# Is sound change adaptive?

Björn Lindblom, Susan Guion, Susan Hura, Seung-Jae Moon & Raquel Willerman

This paper proposes that, in analogy with other phenomena of both biological and cultural change, sound change be analyzed in terms of a two-step process of variation and selection.

Phonetic variations are seen to arise from the ability of speakers to adaptively tune their performance to the various social and communicative needs that they associate with specific speaking situations. During the listener's processing, such variation is a means rather than an end. The focus is normally on 'what' is being said (message), rather than on 'how' it is said (signal). However, occasionally, the 'how' information is also conveyed.

It is proposed that it is this incidental 'how'-mode of perception that provides one of the beginnings of new pronunciations, and that selection, or rejection, by the speech community occurs as a result of an evaluation that language users implicitly perform with respect to, among other things, social, articulatory, perceptual, lexical-systemic dimensions of the new phonetic patterns.

Sound change is taken to be adaptive in the sense that, if forms arise that match the current values of the evaluation criteria better than the old forms, they are more likely to be phonologized.

In many respects, the present account follows the listener-based scenario proposed by Ohala. However, it differs in that misperceptions are de-emphasized as the sole seeds of change, and in that a significant role is played also by the speaker.*

## 1. Introduction

Our choice of title reflects a deliberate attempt to place the search for explanatory accounts of sound change in an evolutionary context. The appeal of the neo-Darwinian theory of evolution is that it offers general principles of "change" whose power and scope have long been recognized among evolutionary biologists.

There is wide consensus that evolutionary change is essentially a two-step process (Mayr 1982). At the first stage, *genetic variation* arises from mutation, recombination and various effects of (a)sexual reproduction.

As a result of this variability, individual members of a given species differ from each other in morphology, physiology and behavior and therefore vary in their ability to survive and reproduce.

*Rivista di Linguistica*, 7.1 (1995), pp. 5-37 *(ricevuto nell'ottobre 1994)*

This ability is put to the test at the second stage, *natural selection*, which favors individuals who are more successful competing for resources and coping with the environment. Biologists have pointed out that the two-step model is deceptively simple (Dawkins 1986:xi), and that it therefore often tends to be underestimated (Mayr 1978). Nevertheless, given sufficient time, the cumulative effects of natural selection are powerful enough to bring about change[1] and to help organisms *adapt* to new problematic conditions, thereby giving evolution its non-random, seemingly 'purposeful' direction.

> Adaptation is the process of evolutionary change by which the organism provides a better and better "solution" to the "problem", and the end result is the state of being adapted. (Lewontin 1978:213).

Are sound changes 'adaptive' in the sense of evolutionary biology? It may be objected that the question is not relevant, since language change is cultural, not biological. It is true that cultural evolution is in many ways different from organic evolution. For instance, it has been characterized as more 'Lamarckian' than 'Darwinian' (Mayr 1978) in that information that has been acquired by one generation, (e.g. practical and theoretical knowledge, systems of religious beliefs and moral values, and language), is passed on to the next with no effect on the genotype. Accordingly, cultural evolution obeys Lamarck's principle of "inheritance of acquired characters". Owing to this direct mode of transmission, it proceeds much faster than organic evolution.

However, these and other important differences notwithstanding, cultural and organic evolution share the same fundamental elements: They are both products of a process of 'selection' from 'variation' in the presence of biasing 'constraints' (Mayr 1978, Boyd and Richerson 1985, Cavalli-Sforza and Feldman 1981). Forms compatible with constraints stand a better chance of being selected than those that do not.

Is sound change 'adaptive'? Phrasing the question in that way we are implying that, right or wrong, the evolutionary model of 'change' is in fact worth taking seriously in the study of historical phonetics and phonology.

From there, and in keeping with the two-step model, we are led to ask: What is the nature of the variation from which the selections are made? By what processes are those selections implemented, and how are they constrained?

The goal of the present paper is to discuss those questions in the light of relevant evidence and to assess, in a preliminary way, the merits of an 'adaptationist' approach to sound change.





## 2. *Sources of phonetic variation*

### 2.1. *The multiple tasks of the speaker*

Speaking is often described as 'goal-directed' activity. Talkers endeavor to choose the grammatical and lexical form of utterances in a manner that serves their communicative intentions. In a significant way, that choice depends on what listeners know and what they do not know. For instance,

(1) He put it there

would be a perfectly well-formed utterance, if, say, "the carpenter" and "the hammer" had just been mentioned in the conversation, and if "there" was accompanied by a pointing gesture. In saying *he put it there*, rather than *the carpenter put the hammer on the table*, the speaker makes the 'presupposition' that, in the current situation, there are certain facts that need not be made fully explicit, since the listener already 'has access to' them.

The study of text comprehension (Hellman 1992) presents a number of suggestive, but perhaps incompletely recognized, parallels to research on the perception and understanding of speech. This work shows that producing a sentence invariably entails making tacit assumptions about what is 'old' and what is 'new' information at the moment of communicating. The text itself tends to be the 'tip of the iceberg'. For successful comprehension to take place, a vast amount of implicit information must be brought to bear. Similarly, research on discourse indicates that the amount of linguistic material talkers use tends to be inversely related to how much they can assume that the hearer already knows. Thus, Levy and McNeill (1992) show that, in narrations, expressions used to refer to characters in a film are long (e.g. character's name) when the character is not presumed to be foregrounded for the listener, but are short and opaque (pronouns, anaphoric expressions) when the character has already been made focal. For further examples see also Gundel, Hedberg and Zacharski (1993).

In the speech domain, Fowler and Housum (1987) compared forms that were used repeatedly in a monologue. They found that the first version was longer than the corresponding second version. This and other studies (Eefting 1991, 1992) suggest that speakers adapt timing and other prosodic attributes to the listener's assumed needs and tend to provide more processing time for 'new' than for 'old' information (Nooteboom 1985, Nooteboom and Eefting 1994).

One phonetic theory has been outlined that is, in many ways, a 'presupposition' account of (intra-speaker) phonetic variation – for short, the H(yper)&H(ypo) theory (Lindblom 1990). It is developed from observations of general characteristics of motor and perceptual systems. One of its cornerstones is experimental evidence indicating that speech intelligibility depends partly on the quality and contents of the signal, partly on the extent to which the signal engages short- and long-term phonetic, lexical, grammatical and other knowledge stored in the native listener's brain (Pickett and Pollack 1963, Pollack and Pickett 1964). During processing, the amount of such signal-independent information fluctuates between and within utterances. As a result, the predictability of a given speech unit does not stay constant either, but varies from one situation to the next (Lieberman 1963), as illustrated below for the word *bush*:

(2) a. The next word is __________
    b. A bird in the hand is worth two in the __________

A consequence of this analysis is that, since it is not solely responsible for the formation of the speech percept, the signal need not contain 'all' the information, but just enough for that percept to emerge in interaction with the stored knowledge. Minimally, the talker needs to ensure that the linguistic units have sufficient discriminatory power for making the correct lexical identifications, not necessarily that they be invariant (see Lindblom, Brownlee, Davis and Moon (1992) for further comments on this point).

H&H theory highlights two characteristics of action systems that make the production of such 'sufficiently informative' signals possible: Plasticity and economy. The former results from a very general motor principle, 'output-oriented' control, which enables an animal to attain the same motor goal under conditions that are usually drastically different from time to time (Sherrington 1941, Lashley 1930/60, 1951, Bernstein 1967, Granit 1977, Fukson, Berkinblit and Feldman 1980). The latter characteristic is evidenced in numerous studies showing that animals style their movements in accordance with the minimal expenditure of energy compatible with the task, although, in principle, they have a range of behaviors to choose from (Alexander 1988, Hoyt and Taylor 1981, Williams, Friedl, Fong, Yamada, Sedivy and Haun 1992). In the absence of constraints (prey slipping away, predator getting closer ...), motor systems tend to default to low-cost forms of behavior.

On the basis of such evidence the main argument of the theory is developed as follows. The (ideal) speaker makes a running estimate of the listener's need for explicit signal information on a moment-to-moment basis and then adapts the production of the utterance elements (words, syllables or phonemes) to those needs. This occurs along a continuum with more forcefully articulated 'hyper'-forms at one end and





less energetic 'hypo'-forms at the other. As the performance level increases from hypo to hyper, both the duration and the amplitude of articulatory gestures tend to increase, whereas their temporal overlap tends to decrease. The framework of articulatory phonology (Browman and Goldstein 1986, 1990a, 1990b) is well suited for the specification of such transforms. As a result, the context-dependence of articulatory and acoustic patterns is minimal in hyper-speech and maximal in hypo-speech. As further consequences, coarticulation and reduction are typical of the hypo-mode. And the vowels and consonants of hyper-speech are expected to be closer to their target values in hyper-speech.

According to this account, idealized for heuristic purposes, phonetic variation occurs along a single continuum. The exact point at which a given phonetic form is produced is determined by the speaker's assumptions about the informational needs of the listener and by his own tacit demand for articulatory simplification. It is in this sense that H&H theory can be said to give a 'presupposition' account of phonetic variation.

What is the evidence? For a comprehensive answer we refer to other sources (Lindblom 1990, Lindblom 1993). Here we shall concentrate on 'clear speech' whose very existence and acoustic nature lend support to the H&H theory. Of particular relevance is a study of the effect of clear speech on 'formant undershoot' (Moon 1991).

In this work five American English speakers produced words with one, two and three syllables in which the initial stressed syllable was [wil], [wɪl], [wɛl] or [weɪl]. They were asked first to produce the test words in isolation at a comfortable rate and vocal effort ('citation-form speech'), and then to pronounce the same words as clearly as possible ('clear speech'). As expected for English (Ladefoged 1993), the differences in word length were associated with a fairly wide range of vowel durations. Large displacements (undershoot effects) were observed, especially at short durations, for F2 sampled at the vowel midpoint.

It was possible to describe all the measurements[2] with the aid of a decaying exponential:

(3) $$F_{20} = k \cdot (F_{2i} - F_{2t}) \cdot e^{-aD} + F_{2t}$$

which says that the measured value, $F_{20}$, will approach its 'target', $F_{2t}$, more and more closely, as vowel duration D gets longer and longer. When it is short, a term proportional to $(F_{2i} - F_{2t})$, the difference between the formant frequency values for the consonant-vowel boundary ($F_{2i}$ or 'locus') and for the vowel ($F_{2t}$ or 'target'), will become significant and shift the formant away from the target value. If, as was the case with the [w_l]-frame, that term is large, a substantial displacement, or undershoot effect, would be expected, and was indeed found by Moon (1991), at short durations.

According to (3), the degree of undershoot depends on two factors: vowel duration and context (more specifically, the vowel's acoustic similarity to its consonant environment, the 'locus'-'target' distance). However, undershoot effects were less marked in clear speech. Speakers achieved this by increasing durations and speeding up the F2 transition from [w] into the following vowel. In some instances they also chose to increase the F2 target value. Moon interpreted his findings to suggest that speakers responded to the request for 'clear speech' by articulating more energetically and thus compensating for undershoot. This interpretation was confirmed by the results of intelligibility tests which presented the [w_l] sequences in noise at various S/N levels and which showed that the clear speech tokens were more intelligible than the corresponding citation forms.

It is evident from Moon's study as well as from other work (Chen 1980, Chen, Zue, Picheny, Durlach and Braida 1983, Clark, Lubker and Hunnicutt 1987, Lively, Pisoni, Summers and Bernacki 1993, Summers, Pisoni, Bernacki, Pedlow and Stokes 1988) that speaking clearly is not merely invoking identical, but louder 'normal' speech patterns. The facts of undershoot compensation imply an adaptive process with a much more far-reaching output-oriented articulatory reorganization.

Why should there be such a thing as clear speech? Why should it exhibit the phonetic properties it does? From the vantage point of H&H theory, clear speech owes both its existence and its phonetic characteristics to the way the production and perception of speech work. Clear speech is a manifestation of the adaptive organization of speech production which is evident in the interplay between the goals of satisfying the listener's informational needs on the one hand and of optimizing articulatory energetics on the other. Accordingly, clear speech is predicted by H&H theory.

### *2.2. Speech percepts: Role of signal-independent information*

We next consider how phonetic variations are processed perceptually.

Speaker /ut/ — distorted by voical tract into — [yt] - - - - - - - - - heard as- - - - - - - - -> [yt] — reconstructed as — Listener /ut/

**Figure 1.** Scenario 1





The listener plays a central role in the theory of sound change proposed by Ohala (Ohala 1974, 1978, 1981, 1983, 1990). His basic observation is that sound changes originate in listener misperceptions.

Their source is compared to a 'copying error' not unlike a 'scribal error', or a 'mutation' arising from a mistake in genetic copying. For a recent summary of the theory and its empirical bases see Ohala (1993). Assuming error-free conditions, Ohala represents the process of speech communication as shown in Scenario 1 of Figure 1. In the diagram (from Ohala 1981:181), the speaker aims at producing the syllable /ut/, but, because of coarticulation, the /u/ is fronted by the following /t/ and the result is phonetically more like [yt].

It is assumed that, in the normal situation, the listener expects coarticulation to take place and copes with it by applying 'reconstructive' rules. By identifying the cause of the vowel change, he is able to 'reconstruct' the identity of the intended /ut/ syllable.

Speaker /ut/ — distorted as — [y(t)] - - - - - - - - - heard as- - - - - - - - -> [y] — interpreted as — Listener /y/ — Listener-Turned Speaker — produced as — [y]

**Figure 2.** Scenario 2

Now suppose that, owing to disturbing noise or other factors (such as coarticulation and and other effects of "vocal tract constraints" (Ohala 1983)), context *cannot* be taken into account. The process would then occur as in Scenario 2 (Figure 2).

In this example, the context-based normalization rules have nothing to work on, and, therefore, no 'reconstruction' can take place. The listener's percept is determined by the surface information alone. Incorrectly, the vowel in question is identified as [y]. Ohala comments:

> *When this listener turns speaker* he will in his most careful, minimally coarticulated pronunciation say simply [y]. Thus a sound change would have occurred – or, if one wishes, a mini-sound change, since it would so far only involve one speaker-hearer. However, if this person's speech is copied by other speakers this mini-sound change could become a regular sound change, i.e. a characteristic of a well-defined speech community. (I assume without further argument that the initiation of

such sound changes is accomplished by the phonetic mechanism just described; their spread, however, is done by social means, e.g., borrowing, imitation, etc.) (Ohala 1981:184; italics ours).

The interchange between speakers and hearers is classified into three types: A. Correction; B. Hypo-correction; and C. Hyper-correction. A is Scenario 1, the normal error-free case involving 'reconstructive' rules. B is Scenario 2, the failure to apply such rules which can result in mini-sound changes. C is applying reconstructive rules where it is not called for. This can also lead to sound changes. The speaker aims at producing [yt] and does so without distortions due to coarticulation or other phonetic processes. What the listener receives is the acoustic pattern of [yt], but, in a particular situation, the listener expects back vowels to undergo fronting before consonants with anterior tongue bodies such as [t]. That expectation would interact with the stimulus and lead to the incorrect conclusion that the speaker had produced an /ut/ modified by coarticulation into [yt]. When the form is used by the listener, [y] could be replaced by [u]. If so happens, a mini-change will occur.

A key point in Ohala's account is "when this listener turns speaker" (cf above quote). What happens then? Let us try to picture some of the details of that step.

For one thing, it is evident that, if a listener should decide to pronounce a word that she has misperceived, she could not do so unless she knew what that word was.[3] That appears to be a correct observation, for consider a case where the meaning and non-phonetic information are also lost. There would be a problem: To which lexical item should the misperceived phonetic pattern be assigned?

From such considerations we realize that the type of misperception described in Scenario 2 has to be a *partial* misperception in the sense that lexical access occurs successfully despite a phonetic error. If that were not so, the listener-turned-speaker would not know which word to pronounce and no mini-sound change would be produced.

We also realize that, if misperceptions are partial, errors will be corrected immediately on recognition. In everyday conversational terms: "She meant to say [x], but it came out as [x*]". Consequently, on this analysis, the tight link that Scenario 2 draws between misperception and sound change is considerably loosened. Is there a way of revising the scenario while still preserving its basic insight?

We first follow Ohala in his analysis of perceptual errors which essentially says that speech percepts tend to have a bimodal distribution: There is (i) a context- and knowledge-dependent mode which is content-oriented and (more) focused on *what* is being said, and there is (ii) a signal-oriented mode (more) focused on *how* something is said. In the first mode, the normal mode, the listener is aided by signal context and his native language knowledge. (Such a process is also assumed by H&H, cf above). In the second mode, reference to signal context and language knowledge is for some reason inhibited.

Then, also basically following Ohala, we retain the idea that it is this incidental 'how'-mode of perception that provides the breeding-ground for new pronunciations (mini-sound changes) and that puts speakers-listeners in a state of readiness for phonetic and phonological innovations. However, unlike Ohala, we shall assume that there are several ways in which listeners can gain access to these unnormalized 'how' representations. Therefore we shall play down the role of misperceptions as direct triggers of sound change.

As a first step of the argument, let us examine some additional evidence for the dissociation between a signal-oriented and a content-oriented mode of perception. A good example is found in a study by Williams (1986, 1987). It involved determining the boundary between /ʊ/ and /ɪ/ using three synthetic continua matched for midpoint formant frequencies: /#ʊ#/-/#ɪ#/, /wʊw/-/wɪw/ and /jʊj/-/jɪj/. In previous work (Lindblom and Studdert-Kennedy 1967/1991), it had been found that the /ʊ/-/ɪ/ boundary varied with consonant context. For instance, in the /w_w/ environment, it was much lower than for the /#ʊ#/-/#ɪ#/ condition which makes sense in view of the fact that words like "will" are often reduced and thus show formant undershoot (Moon 1991).

Williams used two procedures: standard formant synthesis and the method of substituting three pure tones for the first three formants (Remez, Rubin, Pisoni and Carrell 1981). Two continua were produced, one with steady-state (SS) formant patterns ranging from /ʊ/ to /ɪ/, the other embedded in /w_w/ frames with the same vowels midpoints as for the SS stimuli but with time-varying (TV) formant transitions. Experimental tasks were blocked. Listeners first heard the SS and TV pure-tone patterns. They were asked to judge them in terms of relative pitch height, as 'low' or 'high' (psychoacoustic task). Then they had to decide whether a stimulus randomly drawn from a continuum was more like the /ʊ/, or the /ɪ/ endpoint (phonetic task). For the three-tone sine-wave stimuli, it was found that the boundary between the pitch categories was identical for all conditions. However, when the same sine-wave stimuli were judged with respect to vowel identity, the TV boundary was significantly lower than that of the SS. That was the result for the voice-excited formant stimuli for which only phonetic judgements were made. It is also in perfect agreement with the Lindblom and Studdert-Kennedy (1967/1991) findings.

From his data, Williams was able to conclude that a change in the instructions caused subjects to use either a pitch criterion or a phonetic 'frame of reference' and that the position of the /#ʊ#/-/#ɪ#/ boundary depended significantly on what those instruction were. Consequently, he was able to rule out that the boundary shifts were 'merely' a peripheral auditory effect. For the purpose of the present discussion, these results demonstrate the two modes of perception just described. The psychoacoustic pitch judgements resulted in percepts that directly paralleled the structure of the acoustic stimuli. The phonetic task, on the other hand, made the listeners focus on the stimuli's similarity to speech which presumably brought into play their familiarity with coarticulation, formant undershoot and other kinds of systematic variability found in natural speech.

As a second step of the argument, let us look some of the reasons why we should want to weaken the connection between misperceptions and sound change.

We begin with a comment on auditory phonetic analysis. What phoneticians do when they make narrow transcriptions of speech is basically to disengage their 'what'-mode and activate their 'how'-mode of perception. Since, to segment and judge phonetic values correctly, even the most skilled phoneticians must know the language they are analyzing and transcribing, the 'what'-mode may never be completely eliminated. Nevertheless, the ability of this group of listeners exemplifies one aspect of the distinction that we have introduced and shows that the 'how' can be accessed without misperception.

Normal listening is different from auditory phonetic analysis, but there are clearly numerous everyday instances where, consciously or subconsciously, the listener can register the 'how' along with the 'what'.

Of relevance here is the distinction between 'propositional' and 'automatic' speaking styles proposed by Bates (Fig. 7.2 of Bates (1979)). Examples of the latter category are nursery rhymes, poems, memorized lines and lists, overlearned phrases, idioms, greetings, formulas and clichés, exclamations, expletives and curses. These are utterance types that are low in 'meaning content'. In H&H terms, the speakers do not have to worry too much about being intelligible. Nevertheless, they may, or they may not, choose to adopt a very casual pronunciation (as a strict intelligibility-based version of the H&H argument would predict). We propose that these situations provide listeners with an opportunity to tune in to the phonetic shape of utterances.

That they in fact do is illustrated by the following observation that shows how a variant with emotional coloring can give rise to a new pronunciation by a mechanism not unlike the one envisioned in Scenario 2. BL:s daughter spent some time in the United States during her early





years. From one of her bedtime stories she had picked up the word 'sneaky'. When, back in Sweden, her family bought a dog, she named it Sneaky. She would call it ['ni:ki], or with strong affection, [ˈnœ̃:kə̃]. This variant, which was an 'emotive transform' produced by strong rounding and protrusion of the lips and by nasalizing heavily, later gave rise to [$^{x}$nøk:ɛ], which won general acceptance as an alternate name among family members and friends. In this development there was no new 'phoneme'. However, there was a sound analysis and a mechanism similar to that of Scenario 2. (Note that the [ˈ...] and the [$^{x}$...] symbols represent the grave accent I and the acute accent II of Swedish words. The [ø] is the short variant of the Swedish /ʉ/ phoneme).

Schematically:

| | | | | |
|---|---|---|---|---|
| | [ˈnœ̃:kə̃]. | - affection (normalization) | —> | [ˈ(s)ni:ki] |
| (4) | | | | |
| | [ˈnœ̃:kə̃]. | (no normalization) | —> | [$^{x}$nøk:ɛ] (lex. 'quantization') |

The new form [$^{x}$nøk:ɛ] makes sense, if we assume that it was incorporated through a process of lexical 'quantization', that is on the basis of its resemblance to the phonetic values and phonotactic sequencing of already existing discrete segments. Note that Swedish does not have phonemically nasalized vowels. (Interestingly, loan phonology offers similar instances of how new foreign phonetic materials are brought into line with the native phonology (Maddieson 1986)). We also see the similarity criterion at work in that the word accent changed and that the syllabic structure went from [ˈV:C] to [ˈVC:], both in conformity with numerous disyllabic Swedish names (Hasse, Acke, Britta, Ulla etc ...).

Other everyday examples come easily to mind.[4] Imagine a mother reproaching her daughter. The younger brother: *You sound funny when you yell at Samantha*, and, imitating his mother's loud and emotionally intense speech: [ˈsæ ˈmæn ˈθæ]. This and similar remarks would not be made unless the speaker had noticed a discrepancy between 'actual' and 'expected' pronunciations, in this case made salient by the sequence of near equal stresses and a lack of clearly reduced vowels in the 'angry' variant.

The above examples are anecdotal, but they nonetheless make a strong point: For listeners to gain access to the surface value of a particular pronunciation variant, they do not necessarily have to make a perceptual error. In fact, it appears reasonable to suggest that any transform of speech[5] merits the phonetician's attention as a potential source of a 'new pronunciation' – provided that: (i) there is a significant change in the phonetic pattern; (ii) that change calls for perceptual normalization; (iii) there is a possibility for that normalization not to be applied.

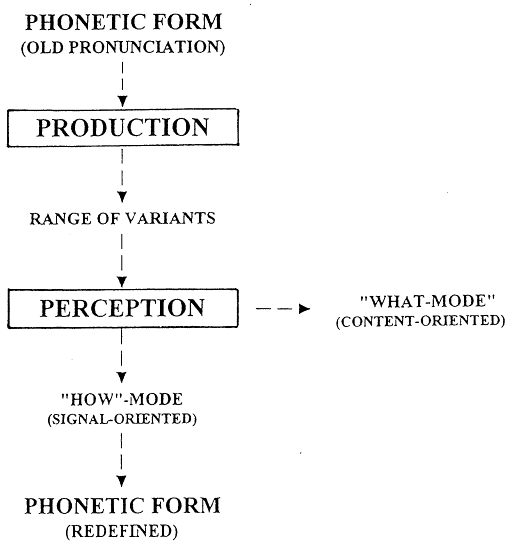


**Figure 3A.** Origins of phonetic variation

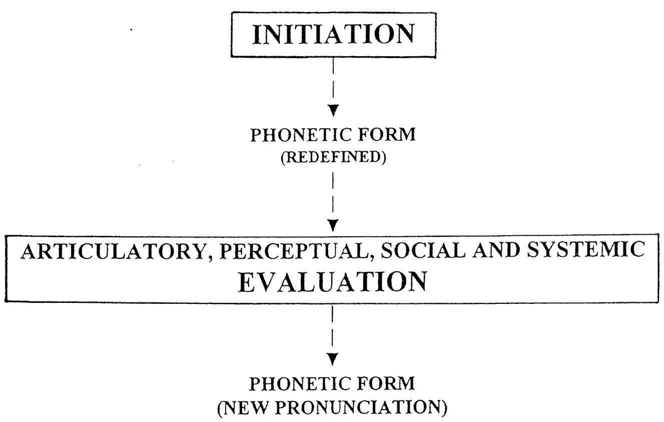


**Figure 3B.** Phonological selection





## *3. Nature of phonological selections*

### *3.1. Combining the frameworks*

In Figures 3A and 3B, the preceding discussion is summarized as we follow the hypothetical trail of a single lexical item on its route from an old to a new pronunciation.[6] Figure 3A illustrates the idea that, as it gets used by the members of the speech community under a great variety of conditions, this form is represented by a large range of phonetic variants. That variation, we assume, results from listener-dependent adaptations which distribute phonetic shapes along the H&H and other dimensions.[7] During perceptual processing, listeners are normally unaware of that variability because the signal interacts with the listener's stored knowledge, and, as a rule, that has the effect of reversing the contextual transforms applied by the speaker. Above we referred to that type of perceptual processing as the 'what'-mode (= Ohala's Scenario 1).

But occasionally, modulation by signal-independent information is somehow inhibited (= Ohala's Scenario 2), or becomes superfluous (because intelligibility demands have already been redundantly satisfied, or are of secondary importance for social or speaker-related physiological or cognitive reasons, cf above). That is when phonetic variants have the opportunity of presenting themselves to listeners in close to 'raw' acoustic/auditory form. In Fig. 3A, this less frequent, but still important, form of processing is represented by an arrow pointing at the 'how'-mode.

Drawing the diagram in this manner, we want to imply that native speakers store in their phonetic memories, not only (lexical) motor-perceptual information on the 'canonical' ('should-be') pronunciation of each item, but also (relatively) unprocessed phonetic patterns captured in sporadic moments of acoustic/auditory 'truth'. In conformity with Ohala's thinking, we propose that it is from this pool that 'new pronunciations', or mini-sound changes, are selected.

Cross-linguistic observations of sound change lends support to combining Ohala's model with the H&H perspective.

### *3.2. Typology of sound changes*

Adopting what he calls a traditional classification (Donegan and Stampe 1979), Kiparsky (1988) distinguishes three main types of phonological processes: "Prosodic processes" affect speech timing and syllable structure, e.g. compensatory lengthenings, consonant geminations, epentheses. "Weakening processes" are exemplified by assimilations, vowel reductions, consonant deletions and lenitions. They are:

> [...] usually context-sensitive and favored in unstressed position, in the syllable coda, and in casual speech. These processes make things *easier to say, but* (in so far they reduce or neutralize contrast) *harder to understand*. (Kiparsky 1988:377); (our italics).

"Strengthening processes ('polarizations')", e.g. vowel shifts and consonant fortitions, have a complementary description. They are:

> [...] usually independent of segmental context and favored in stressed position, in the syllable onset, and in explicit speech. They make things *easier to understand*, and usually *demand extra articulatory effort* from the speaker. (Kiparsky 1988:378); (our italics).

This taxonomy appears entirely compatible with the present distinction between 'hypo' and 'hyper' forms. However, the latter were not derived from phonological data, but from general principles of motor and perceptual organization. Moreover, they describe 'on-line' phonetic properties of speech, whereas 'weakening' and 'strengthening' refer to phonological facts.

What are the steps that link the phonetic hypo-hyper with the phonological weak-strong? Figure 3A provides a way of envisioning those steps for a single lexical item. Let us assume that speakers produce phonetic variants distributed along an H&H dimension, and that, forming part of the listener's experience (by way of the 'how'-mode), this variation can be used to supply the raw materials for mini-sound changes. Further suppose, that the variants are all equally probable to be promoted as 'new' pronunciations and all equally probable to be selected by the speech community. With those assumptions (to be reconsidered in 3.5), a mechanism has been sketched capable, in principle, of statistically either 'weakening' or 'strengthening' old pronunciations.

### *3.3. Extending the model*

Figure 3A is essentially a restatement of Scenarios 1 and 2 of Ohala's model. In extending it, however, the present account differs from his at two important points: First, misperceptions are de-emphasized. Second, the speaker – not only the listener – plays an essential role in shaping the contents of the phonetic variations from which phonological selections are made. If that assumption is not made, the natural, explanatory link between phonological 'strengthening' and 'weakening' processes on the one hand, and phonetic 'hypo' and 'hyper' transforms on the other, would seem to be lost.





Basing mini-sound changes on misperceptions is attractive because it makes their occurrence accidental (non-teleological), and it answers the question how phonological selections come to be initiated. However, as pointed out above, hearing word x wrong and then pronouncing it as heard, presupposes knowing that it was word x that was misperceived. We are forced to conclude that, if listeners-turned-speakers have access to the lexical identity of the word, they must also have access to the correct pronunciation. Accordingly, they are not compelled to produce the mini-sound change, since they have the option of not revising the old pronunciation.

When learning new words, – either as an adult (a technical term, the name of a new acquaintance etc), or during speech development, where there are no 'old pronunciations' – there does seem to be an opportunity for change à la Scenario 2. However, limiting sound change to that possibility appears undesirable, since it excludes competent adults as instigators of sound change. In sociolinguistic accounts, that group has been assumed (Labov 1972, Kroch 1978) to play an important role for the innovation process.

Once the tight coupling between mini-sound changes and misperceptions is abandoned, the answer to how phonological selections are initiated, becomes less clear-cut. An alternative mechanism must be specified both for initiating new forms and for replacing old ones. We shall have little further to say on the initiation process beyond suggesting that speakers sometimes accidentally sample the pool of variants (as defined by the stages of Figure 3A) and aim at 'surface' values rather than 'underlying' representations. Admittedly, this step (labeled 'Initiation' in Figure 3B) needs to be specified more precisely.

As for a replacement mechanism, we hypothesize that it exists in the implicit evaluation that takes place among speech community members, and that is performed with respect to, among other things, the phonetic shape of an utterance. Therefore the model should be extended by letting the probability that a given speaker's phonetic 'mutation' (=redefinition) will be accepted by others be directly related to how well it fits the social, communicative, articulatory and lexico-systemic[6,13] criteria that speech community members tacitly apply to it when they encounter it as listeners, and when, and if, they try it out as speakers. In Figure 3B this stage is labeled "Articulatory, Perceptual, Social and Systemic Evaluation". As a rule, such evaluations are made automatically and without explicit awareness. Only occasionally are they expressed indirectly (*Say that again, please!*), or more overtly (*She just SOUNDS snobbish*).

For the present, we focus on the interaction of articulatory and perceptual factors. It appears clear that the argument developed earlier for the phonetic signal and the (ideal) speaker could, with equal justification, be made for phonology and the (ideal) speech community. If a new articulation makes production easier, its acceptance will depend on perceptual conditions such as its lexical "neighborhood structure" (Luce 1986), – in other words, on how confusable it has become. Similarly, if a pronunciation has been changed in the direction of a more contrastive, acoustically richer shape, its spread might be resisted for articulatory reasons. In other words, also the phonological selections (= the output of "Evaluation" in Figure 3B) ought to be structured, by and large, along the H&H dimension.

Extending the H&H hypothesis in this manner has the consequence of offering an explanation for the dynamic articulatory and perceptual interplay present also at the level of sound structure.

*3.4. Place assimilation in nasals*

Our first illustration of that interplay is the well-known, cross-linguistically general tendency for nasal consonants to adopt the place of articulation of a following stop consonant. Both traditionally and recently, this and other assimilations have been described in ways suggesting that they are articulatorily motivated. For instance, within the framework of 'feature geometry' which uses articulatory features (MacCarthy 1988), this process is handled by copying the place specification of the following stop onto the preceding nasal segment (feature spreading).

Ohala (1990), on the other hand, proposes a perceptual interpretation that he bases on the '$C_2$-dominance' effect, that is the experimental demonstration that, in $VC_1C_2V$-sequences, the place cues of $C_2$ are often strong enough to perceptually overrule those of $C_1$ (Repp 1978, Fujimura, Macchi and Streeter 1978, Schouten and Pols 1983). Place assimilation in nasals would thus tend to occur because of the perceptual weakness of nasals in $C_1$ positions. It is thus a development fully consistent with Ohala's misperception scenario.

Kohler (1990) favors a third account based on the observation that, in German casual speech, nasals and unreleased stops tend to undergo place assimilation, whereas, under comparable conditions, voiceless fricatives do not. He argues that the observed assimilatory patterns result from the motor system's opportunistic propensity to simplify articulations. Importantly, he adds that, whether they get incorporated into the phonology or not is constrained by their perceptual properties. The phonological consequence of this articulatory-perceptual balance is that, since German fricatives are more distinctive, the language has entrusted them with the task of carrying phonemic place distinctions in





positions where nasals and unreleased stops could not be so used reliably. As a result, final consonants are preserved in prefixes such as *Ausfahrt* and *Auffahrt* (and many other similar pairs), but assimilated in cases like "anbringen" [a*m*b-] and "angeben" [aŋg-]. Hura, Lindblom and Diehl (1992) report data from an experiment with American English listeners supportive of Kohler's suggestion. It shows that, in comparable positions, voiceless fricatives tend to be more distinctive than nasals and unreleased stops.

By what unsupervised mechanism do languages reach the 'wise' decision to invoke phonemic place contrasts only where they are sufficiently robust to be maintained perceptually? How do they arrive at sacrificing the place distinctions of perceptually weak consonants by letting them undergo assimilation?

Those adaptations are implemented, we propose, by the tacit and automatic trial-and-error process mentioned above: Utterances are continually being tested on-line, perceptually by the listener and articulatorily by the speaker. Necessary distinctions tend to remain or to be enhanced, and superfluous articulations fall into disuse (Passy 1890:227, Martinet 1955).

*3.5. Role of frequency*

The H&H hypothesis implies that words that are used frequently, and are therefore relatively predictable to the listener, should over time exhibit a distribution of realizations biased towards 'hypo' forms. Are high-frequency words particularly likely to undergo sound change and to show the effects of phonological 'weakening'?

That is an old question that still attracts some attention (Kiparsky 1988). Zipf (1935) observed that there seems to be an inverse relationship between frequency and word length: the greater the frequency, the shorter the word. In agreement with Zipf's observation, Mańczak (1968, 1978) offers many cases of change in German and Russian that he analyzes as conditioned by frequency. He notes (1968:289) that:

> les éléments linguistiques dont la fréquence d'emploi augmente diminuent, en général, de leur volume,

and suggests that many exceptions from neo-grammarian regular sound change should be accounted for in terms of frequency, "the third essential factor of lingustic evolution, in addition to regular sound change and analogical development" (1978:309).

Hooper (1976) investigated patterns of schwa deletion in American English. Words ending in [ˈVC$_0$ᵊri] were studied. Their frequencies were obtained from Kučera and Francis (1967). Eight native speakers of American English (linguistics graduate students) were asked whether they deleted the schwa of the penult usually, sometimes or rarely. The responses indicated that deletions occurred more often in words with higher frequency.

If frequently used words are likely to occur more often in reduced form than less frequent words, they might have fewer carefully articulated counterparts with which to be juxtaposed than less frequently used words. Accordingly, the phonetic properties of such high-frequency variants might be relexified as new canonical form, thus producing a change in a frequently used word while not affecting lower-frequency words of similar structure. In a preliminary study of the phonetic bases of that reasoning, Guion (1994) elicited speech samples from five native American-English speakers reading sentences with tokens of homophonous word pairs of different frequencies (need-knead, night-knight, way-whey, time-thyme). Formant and duration measurements indicated a trend towards a more reduced pronunciation of the more frequent word.

### *3.6. The "Size Principle"*

A related set of facts comes from the study of sound inventories. Maddieson (1984) notes that the typical five-vowel system is [i e a o u], not [i ẽ a̤ o̰ uˤ]. In the latter set, perceptual distances are augmented by the presence of secondary articulations and phonation types. If vowel systems had evolved in response to a demand for "maximum perceptual contrast" (Liljencrants and Lindblom 1972), vowels with secondary elaborations should be favored and would be expected to be more frequent than they are.

A clue to explaining why predicting vowels from maximum intervocalic contrast fails, is offered by Lindblom and Maddieson (1988) who used the UPSID database (Maddieson 1984) to examine the phonetic contents of consonant systems as a function of inventory size. They found that segments with one or several secondary modifications tended to be absent from the smallest systems[8], but to grow more numerous in proportion to the size of the inventory (the 'Size Principle').

Before presenting an interpretation of those findings let us turn to some recent work by Willerman (1994) which adds further information on the relationship between phonetic content and inventory size.

### *3.7. The phonetics of pronouns*

In many languages, grammatical morphemes (as opposed to semantic morphemes) seem to "lack phonetic bulk" (Bolinger and Sears





1981:58). Against the background of such observations, a typological study of the phonetics of pronouns was undertaken (Willerman 1994).

A first question was whether consonants in pronouns can be seen as drawn at random from the overall inventory, or whether they tend to form systematically biased subsets. Thirty-two languages having twenty-six consonants or more were selected. Consonant segments were described in terms of phonetic dimensions similar to those of Maddieson (1984). Pronouns and overall inventories were pooled separately and comparisons were made of how the dimensions were used in the two sets. Relative frequencies of occurrence were calculated for primary and secondary places of articulation, for primary and secondary manner mechanisms and for source dimensions.

It was found that pronouns made significantly less use of the palato-alveolar, retroflex, uvular and pharyngeal places. They invoked secondary articulations more seldom. They had fewer laterals, affricates, trills, clicks, ejectives and aspirated segments. They favored segments with spontaneous (as opposed to compensatory) voicing or voicelessness (Chomsky and Halle 1968). They tended to avoid laryngealization and phonological length.

These findings provide strong evidence against the assumption that pronoun consonants are merely a random subset of the total set of consonants used by a given language.

A second question was whether the pronouns of the corpus had significantly more articulatorily 'simple' segments than what would be expected from the contents of the overall inventories. To address this issue, an independently motivated definition of 'articulatory complexity' was worked out. The metric involves making a binary classification of an arbitrary segment with respect to how demanding it is to produce according to two criteria: bio-mechanical cost and spatio-temporal control. The bio-mechanical measure penalizes extreme articulatory displacements and extreme articulatory rates. It becomes applicable in the treatment of place mechanisms, and it divides segments into those that deviate significantly from neutral[9] and those that do not.[10]

Spatio-temporal complexity depends on whether a secondary process is present[11], and on degree of temporal or spatial precision[12]. Final scores for individual segments are derived by summing all the 'penalties' incurred and are used to assign one of three degrees of complexity (Basic, Elaborated, or Complex) to any given segment.

To give an illustrative example of the analyses performed, we shall examine !Xũ which has 95 consonants, five of which are used in its system of personal pronouns (Snyman 1969). In the inventory as a whole, the proportions of consonants in the three complexity classes are 11.6% (Basic), 26.3% (Elaborated) and 62.1% (Complex). If the five consonants of the pronouns were to show the same proportions, there ought to be .58 Basic, 1.32 Elaborated and 3.11 Complex consonants. Actual numbers are 3, 2 and 0 respectively. A similar trend was present in the sample as a whole. A chi-square calculated for the entire of 32 languages indicated that the discrepancy between observed and expected frequencies of Basic, Elaborated and Complex segments was highly significant (chi-square = 191.66; degrees of freedom = 64; $p < .001$).

On the basis of her findings, Willerman was able to conclude that there is a strong tendency for pronouns to favor consonants with 'elementary' articulations.

Since pronoun systems are comparable in size to the small inventories of UPSID, and since pronouns and consonant inventories make similar use of phonetic dimensions, the two data sets can be given a common interpretation.

For purposes of distinctiveness, a consononant in a small system must signal that it is not a, b, c or d, whereas a consonant in a large system needs to signal that it is not a, b, c, d, e, f, g, h, j or k. In other words, the average amount of information carried by each consonant in a small system is smaller than in a large system. Therefore smaller paradigms entail less competition among units and make them, in relative terms, more predictable. Demands on perceptual distinctiveness could in principle be met by random selection of phonetic values from the set of universally available dimensions. But, as we have seen, languages prefer to exhaust its choices among the 'less complex' segments before it recruits 'more complex' possibilities for additional distinctions. Metaphorically, it is as if languages acquire their phonetic paradigms by scanning the phonetic space from the end of low-cost articulations and by gradually moving towards more costly solutions if need be.

Again, the question arises about the origin of these regular patterns. By what mechanisms do they arise? The answer we propose is that their emergence is a three-step process: (1) At the level of individual speakers, phonetic variation is shaped by a balance between intelligibility and articulatory energetics. Hence it is distributed along the H&H (and other) dimension(s); (2) Also at this level, mini-sound changes are produced by fortuitous reinterpretation of that variation through a mechanisms of "decontextualized" perception; (3) At the level of the speech community, the viability of new pronunciations is tested by a number of behaviorally based criteria, notably intelligibility and articulatory energetics, which, in analogy with the first step, imposes a Hyper-Hypo, or a Strong-Weak, structuring on patterns that successfully meet the selection criteria.

Note that 'hyper' and 'hypo' forms are assumed to arise twice: At the signal level (step 1), and at the phonological level (step 3).





## *4. Discussion*

### *4.1. Phonetic constraints and the teleology issue*

The present account has largely been developed on the basis of Ohala's model of sound change. However, certain extensions and modifications have been introduced. For instance, were we to paraphrase Ohala's "listener as a source of sound change", we would replace 'listener' by 'language user', since, both for a priori and empirical reasons, we must assume that, not only the listener, but also the speaker has a marked influence on phonological change.

Moreover, while Ohala's listener plays his part once, our speaker-listener makes his contribution twice, first as a source of change, then as a member of the collective filter that selects or rejects specific innovations. Since, in most situations, staying intelligible is balanced against simplifying articulation, lexical forms will appear in many physical shapes. We have proposed that this dynamic interaction is manifest both at the signal level (source), and at the level of sound structure (filter).

This source-filter model establishes a natural link between the hyper-hypo phenomena of on-line speech and the strengthening-weakening processes of phonology (section 3.2). It also suggests (the beginnings of) a principled search for behaviorally based, as opposed to 'formal' (non-behavioral), explanations of phonological facts. We have presented a few cases (sections 3.4-3.6) where such explanations seem feasible. There is no reason why they should not be successful also for other data sets.

In fact, one reason for expecting the present model to be even more successful than anything proposed so far, has to do with how it deals with the central, but as yet still unanswered, questions: Are distinctive features (primarily) articulatory (cf Chomsky and Halle 1968)? Or are they (primarily) perceptual (cf Jakobson, Fant and Halle 1952/1968)? The present answer is that they are neither one nor the other, but always inextricably both. This expectation derives from the inseparability of the perceptual-informational and articulatory-energetic constraints assumed to be implicitly at work both at the phonetic signal level and at the level of phonologization (cf Hura et al 1992 for further discussion of this point).

A significant role is accorded to the speaker. Reference to 'articulatory energetics' has been frequently made. However, despite the common-sense appeal and traditional popularity of 'articulatory effort', leading phoneticians (Ladefoged 1990, Ohala 1990) have, in the past, remained singularly unconvinced by such notions. In brief, their criticism is that 'articulatory ease' is "language-specific" (Ladefoged 1990), "immeasurable" (Ladefoged 1990, Ohala 1990) and "teleological" (Ohala 1990).

'Articulatory effort' as perceived by native speakers may indeed be language-dependent (cf Menn (1980, 1983) who presents speech development data that support a language-dependent interpretation of this condition). However, it is not controversial to assume that human speakers all have similar, language-independent vocal tracts and that similar constraints govern those vocal tracts. Hence, before language-specific factors can be usefully pinned down, it seems necessary to begin by trying to identify what languages have in common, namely the 'language-innocent' aspects of 'articulatory ease'. Ladefoged makes a relevant observation, but, as an objection to launching a research program on 'articulatory ease', it is without force.

We should point out that there are fields, e.g. locomotor energetics (Alexander 1988), in which movement-induced energy expenditure is both routinely and successfully measured (Hoyt and Taylor 1981, Williams, Friedl, Fong, Yamada, Sedivy and Haun 1992), and in which the relevance of 'effort' to movement control has been clearly demonstrated. What is the implication of such work for phonetics? It is evident that, if we suppose that 'energetics' does not apply to speech, we would, from a biological point of view, be making a very far-fetched assumption.

Moreover, that finding *direct* ways of measuring articulatory effort is difficult, does not mean that useful *indirect* estimates could not be made from general principles of motor behavior. For example, such an indirect approach is taken in the work by Nelson (1983) and Nelson, Perkell and Westbury (1984) who present an insightful quantification of articulatory effort based on bio-mechanics. A first step in bringing biomechanical and other universal motoric principles into the realm of phonology is the "articulatory complexity metric" of Willerman (1994). Such work demonstrates not only that quantifying 'articulatory ease' is feasible. It points to a productive avenue for seeking quantitative and independently motivated theories of 'markedness'.

On the issue of teleology we agree with Ohala that sound change is accidental. At no level of language use, social or individual, is there an *explicit* specification of a 'goal' towards which a language system is driven during its historical development. In that sense, sound change must be seen as 'non-teleological'. On the other hand, we do not subscribe to the idea that invoking 'articulatory ease' to explain sound change must necessarily be 'teleological'. When speakers introduce an assimilation into their phonology, they do not do so 'therapeutically'. They do not respond to a 'recognized need' to make the sequences undergoing assimilation easier to say. Assimilations 'just happen'. However, it needs to be recognized that, once the new patterns of articulatory short-cuts have been established, bio-mechanical analysis can be applied to show that they do indeed involve reduced articulatory effort. The fact is that, by virtue of their definition, assimilations *are* bio-mechanically easier to produce. Therefore, we should expect that property, in an implicit way, to bias speakers in favor of phonologizing them. Clearly, such a use of 'articulatory ease' does not postulate a 'purpose' behind the assimilatory processes, and, accordingly, is not teleological in Ohala's sense of the term (cf Willerman 1994 for further discussion of this point).





### *4.2. Actual and possible socio-phonetic change*

Is sound change adaptive? In his commentary on a previous discussion of that question (Lindblom 1988), Ohala says no.

> To come back to the comparison with natural selection, then, I have suggested that the evolution of pronunciation, sound change, is similar in that it starts with a kind of natural variation, but is dissimilar in that it makes no assumption that there is substantial ecological competition between pronunciation norms or that most variants are any more adaptive than others.
> [...] Physical (including physiological) principles constrain what variants come into being [...], but after those variants exist there is little evidence for subsequent optimization through competition of languages' sound systems.
> [...] I think most variations occur due to errors in the transmission of pronunciation norms – due to listener mistakes – and thus resemble scribes' errors in copying manuscripts. Like scribal errors, *there is no adaptive value to such variations.* (Ohala 1988:179; our italics).

Two further points on Ohala's position. First, empirically, it is not as easy as Ohala implies to come up with a clear-cut answer, either positive or negative, whether a certain phenomenon is adaptive or not (further discussion below). The facts are complex and have to be collected and synthesized from a number of specialized domains (socio-linguistics, phonetics, historical linguistics etc). Such an issue requires a lot of demanding research and cannot be prejudged.

Second, from an a priori, theoretical viewpoint, it would be remarkable if sound change were radically different from other processes of historical change, both biological and cultural, whose description appears to necessitate according, not an exclusive, but an important role to adaptive mechanisms (further discussion below).

What would be required to substantiate or refute the assumption that sound change is adaptive? What determines the 'survival' of a new pronunciation? What are the characteristics of the 'niche' conquered by a novel sound pattern? For discussion purposes, we here offer the following schematization as a preliminary point of departure.

Suppose that an objective method were available for comparing old and new pronunciations in terms of: (i) their *social value*, (ii) their *articulatory complexity* and (iii) their *perceptual distinctiveness*[13]. Further assume that each of those dimensions takes one of three values: positive, negative or neutral. From those decisions, we obtain an ad hoc, but still useful method for classifying the implicit judgements that language users make on a moment-to-moment basis of old and new forms in everyday communication. It would imply that there would be 27 possible outcomes of the evaluations as specified by a $3^3$-matrix. Which combinations of parameter values lead to sound changes? Are all cells of that matrix equally likely to be linked to sound changes?

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

Do the collective evaluations by the speech community also assess the articulatory and perceptual value of new forms? If they do not, by what other mechanisms could articulatory 'weakenings' and perceptual 'strengthenings' arise (section 3.2)? What other processes would automatically guarantee that assimilations would remain 'functional', that is be limited to articulatory simplifications that do not jeopardize systemic distinctiveness (3.4)? What similarly unsupervised course of events would create segment inventories with size-dependent phonetic content in which articulatory costs are carefully balanced against the extravagance of perceptual benefits (3.6 and 3.7)?

Whereas Ohala maintains that his misperception mechanism does offer an alternative, we would argue that that will not be enough: The cross-linguistic systematicities strongly suggest that phonetic forms are put to both articulatory and perceptual tests by speakers-listeners and that, in a significant way, such evaluations determine the phonetic shape of sound patterns.

The $3^3$-matrix offers a way of formulating that interpretation somewhat more precisely: Some parameter combinations (notably sets of positive and non-negative values) are more likely to be associated with sound change than others (sets of negative and non-positive scores). Clearly, that hypothesis, the *socio-phonetic adaptation hypothesis*, is testable and can be further developed with respect to the precise





specification of the relevant social, articulatory, perceptual and other variables. In parallel, its competitor, Ohala's 'scribal error' or *misperception scenario*, should also be considered. It claims that phonetics should be invoked to account for how sound changes originate, but not for how they get selected: A scenario with phonetically motivated 'mutations' but with 'selections' that are phonetically neutral.

### *5. Concluding remarks*

We shall return to our point of departure to make an important caveat.

For the biologist, is judging the adaptive value of a given phenomenon a relatively unproblematic matter? No, not at all. One difficulty is that, in many cases, the 'unit of selection' is hard to define. Here an illustrative example of that problem.

Why do adult humans have chins, whereas human infants and chimpanzees do not? What is the adaptive significance of the 'human chin'? The answer that biologists have eventually come up with is that it has no adaptive value at all (Lewontin 1978). It is simply a consequence of the fact that two growth fields, the dentary (that includes the jaw bone) and the alveolar (with the teeth), do not develop at identical rates in humans. The chin is a fortuitous by-product, – an epiphenomenal "spandrel" in the sense of Gould and Lewontin (1979) – for which an adaptive explanation would be totally inappropriate.[14]

The history of the human chin, and many other developments, teach us that *not all evolutionary change is adaptive*. As we strongly argue in favor of taking the adaptationist program seriously in historical phonetics, it is very important to stress that fact, since, in so doing, we run a risk of promoting 'naive adaptationism' (Gould & Lewontin 1979).

Proposing adaptationist accounts of linguistic phenomena, we should not expect the situation to be any better than for the biologist. Why should phoneticians nonetheless be interested in using an approach that is associated with so many difficulties? Why import the problems of other fields?

We offer two major reasons. First, the evidence supports the adaptive model of sound change. Second, students of sound change currently lack the strong reasons needed to say with impunity that they have nothing to learn from evolutionary theory. Dismissing the general principles of change that have been uncovered and tested during more than a century's study presupposes an understanding of why they fail for phonological change. It is not controversial to claim that, in present-day linguistics, there is as yet no such general understanding.

Among biologists, even the most sceptical recognize adaptation as a valid mechanism (Lewontin 1978). There is a massive body of facts that can be successfully understood in terms of adaptation. And, for methodological reasons,

> [...] biologists are forced to the extreme adaptationist program because the alternatives [e. g. allometry, pleiotropy, random gene fixations and indirect selection],[15] although they are undoubtedly operative in many cases, are untestable in particular cases.

Furthermore,

> [...] to abandon the notion of adaptation entirely, to simply observe historical change [...], with no functional explanation, would be to throw out the baby with the bathwater. (Lewontin 1978:230).

Is sound change a very special (non-adaptive) case of biological and cultural evolution? The burden of proof would seem to be on those who are inclined to say that it is.

## *Addresses of the Authors*

Björn Lindblom: Dept. of Linguistics, Stockholm University, S-106 91 Stockholm, Sweden

Susan Guion: Dept. of Linguistics, University of Texas, 78712 Austin, Texas, USA

Susan Hura: Dept. of Audiology & Speech Sciences, 1353 Heavilon Hall, Purdue University, West Lafayette IN 47907-1353, USA

Seung-Jae Moon: Dept. of English, College of Liberal Arts, AJOU University, Suwon 442-749, Korea

Raquel Willerman: Dept. of Speech & Hearing Sciences, University of Washington, Seattle, Washington, USA

## *Notes*

* The authors are indebted to John Ohala whose seminal work and constructive criticism (cf. especially Ohala (1988)) provided much of the inspiration for the discussion leading to the present article. This work was supported by Projet Sciences Européen (ERB4002PL910339), a grant from the Advanced Research Program of the Texas Board of Coordination and grant No. BNS-9011894 from the National Science Foundation. The authors are grateful to René Carré, Randy Diehl, Carol Fowler, Kicki Hellman, Shinji Maeda and Michael Studdert-Kennedy for helpful comments on early versions of the paper.





1 See e.g. Dawkins (1983) and (1986, especially the "Biomorph" experiments).

2 The total set of data comprised four vowels, three word lengths, and five repetitions of each token spoken by five speakers in two styles (see also Moon & Lindblom 1994).

3 By "knowing the identity of the word", we mean that, although the listener hears it wrong, she is able to "infer" its meaning and non-phonetic properties from signal context and knowledge of her native language.

4 Picture an emcee introducing a famous performer at a concert. At the end of his introduction, he might say: *The Academy of Contemporary Jazz brings you ...* . Although the name of the artist would be known to almost everyone in the audience, an overarticulated and emphatic pronunciation would normally be used. Possibly, an absent-minded listener might fail to normalize for the emcee speaking style and subconsciously register the exaggerated phonetic variants at their face value.

5 Such as becoming emotional (physiological), talking to oneself (cognitive), adapting one's voice to room acoustics (communicative), speaking publicly (social), etc.

6 A striking characteristic of sound change is that it is generally not restricted to individual lexical items but tends to be generalized across large domains. This is sometimes captured by the phrase *phonemes change* which highlights the tendency for change to spread, in a lockstep manner, throughout the lexicon to all contextually relevant instances of a phoneme. It is used in contrast with "every word has its own history" (Bloomfield 1933) which refers "to sporadic cases in which change gives rise to stray, lexically idiosyncratic forms" (Kiparsky 1988). This significant aspect is linked to the role of 'systemic' constraints on sound change, but will not be discussed in the present context (for further comments see chapter 9 of Lindblom, MacNeilage and Studdert-Kennedy (under review)).

7 Although they are, we assume, major ones in many socio-linguistically important situations, staying intelligible and simplifying articulation are not the only factors shaping on-line speech patterns. That is made evident by, for instance, speech addressed to infants, or Baby Talk (Ferguson 1977). A primary task of this speaking style is to maintain contact with the child (the 'phatic' function of Jakobson (1960)) and to keep her/him happy ('emotive' function of Jakobson (1960)). To that end, BT tends to amplify prosodic aspects at the expense of vowel and consonant cues (Davis and Lindblom 1994) and to be rich in emotionally positive information (Fernald 1984).

8 For instance, Pawaian with ten consonants has [p t k m n r j w s h] rather than say [p' t' k' m̥ n̥ r̥ j w̥ sʰ ɦ].

9 E.g. retroflexes, uvulars and pharyngeals.

10 E.g. labials, dentals and alveolars.

11 E.g. labialization, palatalization etc.; an ejective, or implosive airstream.

12 Temporal precision as in pre- and post-aspirated, or pre- and post-nasalized stops. Spatial precision as for turbulence production in fricative constrictions as compared with constrictions in vowels and stops.

13 Conceivably, several other dimensions should also be considered, e.g. *systemic compatibility* (degree of alignment with existing lexico-phonological structures).

14 This conclusion does not invalidate adaptive explanations. It simply changes the task into finding one for the different growth rates instead.

15 Allometry = different growth rates in different parts of an organism. Pleiotropy = change in a gene can have many different physiological and developmental effects.

## References

ALEXANDER R. McN. (1988), *Elastic mechanisms in animal movement*, Cambridge, Cambridge University Press.

BATES E. (1979), *The emergence of symbols*, Academic Press, New York.

BERNSTEIN N. (1967), *The co-ordination and regulation of movements*, Oxford,Pergamon.

BLOOMFIELD L. (1933), *Language*, New York, Holt.

BOLINGER D. & SEARS D. (1981), *Aspects of language*, Harcourt, Brace & Jovanovich, New York.

BOYD R. & RICHERSON P. (1985), *Culture and the evolutionary process*, Chicago, Chicago University Press.

BROWMAN C. & GOLDSTEIN L. (1986), "Towards an articulatory phonology", *Phonology Yearbook* 3:219-252.

BROWMAN C. & GOLDSTEIN L. (1990a), "Gestural specification using dynamically-defined articulatory structures", *Journal of Phonetics* 18:299-320.

BROWMAN C. & GOLDSTEIN L. (1990b), "Tiers in articulatory phonology, with some implications for casual speech", in BECKMAN M. & KINGSTON J. (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, Cambridge, Cambridge University Press.

CAVALLI-SFORZA L.L. & FELDMAN M.W. (1981), *Cultural transmission and evolution*, Princeton, Princeton University Press.

CHEN F.R. (1980), *Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level*, Master's Thesis, MIT, Cambridge, Mass.

CHEN F.R., ZUE V.W., PICHENY M.A., DURLACH N.I. & BRAIDA L.D. (1983), "Speaking clearly: Acoustic characteristics and intelligibility of stop con-sonants", 1-8 in *Working Papers II*, Speech Communication Group, MIT.

CHOMSKY N. & HALLE M. (1968), *The sound pattern of English*, New York, Harper & Row.

CLARK J.E., LUBKER J.F. & HUNNICUTT S. (1987), "Some preliminary evidence for phonetic adjustment strategies in communication difficulty", in STEELE & THREADGOLD (eds.), *Language topics: Essays in honour of Michael Halliday*, Amsterdam, Benjamins.

DAWKINS R. (1983), "Universal Darwinism", in BENDALL D.S. (ed.), *Evolution from molecules to men*, Cambridge, Cambridge University Press:403-425.

DAWKINS R. (1986), *The blind watchmaker*, Burnt Mill, Harlow, England, Longman.

DONEGAN P. & STAMPE D. (1979), "The study of natural phonology", in DINNSEN D. (ed.), *Current approaches to phonological theory*, Bloomington, Indiana University Press.

EEFTING W. (1991), "The effect of 'information value' and 'accentuation' on the duration of Dutch words, syllables and segments", *J. Acoust. Soc. Am.* 89:412-424.

EEFTING W. (1992), "The effect of accentuation and word duration on the naturalness of speech", *J. Acoust. Soc. Am.* 91:411-420.

FERGUSON C.A. (1977), "Baby talk as a simplified register", in SNOW C.E. & Ferguson C.A. (eds.) (1977), *Talking to children: Language input and acquisition*, Cambridge, Cambridge University Press:219-236.





FERNALD A. (1984), "The perceptual and affective salience of mothers' speech to infants", in FEAGANS L., GARVEY C. & GOLINKOFF R. (eds.), *The origins and growth of communication*, New Brunswick, Ablex:5-29.

FOWLER C.A. & HOUSUM J. (1987), "Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction", *J. Mem. Lang.* 26:489-504.

FUJIMURA O., MACCHI M.J. & STREETER L.A. (1978), "Perception of stop consonants with conflicting transitional cues: A cross-linguistic study", *Language and Speech* 21:337-346.

FUKSON O.I., BERKINBLIT A.G. & FELDMAN A.G. (1980), "The spinal frog takes into account the scheme of its body during the wiping reflex", *Science* 209:1261-1263.

GOULD S.J. & LEWONTIN R.C. (1979), "The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme", *Proceedings of the Royal Society of London* 205:281-288.

GRANIT R. (1977), *The purposive brain*, Cambridge, MA, MIT Press.

GUION S.G. (1994), "Word frequency effects among homynyms", manuscript, Department of Linguistics, University of Texas at Austin.

GUNDEL J.K., HEDBERG N. & ZACHARSKI R. (1993), "Cognitive status and the form of referring expressions in discourse", *Language* 69:274-307.

HELLMAN C. (1992), *Implicitness in discourse*, Doctoral dissertation, Stockholm University.

HOOPER J.B. (1976), "Word frequency in lexical diffusion and the source of morphological change", in CHRISTIE W. Jr. (ed.), *Current progress in historical linguistics*, Amsterdam, North Holland.

HOYT D. & TAYLOR C. (1981), "Gait and the energetics of locomotion in horses", *Nature* 292:239-240.

HURA S.L., LINDBLOM B. & DIEHL R. (1992), "On the role of perception in shaping phonological assimilation rules", *Language and Speech* 35:59-72.

JAKOBSON R. (1960), "Linguistics and poetics", in SEBEOK T. (ed.), *Style in language*, Cambridge, Mass, MIT Press: 350-377.

JAKOBSON R., FANT G. & HALLE M. (1952/69), *Preliminaries to speech analysis*, Cambridge, MA, MIT Press.

KIPARSKY P. (1988), "Phonological change", in NEWMEYER F.J. (ed.), *Linguistics: The Cambridge Survey, Vol I, Linguistic Theory: Foundations*, Cambridge, Cambridge University Press:363-415.

KOHLER K. (1990), "Segmental reduction in connected speech in German: Phonological facts and phonetic explanations", in HARDCASTLE W.J. & MARCHAL A. (eds.), *Speech production and speech modeling*, Kluwer, Dordrecht:69-92.

KROCH A. (1978), "Toward a theory of social dialect variation", in ALLEN H. & LINN M. (eds.), *Dialects and language variation*, New York, Academic Press.

KUČERA H. & FRANCIS N. (1967), *Computational analysis of present-day English*, Providence, Brown University Press.

LABOV W. (1972), *Sociolinguistic patterns*, Philadelphia, University of Pennsylvania Press.

LABOV W. (1981), "Resolving the neogrammarian controversy", *Language* 57:267-308.

LADEFOGED P. (1990), "Some reflections on the IPA", UCLA *Working Papers in Phonetics* 74:61-76.

LADEFOGED P. (1993), *A course in phonetics*, Hartcourt, Brace & Jovanovich, 3rd edition.

LASHLEY K.S. (1930/1960), "Basic neural mechanisms in behavior", *Psychological Review* 37:1-24. Also in BEACH F.A., HEBB D.O., MORGAN C.T. & NISSEN H.W. (eds), *The neuropsychology of Lashley*, McGraw-Hill, New York, 1960:191-209.

LASHLEY K.S. (1951), "The problem of serial order in behavior", in JEFFRESS L.A. (ed.), *Cerebral mechanisms in behavior*, New York, Wiley:112-146.

LEWONTIN R.C. (1978), "Adaptation", *Scientific American* 239(3):213-230.

LIEBERMAN P. (1963), "Some effects of semantic and grammatical context on the production and perception of speech", *Language and Speech* 6:172-187.

LILJENCRANTS J. & LINDBLOM B. (1972), "Numerical simulation of vowel quality systems: The role of perceptual contrast", *Language* 48:839-862.

LINDBLOM B. (1989), "Phonetic invariance and the adaptive nature of speech", in BOUMA H. & ELSEDOORN B. (eds.), *Working models of human perception*, London, Academic Press:139-173.

LINDBLOM B. (1990), "Explaining phonetic variation: A sketch of the H&H theory", in HARDCASTLE W.J. & MARCHAL A. (eds.), *Speech Production and Speech Modeling*, Kluwer, Dordrecht:403-439.

LINDBLOM B. (1993), "On phonetic variation", chapter 8 in LINDBLOM B., MACNEILAGE P. & STUDDERT-KENNEDY M. (under review).

LINDBLOM B. & STUDDERT-KENNEDY M. (1967/1991), "On the role of formant transitions in vowel recognition", *J. Acoust. Soc. Am.* 42:830-843. Also in MILLER J.L., KENT R.D. & ATAL B.S. (eds.), *Papers in Speech Communication: Speech Perception*, Acoust Soc Am, New York:501-514.

LINDBLOM B. & MADDIESON I. (1988), "Phonetic universals in consonant systems", in HYMAN L.M. & LI C.N. (eds.), *Language, speech and mind*, London and NewYork, Routledge:62-78.

LINDBLOM B., BROWNLEE S., DAVIS B. & MOON S.-J. (1992), "Speech transforms", *Speech Communication* 11:357-368.

LINDBLOM B., MACNEILAGE P. & STUDDERT-KENNEY M. (under review), *Evolution of spoken language*, Cambridge, Cambridge University Press.

LIVELY S.E., PISONI D.B., SUMMERS V.W. & BERNACKI R.H. (1993), "Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences," *J. Acoust. Soc. Am.* 93:2962-2973.

LUCE P.A. (1986), *Neighborhoods of words in the mental lexicon*, Doctoral dissertation, Department of Psychology, Indiana University.

MACCARTHY J.J. (1988), "Feature geometry and dependency: A review", *Phonetica* 43:84-108.

MADDIESON I. (1984), *Patterns of sound*, Cambridge University Press, Cambridge.

MADDIESON I. (1986), "Borrowed sounds", in FISHMAN J.A., TABOURET-KELLER A., CLYNE M., KRISHNAMURTI BH., ABDULAZIZ M. (eds.), *The Fergusonian impact*, Vol. 1, *From phonology to society*, Berlin, Mouton de Gruyter:1-16 .

MAŃCZAK W. (1968), "Le développement phonétique irrégulier du à la fréquence en Russe", *Lingua* 21:287-293.

MAŃCZAK W. (1978), "Irregular sound change due to frequency in German", in FISIAK J. (ed.), *Trends in linguistics: Studies and monographs*, vol. 4: *Recent developments in historical linguistics*, The Hague, Mouton.

MARTINET A. (1955), *Économie des changements linguistiques*, Bern, Francke.

MAYR E. (1978), "Evolution", *Scientific American* 239:39-47.

MAYR E. (1982), *The growth of biological thought*, Cambridge, Mass, Harvard University Press.

MENN L. (1980), "Phonological theory and child phonology", in YENI-KOMSHIAN G.H., KAVANAGH J.F. & FERGUSON C.A. (eds.), *Child phonology*, vol. 1, New York, Academic Press:23-41 .

MENN L. (1983), "Development of articulatory, phonetic and phonological capabilities", in BUTTERWORTH B. (ed.), *Language production*, vol. II, London, Academic Press.

MOON S.-J. (1991), *An acoustic and perceptual study of undershoot in clear and citation-form speech*, unpublished doctoral dissertation, University of Texas at Austin.

MOON S.-J. & LINDBLOM B. (1994), "Interaction between duration, context and speaking style in English stressed vowels", *J. Acoust. Soc. Am.* 96:40-55.

NELSON W. (1983), "Physical principles for economies of skilled movements", *Biol. Cybernetics* 46:135-147.

NELSON W., PERKELL J.S. & WESTBURY J.R. (1984), "Mandible movements during increasingly rapid articulations of single syllables: Preliminary observations", *J. Acoust. Soc. Am.* 75:945-951.

NOOTEBOOM S.G. (1985), "A functional view of prosodic timing in speech", in Michon J.A. & Jackson J.J. (eds.), *Time, mind and behavior*, Berlin, Springer:242-252.

NOOTEBOOM S.G. & EEFTING W. (1994), "Evidence for the adaptive nature of speech on the phrase level and below", *Phonetica* 51:92-98.

OHALA J.J. (1974), "Phonetic explanation in phonology", in BRUCK A., FOX R.A. & LAGALY M.W. (eds.), *Papers from the parasession on natural phonology*, Chicago, Chicago Linguistic Society:251-274.

OHALA J.J. (1978), "Phonological notations as models", in DRESSLER W.U. & MEID W. (eds.), *Proceedings of the 12th International Congress of Linguists*, Innsbruck, Innsbrucker Beiträge zur Sprachwissenschaft:811-816.

OHALA J.J. (1981), "The listener as a source of sound change", in MASEK C., HENDRICK R.A. & MILLER M.F. (eds.), *Papers from the parasession on language and behavior*, Chicago, Chicago Linguistic Society:178-203.

OHALA J.J. (1983), "The origin of sound patterns in vocal tract constraints", in MACNEILAGE P.F. (ed.), *The production of speech*, New York, Springer Verlag:189-216.

OHALA J.J. (1988), "Discussion of Björn Lindblom's 'Phonetic invariance and the adaptive nature of speech', in BOUMA H. & ELSEDOORN B. (eds.), *Working models of human perception*, London, Academic Press:175-183.

OHALA J.J. (1990), "The phonetics and phonology of aspects of assimilation", in BECKMAN M. & KINGSTON J. (eds), *Papers on Laboratory Phonology I: Between the grammar and the physics of speech*, Cambridge, Cambridge University Press:258-275.

OHALA J.J. (1993), "The phonetics of sound change", in JONES C. (ed.), *Historical Linguistics: Problems and perspectives*, London, Longman: 237-278.

PASSY P. (1890), *Étude sur les changements phonétiques*, Paris, Firmin-Didot.

PICHENY M.A., DURLACH N.I. & BRAIDA L.D. (1986), "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech", *Journal of Speech and Hearing Research*, 29:434-446.

PICKETT J.M. & POLLACK I. (1963), "Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt", *Language and Speech* 6:151-164.

POLLACK I. & PICKETT J.M. (1964), "Intelligibility of excerpts from fluent speech: Auditory vs structural context", *Journal of Verbal Learning and Verbal Behavior* 3:79-84.

REMEZ R.E., RUBIN P.E., PISONI D.B. & CARRELL T.D. (1981), "Speech perception without traditional speech cues", *Science* 212:947-950.

REPP B.H. (1978), "Perceptual integration and differentiation of spectral cues for intervocalic stop consonants", *Perception & Psychophysics* 24:471-484.

SCHOUTEN M.E.H. & POLS L.C.W. (1983), "Perception of plosive consonants", in VAN DEN BROECKE M., VAN HEUVEN V. & ZONNEVELD W. (eds.), *Sound structures: Studies for Antonie Cohen*, Dordrecht, Foris:227-243.

SHERRINGTON C.S. (1941), *Man on his nature*, London, MacMillan.

SNYMAN J.W. (1969), *An introduction to the !Xū language*, Balkema, Cape Town.

SUMMERS W.V., PISONI D.B., BERNACKI R.H., PEDLOW R.I. & STOKES M.A. (1988), "Effects of noise on speech production: Acoustic and perceptual analyses", *J. Acoust. Soc. Am.* 84:917-928.

WILLERMAN R. (1994), *The phonetics of pronouns*, Doctoral dissertation, Department of Linguistics, University of Texas at Austin.

WILLIAMS D. (1986), *Role of dynamic information in the perception of coarticulated vowels*, Ph. D. dissertation, University of Connecticut.

WILLIAMS D. (1987), "Judgements of coarticulated vowels are based on dynamic information", in *Perceiving-Acting Workshop Review*, vol. II, No. 2, University of Connecticut:16-20.

WILLIAMS T.M., FRIEDL W.A., FONG M.L., YAMADA R.M., SEDIVY P. & HAUN J.E. (1992), "Travel at low energetic cost by swimming and wave-riding bottle nose dolphins", *Nature* 355:821-823.

ZIPF G.K. (1935), *The psycho-biology of language*, Boston, Houghton Mifflin Co.