Speech Communication 13 (1993) 155–161
North-Holland

Sound change as nature's speech perception experiment

John J. Ohala

Department of Linguistics, University of California, Berkeley, CA 94720, USA

Received 10 June 1993
Revised 15 June 1993

Abstract. Variation in pronunciation observed in speakers today parallels in many details the documented variation in pronunciation over the centuries (sound change). It is reasonable to conclude that there is some necessary link between the two. argue that diachronic variation emerges for the most part from synchronic variation thus: universal and timeless physical constraints on speech production and perception leads listeners to misapprehend the speech signal. Any such misapprehension that leads the listener to pronounce things in a different way is potentially the beginning of a sound change. If we study sound change we can gain insights into how speech is produced and perceived. I exemplify this point by considering a variety of sound changes that involved voiceless fricatives: so-called spontaneous nasalization, s-aspiration, and nasal effacement. They suggest that one cue to this class of sounds is a special voice quality on that portion of vowels immediately abutting the fricative.

Zusammenfassung. Die heute bei Rednern beobachteten Variationen der Aussprache sind in zahlreichen Punkten parallel zu den im Laufe der Jahrhunderte übertragenen Variationen der Aussprache (Lautwandel). Daraus kann man schließen, daß zwangsweise eine bestimmte Beziehung zwischen den beiden besteht. Ich behaupte, daß die phonetischen Variationen zum Großteil aus synchronen Variationen entstanden sind, d.h. universelle und zeitlose, physische Auflagen bei der Erzeugung und Erkennung der Sprache führen bei den Hörern zu einer Fehlerfassung der Sprache. Jedes dieser Mißverständnisse, die den Hörer dazu veranlaßt, die Dinge unterschiedlich auszudrücken, ist der potentielle Anfang einer Lautwandel. Die Untersuchung der Lautwandel ermöglicht daher bestimmte Aufklärungen über die Art und Weise, wie die Sprache erzeugt und erkannt wird. Ich zeige dies an einem Beispiel durch Untersuchung eine ganze Reihe von Lautwandelen bei stummen Reibelauten, was man spontane Nasalisierung, die S-Atmung und das nasale Löschen nennt. Diese Entwicklungen lassen vermuten, daß bei dieser Kategorie von Tönen die spezifische Stimmqualität des Vokals direkt hinter dem Reibelaut einen Anhaltspunkt darstellt.

Résumé. Les variations de prononciation observées aujourd'hui chez les locuteurs sont parallèles, sur de nombreux points, aux variations de prononciation reportées au cours des siècles (le changement phonétique). Il est raisonnable de conclure qu'il existe une certaine relation nécessaire entre les deux. Je soutiens que les variations diachroniques émergent, pour la plupart, à partir des variations synchroniques, et donc que des contraintes physiques universelles et atemporelles de la production et de la perception de parole conduisent les auditeurs à mal appréhender le signal de parole. Chacune de ces méprises qui conduisent l'auditeur à prononcer des choses d'une façon différente est potentiellement le début d'une changement phonétique. L'étude de le changement phonétique permet donc bénéficier de certains éclairages sur la façon dont la parole est produite et perçue. Je montre cela sur un exemple, en étudiant toute une série des changements phonétiques portant sur des fricatives sourdes: ce que l'on appelle la nasalisation spontanée, la s-aspiration et l'effacement nasal. Ces changements suggèrent que, pour cette classe de sons, la qualité de voix spécifique sur la portion de la voyelle qui suit immédiatement la fricative constitue un indice.

Keywords. Sound change; speech perception; fricatives; nasals; phonetics; phonology.

1. Introduction

The quest to understand the structure and the workings of speech has been the unifying character of the phonetic sciences for over two millennia. Whether we are interested in alleviating the problems encountered in the speech clinic, in making it possible for humans to communicate with and control machines using speech, in facilitating the teaching of pronunciation to the language-learner, or in trying to understand the causes of changes in pronunciation over the centuries, we all ultimately need to know the particular and the general ways in which messages are conveyed via speech. Unfortunately given the inevitable necessity of specialization, researchers in the various disciplines are sometimes unable to study and take advantage of data, methods, and concepts in other phonetic sciences. Professor Hiroya Fujisaki, whom we honor with this issue, has worked hard to attempt to re-unify the phonetic sciences. His creation of the bi-annual International Conference on Spoken Language Processing (ICSLP), held in Kobe in 1990, in Banff in 1992, and scheduled for Yokohama in 1994, is one of the prime international conference series bringing together researchers from the various disciplines studying spoken language. In the interest of promoting interdisciplinary studies, one of Hiroya Fujisaki's lifelong goals, I offer in this paper some ideas on how a study of sound change can give insights into certain aspects of speech perception.

Sound change, an object of study in linguistics and philology for the past two centuries [1], may seem like an unlikely subject domain to contribute to our understanding of speech perception, a subject usually studied in laboratories and clinics. Of course, lab-based studies are absolutely essential and linguistic work on diachronic variation in pronunciation does not substitute for it. But an informed reading of the record of sound change can provide useful hints about speech perception (and speech production) which can then be followed up in lab-based studies.

2. The link between sound change and the mechanisms of speech

The link between sound change and speech perception (and production) is both inductive and deductive. Inductively, a close examination both of common sound changes [2] and the phonetic details of speech, especially how the "same" units of speech vary in different contexts and in different modes of speaking, reveals many unmistakable parallels. This variation can be found both in their mode of production and the manner in which they are perceived. For example, the phonological histories of a number of unrelated languages reveal that velar stops in the environment of round vowels or glides are subject to change to labial stops; see Table 1. Parallel to this is the observation that the acoustics of velar stops, both their bursts and formant transitions, resemble the bursts and transitions of labials. Finally, speech perception studies show that listeners often confuse syllables such as /ku/ and /pu/ more often than syllables like /ka/ and /pa/ (Winitz et al., 1972).

Faced with such parallels the conclusion is inescapable that there is some connection between speech variation in diachronic and synchronic domains (Durand, 1955).

The other link between these domains is deductive or theoretical. In various papers I have offered an account of sound change which is based on phonetic principles. Here is a brief summary: In speaking the speaker attempts to implement a string of phonetic units, *A B C* (refer to Figure 1). A given speech unit, *A* in Figure 1, may produce phonetic events which are distributed over time and over the frequency spectrum. For example, cues for the voicing of a

[1] Interest and speculation in sound change (or more generally, variation in language) can be found in the Bible (the "tower of Babel" story) and in the writings of Plato (*Cratylus*; c. 4th century BPE) and continued up to the 18th century. However, by most accounts the first serious scholarly study of it started in the late 18th and early 19th centuries (Robins, 1967).

[2] Changes in pronunciation can occur for many reasons, some of which have little or nothing to do with the phonetics: spelling pronunciations (e.g., the re-introduction of /l/ in *soldier* after the pronunciation had been stable as ['sodʒɚ]), analogy (e.g., the change, among some speakers, in the past tense of *bring* from *brought* to *brang* on the analogy of *sing*, *sang*), etc. Sound changes due to such causes tend to be language- and culture-specific. Only sound changes that are found independently in many languages – and thus due to universal factors – are likely candidates to tell us something about how speech is produced and perceived.

Table 1
Examples of labialized velars having changed to simple labials

Indo-European	*Classical Greek*	
★ekwōs	hippos	"horse"
★g^{w}iwos	bios	"life"
Proto-Bantu	*West Teke*	
★-kumu	pfuma	"chief"
Proto-Yuman	*Yuma*	
★imalik^{w}i	mal^{y}pu	"navel"

For sources of data, see (Ohala and Lorentz, 1977).

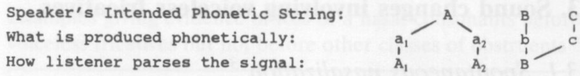

Fig. 2. A variant on Fig. 1. Here the transmission of the pronunciation from speaker to listener fails because the listener misparses the phonetic string. See text for further details.

medial stop might include the duration of the preceding vowel and an F_0 perturbation on the following vowel. The listener must know to parse these together. In the vast majority of cases the listener parses the signal correctly and infers the speaker's intended pronunciation. But sometimes the listener may misparse the signal (see Figure 2). The signal would be misapprehended if two events associated with one unit, a_1, a_2, were parsed separately as A_1, A_2. For example, it is thought that this is how certain new tones originated in the daughter dialects of Middle Chinese and certain Southeast Asian languages: the naturally occurring F_0 perturbation accompanying obstruents of different voicing were taken as independent aspects of the signal and, in the process, exaggerated into distinct lexical tones (Hombert et al., 1979). Such an error can be called a "dissociation" parsing error: two events that should have been associated were not. Another possible misapprehension of the signal would occur if events stemming from two separate speech units, say *B* and *C*, are erroneously parsed together by the listener. In Figure 1 the listener was correct to parse a_1 and a_2 together but in Figure 2 errs in parsing *b* and *c* together. Errors of this sort, which can be referred to as false association, can easily occur in cases where the events *b* and *c* resemble those that in other instances should be parsed together. For example, in English *sword* [sɔrd] has lost its original /w/ glide presumably because listeners thought it was a predictable feature of the rounded vowel.

A misparsing of the speech signal is a potential sound change: it affects the way the listener himself speaks and this changed pronunciation may be imitated by other speakers. My point in this paper is this: the documentary evidence we have of the pronunciation of words at different periods of history, especially when there is evidence that the pronunciation has changed, gives us, in essence, information about the units intended by the Speaker and the units extracted by the Listener (as schematized in Figures 1 and 2). From this it may be possible to deduce the intermediate level, the physical events in the speech signal which are used as cues by the listener. Of course we already have a great deal of knowledge about this level from painstaking speech perception experiments and from detailed acoustic analyses – and this effort owed little or nothing to historical phonology. However, we certainly do not know everything we need to know about the cues for speech units and I think a study of sound change could assist us. The specific way that the record of sound change can help us is this: by finding out the sound changes that a certain segment is subject to or those changes for which it provides an environment, it may be possible to get some idea of the multiple time- and frequency-distributed cues for it – which cues evidently were misparsed by listeners at some point in history. I want to demonstrate this in the case of some sound changes which involve voiceless fricatives.

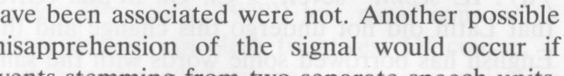

Fig. 1. Schematic representation of stages involved in transmitting a pronunciation from speaker to listener. The listener correctly parses the phonetic string produced by the speaker.

3. Sound changes involving voiceless fricatives

3.1. Spontaneous nasalization [3]

One common and well-recognized sound change is the development of nasalized vowels from vowels in the environment of nasal consonants, e.g., Latin *ūnus* "one" > French *un* [œ̃]; Sanskrit (Skt) *danta* "tooth" > Hindi [dãt]. But there are some cases where a nasalized vowel arises and there is no nasal consonant in the history of the word to explain it, e.g., Skt *sarpa* "snake" > Hindi [sãp], Skt *akṣii* > Hindi [ãkʰ]; Chinese *patʃjo* "plantain" > (via borrowing) North Korean dialect /pʰantʃjo/; dialectal English [hæ̃f] "half". This change has been called "spontaneous nasalization". A recurring class of segments in words showing spontaneous nasalization are those characterized by high air flow, including voiceless fricatives (especially [h]), affricates, and voiceless aspirated stops (but not voiceless unaspirated stops). There is no plausible articulatory or aerodynamic reason why such segments would cause the soft palate to lower; on the contrary, their production demands a tightly closed velopharyngeal port. However, another explanation was offered based on a somewhat complex interaction of articulatory and acoustic-auditory facts (Ohala, 1975, 1980).

(a) High airflow segments such as voiceless fricatives and aspirated stops are known to have a wider-than-normal glottal opening. In the case of voiceless fricatives this presumably occurs in order to maintain the high cross-constriction pressure differential needed to produce the audible turbulence.
(b) The abductory gesture of the vocal cords needed for the wider-than-normal glottal opening spreads onto the abutting margins of adjacent vowels (although the vowels remain voiced) (Klatt et al., 1967).
(c) This slightly open glottis during voicing alters the acoustics of the vowel by lowering the amplitude and increasing the bandwidth of F_1 (probably by increasing the coupling between the oral and subglottal cavities) (Fant, 1973; Fujimura and Lindqvist, 1971).
(d) These acoustic alterations mimic the effects of nasalization; what sounds like nasalization to the listener could be imitated as actual physiological nasalization.

All points of this hypothesis except the last were documented in the existing phonetic literature. Two experimental studies involving English, Mexican Spanish, and Hindi have shown that single period vowels excised from the portions of vowels immediately adjacent to voiceless fricative and then iterated into 300–500 ms vowels were judged to be nasal by listeners (Ohala, 1983; Ohala and Ohala, 1993). So apparently spontaneous nasalization arises because voiceless fricatives impose on adjacent vowels a special type of glottal state whose effects mimic nasalization and were thus likely to be reinterpreted as and substituted by physiological nasalization.

3.2. s-Aspiration

In many New World dialects of Spanish there was a process called s-aspiration by which post-vocalic /s/'s changed to [h], e.g., Colombian Spanish *los libros* [loh libroh] "the books". (Although Spanish has other voiceless fricatives, /s/ is the only one that can appear in tauto-syllabic final position.) In its development from Proto-Indo-European (IE) some Classical Greek (Cl. Gk) dialects also show the change of /s/ into /h/: IE *septḿ̥* "seven" > Cl. Gk hɛptá. (Given that Latin did not undergo this change and that English has borrowed some words with the same IE root from both Latin and Greek, English has many cognate words and morphemes showing the *h–s* alternation: *super-/hyper-*, *sex-/hex-* "six", *semi-/hemi-* "half".) A common explanation for loss of /s/ in such cases is that it "weakened" gradually to an [h]. Widdison (1991) proposed a different account. He suggested that the above-mentioned changes that [s] (and other voiceless fricatives) are known to impose on the margins of adjacent vowels may itself sound sufficiently [h]-like. If so, then the change of Vs > Vh would consist just of loss of the final [s], not, strictly speaking, its change to [h]; an [h]-like termination

[3] This discussion of spontaneous nasalization draws heavily from the work and insights of Manjari Ohala and Mariscela Amador (see (M. Ohala, 1983, pp. 78, 90; J. Ohala and Amador, 1981).

of the vowel would have been present all along. Widdison tested this by cross-splicing vowels between Spanish words like *pasta* and *pata* as spoken by a Mexican Spanish speaker. Crucially, Mexican Spanish does not exhibit s-aspiration. The vowel before the /s/ in *pasta* (and other comparable words) was sufficient to change *pata* perceptually into *pasta* 33% of the time. Although 33% was not the majority judgment of the listeners, the success of the vowel cross-splicing to effect a change in percept far exceeded that of other control splicings. A subsequent test showed that it was not the case that listeners identified the [h]-like vowel offset as an [s] but rather that they recognized it as the sound that is commonly substituted for an [s] by other Spanish dialects. (Such awareness of the pronunciation variants of other dialects is comparable to speakers of Midwest American English, which retains /Vr/, being able to understand other English dialects' post-vocalic r-deleted words by recognizing the vowel length, diphthongization, or vowel quality changes that have replaced the /Vr/ sequence.)

Spontaneous nasalization and s-aspiration, then, appear to stem from a misparsing of the same phonetic event, the peculiar transition between them and voiceless fricatives. In one case this transition was re-interpreted as vowel nasalization; in the other, as an [h]-like element.

3.3. Nasal effacement

In its development from Proto-Germanic, English (and other West and Northern Germanic dialects (Prokosch, 1938)) lost nasal consonants before certain voiceless fricatives. Of course, it must be acknowledged here that *any* voiceless obstruent seems to be a possible environment for loss of a preceding nasal, e.g., in American English, *pink* is often [pɪ̃k], but it seems that voiceless fricatives provided an even more conducive environment than other voiceless obstruents for such loss. Voiceless fricatives are implicated in the same process independently in other languages. Table 2 presents some examples. This process is often labeled "nasal effacement". At first glance this is a puzzling development because although there is a certain articulatory antagonism between nasals and voiceless fricatives (namely, they have opposite requirements for soft palate position), the same holds true for nasals and any obstruent, voiceless or voiced, and fricative or not. But if there is not a plausible articulatory motivation, could there be a perceptual one? There is a sense in which this sound change is the reverse of that discussed above under "spontaneous nasalization". There it was voiceless fricatives that provided an environment for introducing nasalization; here voiceless fricatives provide the environment for eliminating already-present nasal consonants. This "oppositeness" suggests the following hypothesis. The correct parsing of the peculiar acoustic quality of the vowel margin abutting on voiceless fricatives is that although it may sound nasal it is not, it is just a predictable consequence of, and thus should be parsed with, the voiceless fricative. Spontaneous nasalization arose due to a misparsing which dissociated that event from the voiceless fricative. In the case of nasal effacement we may have a case where there was an original nasal element and this was misparsed as a predictable consequence of the voice-

Table 2
Examples giving evidence of loss of a nasal consonants before voiceless fricatives but not before other classes of obstruents [a]

English	*German*
mouth	Mund
us	unser
fif-(ty)	fünf
goose [b]	Gans
but:	
hound	Hund
thank	danken
Latin	*Italian*
institutus	istituto
instinctus	istinto
constringěre	costringere
Ila (a Southern Bantu language)	
valan + *verb stem* =	
valambala	"they go past me"
valandetela	"they bring for me"
but:	
valãsempula	"they carry for me"
valã∫ija	"they leave me behind"

[a] In the case of English, cognate German words are provided because this sister language of English preserves the prefricative nasals of the common parent language.

[b] But cf. *gander*, where the original nasal is preserved when it appears before a voiced stop.

less fricative, i.e., an *erroneous association* was made, in the term used above.

This was tested in a perceptual experiment conducted by Grazia Busà and myself (Ohala et al., 1992). In essence the experiment sought to determine whether the presence of a nasal consonant is harder to detect when it appeared before voiceless fricatives as opposed to other segment types. Accordingly a naturally spoken syllable [gɛn] was obtained (I was the speaker) and the nasal consonant truncated in eight 10 ms steps from 70 ms (which counted as a "full" nasal) to 0 ms (which counted as no nasal). Then naturally spoken syllable-final [t] [z] [s] and [θ] were appended to each of the eight versions of the syllable in order to produce at the endpoints of the continua the nonsense words [gɛnt] [gɛt], [gɛnz] [gɛz], [gɛns] [gɛs], [gɛnθ] [gɛθ]. Using the endpoint stimuli we trained native speakers of American English and Northern Italian to differentiate them as having a nasal consonant or not having a nasal consonant. Ten speakers of each language reached criterion (making correct judgments on all eight stimuli) and these then were presented with a randomized list consisting of two repetitions of all 36 stimulus tokens. The answer sheets specified two choices for each stimulus, e.g., genth geth, and subjects just had to circle the answer they felt was correct. The results showed that on the average, for all twenty subjects there were fewer "nasal present" judgments when [s] and [θ] terminated the syllable than when [t] did. For the American English subjects [z] behaved identically to [t]; for the Italian subjects [z] was intermediate in its behavior between [s], [θ] on the one hand and [t] on the other. The 50% cross-over (from majority "nasal present" judgments to majority "no nasal" judgments) occurred at about 32 ms of nasal duration in the case of [s], [θ] and at about 20 ms duration in the case of [t]. Though these values are small, they are in the range of durations of nasals in such final clusters.

The results are consistent with the hypothesis that the differential ability of listeners to detect the presence of a nasal consonant before different obstruent types stems from their expectations of the effect that these consonants would produce on the margins of abutting vowels. Clearly more research is necessary before we can be highly confident about this conclusion. However, at the very least the results suggest that the cause of the sound change "nasal effacement" is to be sought in the perceptual domain.

4. Summary

The sound changes reviewed above represent failures in speech perception. But these failure offer a *prima facie* case that listeners use a distinctive transition with adjacent vowels as a cue for voiceless fricatives. This transition involves a special mode of vocal cord vibration that has a relatively open glottis and that produces acoustic effects which mimic nasalization. In most cases of speech perception this event is correctly parsed with the voiceless fricatives. Sound changes – change in the pronunciation norm for words – may occur when this cue is misparsed: either when it is dissociated from the fricatives or the erroneous association of an original independent nasal element with an adjacent voiceless fricative.

The inventory of possible distinctive speech sounds is very large even if any given language uses only a small subset of them. It is generally recognized that each speech sound has multiple cues. Ultimate validation of what constitutes a cue for listeners will have to be done using laboratory techniques. But how are these multiple cues to be discovered? How will we know where to focus our efforts in the laboratory? I suggest in this paper that we can gain useful hints on what some of these cues are by looking at the record of sound changes which languages have undergone. Sound changes are perceptual errors which give hints as what cues listeners use when speech perception is successful.

References

M. Durand (1955), "Du rôle de l'auditeur dans la formation des sons du language", *Journal de Psychologie Normale et Pathologique*, Vol. 52, pp. 347–355.

G. Fant (1973), "The acoustics of speech", in *Speech Sounds and Features*, ed. by Gunnar Fant (MIT Press, Cambridge, MA), pp. 3–16.

O. Fujimura and J. Lindqvist (1971), "Sweep-tone measurements of vocal-tract characteristics", *J. Acoust. Soc. Amer.*, Vol. 49, pp. 541–558.

J.-M. Hombert, J.J. Ohala and W.G. Ewan (1979), "Phonetic explanations for the development of tones", *Language*, Vol. 55, pp. 37–58.

D.H. Klatt, K.N. Stevens and J. Mead (1968), "Studies of articulatory activity and airflow during speech", in *Sound Production in Man*, ed. by A. Bouhuys, *Ann. N.Y. Acad. Sci.*, Vol. 155, pp. 42–54.

J.J. Ohala (1975), "Phonetic explanations for nasal sound patterns", in *Nasálfest: Papers from a Symposium on Nasals and Nasalization*, ed. by C.A. Ferguson, L.M. Hyman and J.J. Ohala (Language Universals Project, Stanford, CA), pp. 289–316.

J.J. Ohala (1980), "The application of phonological universals in speech pathology", in *Speech and Language: Advances in Basic Research and Practice*, Vol. 3, ed. by N.J. Lass (Academic Press, New York), pp. 75–97.

J.J. Ohala (1983), "The phonological end justifies any means". in *Proc. XIIIth Int. Cong. of Linguists, Tokyo, 29 August – 4 September 1982*, ed. by S. Hattori and K. Inoue (ICL Committee, Tokyo), pp. 232–243.

J.J. Ohala (1992), "What is the input to the speech production mechanism?", *Speech Communication*, Vol. 11, Nos. 4–5, pp. 369–378.

J.J. Ohala and M. Amador (1981), "Spontaneous nasalization", *J. Acoust. Soc. Amer.*, Vol. 68, pp. S54–S55 (Abstract).

J.J. Ohala and J. Lorentz (1977), "The story of [w]: An exercise in the phonetic explanation for sound patterns", *Berkeley Ling. Soc., Proc., Ann. Meeting*, Vol. 3, pp. 577–599.

J.J. Ohala and M. Ohala (1992), "Nasals and nasalization in Hindi", *Proc. 3rd Internat. Symp. on Language and Linguistics: Pan-Asiatic Linguistics, Chulalongkorn Univ., Bangkok, Thailand, 8 – 10 January 1992*, Vol. 3.

J.J. Ohala, M.G. Busà and K. Harrison (1992), "Phonological and psychological evidence that listeners normalize the speech signal", *Proc. Internat. Conf. on Spoken Language Processing, Banff, 12 – 16 October 1992*, ed. by J.J. Ohala, T. Nearey, B. Derwing, M. Hodge and G. Wiebe (University of Alberta, Edmonton, AB), pp. 1303–1306.

M. Ohala (1983), *Aspects of Hindi Phonology* (Motilal Banarsidass, Delhi).

E. Prokosch (1938), *A Comparative Germanic Grammar* (The Linguistic Society of America, Philadelphia, PA).

R.H. Robins (1967), *A Short History of Linguistics* (Longmans, London).

K. Widdison (1991), The phonetic basis for s-aspiration in Spanish, Doctoral dissertation, University of California, Berkeley.

H. Winitz, M.E. Scheib and J.A. Reeds (1972), "Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech", *J. Acoust. Soc. Amer.*, Vol. 51, pp. 1309–1317.