Title
Languages' Sound Inventories: The Devil in the Details

Journal
UC Berkeley PhonLab Annual Report, 4(4)

ISSN
2768-5047

Author
Ohala, John J

Publication Date
2008

DOI
10.5070/P76s79b30w

Languages' sound inventories: the devil in the details

John J. Ohala

1. Introduction

In this paper I am going to modify somewhat a statement made in Ohala (1980) regarding languages' speech sound inventories exhibiting the 'maximum use of a set of distinctive features'. In that paper, after noting that vowel systems seem to conform to the principle of maximal acoustic-perceptual differentiation (as proposed earlier by Björn Lindblom), I observe:

> ... it would be most satisfying if we could apply the same principles to predict the arrangement of consonants, i.e., posit an acoustic-auditory space and show how the consonants position themselves so as to maximize the inter-consonantal distance. Were we to attempt this, we should undoubtedly reach the patently false prediction that a 7 consonant system should include something like the following: ɗ, k', t͡s, ɬ, m, r, ǀ . Languages which do have few consonants, such as the Polynesian languages, do not have such an exotic inventory. In fact, the languages which do possess the above set (or close to it) such as Zulu, also have a great many other consonants of each type, i.e., ejectives, clicks, affricates, etc. Rather than maximum differentiation of the entities in the consonant space, we seem to find something approximating the principle which would be characterized as "maximum utilization of the available distinctive features". This has the result that many of the consonants are, in fact, perceptually quite close – differing by a minimum, not a maximum number of distinctive features.[i]

Looking at moderately large to quite large segment inventories like those in English, French, Hindi, Zulu, Thai, this is exactly the case. Many segments are phonetically similar and as a consequence are confusable.

Some data showing relatively high rates of confusion of certain CV syllables (presented in isolation, hi-fi listening condition) (from Winitz et al. 1972) are given in Table 1.

Table 1. Confusion matrix from Winitz et al. (1972). Spoken syllables consisted of stop burst plus 100 msec of following transition and vowel; high-fidelity listening conditions. Numbers given are the incidence of the specified response to the specified stimulus.

	Response:	/p/	/t/	/k/
Stimulus:	/pi/	.46	.38	.17
	/pa/	.83	.07	.11
	/pu/	.68	.10	.23
	/ti/	.03	.88	.09
	/ta/	.15	.63	.22
	/tu/	.10	.80	.11
	/ki/	.15	.47	.38
	/ka/	.11	.20	.70
	/ku/	.24	.18	.58

I do actually believe that the degree of auditory distinctness plays some role in shaping languages' segment inventories—especially when auditory distinctness is low. Sound change, acting blindly (i.e., non-teleologically), weeds out similar sounding elements through confusion which results in mergers and loss. The loss in some dialects of English of /θ/ and /ð/ and their merger with either /f/ and /v/ (respectively) or with /t/ and /d/ (respectively) is a probable example. I also believe it is sound change, again acting blindly, which is largely responsible for the introduction of new series of segments which involve re-use of some pre-existing features. In some cases there is historical evidence of this. Proto-Indo-European had only three series of stops: voiced, voiceless, and breathy-voiced (i.e., among labials: /b/, /p/, /b̤/. The voiceless aspirated series, /pʰ/, exemplified in Sanskrit and retained in many of the modern Indo-Aryan languages (like Hindi) developed by sound change from the (simple) voiceless series. And a fifth (!) series of stops, the voiced implosives, /ɓ ɗ/ etc, in Sindhi developed from geminated versions of the (simple) voiced stops.

Similarly we know that the nasal vowels in French and Hindi developed out of the pre-existing oral vowels plus following nasal consonant (with the nasal consonant lost). (E.g., French *saint* [sɛ̃] < Latin *sanctus* "holy"; Hindi *dant* "tooth" [dɑ̃t] < IE *dont*-, *dent*- "tooth".) It is also relevant to my case that historically French once had as many nasal as oral vowels and then over the centuries reduced the nasal vowel inventory due to, I have argued, auditory similarity (Ohala and Ohala, 1993).

But the point that I want to revise or distance myself from somewhat is the idea that re-use of distinctive features always results in a cost-minimal augmentation, vis-à-vis the introduction of segments that are distinguished by virtually all new distinctive features.

I suppose the basic message I am emphasizing here is that the apparent symmetry found in many languages' segment inventories (or possibly the symmetry imposed by the analyst who put segments in matrices where all rows and columns are uniformly filled) obscures a more complicated situation. There is a great deal of what is referred to as allophonic variation, usually lawful contextual variation. What this means is that the neat symmetrical matrices of speech sound inventories are really abstractions. The complications – the devilish details referred to in my title – have been 'swept under the rug'! Can we ignore this variation when speculating about common cross-language tendencies in the form of languages' segment inventories? I say 'no' since in many cases the same principles are at work whether they lead to apparent symmetry or asymmetry, and I'll give some examples in the following sections.

2. Some examples of devilish details

2.1. [p] is weaker than other voiceless stops.

Among languages that have both voiced and voiceless stops, the voiceless bilabial [p] is occasionally missing, e.g., Berber, and this gap is much more common than a gap at any other place of articulation among voiceless stop series.[ii] In Japanese the /p/ has a distribution unlike other voiceless stops: it doesn't occur in word-initial position except in onomatopoeic vocabulary (e.g., /pat͡ʃapat͡ʃa/, 'splash') or medially except as a geminate (e.g., /kapːa/ 'cucumber sushi') or in a few other medial environments. Phonetically in English and many other languages the burst of the /p/ has the lowest intensity of any of the voiceless stops. The reason, of course, is that there is no downstream resonator to amplify the burst. We should see that the latter phonetic fact is the unifying principle underlying all these cited patterns. (And this is, in part, the reason why the sequence [pi] is confused with [ti], as documented in Table 1.)

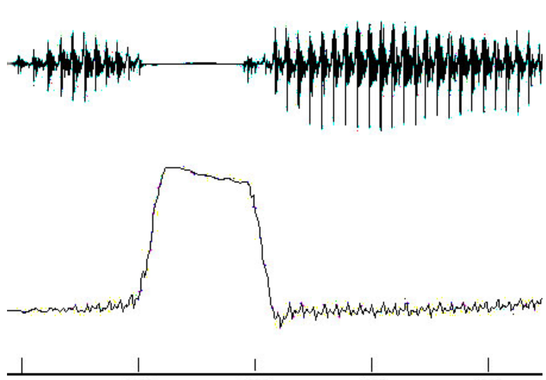
2.2. Voicing in stops and place of articulation

Among voiced stops, the velar, [g], is often missing in languages stop inventory even though they may have a voicing contrast in stops articulated at more forward places of articulation, e.g., in Thai, Dutch, and Czech (in native vocabulary). In some languages, morphophonemic variations involving the gemination of voiced stops shows an asymmetry in their behavior depending on how far front or back the stop is articulated. E.g., in Nubian (see Table 2), the geminate bilabial stop retains its voicing; those made further back become voiceless.

Table 2. Morphophonemic variation in Nubian (from Bell, 1971)

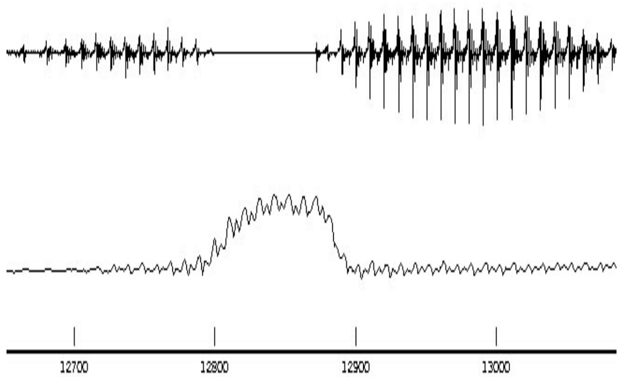
Noun stem	Stem + "and"	English gloss
/fab/	/fabːɔn/	father
/sɛgɛd/	/sɛgɛtːɔn/	scorpion
/kad͡ʒ/	/kat͡ʃːɔn/	donkey
/mʊg/	/mʊkːɔn/	dog

The usual descriptions of the allophonic variation of "voiced" stops in English (/b d g/) is that they are voiceless unaspirated in word-initial position but voiced between sonorants. In my speech, however, and that of another male native speaker of American English, I have found that /g/ is voiceless even intervocalically. See Figure 1 (subject DM) which gives the waveform and accompanying pharyngeal pressure (sampled with a thin catheter inserted via the nasal cavity). The utterance, targeted as /əˈgɑ/ is manifested as [əˈkɑ]. However, as shown in Figure 2, when the pharyngeal pressure was artificially lowered (by suction applied via a second catheter inserted in the other side of the nasal cavity), The /g/ was voiced!



[ə ˈg̊ ɑ]

Figure. 1. Acoustic waveform (top) and pharyngeal pressure (bottom) of the utterance /əˈgɑ/ spoken by subject DM, a male native speaker of American English. Condition: no venting of pharyngeal pressure. Phonetically the realization of the intervocalic stop was voiceless.



[ə ˈg ɑ]

Figure. 2. Acoustic waveform (top) and pharyngeal pressure (bottom) of the utterance /əˈgɑ/ spoken by subject DM, a male native speaker of American English. Condition: artificial venting of pharyngeal pressure. Phonetically the realization of the intervocalic stop was voiced (evident in the pressure signal).

All of these patterns, from the absence of [g] in Thai to the voiceless realization of /g/ intervocalically in (at least some) American English speakers are manifestations of the same universal aerodynamic principle: the possibility of voicing during stops requires a substantial pressure drop across the glottis and this depends partly on the volume of the cavity between the point of articulation and the larynx and more importantly on the possibility of passive expansion of that cavity (through inherent tissue compliance) in order to 'make room' for the incoming air flow. Velars and back-articulated stops have less possibility to accommodate the incoming airflow and so voicing is threatened.

3. On the various cues for obstruent "voicing" in English.

Lisker (1986) listed several features in addition to presence/absence of voicing or differences in VOT by which the so-called 'voicing distinction' in English obstruents is differentiated perceptually.

3.1. F0 perturbations on vowels following stops

The vowels immediately after voiced and voiceless obstruents show a systematic F0 variation. Figure 3 shows data from Hombert et al. (1979). These curves represent unnormalized averages of 100 msec of the F0 contours following /b d g/ (lower curve) and /p t k/ (upper curve) from 5 speakers of American English. Each curve is the average of 150 tokens. (Given that /p t k/ have a positive VOT whereas /b d g/ has VOT close to zero, the onset of the curves are phase shifted with respect to moment of stop release.)[iii] Such F0 differences can be explained as being mechanically caused due to differences in vocal cord state, i.e., they seem not to be purposeful on the part of speakers; see Ohala et al. 2004). Nevertheless Fujimura (1971) has presented evidence that such F0 contours are used by native speakers of English to differentiate this contrast when all other cues have been neutralized. Does this mean that English is a tone language? We would probably answer 'no' since the speaker doesn't have to separately produce and control the tension of the laryngeal muscles to implement these F0 differences. So it is English listeners, if not the speakers, that have the added complexity in their perceptual task of recognizing F0 differences just as native speakers of tone languages do. It is not much of a simplification of the sound system of a language if the language users (in their role as listeners) have to have skill in categorical recognition of short-term F0 contours in addition to recognizing voicing itself or VOT differences.

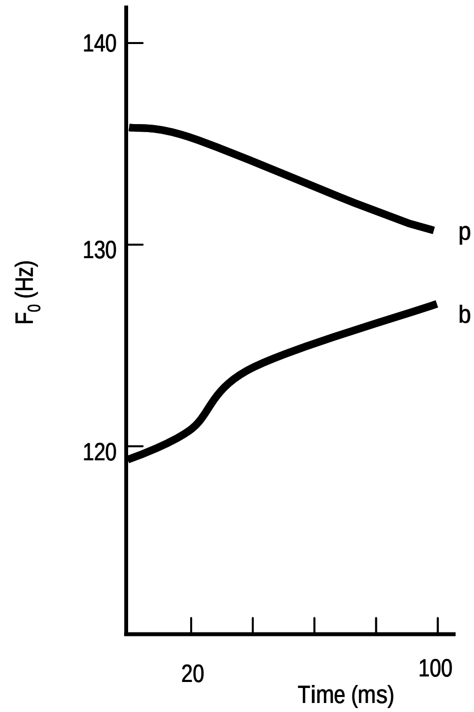


Figure 3. Average fundamental frequency values of vowels following English stops (data from five speakers). The curves labelled [p] and [b] represent the values associated with all voiceless and voiced stops, respectively – regardless of place of articulation. The zero point on abscissa represents the moment of voice onset; with respect to stop release, this occurs later in real time in voiceless aspirated stops (from Hombert et al., 1979).

3.2. Secondary cues to voicing in coda obstruents

As is well known, the class of supposedly voiced and voiceless obstruents in coda position are reliably differentiated by vowel duration, longer duration of the vowel before ‘voiced’ obstruents than before voiceless ones (by ratios of up to 3:2). Since this ratio is so large and there is no apparent “mechanical” cause of this difference, as Lisker (1974) concluded, this means that in this case both speaker and listener have to have distinctive vowel length in their grammars.

3.3. Vowel-influenced variations in VOT

Several studies have shown that the positive VOT of the voiceless aspirated stops in English show vowel-specific variations (Lisker and Abramson 1964, 1967; Ohala 1981a): VOT is longer before (actually when the stop is coarticulated with) high close vowels than before open vowels. The higher resistance offered by the close vowels delays the decay of P_o and thus the onset of voicing. These variations are probably an automatic consequence of differences in degree of aerodynamic resistance to the exiting airflow. The higher resistance offered by the close vowels delays the decay of P_O and thus the onset of voicing. Figure 4 (from Ohala, 1981a) gives data from English and Japanese (from unpublished studies by Robert Gaskins and Mary Beckman, respectively). The English data provide further evidence on the tendency of back-articulated “voiced” stops to be voiceless since here, even the so-called “voiced” velar has a positive VOT. Lisker and Abramson (1967) have shown that listeners are sensitive to these vowel-specific variations: cross-over points in the identification of the two categories of stops when presented in a VOT continuum also vary with the quality of the following vowel. This phonetic detail therefore must be part of the English-speaking listener's knowledge about the sound pattern of the language.

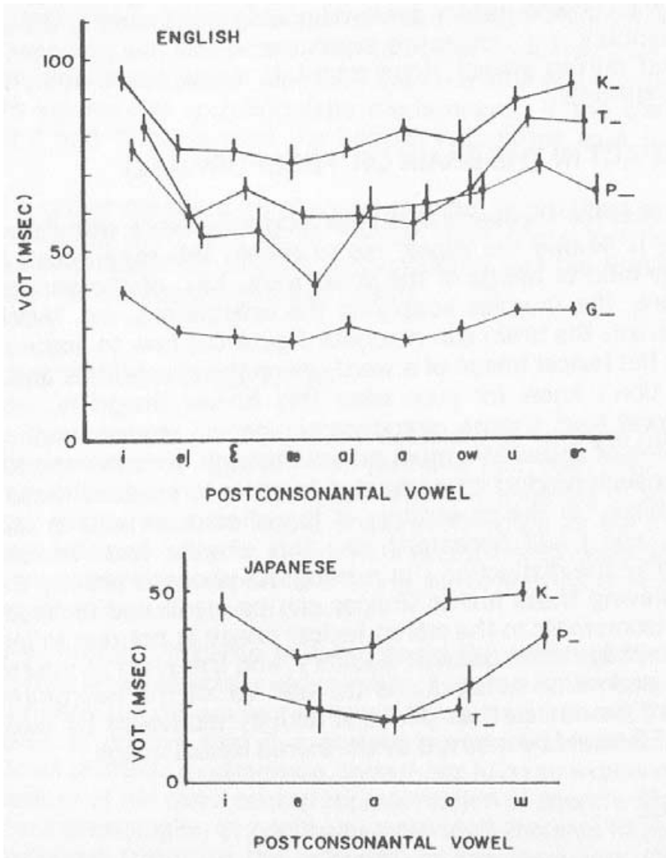


Figure 4. VOT variation for stops as a function of following vowel in English (a) and Japanese (b). (from Ohala, 1981a)

I could add many other examples where there are ***numerous*** acoustic features characteristic of specific consonant-vowel sequences or at least specific classes of sounds in the context of other specific classes. The net result of this is to add complexity to the signalling system of language that goes beyond what is implied by simply adding another row or column to the language's phoneme inventory.

4. Conclusion

If we conceive our task as phonologists as one of characterizing and understanding the function of speech to serve as a medium of communication then we want to know the implications for this function of the differences between the phonological system of, say, Rotokas with its 11 phonemes and !Xu with its 141 phonemes. Just by listing the segmental inventories in the traditional articulatory matrix does not tell the whole story. Adding new columns or rows *can* complicate the task of the language's speakers. The evidence that experimental phonetics has uncovered about so-called 'secondary distinctive features' in virtually every language whose sound system has been studied in some detail, especially the findings that these features may be different in different contexts, makes it clear that a language's phonological complexity is itself a complex issue.

In the model of sound change that I have proposed (Ohala 1981b, 1993) the so-called 'secondary' features are very important for understanding why change takes place and why it takes a particular direction. A feature that was secondary can become one of the primary distinctive features of a phonological contrast if the primary feature(s) are not detected or are misinterpreted. The important element in this model is that some of the elements of the "after" state were already present in the "before" state, without being explicitly listed in the inventory.

Thus we need to pay more attention to the devilish details in the implementation of phonological contrasts. It may help us to understand better both sound change and the communicative function of speech.

References

Bell, H.
1971 The phonology of Nobiin Nubian. *African Language Review* 9.115-139.

Fujimura, O.
1971 Remarks on stop consonants: synthesis experiments and acoustic cues. In: L. L. Hammerich, R. Jakobson and E. Zwirner (eds.), Form and substance: phonetic and linguistic papers presented to Eli Fischer-Jørgensen. 221-232. Copenhagen: Akademisk Forlag.

Hombert, J.-M., Ohala, J. J., & Ewan, W. G.
1979. Phonetic explanations for the development of tones. *Language* 55.37-58.

Lisker, L.
1974 On 'explaining' vowel duration variation. *Glossa* **8**, pp. 233-246.

Lisker, L.
1986 'Voicing' in English: A catalogue of acoustic features signalling /b/ vs. /p/ in trochees. *Language and Speech* 29. 3-11.

Lisker, L., & Abramson, A.
1964 A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20,* 384-422.

Lisker, L., & Abramson, A.
1967 Some effects of context on voice onset time in English stops. *Language and Speech, 10,* 1-28.

Ohala, J. J.
1974 Experimental historical phonology. In: J. M. Anderson & C. Jones (eds.), *Historical Linguistics II, Theory and description in phonology: Proceedings of the 1st International Conference on Historical Linguistics*, Edinburgh. 353-389. Amsterdam: North-Holland.

Ohala, J. J.
1980 Moderator's summary of symposium on 'Phonetic universals in phonological systems and their explanation.' *Proc., 9th Int. Cong. of Phonetic Sciences.* Vol. 3. Copenhagen: Institute of Phonetics. 181 - 194.

Ohala, J. J.
1981a Articulatory constraints on the cognitive representation of speech. In: T. Myers, J. Laver, and J. Anderson (eds.), *The cognitive representation of speech.* Amsterdam: North Holland. 111-122.

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

Ohala, J. J.
1993 The phonetics of sound change. In Charles Jones (ed.), *Historical Linguistics: Problems and Perspectives.* London: Longman. 237-278.

Ohala, J. J., A. Dunn, & R. Sprouse.
2004 Prosody and Phonology. Speech Prosody 2004, Nara, Japan.

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.* [Phonetics and Phonology Series, Vol. 5] San Diego, CA: Academic Press. 225-249.

Sherman, D.
1975 Stop and fricative systems: a discussion of paradigmatic gaps and the question of language sampling. *Stanford Working Papers in Language Universals*. 17.1-31.

Winitz , H ., Scheib , M . E . & Reeds , J . A .
1972 Identification of stops and vowels, *Journal of the Acoustical Society of America*, 51 (4) , 1309-1317

Notes

[i] The IPA symbols in this quote conform to current conventions, not those in 1980.
[ii] For a survey of gaps in consonant inventories, see Sherman (1975).
[iii] See also Ohala (1974) for similar data on F0 following the release of /s/ and /z/.