Discussion of Björn Lindblom's 'Phonetic Invariance and the Adaptive Nature of Speech'

John J. Ohala*

1 Introduction

I compliment Björn Lindblom on a masterful well-documented presentation that consitutes, I think, a conceptual breakthrough in the troublesome problem of invariance. He has boldly presented a position which challenges prevailing wisdom and the foundations of some well-known research paradigms. The existence of clearly-defined competing hypotheses will promote well-focussed research which will help to resolve this conflict.

His account of the origin of speech variability is explicitly a biological one and uses terms and concepts common in biology such as 'adaptation' and 'evolution'. This is not just a metaphor; he quite correctly treats speech as a biological activity and believes that its shape and behaviour must be governed by and therefore explainable by biological principles, i.e. by reference to physical and physiological constraints impinging on speech (see also Zipf, 1935).

I am convinced by much of his argument and will offer here a few brief remarks which supplement some of the points he made, in particular the connection between the biologist's and linguist's treatment of variation. In a few instances, I will suggest some qualifications to his scheme.

2 Parallels between biology and linguistics

Preliminaries

Oversimplifying a bit, the biologist must explain two types of variability: the 'plastic' behaviour and forms of organisms in the face of differing environmental pressures, e.g. temperature regulation by means of perspiration,

* Phonology Laboratory, Department of Linguistics, University of California, Berkeley, California 94720, USA.

WORKING MODELS OF HUMAN PERCEPTION
ISBN 0-12-238050-9

adjusting the thickness of fur or feathers – this can be called short-term adaptation – and the variable genetically-determined shapes and behaviours of organisms over time due to natural selection – this is long-term adaptation. The first may be characterized as purposeful or teleological in that it is controlled by the error between a pre-set goal and the detected current state of an organism. Variable behaviour of this type can be continuous as it is, for example, in the case of thermoregulation. Lindblom's adaptive variability is of this type, where, as the evidence he has reviewed shows, the variable behaviour is determined by the speaker's estimation of the decoding capabilities of the listener (see Nooteboom(1983) for further arguments and evidence towards this point). The second type of variation in biology, caused by natural selection, is not purposeful or teleological; it operates blindly. More adaptive organisms simply propagate better than less-well-adapted ones. Does this have an analogue in speech? Before committing ourselves on this point, let us take a closer look at the mechanisms Darwin proposes which underlie natural selection.

Darwinian evolution

Briefly, Darwin's theory explains the origin of species on the basis of three principles:

1. There exists 'natural' genetic variation in offspring vis-à-vis their parents. This is due to the random shuffling of the genes from both parents in the case of sexual reproduction and due to imperfect copying or other distortion (mutation) of the genetic code in sexual or asexual reproduction.
2. The demand for resources needed to sustain life (territory, food) exceeds the supply.
3. Those individuals whose variable genetic endowment ('genotype') makes them more competitive in garnering the available resources will reproduce in greater numbers than other less competitive.

From these axioms is derived the main theory of evolution that, in time, these adaptive variations will breed true and will no longer be random; different species will result. Genetically-maintained variations differ from the plastic adaptations not only in being non-teleological in origin but also by being discrete. An okapi is born with a definite limit on the extent to which it can stretch its neck; if it wants to stretch its neck further, it has to become its cousin, the giraffe!

Sound change

It would seem that the closest analogue in speech to the variation that gives rise to different species is *sound change*. Sound change, of course, is the change in pronunciation norms from one generation to the next, e.g. the differences between Chaucer's 14th century pronunciation [wɪf], 'wife', and the current English pronunciation of [waɪf]. When a sound change affects one regional or social speech community but not another, it can give rise to dialect differences, e.g. British and American English [streɪt] 'straight' vs Australian [strʌɪt] or, of course, to different languages. When it affects a given morpheme in one phonological context but not in another it can lead to alternation such as [lɔŋ], 'long', and [lɛŋkθ], 'length', or to complete merger and homophony as when 'less than' merges with 'lesson' in the example Lindblom cited (similar examples of homophony abound and constitute the stuff out of which jokes and puns are made). Sound change, as he mentioned, can also affect a language's segment inventory: if the sound change causes a complete merger of two previously distinct sounds the inventory is reduced, as has happened to Western American English in the case of the vowels [ɔ] and [ɑ], leading to homophony of words such as 'caught' and 'cot'. Augmentation of segment inventories also occurs when what were once predictable phonetic variants of a given sound become distinctive or non-predictable, as happened in the history of English some nine centuries ago: the previously predictably voiced variants of fricatives – occurring only in intervocalic position – became distinctive, thus leading to morpheme alternations such as 'wife' – 'wives', 'waft' – 'wave', etc.

Differences between sound change and biological evolution

Although sound change is similar in many respects to the evolution of species, I think there are some subtle but important differences between the two phenomena. In what follows, I present a brief summary of my own work on the mechanisms of sound changes. These matters are controversial, but I have attempted to provide empirical support for my claims by reference to a variety of laboratory studies (see Ohala, 1974, 1978, 1981, 1983*abc*, 1985, in press; Ohala and Lorentz, 1977; Ohala and Riordan, 1979). Current accounts of the causes of sound change, like the theory of evolution (of species), assume the existence of natural variation in pronunciation[1].

[1]It must be recognized that pronunciation may change due to a multiplicity of factors, some of them non-phonetic, e.g., paradigm regularization, spelling pronunciations, etc.

Some of this variation becomes 'fixed' or lexicalized into sound changes. What is the source of this variation? It may be possible, as Lindblom claims, that some of this variation occurs due to the kind of speaker adaptation he has described and that some effects of sound change thus represent a 'fossil record' of adaptive variation, but this has not yet been demonstrated empirically. I rather think that the majority, if not all, of such variation – like the variations in genotypes – is mechanical and non-purposeful.

There is good evidence for this latter view. First, some variation in pronunciation almost certainly is not under the full control of the speaker but rather crops up because of physical constraints of the speaking mechanism. For example, it is aerodynamic constraints that give rise to noisy releases of stops before high close glides and vowels. This can be misconstrued by listeners as intentional and lead to sound changes such as [ækt ʃuəl] 'actual' from [ækt+juəl]. The fortuitous character of some variations seems to be demonstrated by the fact that listeners usually factor them out: Mann and Repp (1980) showed that listeners show a crossover between [s] and [ʃ] at a lower frequency when the following vowel is [u] in comparison to following [ɑ], presumably because they are aware that the assimilated labial rounding during the [s] would fortuitously lower its centre frequency (for other examples, see Ohala, 1981; Beddor, Krakow and Goldstein, 1986; Ohala and Feder, 1987). (Of course, it must not be assumed that listeners always succeed in factoring out such distortions in speech. Indeed, the small fraction of cases where they fail can also be a source of sound change; see below). Furthermore, a significant fraction of the variation can be shown to occur not in the mouth of the speaker but in the ears of the listener (Sweet, 1900, p.21-22; Jonasson, 1971; Ohala, 1981, 1983*ab*) through what must be innocent misapprehensions, that is, neither purposeful nor adaptive in any sense. For example, labialized velars [kʷ gʷ xʷ] are confused with simple labial consonants [p b f], thus giving rise to Modern English [læf] 'laugh' with [f] where the conservative spelling reveals there was once a velar fricative [χ] (labialized by virtue of the preceding labial glide). Also, a completely new series of consonants, e.g. the palatalized series in Slavic languages, can presumably arise because of what may be called 'parsing' errors by the listener: a mistake in assigning the 'sharpness' (high F_2 transition) to an adjacent consonant rather than to a vocalic segment.

However, if we focus on sound changes of the type that are found in many different languages, though they be typologically, geographically, chronologically, and genetically separated, we can be fairly sure we are dealing with those caused by universal physical factors.

That it is listeners' misapprehensions which underlie a major fraction of sound changes is evidenced by the fact that listeners' confusions in lab-based listening tests parallel sound changes to a degree that cannot be due to chance: in the nature of the change, in the environments that promote it, in the asymmetrical directionality of the change (if any). For example, in both sound change and listening tests we find [p] changed to (confused with) [t] primarily in the environment of following [i] (or if the labial stop is palatalized); in addition, this change/confusion is asymmetrical since [t] rarely changes to [p] (Ohala, 1978, 1983*a*).

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. I should clarify this last statement since it may seem to contradict an earlier observation that some sounds or sound combinations are more subject to confusion than others. The point is that physical (including physiological) principles constrain what variants come into being by influencing the fortuitous distortions in the speech of speakers and the hearing and articulatory capabilities of misapprehending listeners, but after those variants exist there is little evidence for subsequent optimization through competition of languages' sound systems. (Prague School phonologists would not agree with this, cf. Jakobson, 1972 [1931]). I think most variations occur due to errors in the transmission of pronunciation norms – due to listeners' mistakes – and thus resemble scribes' errors in copying manuscripts. Like scribal errors, there is no adaptive value to such variations. Further support for this view comes from the observation that over the time span that linguists have been able to investigate the history of languages, c. six millennia, – in which time many languages' phonologies, including their segment inventories, have undergone remarkable changes – there has been no detectable improvement in the communicative capacity of speech.

Now I must qualify this claim by admitting that, although the variations that become fossilized in sound change do not *per se* make speech adaptive, it may nevertheless be true that if one variant is adopted by or associated with a prestigious speaker or group, that might make it propagate more widely than others. This is a sociolinguistic fact: pronunciation norms may be adaptive (benefit the speaker) if they impart some desired social status. This dimension, however, really lies outside the strictly phonetic domain I have been discussing and so does not contradict my claim.

On Lindblom's analysis of universal tendencies in segment inventories

For this reason, I am not fully convinced that Lindblom's analysis of the Maddieson data, where large segment inventories show more 'marked' segments than small ones, necessarily represents a fossilization of adaptive variation, i.e. the purposeful variation speakers implement as a function of listeners' needs. It is as likely, I think, that such augmentation occurs mechanically and inadvertently due, e.g., to listeners' errors in parsing the speech signal. We should also keep in mind that the addition of 'marked' consonants does not necessarily increase the overall distinctiveness of consonants in proportion to their added number. Put more explicitly: the increment of distinctiveness between one unmarked segment and another, e.g., [p] and [b], is greater than the distinctiveness between an unmarked segment and the marked segment that developed from it, e.g., [p] and [p']. There is evidence that the palatalized consonants in Russian (which count as 'marked') are much more subject to confusions than are the plain (unmarked) consonants under various listening conditions (A. Stern, personal communication).

Speech style not necessarily correlated with distinctiveness

These points also have some bearing on the *discrete* style-dependent phonological recodings of pronunciation mentioned by Lindblom, e.g., 'native' as [neitʰɪv] in more formal, careful speech or [neɪɾɪv] in more casual style. The existence of these alternants is due to sound change; the one form is historically derived from the other but is no longer so derived by the speaker; they are, as Lindblom implied, just style-dependent variant forms each stored separately in the speaker's lexicon. Not all such alternants correlate with greater vs lesser distinctness as a function of the listener's greater vs lesser need for it, respectively. The style-dependent variants of the English present participle suffix [-ɪŋ] (formal) and [-ɪn] (informal) do not, as far as I know, differ in terms of distinctness.

3 The lesson of adaptive variability

The notion of adaptive variability is an important one for many practical tasks facing us currently, e.g., trying to construct robust devices for the automatic speech recognition (ASR). I have tried to strengthen Lindblom's

claim by removing questionable evidence offered in support of it. If sound change has no direct connection to adaptive variability does it nevertheless have some relevance to tasks such as ASR? I think it does, in a number of ways. First, given the vast treasure of data on sound change which linguists have accumulated in a little less than two centuries, it provides us with valuable information on the favoured paths of speech sound variation and confusion. Where human ears have stumbled (apologies for the incongruous metaphor) there will the ASR device stumble, too (see Ohala 1975, 1983*b*, 1985, 1986*a*). Second, a close examination of the record of sound changes, in particular those known as 'dissimilation', shows that the listener is not passive in the process of speech communication; the speaker is not the only one to 'adapt' to the constraints of the speaking situation. Dissimilation is the process whereby two sounds having similar features undergo change such that the shared feature is removed from one of them. An example is Latin [kʷTŋkʷə], 'five', with two labialized velars which became [kTŋkʷə] with the labialization removed from the first velar. (Subsequently this first velar changed to an affricate or fricative, giving Italian [tʃiŋkʷə], Spanish [siŋko], etc.). I have suggested (Ohala, 1981, 1985, 1986*a*) that this came about due to the listener erroneously attributing the labialization on the first velar to the fortuitous spillover of the labialization from the second velar and therefore eliminating it from his own most careful pronunciations. This is an auditory analogue of the visual error we call 'camouflage'. Crucial to this analysis is the idea that the listener tries to 'make sense' of the speech signal by 'parsing' it (see also: Beddor *et al.*, 1986; Ohala, 1986*b*). Undoubtedly the listener does this successfully in most cases, but occasionally makes mistakes that can show up as the sound change dissimilation. Third, a study of linguists' pronouncements on sound change shows that we should be quite cautious in trying to identify those things in speech that we expect to be the 'same'. Linguists are good at identifying the common historical origins of different words, morphemes, sounds – for example, they can show that the roots of '*eques*trian' and '*hippo*potamus' are the 'same' (both from the Indo-European word for horse **ekwōs*, but it is not clear that this notion of 'same' is the one that will do much good in an ASR task. The sameness of the roots in 'equestrian/hippopotamus' is an extreme case, perhaps, and not really one that would mislead anyone engaged in ASR. But there are other, more troublesome cases, e.g., the alleged 'sameness', according to linguists, of the variants of the phoneme /t/ in English, namely [tʰ], [t], [ɾ] and [ʔ] in [tʰɑp], 'top', [stɑp], 'stop', [bɛɾi], 'Betty', and [mɑ̃ʊ̃n̥ʔn̥], 'mountain', respectively. Are these the 'same' sounds or simply different sounds that

had a common historical origin? We would do well to keep in mind that the linguists know historical 'sameness' best and have not yet refined their techniques to identify psychological or neuromotor sameness (however, see Jaeger, 1980, 1986; Ohala, 1983*a*; Derwing and Nearey, 1986).

4 Conclusion

In summary, Lindblom has presented a convincing case that invariance is an elusive concept in speech and is not to be found in the articulatory, acoustic, or even in the auditory domain; rather, it is something that must be constructed by the listener using whatever clues are available, even non-phonetic ones.

References

Beddor, P.S., Krakow, R.A., and Goldstein, L.M. (1986) Perceptual constraints and phonological change: A study of nasal vowel height. *Phonology yearbook 3*, 197-217.

Derwing, B.L. and Nearey, T. (1986) Experimental phonology at the University of Alberta. In: J.J. Ohala and J.J. Jaeger (eds) *Experimental phonology*, Orlando: Academic Press, 187-209.

Jaeger, J.J. (1980) Testing the psychological reality of phonemes. *Language and Speech, 23*, 233-253.

Jaeger, J.J. (1986) Concept formation as a tool for linguistic research. In: J.J. Ohala and J.J. Jaeger (eds) *Experimental phonology*, Orlando: Academic Press, 211-237.

Jakobson, R. (1972) Principles of historical phonology (Translation by A.R. Keiler of original 1931 German). In: A.R. Keiler (ed.) *A reader in historical and comparative linguistics*, New York: Holt, Rinehart and Winston, 121-138.

Jonasson, J. (1971) Perceptual similarity and articulatory reinterpretation as a source of phonological innovation. *Speech Transmission Laboratory Quarterly Progress and Status Report, 1/171*, Stockholm, 30-41.

Mann, V.A. and Repp, B.H. (1980) Influence of vocalic context on perception of the [š] vs [s] distinction. *Perception and Psychophysics, 28*, 213-228.

Nooteboom, S.G. (1983) Is speech production controlled by speech perception? In: M. van den Broecke, V. van Heuven and W. Zonneveld (eds), *Sound structures. Studies for Antonie Cohen*, Dordrecht: Foris Publications, 183-194.

Ohala, J.J. (1974) Experimental historical phonology. In: J.M. Anderson and C. Jones (eds) *Historical linguistics II. Theory and description in phonology*, Amsterdam: North Holland, 353-389.

Ohala, J.J. (1975) How a study of sound change can aid in automatic speech recognition. In: G. Fant (ed.) *Speech Communication, volume 3: Speech perception and automatic recognition*, Stockholm: Almqvist and Wiksell, 299-302.

Ohala, J.J. (1978) Southern Bantu vs the world: the case of palatalization of labials. *Berkeley Linguistic Society, Proceedings of Annual Meeting 4*, 370-386.

Ohala, J.J. (1981) The listener as a source of sound change. In: C.S. Masek, R.A. Hendrick, and M.F. Miller (eds) *Papers from the Parasession on Language and Behavior*, Chicago: Chicago Linguistic Society, 178-203.

Ohala, J.J. (1983*a*) The phonological end justifies any means. In: S. Hattori and K. Inoue (eds) *Proceedings of the XIIIth International Congres of Linguists, Tokyo, 29 August - 4 September 1982*, Tokyo: Sanseido Shoten, 232-243.

Ohala, J.J. (1983*b*) Modern applied linguistics. In: *Proceedings of the Arab School on Science and Technology, First Fall Session: Applied Arabic Linguistics and Signal and Information Processing*, Damascus, Syria, 51-62.

Ohala, J.J. (1983*c*) The direction of sound change. In: A. Cohen and M.P.R. van den Broecke (eds) *Abstracts of the 10th International Congres of Phonetic Sciences*, Dordrecht: Foris Publications, 253-258.

Ohala, J.J. (1985) Linguistics and automatic speech processing. In: R. De Mori and C.-Y. Suen (eds) *New systems and architectures for automatic speech recognition and synthesis*, (NATO ASI Series, Series F: Computer and System Sciences, Vol. 16), Berlin: Springer-Verlag, 447-475.

Ohala, J.J. (1986*a*) Phonological evidence for top-down processing in speech perception. In: J.S. Perkell and D.H. Klatt (eds) *Invariance and variability in speech processes*, Hillsdale, NJ: Lawrence Erlbaum Associates, 386-397.

Ohala, J.J. (1986*b*) Against the direct realist view of speech perception. *Journal of Phonetics, 14*, 75-82.

Ohala, J.J. (in press) The phonetics and phonology of aspects of assimilation. In: a volume edited by M. Beckman and J. Kingston.

Ohala, J.J. and Feder, D. (1987) Listeners' identification of speech sounds is influenced by adjacent 'restored' phonemes. In: *Proc. 11th International Congress of Phonetic Sciences, August 1-7, 1987, Tallinn, Estonia, USSR, Vol. 4*, 120-123.

Ohala, J.J. and Lorentz, J. (1977) The story of [w]: an exercise in the phonetic explanation for sound patterns. *Berkeley Linguistic Society, Proceedings of Annual Meeting 3*, 577-599.

Ohala, J.J. and Riordan, C.J. (1979) Passive vocal tract enlargement during voiced stops. In: J.J. Wolf and D.H. Klatt (eds) *Speech communication papers*, New York: Acoustical Society of America, 89-92.

Sweet, H. (1900) *The history of language*. London: J.M. Dent.

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