Languages' sound inventories: the devil in the details.

John J. Ohala
University of California, Berkeley

I'm going to say something a bit contrary to Ohala 1979 regarding languages' speech sound inventories exhibiting the 'maximum use of a set of distinctive features'.

Björn Lindblom referred to this paper. Just to set the scene for my remarks here I will briefly go over this statement again.

After noting that vowel systems seem to conform to the principle of maximal acoustic-perceptual differentiation, I say …

be most satisfying if we could apply the same principles to pre-
dict the arrangement of consonants, i.e., posit an acoustic-aud-
itory 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 fol-
lowing set:

d, k', ts, ł, m, r, ʔ,

Languages which do have few consonants, such as the Polynesian
languages, do not have such an exotic consonant 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 max-
imum 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.

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)

Sound change is drawn from a pool of synchronic variation   185

Table 1. Confusion matrix from Winitz et al. (1972). Spoken syllables consisted of stop bursts plus 100 msec of following transition and vowel; high-fidelity listening conditions.

Heard:	/p/	/t/	/k/
Spoken: /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 think auditory distinctness plays some role in the introduction and/or retention of speech sounds in a segment inventory—but it is *lack* of auditory difference. Sound change (which I thinks acts blindly) weeds out similar sounding elements through confusion which results in mergers or 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 whole new series of segments which involve re-use of 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. A fifth (!) series of stops, the voiced geminates, /ɓ ɗ / etc, in Sindhi developed from geminated versions of the (simple) voiced stops.

Similarly we know that the nasal vowels in French and Portuguese and Hindi developed out of the pre-existing oral vowels plus following nasal consonant (with the nasal consonant lost). (French *saint* [sã]; Hindi *dant* "tooth" [dãt].) (It is also relevant to my case that historically French had as many nasal as oral vowels and then over the centuries reduced the nasal vowel inventory due, I have argued, auditory similarity.)

The point that I want to revise or distance myself from is the idea that re-use of distinctive features always results in a cost-minimal augmentation, (vis-à-vis the introduction of segments that require all completely new distinctive features).

I suppose the basic message I am emphasizing today 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 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.

I'll give some examples:

Among languages that have both voiced and voiceless stops, the voiceless blabial [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.

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 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 burst 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 patterns.

Among voiced stops, the velar, [g], is often missing. Examples of languages where this is true includes Thai, Dutch, and Czech (among native vocabulary).

In some languages, e.g., Nubian, 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.

Table 9-7. Morphophonemic Variation in Nubian

| Noun stem	Stem + "and"	English gloss
/fab/	/fab:ɔn/	father
/seged/	/seget:ɔn/	scorpion
/kadʒ/	/katʃ:ɔn/	donkey
/mʊg/	/mʊk:ɔn/	dog

Note. From Bell (1971) and Ohala and Riordan (1979).

The usual descriptions of 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, I have found that /g/ is voiceless even intervocalically, thus /əˈgoʊ/ is [əˈkoʊ].

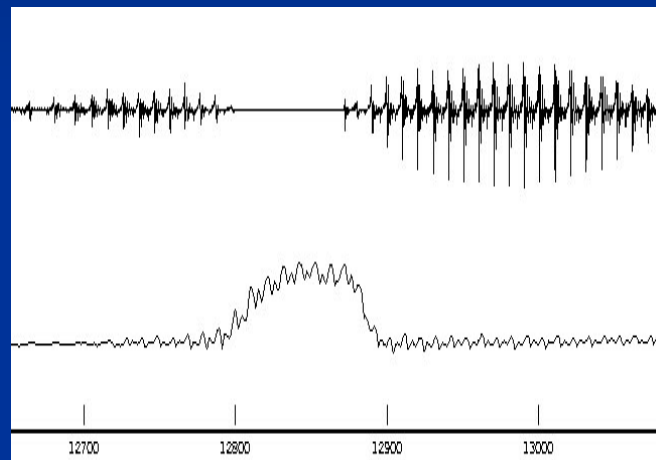Experiment involving artificial venting and suction of oral air pressure.
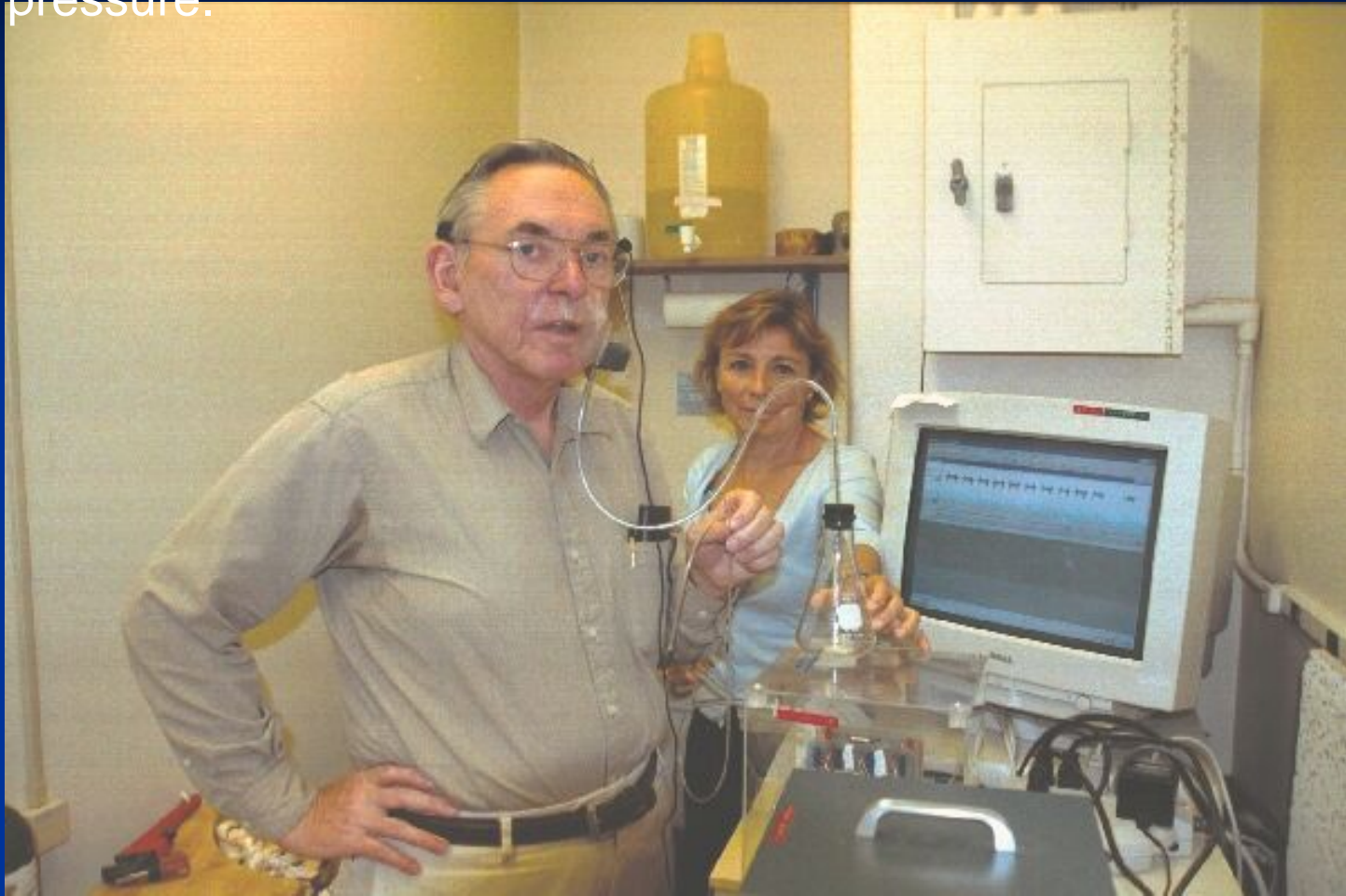
Author with collaborator, Prof. Maria-Josep Solé (Universitat Autònoma de Barcelona)

Subject DM:
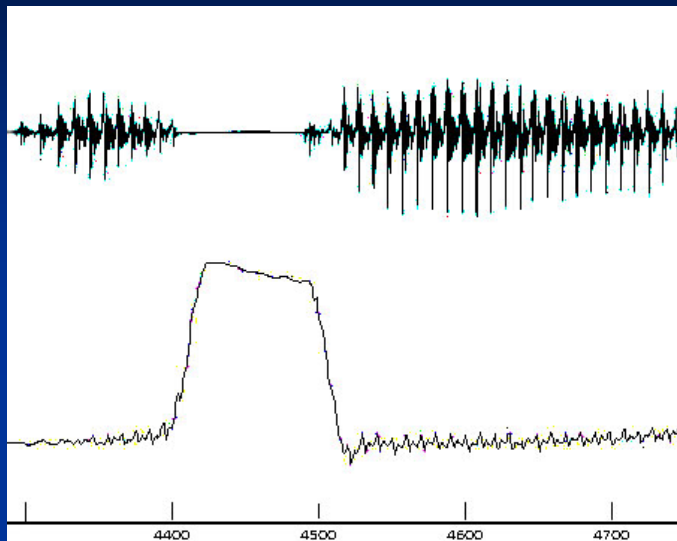
[ə 'g̊   ɑ]

But, with suction applied:

[ə 'g   ɑ]

The /g/ was voiced!

All of these patterns—the absence of [g] in Thai—to the voiceless realization of /g/ intervocalically in my speech 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 on the volume of the cavity between the point of articulation and the larynx and the possibility of expanding that cavity in order to 'make room' for the incoming air flow. Velars and back-articulated stops have less volume & so voicing is threatened.

And let's recall some other supposedly predictable allophonic variants:

The vowels immediately after voiced and voiceless obstruents show a systematic F0 variation.

Voiced fricatives require an even greater pressure drop across the glottis than do stops. Voicing requires $P_s > P_o$ but frication requires

$P_o > P_{atmos}$. One requires Po as low as possible but the other requires Po as high as possible. Both constraints cannot be met at the same time. The result is that voiced fricatives are frequently devoiced. In English this happens often in word- and utterance-final position.

But as is well known, the class of allegedly voiced obstruents are reliably differentiated in final obstruents by vowel duration, longer duration of the vowel before 'voiced' obstruents than before voiceless ones (by ratios of up to 3:2). So then the English phonological system should include distinctive vowel duration.

Lisker (1986) listed several (8 or more?) features in addition to presence/absence of voicing or differences in VOT by which the 'voicing distinction' is maintained.
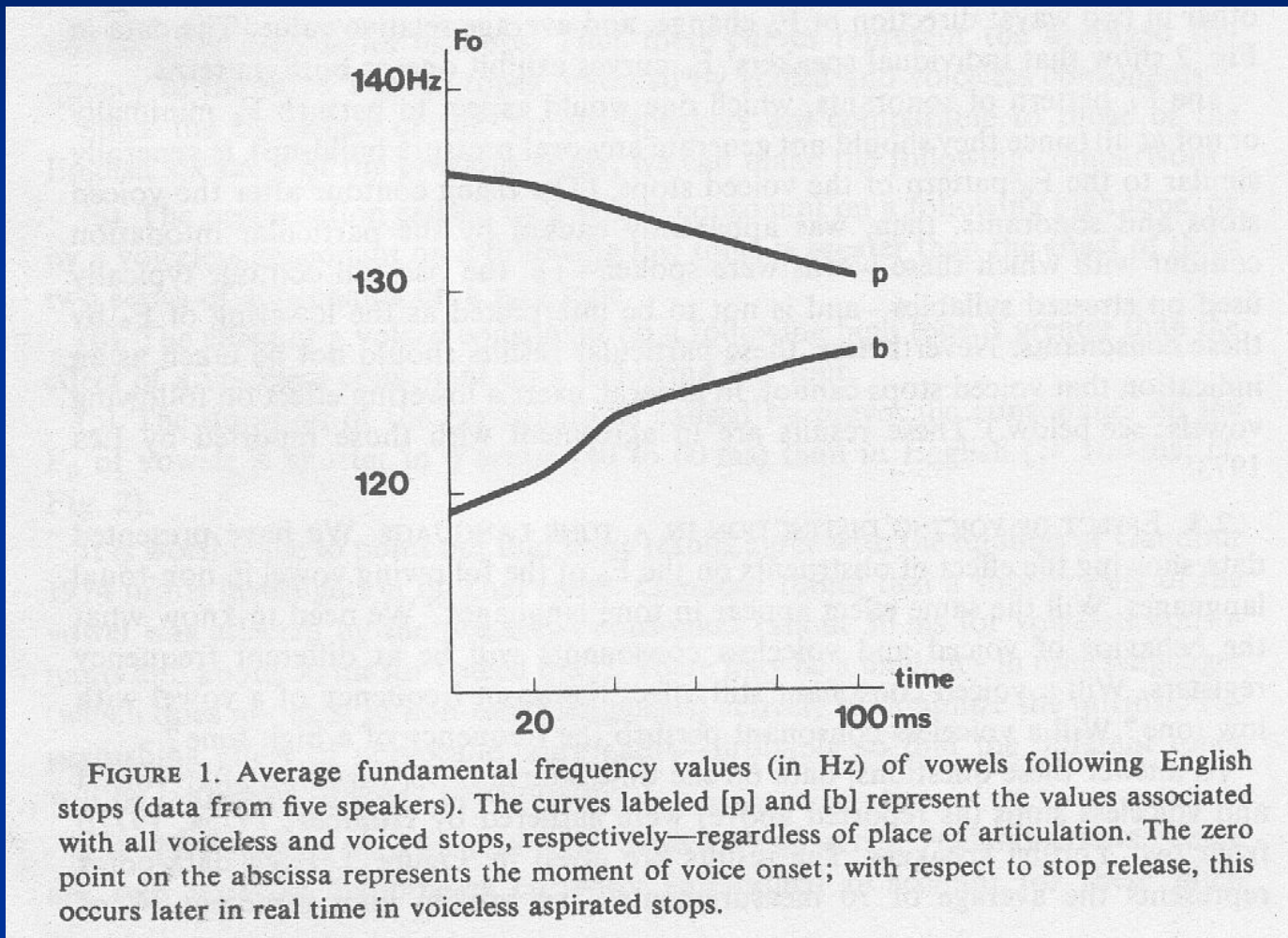
These include stop closure duration, F0 differences on adjacent vowels, etc.

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

Some data on the F0 differences on vowels:

The vowels immediately after voiced and voiceless obstruents show a systematic F0 variation.

From the UC Berkeley dissertation of J-M Hombert: unnormalized averages from about 100 tokens:

FIGURE 1. Average fundamental frequency values (in Hz) of vowels following English stops (data from five speakers). The curves labeled [p] and [b] represent the values associated with all voiceless and voiced stops, respectively—regardless of place of articulation. The zero point on the abscissa represents the moment of voice onset; with respect to stop release, this occurs later in real time in voiceless aspirated stops.

Does this mean that English is a tone language?

We'd 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. They arise automatically and "free" as a secondary consequence of implementing the VOT differences.

BUT: Osamu Fujimura and others have shown that listeners are sensitive to these F0 differences and can use them to differentiate stimuli – especially when the VOT differences have been artificially neutralized.

So English listeners, if not the speakers, 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 VOT differences.

But it is very common to find that listeners have to be attuned to what are called 'secondary cues' (ha! Sometimes they are the primary cue when other cues are neutralized!)

VOT shows systematic vowel-specific variations. (Japanese data from a study by Mary Beckman).

These variations are probably an automatic consequence of differences in degree of aerodynamic resistance to the exiting airflow. But it has been shown that English listeners are sensitive to these differences.
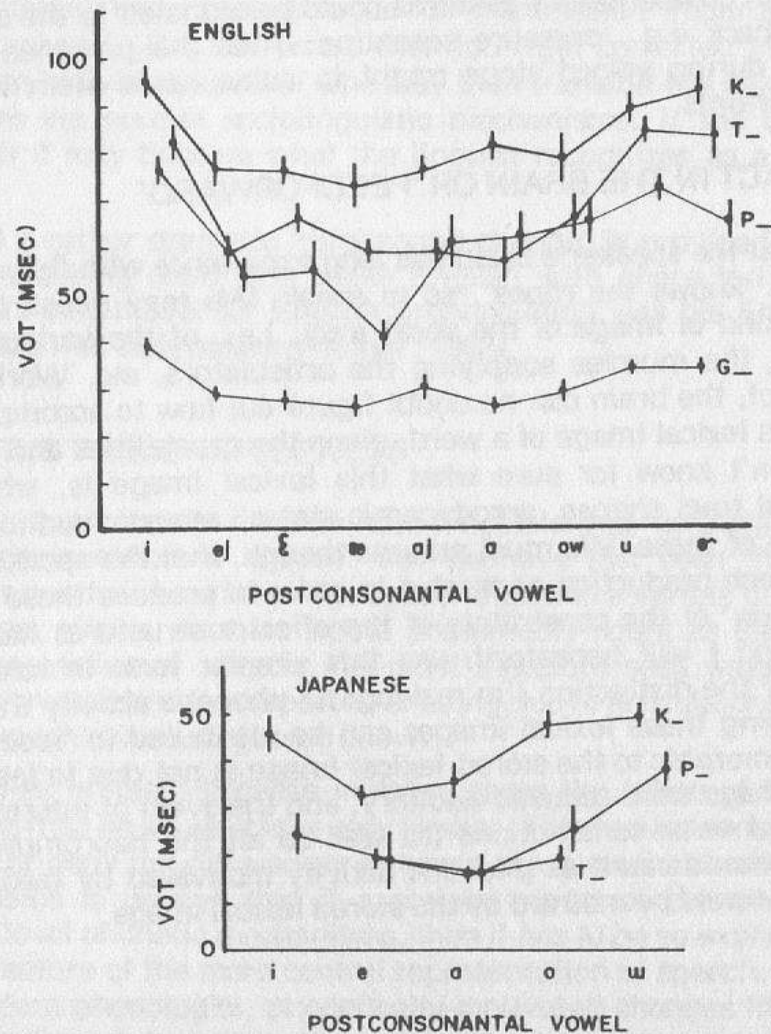

FIGURE 2   *Voice onset time (VOT) for English (top) and Japanese (bottom) stops in the environment of different vowels.*

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 signaling system language that goes beyond what is implied by simply adding another row or column to the language's phoneme inventory.

Implications:

If we conceive our task as phonologists is to characterize and understand the function of speech as serving 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 and Xhosa.

Just by listing the segmental inventories in the traditional articulatory and/or acoustic features of the two languages may not tell the whole story.

Adding new columns or rows can have two opposite results which will be contextually determined: better or worse perceptual differentiation could be different in word-initial position vs. word-final position, different before [i] than before [a] etc.

We might find that adding an extra row or column to the inventory forces listeners to find new acoustic-auditory features to attend to.

Certainly the so-called 'secondary' features are very important for understanding diachronic change (a topic I've addressed in a number of papers). A feature that was secondary can, through listener mis-parsing, become one of the primary distinctive features of a phonological contrast.

Conclusion:

We need to pay more attention to the fine details in the implementation of phonological contrasts. It may help us to understand the communicative function of speech sounds and help us understand how phonological change occurs diachronically.

Merci!

Or possibly … Mercy!