ISCA Archive
http://www.isca-speech.org/archive

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

Towards a universal, phonetically-based, theory of vowel harmony.

John J. Ohala

University of California,
Berkeley, CA 94720
e-mail: ohala@cogsci.berkeley.edu

ABSTRACT

Vowel Harmony (VH), phonological co-occurrence constraints between the features of vowels in polysyllabic words, is a fossilized remnant of an earlier phonetic process involving vowel-to-vowel assimilation. An understanding of the workings and constraints of VH can give us clues on how such inter-syllabic assimilations work and would be important for spoken language processing, e.g., motivating speech recognition using triphones. A universal, phonetically-based, theory of vowel harmony should explain why only certain distinctive features of vowels are subject to harmony, how "neutral" vowels arise, what intervening phonetic contexts favor or disfavor harmony.

I. INTRODUCTION

Vowel harmony is a phonological constraint found in certain languages whereby vowels within a certain domain -- usually a word -- are constrained to co-vary in certain of their parameters. For example, in Hungarian all vowels within a word must agree in the front/back parameter ("palatality") (with certain exceptions), e.g., /doboz/ "box" ~ /dobozotok/ "your box"; /ke:z/ "hand" ~ /ke:zøtøk/ "your hand". Umlaut, common in Germanic languages, is a restricted form of vowel harmony, whereby an original back rounded vowel became fronted in the environment of a palatal vowel or glide in the immediately following syllable (typically a suffix). E.g., Old High German /wu:rm/ "worm" but /wy:rmə/, earlier /wyrmi/ "worms". (The English morphological variants *foot/feet, goose/geese, mouse/mice* are fossilized residues of past umlaut.)

An understanding of vowel harmony is potentially important for speech technology because of its probable origin: it is the fossilized or phonologized result of purely phonetic and non-distinctive between-vowel assimilations. Such "long distance" assimilations between vowels, even those separated by one or more consonants, have been discovered in a number of languages [12] and probably occur in most languages to some degree. An understanding of the principles leading to vowel harmony in languages could lead to an understanding of the factors causing and constraining between-vowel assimilations and this, in turn, would allow us to make principled decisions on the extent to which "distant" contextual effects need to be accommodated in speech analysis, synthesis, recognition, and coding. For example, *triphones* have been advocated by some researchers precisely because they incorporate phonetic variation of the type found in an exaggerated form in vowel harmony [6]. In this paper I briefly review certain meta-patterns in vowel harmony and attempt to link these with explanatory phonetic principles.

II. RELATING DIACHRONIC TO SYNCHRONIC PATTERNS

It is necessary to clarify the relation between synchronic (phonetic) variation in speech production and diachronic (phonological) variation. Although both forms of variation show many parallels as to the segments involved, the direction of change, and the contextual environments in which they occur, they are not the same thing. Whereas the magnitude of the phonetic variation may vary continuously as a function of the strength or proximity of the contextual environment; the phonological variation is typically of an 'all-or-none' sort. To explain how this can come about I have given the following account of sound change.

In speaking the speaker attempts to implement a string of phonetic units, A B C D. Refer to Fig. 1a. Given speech unit, A B in Fig. 1a, may produce phonetic events, a_1, a_2, b_1, b_2, etc. which are distributed over time and over the frequency spectrum. For example, traces of a given vowel might be found not only in the syllable where it has its maximum realization but also in the nearest margins of vowels in abutting syllables [12]. In the vast majority of cases the listener (somehow) parses the signal correctly and infers the speaker's intended pronunciation. But occasionally a listener may misparse the signal. See Fig. 1b. The signal would be misapprehended if two events associated with one unit, a_1, a_2, were parsed separately as A_1, A_2. Presumably, this is how vowel harmony arose -- as well as most other assimilative sound changes --, that is, when the perturbation on a vowel caused by an adjacent vowel was parsed as independent of the perturbing vowel. Such an error can be called a 'dissociation' parsing error: two events that should have been associated were not.

A misparsing of the speech signal is a potential sound change: it creates a new pronunciation norm that the listener follows when he speaks and this changed pronunciation may be imitated by other speakers.

Another possible misapprehension of the signal would occur if events stemming from two separate speech units,

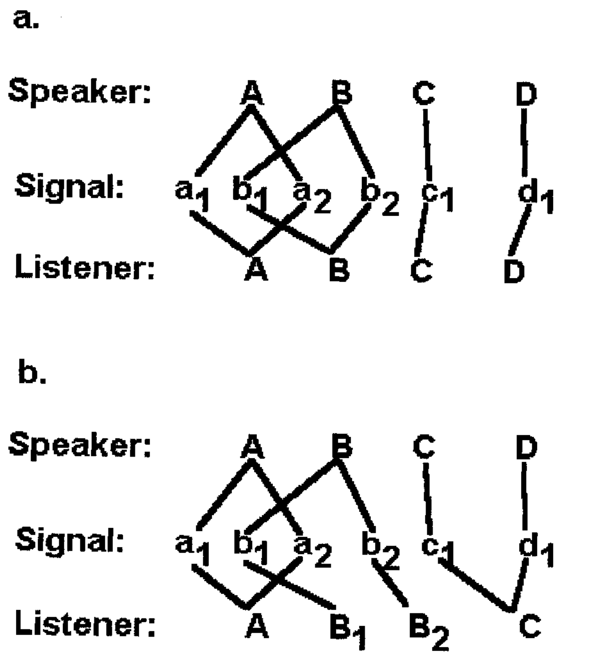


Fig. 1. How the listener might parse the speech signal.

say C and D, are erroneously parsed together by the listener. In Fig. 1a the listener was correct to parse a_1 and a_2 together but in Fig. 1b errs in parsing c_1 and d_1 together. Errors of this sort, which can be referred to as 'false association', can easily occur in cases where the events c_1 and d_1 resemble those that in other instances should be parsed together. For example, in English *sword* [s⊃rd] has lost its original /w/ glide presumably because listeners thought it was a predictable feature of the rounded vowel and thus parsed it with that vowel instead of giving it independent status.

Attempting to interpret phonological patterns created by sound change in the distant past for what they might reveal about universal phonetic tendencies has some pitfalls: First, after they have been shaped by phonetic factors and introduced into the language, several centuries of additional effects (some phonetic, some non-phonetic) may intervene to change them in ways which may obscure their origins. Second, a sound pattern found in a given language may arise due to peculiar phonetic properties of that language which may not be recoverable from the available phonological descriptions. Both of these problems may be avoided by looking for patterns found independently in many languages, i.e., the methodological principle of "safety in numbers".

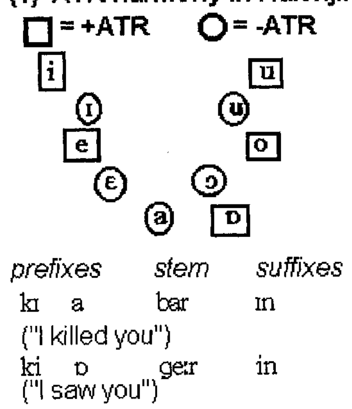
III. WHAT HAS TO BE EXPLAINED.

2.1 Types of vowel harmony.

Palatal harmony is widely attested in the Uralic language family (including, e.g., Finnish and Hungarian) and the Altaic family (including, e.g., Turkish and Mongolian). An example of an inflectional process in Hungarian that exhibits vowel harmony was given above. In addition, monomorphemic words in these languages also exhibit this harmony in that vowels in a word must be either all front or all back, e.g., Turkish /sekiz/ "eight", /dokuz/ "nine" [17].

Cross-height harmony or **'ATR'** (for 'advanced tongue root') or **Horizontal harmony** is found areally in a wide region throughout East and West Africa in languages belonging to the Niger-Congo (e.g., Igbo, Mande), Nilo-Saharan (e.g., Kalenjin, Acholi, Dinka, Luo), and Afro-Asiatic (e.g., Somali) language families. In these languages there are pairs of vowels with similar height, frontness, and rounding which nevertheless have distinct quality said to be differentiated by one member having advanced tongue root (+ATR) (or wide pharynx) and the other retracted tongue root (-ATR) (or constricted pharynx. The +ATR vowels are often transcribed as peripheral vowels, e.g., i, e, u, whereas for the -ATR vowels the symbols for more central vowels are employed, e.g., ɪ, ϵ, ᴜ, although there is often much qualification and puzzlement expressed over the exact nature of the phonetic difference. The harmonic sets and an example of cross-height harmony in Kalenjin are given in (1) (from [2]). Here it is the ±ATR value of the stem vowel that determines the quality of the affix vowels. There is substantial evidence that cross-height harmony originated as (and in many languages is still maintained as) a voice quality harmony, the +ATR set having a breathy quality and those of the -ATR set a creaky quality [2, 3].

(1) ATR harmony in Kalenjin



Rounding harmony is found also in the Uralic and Altaic language family and co-exists with palatal harmony. It is often more restricted in its application. In Turkish, for example, in addition to palatal harmony, affix vowels also harmonize with the rounding of the stem vowel only if it (the suffix vowel) is a high vowel [17]. See (2).

Height harmony occurs widely in Bantu and sporadically in other languages around the world. In Swahili, for example, there are many suffixes which have two forms, -*iwa*/-*ewa*, -*uza*/*oza*, etc. where /i, u, a/ in the penultimate vowel of the stem take suffixes with /i, u/ but /e,o/ take suffixes with /e, o/ [15].

The next two types of harmony are different from those listed above in that they are generally characterized as due to spreading of *consonantal* rather than *vocalic* features. However, they are included here because they are strongly manifested on vowels in adjacent syllables.

(2) Palatal and Rounding Harmony in Turkish.

noun	*accusative suffix*	
it	i	"dog"
gyl	y	"rose"
kɯz	ɯ	"girl"
tuz	u	"salt"

noun	*dative suffix*
it	e
gyl	e
kɯz	a
tuz	a

Nasal harmony is found sporadically in many languages around the world (e.g., in South America, Mexico, Indonesia). It is generally triggered by a nasal consonant. Furthermore there is a definite trend for the direction of spread of nasalization to be perseveratory. Examples from Sundanese are given in (3a) [14]. Here the presence of a nasal consonant is seen to create nasalization on all following segments unless blocked by an oral obstruent.

(3) Nasal Harmony in Sundanese

(a)	ɲãĩãn	*to wet*	bɤŋhãr	*to be rich*
	ɲãhõkɤn	*to inform*	mĩʔãsih	*to love*
(b)	mõẽkɤn	*to dry*	mãroẽkɤn	*to dry (pl.)*

Pharyngeal harmony is found principally in languages with a pharyngealized series of consonants, e.g., the Semitic languages (e.g., Arabic). In these cases vowels in adjacent syllables within a word are lowered or backed, i.e., [i] [æ] [u] become [e] [ɑ] [ʊ *or* o], respectively.

2.2 Limitation on types of harmony

Why should there only be these few types of harmony? Although *assimilation*, the seed from which vowel harmony develops, is an extremely common phonetic and phonological phenomenon, it must be acknowledged that many aspects of it are poorly understood. No doubt inertial constraints of the articulators play some role as do perceptual factors [9]. One of the least understood factors influencing assimilatory behavior is the way the motor program is implemented and why word boundaries, for example, should generally block assimilation [5, 7]. It seems clear, however, that of the various distinctive features characteristic of vowels, it is those which we may call *paradigmatic* such as height, frontness, rounding, voice quality, nasalization, etc. which are capable of spreading to adjacent syllables and thus leading to vowel harmony but not those features which are *syntagmatic* such a vowel length and diphthongization. The latter features are dynamic and are manifested as a vowel unfolds in time; they are not "states" of the vowel as the paradigmatic features are.

2.3 Neutral Segments

In many languages with vowel harmony one or more vowels are said to be "neutral" in that they do not fully participate in the harmony. For example, Finnish has the back harmonic vowels /a, o, u/ and front harmonic vowels /æ, ø, y/ and the neutral vowels /i, e/. Within a polysyllabic word stem the vowels must be from the same harmonic set except that the neutral vowels can appear with vowels of either harmonic set. In ATR harmony the most common neutral vowels are low back ones like [a] [16]. It may seem strange that in vowel harmony systems using [±palatal], the most palatal vowels, those with the highest F2, would be neutral vowels. Although vowels like palatal [i, e] would have the greatest fronting effect on adjacent vowels, listeners would be most aware of such a phonetically mechanical effect and thus be able to parse it out of the signal. Listeners' expectations of the V1-V2 interaction when one of the vowels is the F2-extreme [i] or [e] could either prevent incipient vowel harmony in these cases or neutralize it once it had become phonological through a false association parsing error as described above [10].

2.4 Environments That Favor or Inhibit Harmony

In many languages with vowel harmony the spreading harmonic feature is favored or blocked by specific phonetic segments.

Nasal harmony is blocked by oral obstruents articulated further forward than the point where the nasal and oral cavities are joined. Nasal harmony passes through glottal and pharyngeal obstruents because they are articulated further back than this point. Straightforward aerodynamic constraints account for this pattern: velopharyngeal closure (thus blocking spreading nasalization) is required by any segment, i.e., obstruents, whose production requires a buildup of air pressure behind the most forward point of constriction and where this air pressure would be vented by an open velopharyngeal valve [11]. There are apparent exceptions to this principle: In Tereno the glottal obstruent /h/ blocks the nasalization but this /h/ apparently derives from an earlier oral stop. In Southern Barasano nasalization is manifested on all eligible segments (non-obstruents) in a word and no segment blocks the nasalization [13]. In Mentu Land Dayak a spreading nasalization is blocked by some nasals consonants but there is evidence that these nasals derive from earlier pre-nasalized stops (i.e., which were oral at the end of the consonantal closure) [11]. In such cases of apparent exceptions it is clear that the process of nasal harmony has been completely morphologized: for the most part the nasalization is a property of specific words and morphemes and is no longer motivated or constrained by the physical phonetic factors that gave rise to it. In fact, even in Sundanese, the morphological character of the nasal harmony is revealed by the fact that when an oral infix /-ar-/ is inserted into a word with nasal harmony the word fragment to the right retains nasalization even when it is no longer contiguous to the nasalized segments on the left; see (3b).

Anderson [1] has shown that [+palatal] harmony may be weakened by [+grave] consonants, i.e., those with low F2 (labials and back velars). Since [+palatal] involves high F2, this inhibition of palatal harmony is explicable on phonetic grounds.

Rounding harmony, as noted above in the case of Turkish, may be absent on low vowels. In general, even aside from those languages with vowel harmony, less use is made of the [±round] distinction on low than on non-low vowels. The reason for this may be seen by examining the nomograms computed by Fant [4] giving formant frequencies for vocal tracts with different places and degrees of constriction (his Figs. 1.4-11a and 1.4-11b): for an oral constriction of 0.65 cm^2 -- representative of high, close vowels -- differing degrees of oral constriction and protrusion can modulate F2 a maximum of 1600 Hz; for an oral constriction of 2.6 cm^2 -- characteristic of a more open vowel -- this value reduces to 900 Hz. Similar reductions apply to F3. Evidently distinctive rounding requires a characteristic modulation of the higher formants.

However in Finnish and Turkish rounding of vowels is promoted by an adjacent labial consonant, e.g., Turkish /kaza/ "goose" (dative) vs. /karpuz/ "watermelon" [1, 17] -- because a labial consonant by lowering F2 and F3 produces an effect on the vowel that is similar to rounding.

Pharyngealization is blocked or weakened by a high, front vowel [i]. On the other hand, distinctive [±pharyngealization] is not supported on labial consonants (though it may be present non-distinctively) [5]. Both of these patterns are explicable by the fact that pharyngealization involves high F1 and low F2, just the opposite of the pattern for [i]. Labials create a lowered F2 which is too similar to that caused by pharyngealization and thus would make it difficult for a listener to correctly parse the signal: they would be likely to make a 'false association' parsing error (see above) by interpreting the lowered F2 due to pharyngealization as a predictable consequence of the labial articulation [8].

IV. CONCLUSION

Much more research -- both phonetic and phonological -- is needed to extract generalizations about patterns of vowel harmony and explaining them by reference to universal phonetic factors. Achieving this goal may have applications in speech technology by, for example, motivating triphone units in speech recognition.

REFERENCES

[1] Anderson, L.B. 1979. Using asymmetrical and gradient data in the study of vowel harmony. In R. Vago (ed.), *Issues in vowel harmony*. Amsterdam: Benjamins. 271-340.

[2] Antell, S. A., Cherono, G. K., Hall, B. L., Hall, R. M. R., Myers, A., Pam, M. D. 1973. Nilo-Saharan vowel harmony from the vantage point of Kalenjin. *Research Notes* (Dept of Linguistics and Nigerian Language, Ibadan Univ). 1-58.

[3] Denning, K. (1989). *The diachronic development of phonological voice quality, with special reference to Dinka and the other Nilotic languages*. Stanford Univ. Doc. Diss.

[4] Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

[5] Ghazeli, S. (1977). *Back consonants and backing: coarticulation in Arabic*. U. Texas at Austin Doc. Diss.

[6] Haton, J.P. (1991). Knowledge-based acoustic-phonetic decoding of speech: a case-study with the APHODEX project. *Proc. 12th ICPhS*, Aix-en-Provence 5.438-441.

[7] McLean, M. (1973). Forward coarticulation of velar movement at marked junctural boundaries. *J. of Speech & Hearing Research* 16.286-296.

[8] Ohala, J.J. 1985. Around *flat*. In: V. Fromkin (ed.), *Phonetic linguistics. Essays in honor of Peter Ladefoged*. Orlando, FL: Academic Press. 223-241.

[9] Ohala, J.J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston & M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the grammar and the physics of speech. Cambridge: Cambridge University Press*. 258-275.

[10] Ohala, J.J. (In press). Hierarchies of environments for sound variation; plus implications for 'neutral' vowels in vowel harmony. *Acta Linguistica Hafniensia*.

[11] Ohala, J. J. & Ohala, M. (1993). The phonetics of nasal phonology: theorems and data. M. K. Huffman & R. A. Krakow (eds.), *Nasals, nasalization, and the velum*. San Diego, CA: Academic Press. 225-249.

[12] Öhman, S.E.G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *J. Acoust. Soc. Am.* 39.151-168.

[13] Piggot, G.L. (1992). Variability in feature dependency: the case of nasality. *Natural Language & Linguistic Theory* 10.33-77.

[14] Robins, R.H. (1962). Vowel nasality in sundanese, a phonological and grammatical study. In *Studies in linguistic analysis*. Oxford: Blackwell's. 87-103.

[15] Tucker, A.N. & Ashton, E. O. (1942). Swahili phonetics. *African Studies* 1.77-103.

[16] Ultan, R. (1973). Some reflections on vowel harmony. *Working Papers on Language Universals* [Stanford Univ.] 12.37-67.

[17] Zimmer, K. E. (1969). Psychological correlates of some Turkish morpheme structure constraints. *Language* 45.309-321.