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

Daniel Silverman
Department of Linguistics, and the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign
dan@uiuc.edu
I offer an evolutionary approach to the English alveolar stop allophony pattern—with the lenis value evolving towards context-dependent naturalness, and the fortis value evolving towards context-dependent contrastiveness (provided cue expression is sufficient to avoid neutralization)—arguing that patterns of allophony may be the long-term product of an interaction among phonetic, cognitive, and functional forces.

keywords:
Evolutionary phonology
English phonology
Allophony
Probability matching
Sound change
Language acquisition
On the phonetic and cognitive nature of alveolar stop allophony in American English

“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 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, from voicelessness to aspiration in word-initial stops: *#p > #pʰ); as *b begins to naturally devoice, those tokens of *p-initial words which contain more aspiration are communicated more reliably to listeners, who in turn are 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) the tendency for speakers to copy the variability they perceive—sounds may change their context-specific characteristics. 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. So, the small phonetic variations in which allophones naturally engage are a means by which they take on new properties. However, 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, which 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 1.
Word-initially (1a), we see a primary aspiration contrast. The lenis stop is typically a plain voiceless stop or is only slightly prevoiced (Lisker and Abramson 1964, Flege 1982), while the fortis stop is aspirated. Syllable- and word-finally (1b) the lenis stop is also typically realized as a plain voiceless or partially devoiced stop (Ladefoged 1975), 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 (1c) the lenis stop is usually fully voiced, while the fortis stop is voiceless and aspirated. Word-internal unstressed syllable-initially (1d), there is complete or near-neutralization: both the lenis and fortis stops are tapped, with or without a vowel length distinction on the preceding vowel. Finally, following [s] (2e) there is no contrast between the fortis and lenis stop. There are various additional aspects of the fortis-lenis distinction, including differences in F1 cutback durations, differences in F0 perturbations, and differences in closure durations. Kingston and Diehl (1994) provide a very useful overview of these phonetic characteristics.

In 2 are some examples of alternations among all the different allophones presented in 1. These alternations provide evidence that these distinct phonetic forms are indeed allophonically related.

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 towards the more natural realization. This discussion of phonetic “naturalness” is necessary in order to motivate a major claim in Section 3, that phonetically natural sound changes may induce less natural ones.

2.1 Word-initially
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 supraglottal pressure rise in tandem, resulting in an insufficient pressure drop until oral release finally induces voicing. Even with the vocal folds fully adducted well before stop release, these researchers estimate that perhaps only 30-40 ms 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- (including word-) finally
In final position (1b), Westbury and Keating calculate that, regardless of the positioning of the vocal folds, transglottal flow may naturally persist for the first portion of the closure (about 35 ms 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 ([ð⁰̶̰⁰̶]), 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’ 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 his presented synthesized stimuli, 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 of tongue-to-roof contact (Lindblom, Pauli, and Sundberg 1975). This, in turn, results in shorter and less pronounced formant transitions, comparatively similar to the formant structure in to 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-initially
Stressed-syllable-initially (1c), 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] vs. [tʰa]. The lenis value is the natural value, as voicing is natural for intervocalic stops: their short closure duration allows for vocal fold vibration to continue from one vowel to the next.

2.4 Word-internal unstressed syllable-initially
Medially (intervocically) before a stressless vowel (1d), Westbury and Keating 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
their decreased aerodynamic and articulatory force, and their shorter duration than stressed domains. Here, the oral closure is typically short enough—a tap—so that trans-glottal flow does not markedly dissipate, which establishes a natural environment for voicing. Effective cueing of the distinct values is more difficult under these circumstances. Not surprisingly, in such cue-deprived contexts, the contrast only barely survives (often in vowel length on the preceding (stressed) vowel); in some dialects, the contrast is lost, and only the natural tap survives.

2.5 Following s
After [s] (1e), the fortis-lenis distinction does not exist. Kim (1970), and Yoshioka, Lofqvist, and Hirose (1981) report that in contexts such as [sp, 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 on to 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

The table in (3) offers a summary of the distribution of alveolar stop allophones in American English.

INSERT FIGURE 3 ABOUT HERE

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

Given the potential for cueing in cases 2a, 2b, and 2c, 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 2d and 2e however, cueing potential becomes an overriding (or almost overriding) factor. In these contexts, the contrast is often lost, as in 2d, or the contrast simply does not exist, as in 2e. In these contexts, neutralization or merger is towards the natural, or easier realization.

I should emphasize that characterizing particular speech sounds as “easy” and “natural,” or “difficult” and “unnatural,” does not mean that this ease or difficulty has a significant influence on the speech of any particular speaker or speech community. Rather, as I argue in the next

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. Kingston (1985, 1990) and Silverman (1995) explore these issues in detail.
section, speakers are excellent at copying the speech sounds which they encounter, whether we might characterize them as easy to produce or not. Instead, ease or difficulty of production may only have very slow-acting effects on speech production, observable only across generations of speakers.

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’s 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, Linker, and Huffman 1983, Gurevich in prep.). Based on his survey of over three hundred languages, Maddieson (1984:31) states “…[L]anguages nearly always include a plain voiceless series of stops. If there is only one series it is of this kind.” However, although not discussed by Maddieson, 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 stops in VCV contexts). Final stops are often voiceless, as in Basque, Bulgarian, Cantonese, Choctaw, Dutch, Efik, Ewondo, Finnish, Gaelic, German, Polish, Russian, Zoque, Korean, Nama, Thai, Tikar, 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”; Hockett (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), “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 (1886:43ff.), 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 75% of their word-initial lenis stops without voicing, 20% with minimal voicing, and 5% 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 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 (1990:352ff.) reports on a study in which rats in a T Maze were rewarded with food 75% of the time at one end, 25% of the time at the other. When provided with feedback, rats matched the probability of reward—running to the one end 75% of the time, the other end 25% of the time—despite the fact that they would receive more rewards if they ran to the one end 100% 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 (1994:583ff.), “It is not a hypothesis that children do probability matching [during language learning]. It is simply a description of the observed facts...”

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,

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.
learners’ articulatory talents are harnessed in service to *copying or imitating*, not *modifying* (improving upon or otherwise) the ambient speech pattern.

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, Kelly and Martin (1994), Miller (1994), Pierrhumbert (1994, 1999, 2000) Steels (2000), Bybee (2001), Jurafsky, Bell, Gregory, and Williams (2001), Munson (2001). 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 (for example, Gluck and Bower 1988, Kruschke 1992, Goldinger 1997, 1998, Johnson 1997, and Pierrhumbert 2001).

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 instead match their productions to something else. Let’s suppose for the moment that probability matching is indeed largely perfect, but instead of learners matching ambient *productions*, their own productions match their *perceptions* of these ambient productions. Since perception is demonstrably imperfect, then reproduction is imperfect as well. I’m 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 due to the imperfection of the *system* of communication—not the *medium* of communication—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-initially

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. Upon researching the issue however, it becomes apparent that the existing historical record is frustratingly scanty regarding phonetic descriptions of English stop allophony (for example, none of Robertson 1934, Emerson 1935, 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 gets some support from two separate though related findings. First, Abramson and Lisker (1985) report that while voice onset time (VOT) is the primary determinant of the fortis-lenis distinction in initial position, category ambiguities (at around 20 ms VOT) can be partially resolved by synthetically manipulating fundamental frequency at stop release. It is well established that F0 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 F0 at release may induce the perception of the lenis stop, while a raised F0 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. All these findings (Lisker and Abramson’s, Flege’s, Caisse’s) suggest that speakers might employ the appropriate articulatory posture to lower F0, 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 ms into the following vowel, well surpassing their perceptual limen, which Hombert determines to be around 40 ms. As Hombert notes, the English pattern cannot be accounted for by proximal 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 in these tone languages 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 anti-functional consequences.3

What I’m 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, when pitch perturbations became an increasingly important cue. Probability matching is probably not perfect—even when matched to perception, not production. Slowly, the natural tug towards devoicing stops in initial position may gain ground. 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 perceived unambiguously by listeners. Due to probability matching, these variants were more likely to be employed by listeners, and thus the system moved towards its present-day state. Listeners are imperfectly copying the speech around them, not trying to improve upon it. 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.

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.
I’ve so far suggested that phonetically natural aerodynamic forces may have effected changes in the realization of the lenis value, and that probability matching may have led to a re-conventionalization of the pattern. 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 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 still, devoicing was gaining ground. Specifically, among elders’ “voiced” category tokens, phonetically voiced tokens (see 4a) 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 (4b). 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 (4c). 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), “There 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 sensori-motor 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 sensori-motor constraints select those that can be re-produced and recognised.” In Figure 4, bolded cells highlight those forces which effect the change toward voicelessness, whereas the non-bolded cells contain those forces which maintain voicing.

INSERT FIGURE 4 ABOUT HERE

Consider how such a gradual devoicing of the English “voiced” series affects the “voiceless” series. As the “voiceless” series also engaged in variations that were largely matched from generation to generation, we might 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 VOT forward in the present-day lenis series, functional forces may have pushed VOT forward in the present-day fortis series. Specifically, those “voiceless” variants which possessed a slightly longer VOT came to be functionally beneficial as the “voiced” series began to creep towards voicelessness: exactly those tokens which were more distinct from the “voiced” series were successful in
conveying lexical distinctions to 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 5). That is, when provided with a clear phonetic contrast among phonologically distinct forms, probabilities were matched accordingly, while phonetically indistinct forms, being more likely misinterpreted by listeners, were factored out of the pool of relevant tokens. In short, it is the communicative success of unambiguous aspirated tokens which might have pushed the “voiceless” series toward longer VOT. To paraphrase Darwin, it may metaphorically be said that probability matching involves daily and hourly scrutinizing every variation, even the slightest; rejecting those tokens which are bad, preserving those which are good (1859:88). As in Figure 4, bolded cells in Figure 5 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 (6).

Let’s first say that the lenis value is slowly creeping toward a voiceless state at a rate of 3% per generation, due to natural aerodynamic forces. Entering the sound change midstream, we take a 1000 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 to the listeners of Generation X. Of the voiceless tokens, let’s assume that 95% of them (238) are “supported,” and thus pooled with the lenis category; 5% (12 tokens) are misperceived, and not pooled (some of these may be misperceived as belonging to the fortis category, but for simplicity’s sake, let’s assume that they are all thrown out). Combining the 750 voiced tokens and the 238 supported voiceless tokens gives us 988 lenis tokens, 76% of which are voiced, and 24% of which are voiceless. These are the probabilities that Gen X will match in their own productions.

Now we iterate the process with another random sample of 1000 tokens. Again allowing for a 3% drift towards voicelessness, Gen X produces 730 voiced tokens, and 270 voiceless tokens, 5% of which are unsupported. This yields 987 tokens perceived as lenis, 74% 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’s assume that this category is also at a stage in which 75% of the tokens are voiceless unaspirated (750 out of a random sample of 1000), and 25% are aspirated. All these aspirated tokens will be transmitted successfully to Gen X listeners. Let’s again suppose that 5% (38) of the unaspirated tokens are “unsupported,” and so thrown out of the fortis category. This leaves 712 voiceless tokens, pooled with the 250 aspirated tokens, perceived by Gen X as belonging to the fortis category. Now Gen X matches the probability of perceived occurrence, producing 74% (deriving from 712 perceived out of 967) voiceless realizations, and 26% (deriving 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 might induce, may 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 (5). 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 has taken 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 for listeners, who might then miscategorize these words as belonging to the category with which it is confused.

There are, of course, any variety of different weights we might supply that would model different rates and directions of change, or model stability. 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 accounting for sound changes that lead to new patterns of allophony would not seem overly rash. As Labov writes (1994:583ff.), “We 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- (including word-) finally
In English, recall that the fortis-lenis contrast survives syllable-finally, but again, 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 non-alternating 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, Diehl, and Wright 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” (p.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 that vocal fold vibration will be extinguished. But as diachronic coda devoicing begins, 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 vowel/closure duration ratio difference between the categories which is observed in English today. As Diehl and Kluender write, “Is 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*” (p.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), this phonetic character and structure, which was apt to be of very trifling importance, as a consequence of diachronic coda devoicing 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-initially

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, or may allow for a greater number of contrasts, so we find natural voicing for the lenis value and unnatural aspiration for the fortis value. Paraphrasing Darwin once again (1859:145), an abundance of cueing potential in stressed contexts 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 contexts to support the greatest divergence of values in terms of both number and phonetic quality.

### 3.4 Word-internal unstressed syllable-initially

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: *

\[
\begin{align*}
\text{VdV} & > \text{VtV}; \\
\text{VtV} & > \text{VrV}.
\end{align*}
\]

Recall that some dialects have fully merged the fortis-lenis contrast in this context, in the direction of the tap: *

\[
\begin{align*}
\text{VdV} & > \text{VrV}; \\
\text{VrV} & > \text{VrV}.
\end{align*}
\]

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 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, Sanskrit having been an example (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. 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 point in history, the system was in as much phonetic disarray as it’s 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 (see also Croft 2000 for a Darwinian approach to language 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 more likely to be, in turn, reproduced as these listeners become speakers.
References:


1. American English alveolar stops:

<table>
<thead>
<tr>
<th></th>
<th>lenis:</th>
<th></th>
<th>fortis:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>form:</td>
<td>example:</td>
<td>form:</td>
<td>example:</td>
<td></td>
</tr>
<tr>
<td><strong>(a)</strong> word-initially:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t, d</td>
<td>‘ˈtak ˈdak dock</td>
<td>tʰ, ˈtʰʌp top</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(b) Syllable- (including word-) finally:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t, d</td>
<td>‘nɒt ˈnɔd nod</td>
<td>Ṿ(ʔ)RP Ṿʔ ‘nɒʔ(ʔ)RP ‘nɔʔ knot</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(c) word-internal stressed-syllable-initially:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>əˈdɔpt adopt</td>
<td>tʰ</td>
<td>əˈtʰʌp atop</td>
<td></td>
</tr>
<tr>
<td><strong>(d) word-internal unstressed syllable-initially:</strong></td>
<td>Ṿf</td>
<td>ˈɑɹt, ˈɑɹt odder (neutralized)</td>
<td>r</td>
<td>ˈɑɹt otter</td>
</tr>
</tbody>
</table>
| **(e) following s:** | form: t | example: stop(⟨⟩) | stop (non-contrastive)

2. Allophonic alternations:

<table>
<thead>
<tr>
<th></th>
<th>d - r</th>
<th>addiction</th>
<th>addict</th>
</tr>
</thead>
<tbody>
<tr>
<td>t/d - r</td>
<td>bud</td>
<td>budding</td>
<td></td>
</tr>
<tr>
<td>t - d</td>
<td>do</td>
<td>redo</td>
<td></td>
</tr>
<tr>
<td>t - r</td>
<td>disperse</td>
<td>redisperse</td>
<td></td>
</tr>
<tr>
<td><strong>Fortis:</strong></td>
<td>Ṿ(ʔ)RP - Ṿf</td>
<td>hiť</td>
<td>hitting</td>
</tr>
<tr>
<td>ʃ̣h - r</td>
<td>əˈtɔmic</td>
<td>atọm</td>
<td></td>
</tr>
<tr>
<td>Ṿ(ʔ)RP - ʃ̣h</td>
<td>dictate</td>
<td>dictaṭorial</td>
<td></td>
</tr>
</tbody>
</table>

3. American English alveolar stop allophony:

<table>
<thead>
<tr>
<th>Context ⇒</th>
<th><strong>(a)</strong> word-initially:</th>
<th><strong>(b)</strong> Syllable- (including word-) finally:</th>
<th><strong>(c)</strong> word-internal stressed syllable initially:</th>
<th><strong>(d)</strong> word-internal unstressed syllable initially:</th>
<th><strong>(e)</strong> following s:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenis ⇒</td>
<td>d/t (more natural)</td>
<td>t/d (more natural)</td>
<td>d (more natural)</td>
<td>Ṿf (more natural)</td>
<td>t (more natural)</td>
</tr>
<tr>
<td>Fortis ⇒</td>
<td>ʃ̣h (less natural)</td>
<td>Ṿ(ʔ)RP - Ṿʔ (less natural)</td>
<td>ʃ̣h (less natural)</td>
<td>Ṿf (more natural)</td>
<td>t (more natural)</td>
</tr>
</tbody>
</table>
4.

- **Elders' tokens** — “voiced” category
  - (a) Voiced tokens
  - (b) “Supported” devoiced tokens
  - (c) “Unsupported” devoiced tokens

- These are pooled with the “voiced” category by learners, keeping this category closer to voiced.

- These are factored out of the “voiced” category by learners, keeping this category closer to voicelessness.
5.

(a) aspirated tokens

these are pooled with the "voiceless" category by learners, moving this category closer to aspirated

(b) "unsupported" unaspirated tokens

these are factored out of the "voiceless" category by learners, moving this category closer to aspirated

(c) "supported" unaspirated tokens

these are pooled with the "voiceless" category by learners, keeping this category closer to unaspirated
… random sample: [1000 “lenis” tokens] [1000 “fortis” tokens]

|                | [750 voiced] | [250 voiceless] | [750 unaspirated] | [250 aspirated] |
|----------------|--------------|----------------|
| 78% - 3%       |              |                |                  |
| Gen X:         | [750 perceived] | [238 perceived] | [12 misperceived] |
|                | 76%          | 24%            | ∅                |

… random sample: [1000 “lenis” tokens] [1000 “fortis” tokens]

|                | [730 voiced] | [270 voiceless] | [740 unaspirated] | [260 aspirated] |
|----------------|--------------|----------------|
| 76% - 3%       |              |                |                  |
| Gen Y:         | [730 perceived] | [257 perceived] | [13 misperceived] |
|                | 74%          | 26%            | ∅                |

… random sample: [1000 “lenis” tokens] [1000 “fortis” tokens]

|                | [710 voiced] | [290 voiceless] | [730 unaspirated] | [270 aspirated] |
|----------------|--------------|----------------|
| 74% - 3%       |              |                |                  |
| Gen Z:         | [710 perceived] | [276 perceived] | [14 misperceived] |
|                | 74%          | 26%            | ∅                |

…culminating in present-day devoiced stops

…culminating in present-day aspirated stops
6. Natural selection in species, and in phonology

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