indistinct forms, being more likely to be misinterpreted by learners, were factored out of

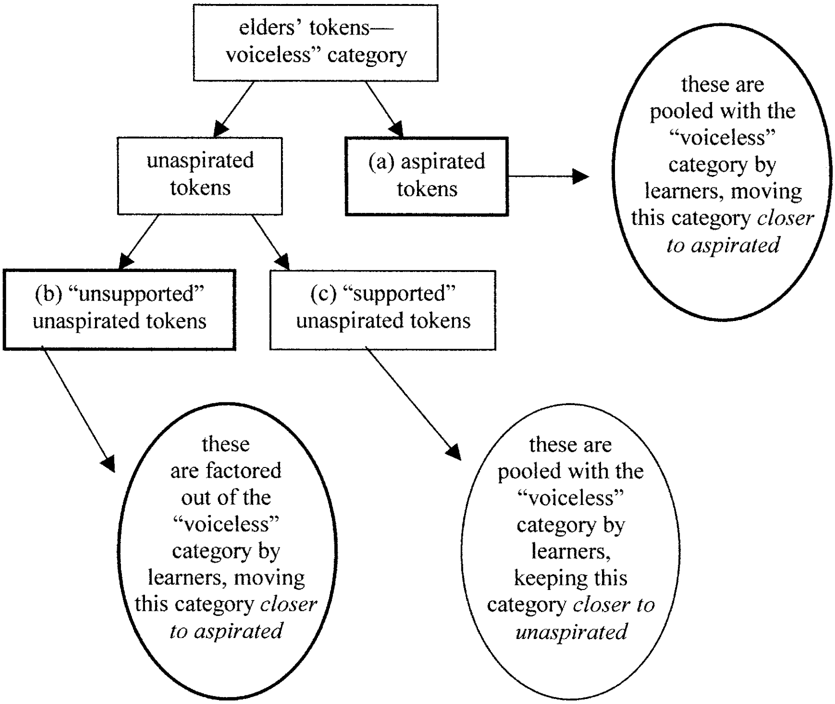

Figure 2. *Probability matching and aspiration*

the pool of relevant tokens. In short, it is the communicative success of unambiguous tokens which might have pushed the "voiceless" series toward later voice onset times. To paraphrase Darwin, it may metaphorically be said that probability matching involves daily and hourly scrutinizing of every variation, even the slightest; rejecting those tokens which are bad, preserving those which are good (1859: 88). As in Figure 1, the heavily outlined cells in Figure 2 highlight the *change* towards aspiration.

We can now consider a hypothetical schematic timetable which captures the main forces argued to be at work in the diachrony of the word-initial values. Consider the chart in Figure 3.

Let us first say that the lenis value is slowly creeping toward a voiceless state at a rate of three percent per generation, due to natural aerodynamic forces. Entering the sound change midstream, we take a thousand-token sample from Generation W's lenis tokens. Of these tokens, 750 are voiced, while 250 are voiceless. All the voiced tokens are transmitted successfully

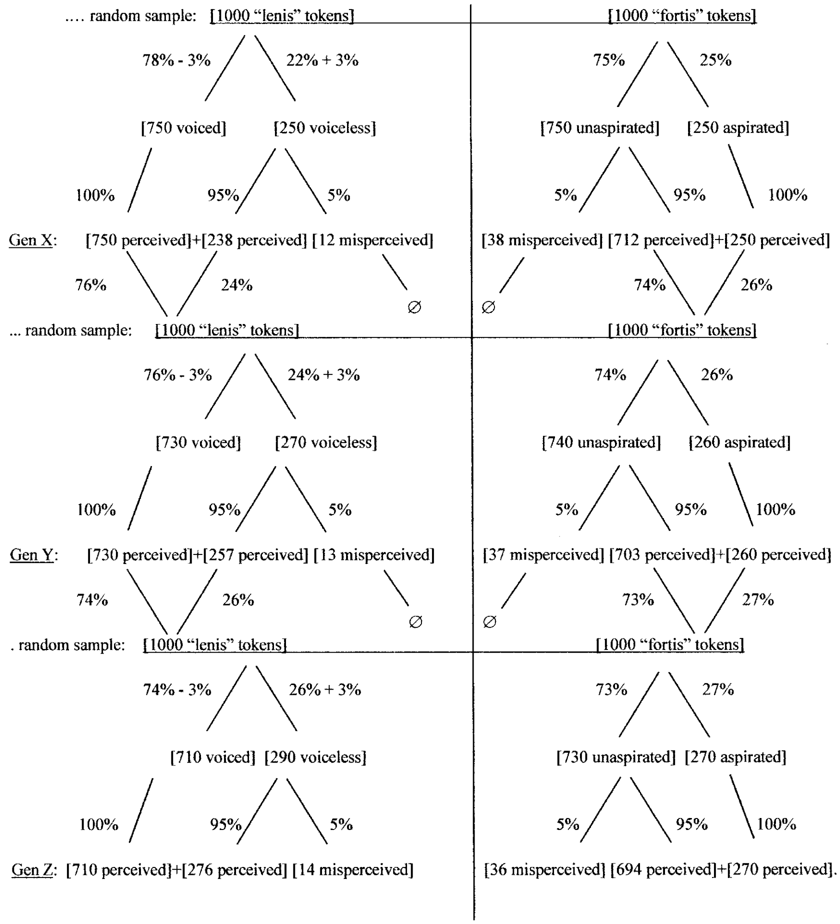

Figure 3. *A hypothetical diachrony*

to the listeners of Generation X. Of the voiceless tokens, let's assume that 95 percent of them (238) are "supported", and thus pooled with the lenis category; five percent (twelve tokens) are misperceived, and not pooled (some of these may be misperceived as belonging to the fortis category, but for simplicity's sake let us assume that they are all thrown out). Combining the 750 voiced tokens and the 238 supported voiceless tokens gives us 988 lenis tokens, 76 percent of which are voiced, and 24 percent of which are voiceless. These are the probabilities that Generation X will match in their own productions.

Now we iterate the process with another random sample of 1000 tokens. Again allowing for a three percent drift towards voicelessness, Generation X produces 730 voiced tokens, and 270 voiceless tokens, five percent of which are unsupported. This yields 987 tokens perceived as lenis, 74 percent of which are voiced. As the generations proceed, slowly, slowly, the lenis category may undergo a sound change from voiced to voiceless.

But while all this is going on with the lenis category, what is happening to the fortis category? For the sake of argument, let us assume that this category is also at a stage in which 75 percent of the tokens are voiceless unaspirated (750 out of a random sample of 1000) and 25 percent are aspirated. All these aspirated tokens will be transmitted successfully to Generation X listeners. Let us again suppose that five percent (38) of the unaspirated tokens are "unsupported" and thus thrown out of the fortis category. This leaves 712 voiceless tokens pooled with the 250 aspirated tokens perceived by the members of Generation X as belonging to the fortis category. Now Generation X matches the probability of perceived occurrence, producing 74 percent (i.e., 712 perceived out of 967) voiceless realizations, and 26 percent (derived from 250 perceived out of 967) aspirated realizations. And so the process continues.

This scenario demonstrates how very minor phonetic tendencies, coupled with the ambiguities they may induce, could eventually have far-reaching consequences for the system of contrasts. Also, consider the consequences of the lenis and fortis categories evolving in imperfect tandem over the generations, as they do in our example in Figure 3. The possibility that the categories might shift at slightly different rates allows for the possibility of diachronic mergers or perhaps near-mergers. In contexts where cueing potential is insufficient to allow the survival of both categories, such mergers are especially possible. But in initial position at least, where stop releases allow for an array of laryngeal modifications, there is ample opportunity for the two categories to maintain a healthy acoustic distance from each other. The lenis/fortis contrast survives, but takes on new phonetic characteristics. If we further assume, along with Bybee (e.g., 2001) that sound changes may affect frequent words before infrequent words, we may further account for individual category shifts for particular words: if some words are changing at a faster rate than others, then they might be expected to induce greater confusion in listeners, who might then miscategorize these words into the category with which the confusion arose.

There are, of course, any variety of different weights we might supply that would model different rates and directions of change or stability. Proper weighting of the proposed forces at work must, of course, be determined empirically. Also, we have not yet considered the possibility that certain social trends might favor one realization over another, or any number of other conceivable variables. We can never isolate every variable that influences sound change, internal or otherwise, but we probably can get a handle on some of the major forces at work and their diachronic interplay. Given the broad application of probability matching in explaining certain animal and human behaviors, proposing an explanatory role for this phenomenon in an account of sound changes that lead to new patterns of allophony would not seem overly rash. As Labov writes (1994: 583–598),

> [w]e should not be embarrassed if we find that systematic readjustments in ... language are governed by the same cognitive faculty that governs the social behavior of mallard ducks ... We are products of evolving history, not only our own but that of the animal kingdom as a whole, and our efforts to understand language will be informed by an understanding of this continuity with other populations of socially oriented animals.

3.2. *Syllable-final (including word-final) allophones*

In English, recall that the fortis/lenis contrast survives in syllable-final contexts, but not in terms of voicing. Instead, it is largely transferred to a length contrast on the preceding vowel. As noted, the absence of voicing in syllable-final stops is quite common cross-linguistically. Many languages neutralize contrasts in this position (e.g., Russian, Dutch), while many others have nonalternating voiceless sounds here (e.g., Thai). For example, while English maintains a robust contrast in syllable-final stops in preceding vowel length, Dutch undergoes the devoicing process without such recourse, and the contrast is barely in evidence. As discussed in section 1, we know that voicing is difficult to implement and maintain when a consonant or word-ending follows; it is certainly not impossible to maintain a contrast in syllable-final position, only less likely.

But what is the probable origin of the pronounced vowel-length contrast? Diehl and Kluender (1989), based on the results of a number of studies (Denes 1955; Raphael 1971; Kluender et al. 1988), suggest that vowel duration distinctions may serve to enhance the contrast between the two stop categories by making the lenis stop *seem* shorter by increasing the ratio between vowel length and closure length. While the authors claim that subjects "interpret a longer vowel as evidence of [stop] voicing" (1989: 129), this "voicing" presumably does not refer to actual phonetic voicing, but instead refers to the lenis category in general, whether genuine voicing is present or not. Indeed, again, Diehl and Kluender's "voicing" might best be interpreted paleophonetically: phonetically voiced stops are usually shorter in duration than voiceless stops, since the shorter the oral occlusion, the less likely it is that vocal fold vibration will be extinguished. But as diachronic coda devoicing began, those tokens with increased vowel length may have enjoyed a better survival rate, since the increased vowel duration–stop duration ratio served to enhance a contrast that was in the process of losing one of its major cues, that is, the distinction between phonetic voicing and voicelessness during closure. The result is a contrast consisting of the difference in vowel duration–closure duration ratio between the categories which is observed in English today. As Diehl and Kluender write,

> [i]s it a mere coincidence that cues linked by production constraints also turn out to have mutually reinforcing effects? If so, it is remarkable that such "coincidences" seem to be rather common in the speech domain ... [W]e suggest that these apparently fortuitous correspondences actually reflect a kind of *linguistic natural selection* (1989: 135; emphasis added).

Indeed, to paraphrase Darwin again (1859: 89), although vowel length may at one time have been only a minor phonetic exponent of the contrast (as it is in many other languages), as a consequence of diachronic coda devoicing this phonetic character and structure (which was apt to be of very trifling importance) may thus have been acted on and enhanced, as its presence became relevant for the maintenance of the contrast.

3.3. *Word-internal, stressed syllable–initial allophones*

As already suggested, the increased energy and duration of stressed domains make it more likely that contrasts will disperse themselves more widely in the available acoustic space, and/or may allow for a greater number of contrasts. Paraphrasing Darwin yet again (1859: 145), an abundance of cueing potential in this context may lead to divergence of character; for, provided cueing potential is sufficiently great, distinct values can be supported in a context the more these values diverge in structure, of which we see proof by looking at the cross-linguistic tendency for stressed syllables to support the greatest divergence of values in terms of both number and phonetic quality.

3.4. *Word-internal, unstressed syllable–initial allophones*

Recall that some dialects maintain a vowel length contrast in the context of a following tap. All the mechanisms are now in place to understand the possible origin of this context-dependent realization. If *t and *d were previously present, the vowel-length distinction may have played a minor role in cueing the contrast. But as the voicing contrast began to yield to tapping, those tokens which maintained (and increased) the vowel-length distinction were more readily perceived by listeners as belonging to distinct categories, and so were more readily produced: *ˈVdV > ˈVɾV; *ˈVtV > ˈV̆ɾV.

Recall that some dialects have fully merged the fortis/lenis contrast in this context, in the direction of the tap: *ˈVdV > ˈVɾV; *ˈVɾV > ˈVɾV. As in those languages which have lost their laryngeal contrasts in coda position, here too—where laryngeal cues at stop release have been lost as a consequence of stresslessness on the following vowel—the opportunities for cue expression are diminished. Once again, the contrast is not impossible to maintain here, only less likely.

3.5. *Allophones following [s]*

It is certainly not impossible for an aspiration contrast to exist after s-stop clusters. Two lines of linguistic evidence prove this. First, some languages indeed have an [st]/[stʰ] contrast, for example Sanskrit (Whitney 1889). Second, English speakers may produce [stʰ] when a word boundary intervenes between the fricative and the stop: [s#tʰ] (Pétursson 1977). So why is this pattern never found contrastively in English? As already mentioned, Kingston (1990) suggests that the open glottis required by the fricative may be difficult to manipulate beyond the stop closure; difficult, but clearly not impossible. Patterns that might be found across word boundaries are sometimes absent within words, and even more commonly absent within morphemes. The origins of these restrictions have been considered in the work of Joan Bybee (e.g., Bybee 2001). Within-word articulatory routines tend to be more frequent than those found across word boundaries. Bybee proposes that articulatory routines which are more frequent may be more susceptible to simplification over time. So words with *s*-stop clusters may more readily be unaspirated than those *s*-stop clusters which occur before word boundaries. In time, no aspiration contrast may remain in the word-internal condition.

3.6. *Past imperfect*

There is a temptation to assume perfect symmetry in the linguistic past, which has been distorted over time into the asymmetries of today. So we might think that the complex allophonic array in the contemporary English lenis/fortis system historically originated in a context-free voicing versus context-free voicelessness distinction. But just as today's asymmetries will be the future's imperfect past, today's past was probably just as imperfect. The various context-specific sound changes I have proposed have probably had their own unique timelines, such that at any given

Table 3. *Natural selection in species and in phonology*

Evolution of species	↔	Sound change
Genetic mutations	↔	Token-to-token variability
Beneficial mutations	↔	Variants which are more robustly distinct from other contrastive values
Successful adaptation to environmental niche	↔	Successful communication of word meaning
Passing successful mutations to subsequent generations	↔	Perception-based probability matching
Propagation of the mutation	↔	The sound begins to change
Population successfully adapts to its local environment	↔	Context-dependent allophones emerge
Competition for resources with similar species may result in extinction	↔	Lack of cueing potential may yield to neutralization or merger

point in history, the system was in as much phonetic disarray as it is in today, and as it will be in the future.

4. Conclusion

Due to the inherent variability of speech production, those tokens of words with the lenis or fortis alveolar stop which better conveyed linguistically relevant phonetic distinctions may have been more likely to be reproduced in listeners' speech. It may thus be the adaptation of a contrastive value to its context, and its subsequent survival as a functionally beneficial component of the communicative system, which is responsible for the allophonic pattern in evidence today.

In the evolution of species, both comportment and physical form are the products of long-term development by which minor genetic variations advantageous to the survival of the individual increase the likelihood of such genetic traits being passed to offspring. While most mutations, especially "monstrosities" are disadvantageous to the development of a species, some may increase the likelihood of survival of the organism by better adapting it to its environmental niche. It is these traits that are likely to be passed on to offspring. These offspring, in turn, are more likely to survive and procreate. Populations of the same species may be modified in their adaptations to their particular environmental niches. I have applied these classical Darwinian principles to another system—phonology—which also depends on an imperfect process of replication, proposing that a similar mechanism exists for certain sorts of internal sound change. In phonology, contrastive values may be seen as the analog of species, and the phonetic variability inherent to speech production may be seen as the mutations which individual organisms may undergo. Probability matching cannot be directly likened to reproduction, but in certain crucial ways these processes play the same role in their respective systems. Reproduction provides the cross-generational link whereby genetic material is imperfectly transmitted, and probability matching provides the cross-generational link whereby the phonological system is imperfectly transmitted. Due to imperfections in the communicative system, those variants which are better adapted to their context—where adaptation refers to successful communication of word meaning from speaker to listener—are in turn more likely to be reproduced as these listeners become speakers.

Received 23 October 2000
Revision received 22 November 2002

University of Illinois at Urbana-Champaign

Notes

* Author's e-mail address: ⟨dan@uiuc.edu⟩.

1. By contrast, aspirated *affricates* pattern with other aspirated plosives in their being rather common. Here, the presence of a stop release may establish the aerodynamic conditions necessary to achieve salient realizations of both frication and following aspiration. Silverman (1995) explores these issues in detail.
2. Gallistel further explains how the "irrationality" of such behavior from the point of view of the individual organism is only apparent: from a broader evolutionary point of view, in the context of natural, populated settings, the observed behavior is actually beneficial.
3. In contrast, consider the case of Cantonese, where similar circumstances led to the loss of the obstruent voicing contrast, and a multiplication of the tonal inventory. Unlike Cantonese, Mandarin forfeited the phonological contrast in favor of a morphological response: the open class vocabulary endured massive phonological neutralization, but was offset by a morphological compounding process. Languages may respond in differing, though phonetically and functionally constrained ways to similar phonetic and functional conditions.

References

Abramson, Arthur S. and Leigh Lisker
1985 Relative power of cues: F_0 shift versus voice timing. In Fromkin, Victoria (ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. New York: Academic Press, 25–33.

Antilla, Raimo
1972 *An introduction to historical and comparative linguistics*. New York: MacMillan.

Browman, Catherine P. and Louis M. Goldstein
1986 Towards an articulatory phonology. *Phonology Yearbook* 3, 219–252.

1990 On articulatory binding: Comments on Kingston's paper. In Kingston, John and Mary E. Beckman (eds.), *Papers in Laboratory Phonology I.* Cambridge: Cambridge University Press, 445–450.

Bybee, Joan
2001 *Phonology and Language Use*. Cambridge: Cambridge University Press.

Caisse, Michelle
1982 Cross-linguistic differences in fundamental frequency perturbation induced by voiceless unaspirated stops. Unpublished M.A. thesis, University of California at Berkeley.

Cole, Ronald A. and William E. Cooper
1975 Perception of voicing in English affricates and fricatives. *Journal of the Acoustical Society of America* 58, 1280–1287.

Darwin, Charles
1859[1861] *The Origin of Species* [3rd ed.]. London: John Murray.

De Jong, Kenneth
1991 The oral articulation of English stress accent. Unpublished Ph.D. dissertation, Ohio State University.

Denes, Peter B.
1955 Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America* 27, 761–764.

Diehl, Randy and Keith R. Kleunder
1989 On the object of speech perception. *Ecological Psychology* 1, 121–144.

Emerson, Oliver F.
1906 *The History of the English Language*. London: McMillan.

Flege, James E.
1982 Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics* 10, 177–192.

Gallistel, Charles R.
1990 The organization of learning. Cambridge MA: MIT Press.

Gluck, Mark A. and Gordon H. Bower
1988 Evaluating an adaptive network model of human learning. *Journal of Memory and Language* 27, 166–195.

Goldinger, Stephen D.
1997 Words and voices: Perception and production in an episodic lexicon. In Johnson, Keith and John W. Mullennix (eds.), *Talker Variability in Speech Processing*. San Diego: Academic Press; 33–56.
1998 Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105.2, 251–271.

Gurevich, Naomi
2003 Functional constraints on phonetically conditioned sound changes. Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Hombert, Jean–Marie
1978 Consonant types, vowel quality, and tone. In Fromkin, Victoria (ed.), *Tone: A Linguistic Survey*. New York: Academic Press, 77–111.

Hock, Hans H.
1991 *Principles of Historical Linguistics*. Berlin/New York: Mouton de Gruyter.

Hockett, Charles F.
1968 *The State of the Art*. (Janua Linguarum. Series minor, 73.) The Hague: Mouton.

House, A. S., and G. Fairbanks

1953 The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25, 105–113.

Janda, Richard D. and Brian D. Joseph

2001 Reconsidering the canons of sound change: Towards a Big Bang theory. Paper presented at the International Conference on Historical Linguistics, Melbourne.

Jespersen, Otto

1956 *Growth and Structure of the English Language*. Garden City, NY: Doubleday.

Johnson, Keith

1997 Speech perception without speaker normalization: An exemplar model. In Johnson, Keith and John W. Mullennix (eds.), *Talker Variability in Speech Processing*. San Diego: Academic Press, 145–165.

Jones, Daniel

1950 *The phoneme: Its nature and its use*. Cambridge: Heffer.

Jurafsky, Daniel, Alan Bell, Michelle Gregory, and William D. Raymond

2001 Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, Joan and Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 229–254.

Keating Patricia A., Wendy Linker, and Marie Huffman

1983 Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics* 11, 277–290.

Kim, Chin W.

1970 A theory of aspiration. *Phonetica* 21, 107–116.

Kingston, John

1990 Articulatory binding. In Kingston, John and Mary E. Beckman (eds.), *Papers in Laboratory Phonology I*. Cambridge: Cambridge University Press, 406–434.

Kingston, John, and Randy L. Diehl

1994 Phonetic knowledge. *Language* 70, 419–454.

Klatt, Dennis H.

1976 Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59, 1209–1221.

Kluender, Keith R., Randy L. Diehl, and B. A. Wright

1988 Vowel-length difference before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 2, 153–169.

Kruschke, John K.

1992 ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review* 99.1, 22–44.

Kruszewski, Mikolaj

1881[1978] On sound alternation. In Baldi, Philip and Ronald N. Werth (eds.), *Readings in Historical Phonology*. University Park: Pennsylvania State University Press, 64–91.

Labov, William

1994 Principles of linguistic change, internal factors. Oxford: Blackwell.

Ladefoged, Peter

1975 *A Course in Phonetics*. Fort Worth: Harcourt Brace.

Lewis, Anthony M.
2001 The phonetics and phonology of Spanish lenition. Unpublished Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Lindblom, Björn, S. Pauli, and Johan Sundberg
1975 Modeling coarticulation in apical stops. In *Speech Communication* 2, 87–94.

Lisker, Leigh and Arthur Abramson
1964 A cross-linguistic study of voicing in initial stops: Acoustical measurements. *Word* 20, 384–422.

Maddieson, Ian
1984 *Patterns of Sounds*. Cambridge: Cambridge University Press.

Malécot, André
1970 The lenis-fortis opposition: Its physiological parameters. *Journal of the Acoustical Society of America* 47, 1588–1592.

Martinet, André
1988 The internal conditioning of phonological changes. *La Linguistique* 24.2, 17–26.

Miller, Joanne
1994 On the internal structure of phonetic categories: A progress report. *Cognition* 50, 271–285.

Munson, Benjamin
in press Pattern frequency and speech production. *Journal of Speech, Language, and Hearing Research*.

Ohala, John J.
1989 Sound change is drawn from a pool of synchronic variation. In Breivik, Leiv E. and Ernst H. Jahr (eds.), *Language Change: Contributions to the Study of its Causes*. Berlin/New York: Mouton de Gruyter, 173–198.

Paul, Hermann
1886[1970] *Principles of the History of Language* [translated from 2nd ed. by H. A. Strong. College Park: McGroth].

Pétursson, Magnús
1977 Timing of glottal events in the production of aspiration after [s]. *Journal of Phonetics* 5, 205–212.

Pierrehumbert, Janet B.
1994 Knowledge of Variation. In *Papers from the Parasession on Variation, 30th Meeting of the Chicago Linguistics Society*. Chicago: Chicago Linguistics Society, 232–256.
1999 What people know about sounds of language. *Studies in the Linguistic Sciences* 29.2, 111–120.
2001 Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee, Joan and Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 137–157.

Raphael, Lawrence J.
1971 Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America* 51, 1296–1303.

Robertson, S.
1934 *The development of Modern English*. New York: Prentice Hall.

Rothenberg, Martin
1968 *The breath-stream dynamics of simple released-plosive production*. Basel: S. Karger.

Silverman, Daniel
1995 Phasing and recoverability. Ph.D. dissertation, University of California at Los Angeles.

Steels, Luc
2000 Language as a complex adaptive system. In Schoenauer, Marc, Kalyanmoy Deb, Guenter Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo, and Hans-Paul Schwefel (eds.), *Lecture Notes on Computer Science. Parallel Problem Solving from Nature* (PPSN-VI). Paris: Springer-Verlag, 17–26.

Westbury, John R. and Patricia A. Keating
1986 On the naturalness of stop consonant voicing. *Journal of Linguistics* 22, 145–166.

Whitney, William D.
1989 *Sanskrit Grammar*. Cambridge, MA: Harvard University Press.

Yoshioka, Hirohide, Anders Löfqvist, and Hajime Hirose
1981 Laryngeal adjustments in the production of consonant clusters and geminates in American English. *Journal of the Acoustical Society of America* 70–6, 1615–1623.