Phonetic and Other Influences on Voicing Contrasts

Patricia A. Keating
University of California, Los Angeles
E-mail: keating@humnet.ucla.edu

ABSTRACT

What factors influence how languages use voicing and aspiration? Some influences are phonetic: a salient auditory boundary between short and long lag VOT values favors the perception of aspiration contrasts, while a clear gestural distinction favors the production of voicing contrasts. But even within a language, speakers differ in how they produce contrasts. Examination of speaker variation in English suggests that some speakers value uniformity of articulation and/or acoustics over articulatory ease, while others value ease of articulation over acoustic uniformity, even though it results in acoustic alternations. Other examples in the literature show that phonetic factors such as auditory salience and ease of articulation interact with another non-phonetic factor, symmetry. Thus it seems that uniformity, i.e. consistency in phonetic form and avoiding allophony, and symmetry, i.e. the uniform or parallel behavior of members of a class, are as important in shaping sound systems as are phonetic factors.

1. INTRODUCTION

It seems obvious that the production and perception of speech, the medium in which language is most commonly used, must influence the form of spoken language. We readily accept that the phonetic aspects of language are products of Good Design, that our sound systems are as well suited to our production and perception capabilities as are those of other species. That the substance of speech depends on what we can perceive is shown most notably by the use of signed languages by Deaf communities, but also by the specific difficulties in speech acquisition for individuals who lack either auditory or visual input. For example, lack of auditory input makes certain manner distinctions such as voicing and nasality particularly difficult to acquire, while lack of visual input makes certain place distinctions such as [m]-[n] and [f]-[θ], and use of the lips in vowel rounding, difficult to acquire [1]. Thus it seems highly likely that if people's perceptual abilities were somewhat different from what they normally are, speech itself would be somewhat different.

Nonetheless, it is just as obvious, from the existence of many different languages, that perception cannot dictate a single best form for speech. Within an overall scheme of Good Design, there must be many different ways to be good. What the various relevant criteria might be is an important question, then, in working out how perception affects sound systems. I found this question intriguing years ago when I was comparing obstruent voicing contrasts across languages. There was an impressive new claim about auditory processing that seemed to answer the question, why is the English VOT boundary between voiced and voiceless stops where it is? The claim was that there is an auditory discontinuity, a nonlinearity of auditory processing of temporal intervals and other cues, that favors a category distinction around +20 msec or more VOT (e.g. [2,3]). Infant and animal studies supported the salience of this auditory boundary (e.g. [4,5]). So naturally languages would exploit it for phonetic categories, giving the unaspirated vs. aspirated stop contrast we see in English and other Germanic languages, Mandarin and other Chinese languages. So far, then, Good Design. But I was worried about the other languages, the ones that had a phonetic voicing contrast rather than an aspiration contrast. The infant and animal evidence for an auditory boundary for voicing was at best weak (e.g. [6,7]), and at worst in the wrong location. Abramson and Lisker [8] found Spanish listeners discrminating aspirated stops from their own: "stop variants with voicing lag are just easier to discriminate on some psychoacoustic basis" (p.7) . If the auditory boundary favoring aspiration contrasts was such a good idea, why didn't these languages take advantage of it? And would there be consequences of not doing so? This is a particular instance of the general problem that the better the explanation of some phenomenon, the more difficult it is to explain the contrary phenomenon.

2. CASES

In my own experiments [9,10] I found that speakers of Polish, with a phonetic voicing contrast could be led to categorize sounds according to the aspiration boundary. Figure 1 shows the average category boundaries of 20 Polish and 20 English listeners, for three different Voice Onset Time continua. The English listeners had consistent boundaries (no boundary is shown for the –100/+50 ms continuum because they had no voiceless responses – their boundaries must have been well above +20 ms VOT). The Polish listeners' boundaries varied according to the continuum; the more aspirated stimuli in the continuum, the higher their category boundary. It was as if the aspiration boundary remained strong and attractive to them, and with a little encouragement they would use it instead of their own voicing boundary.

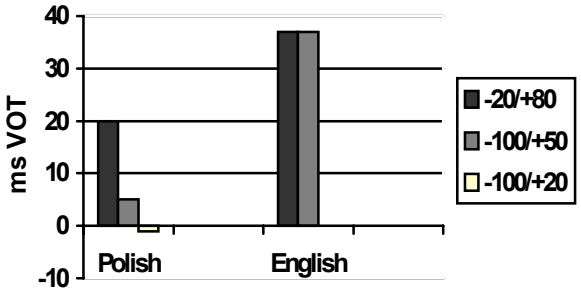

Figure 1: Mean calculated category boundaries for 20 Polish and 20 English listeners, for three VOT continua with different ranges (-20 ms to +80 ms, -100 ms to +50 ms, -100 ms to +20 ms)

I concluded that the aspiration contrast was perceptually easier and hence more "stable" than the voicing contrast. In that case, there must be some compensating advantage of the voicing contrast. I suggested that the advantage lay in its ease and stability in production, in that because speakers only had to turn on voicing sometime during closure, vs. not at all, the timing control needed was simpler than in an aspiration contrast. To some extent this was still a claim about perception, in that variations in timing of voicing would produce differences that wouldn't matter in the perception of this contrast. But it was seen even in production data that the VOT distributions for voiced vs. voiceless stops were clearly separated even across contexts and speaking conditions. Figure 2 from [9] shows one piece of this picture: the clear separation of the VOTs of voiced and voiceless stops in one set of data from Polish.

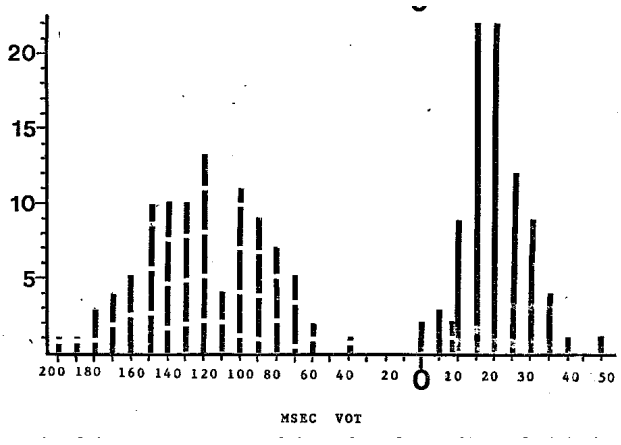

Figure 2: Histogram of VOT values obtained from minimal pair readings of initial /t/ (solid lines) and /d/ (dashed lines) by 24 speakers from Wrocław, Poland, from [9].

In contrast, the VOT distributions for unaspirated and aspirated stops observed in the literature on English overlapped more and varied more across contexts. (And in general, languages with aspiration contrasts show more allophonic variation than do languages with phonetic voicing contrasts [11].) This seems to be because glottal spreading gestures for aspiration participate in prosodic strengthenings and weakenings, while voicing gestures do not.) As a result, the relation between production and auditory boundary is not straightforward. It's not so clear that a hardwired auditory boundary is such a great thing if it's difficult to keep productions clearly on either side of it. Thus I concluded that languages which contrast two categories as [±voice] face a choice, a "trade-off between production and perception stability" [9].

This then is a conflict between purely phonetic criteria.

Going a step further, speakers of aspiration contrasts differ in how they control their productions. In [12], a glottographic and acoustic study, Flege showed that 10 English speakers had 4 different strategies for producing initial /b d g/. Only one speaker did what I would have expected: he kept his glottis adducted, and produced only short-lag stops. That is, he must not have performed any of the voicing-inducing articulations that, in addition to adduction, are necessary for initial voicing (e.g. [13,14]). Another four speakers who also kept their glottises adducted apparently did add such articulations, as they produced fairly consistent prevoicing. Three speakers were variable – they kept their glottises adducted, but produced both short-lag and prevoiced stops. Finally, and surprisingly, two of the 10 speakers kept their glottises abducted for /b d g/, though they adducted in time to prevent aspiration.

Consider now how this diversity plays out in terms of the surface realizations of the contrast. The four speakers who consistently prevoiced presumably paid the price of extra voicing-inducing articulations, but they then gained a degree of surface phonetic uniformity in their voiced stops. Assuming that their medial and final /b d g/ (not studied in this experiment) were also voiced, their /b d g/ were consistently phonetically voiced. The medial stops would be voiced without the need for extra gestures, while the final ones would require the same extra gestures as the initials [15]. The glottal state, and the resulting acoustics, would be consistent, giving a uniform surface phonetic output. Even the voicing-inducing gestures could be consistently deployed, as they would not hurt in medial position; the extra effort would buy even greater articulatory uniformity. In other words, this strategy values uniformity of acoustics, and possibly also of articulation, over articulatory ease. We can formulate this as: **acoustic uniformity /articulatory uniformity > articulatory ease**.

The one speaker who had neither spreading nor prevoicing in initial position expended no extra effort to effect voicing, and therefore probably enjoyed articulatory uniformity, assuming that he also had no spreading in medial (or final) positions. As a result, though, he would have acoustic alternations, because in medial position adduction by itself will generally result in voicing. In other words, this strategy values uniformity of articulation, and articulatory ease, over acoustic uniformity. We can formulate this as: **articulatory uniformity / articulatory ease > acoustic uniformity**.

The three variable-voicing speakers favored uniformity of glottal articulation only. We can formulate this as: **glottal articulatory uniformity > general articulatory uniformity / articulatory ease / acoustic uniformity.**

Finally, the two speakers who used spreading to prevent initial voicing, while still managing to prevent initial aspiration, expended extra effort, both in terms of the presence of the gesture, and its necessary temporal control. If they likewise did this in medial position (thus rather unusually suppressing medial voicing), then they gained articulatory and acoustic uniformity (abduction and voicelessness), though presumably with the result of some neutralization. If they did not do this consistently, then they also would have acoustic alternations; the only benefit, if it is one, then, would be the consistent favoring, across positions, of voiceless unaspirated stops. We can formulate this as: **acoustic uniformity / articulatory uniformity > articulatory ease**.

Thus, it is plausible that 6 of the 10 speakers (the 4 prevoicers and the 2 spreaders) preferred uniformity over articulatory ease. Because the uniformity is arguably (though not necessarily) both articulatory and acoustic, without further study we cannot say whether they have a preference between articulation and acoustics/perception. We can say, though, that they differ from the other 4 speakers, who place a lower value on acoustic uniformity compared to articulation.

It is surely crucial in this example that it doesn't really matter what the speakers do. All of these strategies work in terms of instantiating the English contrast. Yet just when they could all converge on some optimal strategy, if there were one, we see this diversity.

Uniformity is a design criterion in another way. In [15] John Westbury and I noted that ease of articulation alone, measured by the number of articulatory actions, predicts acoustic alternations in voicing across positions in utterance: the easiest utterance-initial stop is voiceless and unaspirated, the easiest utterance-medial stop is mostly voiced, the easiest utterance-final stop is partly voiced, and the easiest medial cluster is voiced-then-voiceless. These differences were relevant to the estimates of ease given above, as the stops in Flege's study [12] were in utterance-initial position. Where we see deviations from these predictions, as indeed we generally do, we again conclude that uniformity is valued over pure ease. For example, in [11] we considered whether languages without a voicing contrast would follow the pattern predicted by ease of articulation. Only one of five such languages in the sample did so; the others preferred uniform voiceless unaspirated allophones.

Furthermore, as Westbury and Keating stressed, these ease of articulation considerations hold only of position in *utterance*, not position in any smaller domains, so would be relevant at the word or syllable level only if a word or syllable constituted an entire utterance. That is, the considerations we outlined do not lead to any predictions about, for example, *word*-final devoicing. Single-word utterances might have some primacy in establishing such patterns, but in general the extension of such patterns from utterances to smaller domains is presumably again a case of acoustic uniformity.

Preference for uniformity amounts to a tendency to avoid allophony, that is, to let a phoneme have a consistent surface realization. Other examples of this preference in American English, pointed out by Bruce Hayes (p.c.), are the tendency to dark /l/ everywhere (presumably a generalization of phonetically motivated allophony), and a more subtle case involving a realignment of allophones. This example concerns the allophone of /e/ before /l/ and /r/ (e.g. *pail, pair*) - roughly [eə], an allophone that most speakers have trouble identifying with any phoneme - for speakers who have a similar vowel in contrast with /æ/ before nasals (speakers who distinguish e.g. *banner* from *ban(n)+er*). For these speakers, /eə/ is a separate surface phoneme, and it is similar in quality to the [eə] allophone of /e/. Such speakers may feel that the vowels in *pail* and *pair* are instances of this phoneme /eə/ rather than of /e/. Through this categorization, these speakers avoid positing a more remote allophone for /e/.

3. DISCUSSION

Uniformity as I use it here is clearly related to the Optimality Theory notion of correspondence, especially Output-Output correspondence or shape invariance, which is concerned with the degree of similarity among surface forms (see e.g. [16]), and I have emphasized this relation by using the term uniformity (as in paradigm uniformity, e.g. [17]). The shape invariance in the cases presented here is invariance of phones rather than morphemes or other forms, and it can be articulatory as well as perceptual. In general, Optimality Theory provides a natural way to express the interaction of a variety of desiderata in shaping sound patterns, but a principle of phone uniformity, or avoidance of allophony, is one that many versions of OT cannot express. That is because typically there are no phonemes *per se*, and thus there can be no concern with how phonemes are realized. Theories developed to allow contrast maximization or dispersion (e.g. [18,19,20]) do refer to the elements being contrasted, making these the versions of OT most accommodating of this principle.

Uniformity, or consistency in form, would seem to be one of the non-phonetic but apparently important functional considerations that shape language. Another is pattern symmetry. This notion dates back to traditional structural phonemics, and has received attention in recent phonetic literature [21,22] in connection with the tendency of vowel systems to be more symmetrical (between front and back vowels) than is predicted only by phonetic factors such as dispersion. Phonological feature representations of vowels capture this symmetry.

Another invocation of symmetry is Hayes's "inductive grounding" [23], the formulation of markedness constraints by a language learner, such that a set of phonetically natural but categorical constraints is available to a learner constructing a grammar. He proposes that "the influence of phonetics in phonology is not direct, but is mediated by structural constraints that are under some pressure toward formal symmetry". As a result of this pressure, constraints that are motivated by ease of articulation do not enforce maximal ease of articulation; instead they maximize the ease available from a relatively simple constraint. His example returns us to ease of articulation of stop voicing: he considers why languages do not seem to ban all and only the hardest voiced stops. Such a constraint might serve to ban, for example, all post-obstruent voiced stops, [d] and [g] in initial position ([b] being a bit easier), and [g] after oral sononants ([b] and [d] being a bit easier in that context). Instead, languages make cruder cuts that eliminate some harder cases and some easier cases as well, such as voiced velars everywhere, or all voiced stops in final position. Symmetry here refers to the inclusion of all members of a class in a constraint, even if the ease of articulation is greater for some members than for others (as is generally the case). As above, the mechanism of such symmetry is the use of phonological features, in this case in the statement of constraints, as features express natural classes rather than individual sounds.

In sum, phonetic factors are important in language, but they appear to be moderated by non-phonetic factors such as uniformity and symmetry.

REFERENCES

[1] A.E. Mills, "The development of phonology in the blind child", in *Hearing by Eye: The Psychology of Lipreading*, B. Dodd and R. Campbell, Eds., pp. 145-161. London: Lawrence Erlbaum Associates, 1987.

[2] D. Pisoni, "Identification and discrimination of the relative onset of two component tones: Implications for the perception of voicing in stops", *Journal Acoustical Society America*, vol. 61, pp. 1352-1361, 1977.

[3] S. Soli, "The role of spectral cues in discrimination of Voice Onset Time differences", *Journal Acoustical Society America*, vol. 73, pp. 2150-2165, 1983.

[4] P. Eimas, E. Siqueland, P. Jusczyk, and J. Vigorito, "Speech perception in infants", *Science*, vol. 171, pp. 303-306, 1971.

[5] P. Kuhl and J. D. Miller, "Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli", *Journal Acoustical Society America*, vol. 63, pp. 905-917, 1978.

[6] P. Eimas, "Speech perception in early infancy", in *Infant Perception*, L. B. Cohen and P. Salapatek, Eds., Vol. II, Ch. 6, 1975.

[7] R.Lasky, A.Syrdal-Lasky, and R.Klein, "VOT discrimination by four and six and a half month old infants from Spanish environments", *Journal Experimental Child Psychology*, vol. 20, pp. 215-225, 1975.

[8] A. Abramson and L. Lisker, "Voice-timing perception in Spanish word-initial stops", *Journal Phonetics*, vol. 1, pp. 1-8, 1972.

[9] P. Keating, *A Phonetic Study of a Voicing Contrast in Polish*, Ph.D. Dissertation, Brown University, 1979.

[10] P. A. Keating, M. J. Mikoś and W. F. Ganong III, "A cross-language study of range of voice onset time in the perception of initial stop voicing," *Journal Acoustical Society America*, vol. 70.5, pp. 1261-1271, 1981.

[11] P. Keating, W. Linker, and M. Huffman, "Patterns in allophone distribution for voiced and voiceless stops," *Journal Phonetics*, vol.11, pp. 277-290, 1983.

[12] J.E. Flege, "Laryngeal timing and phonation onset in utterance-initial English stops," *Journal Phonetics*, vol. 10, pp. 177-192, 1982.

[13] M. Rothenberg, *Breath Stream Dynamics of Simple-Released-Plosive Production,* Bibliotheca Phonetica No. 6. Basel: Karger, 1968.

[14] J.R. Westbury, "Enlargement of the Supraglottal Cavity and Its Relation to Stop Consonant Voicing," *Journal Acoustical Society America,* vol. 73, pp. 1322-1336, 1983.

[15] J.R. Westbury and P.A. Keating, "On the naturalness of stop consonant voicing," *Journal Linguistics*, vol. 22, pp. 145-166, 1986.

[16] R. Kager, *Optimality Theory*, Cambridge UK: Cambridge University Press, 1999.

[17] D. Steriade, "Paradigm uniformity and the phonetics-phonology boundary", in *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, M. Broe and J. Pierrehumbert, Eds., pp. 313-334, Cambridge UK: Cambridge University Press, 2000.

[18] E. Flemming, *Auditory representations in phonology*, Ph.D. dissertation, UCLA, 1995.

[19] E. Flemming, "Contrast and perceptual distinctiveness", to appear in Phonetically-Based Phonology, B.Hayes, R.Kirchner, and D.Steriade, Eds., Cambridge UK: Cambridge University Press, 2001 (ms.).

[20] P. Boersma, *Functional Phonology*, Ph.D. dissertation, Unversity of Amsterdam, 1998.

[21] J.-L. Schwartz, L.-J. Boe, N. Vallee, and C. Abry, "The Dispersion-Focalization Theory of vowel systems," *Journal of Phonetics*, vol. 25.3, pp. 255-286, 1997.

[22] P. Boersma, "Inventories in functional phonology", *Proceedings of the Institute of Phonetic Sciences, Amsterdam,* vol. 21, pp. 59-90, 1997. Also Ch. 16 in [20]

[23] B. Hayes, "Phonetically-Driven Phonology: The Role of Optimality Theory and Inductive Grounding," in *Functionalism and Formalism in Linguistics*, M. Darnell, E. Moravscik, M. Noonan, F. Newmeyer, and K. Wheatly, Eds., Volume I: General Papers, pp. 243-285. Amsterdam: John Benjamins, 1999.