Paper presented at the Congress of Linguistics and Phonetics, Prague, August 30, 1990

MODELLING THE DISTRIBUTION OF CONSONANT INVENTORIES

Paul Boersma
Institute of Phonetic Sciences
University of Amsterdam

The question that I would like to address in this paper is: why are sound systems structured the way they are? I will try to give one possible answer to this question from a functionalist phonetic point of view. This answer will be formalized and tried out on the Germanic consonant shifts as well as on data on consonant manner features from a database of sounds of the languages of the world.

1. Function

I would like to derive the answer to our question from the first function of language, which is: "language is meant to convey information from one person to another as quickly and clearly as possible" (Passy 1890). The *quick* part of this function lies on the part of the speaker, who would rather be lazy than tired; the *clear* part is on the listener, who must understand the message with as little trouble as possible.

This means that a natural sound system must satisfy the following two criteria:
1. The sound system is such that the articulatory effort of the pronunciation of utterances is minimized.
2. The sound system is such that there is a maximum number of distinctions between utterances. Some people have argued that the latter criterion branches into two:
1. Perceptual differences *between* words are maximized.
2. The perceptual salience *within* words is maximized. For instance, sequences like [wu] tend to occur less in the world's languages than is to be expected form the marginal distributions of [w] and [u] (Kawasaki 1982).

2. Sound change

Now, how did these optimum sound systems come about? They have arrived at what they are now by means of long series of sound changes. Let us therefore extend the functionalist hypothesis to sound change. A way to look at it is this. Sound changes constantly propose themselves in the variations that the speech community tries out, whether under external influence or not. A sound change that is proposed in this way, will be actually carried out if it matches the following criteria:

1. Articulatory effort is reduced (within words).
2. The perceptual salience within words improves.
3. Perceptual distinctions between words improve.

It is hard to imagine, though, that all sound changes that have occurred in the history of the world match all of these three criteria. Rather, the criterion for optimization will be a criterion that weighs these three criteria and combines them into one. For instance, we could add and subtract losses and gains from the three criteria and see if we have gained as a net result. If so, the sound change would be granted access to the grammar. If not, the sound change will *not* be carried out. However, there are some problems in this. First, this kind of linear evaluation might not resemble human decision making. Secondly, it is hard to formalize. And thirdly, if languages conformed to this evaluation procedure, they would eventually arrive at an optimum sound system and never change after that. This is not what we see, though, in the languages of the world. All of them keep changing forever. We could assume some temperature in the whole system, so that our optimization criterion just predicts the *probability* that a certain sound change will occur. Alternatively, and that is what my approach is, we could choose a much simpler optimization criterion, namely:

> a sound change is carried through if it matches at least two of the three criteria.

That is, this principle simply amounts to a majority decision. Some properties of this criterion are:

1. We do not have to measure effort and distinction absolutely; our only concern is whether effort, distinctions and salience actually improve or not. Look at it as three demons measuring improvement, and a master who decides at majority vote.
2. There may not exist any optimum system and sound systems may keep changing forever.
3. Every sound change is unidirectional: if a certain sound change would improve the sound system and would therefore be allowed to take place, the reverse change would not be allowed to take place. This is the part of our hypothesis that makes it in principle testable and falsifiable.

3. Formalization

We are now ready to look at an example. The reduction of articulatory effort and the maximization of perceptual salience are the two criteria that can be evaluated without reference to the vocabulary as a whole. These two can conspire to yield CV-syllables. For instance, in [apa] we have only two articulatory movements, the closing and the opening of the lips. Pharyngeal adjustments, e.g., do not have to be made, as the tongue stays in the position needed for [a] during the whole of the word. Adjustments of the state of the glottis do not have to be made either, at least not if there is a high tone on [a]. We can conclude, then, that there are only two articulatory *contours* in [apa].

As a first measure of articulatory effort, then, there is the number of articulatory contours, which is very small in [apa]. Words like [ampa] and [api] feature four articulatory contours and are therefore more difficult to pronounce than [apa]. This bias is reflected throughout the languages of the world. For instance, most languages that allow words like [api], also allow words like [apa], whereas the reverse is much less common.

As a first measure of perceptual salience, there is the degree of perceptive contours. Perceptually, we find in [apa] not only labial contours, but voicing, sonorancy, continuancy and noise contours as well. Thus, the number of perceptual contours in [apa] is large, and the number of articulatory contours is small. Therefore, CV-syllables are preferred if we neglect the third criterion of cross-lexical distinctions.

4. The Germanic consonant shifts

We could go on like this, but there are millions of possible combinations of articulations and we have to restrict ourselves. Let us look therefore at an example that explains the Germanic consonant shifts in word-initial position. Here is a quick review of these historical sound changes:

"Indo-European"	*p	(*b)	*bh	*t	*d	*dh	*k	*g	*gh
"Latin"	pel		fol	tri	dw-	(fa-)	kan	gel	(host)
"Old Germanic"	fel	pool	bal	θri	tw-	doo-	xund	kald	gast
"Alemannic"	fel	pfuol	pal	(dri)	tsw-	tuo-	(hund)	xalt	kast
	skin	*pool*	*leaf/ball*	*three*	*two*	*do*	*dog*	*ice/cold*	*guest*

The endings of the words have been suppressed.

Imagine a universal sound inventory that does not consist of millions of sounds, but just these six: {p, p^h, b, f, v, a}. Imagine further a language with a vocabulary that contains only three words and with strong word-initial stress. The three words of this language can accordingly be chosen from the universal word inventory {*pá*, *p^há*, *bá*, *fá*, *vá*}, which consists of only five words. This makes a total number of ten possible systems, eight of which are shown in figure 1. The arrows in figure 1 represent the possible directions of sound change. To arrive at these directions, we have to assess the hierarchies of articulatory effort, perceptual salience and perceptual distinctiveness of the ten sound systems.

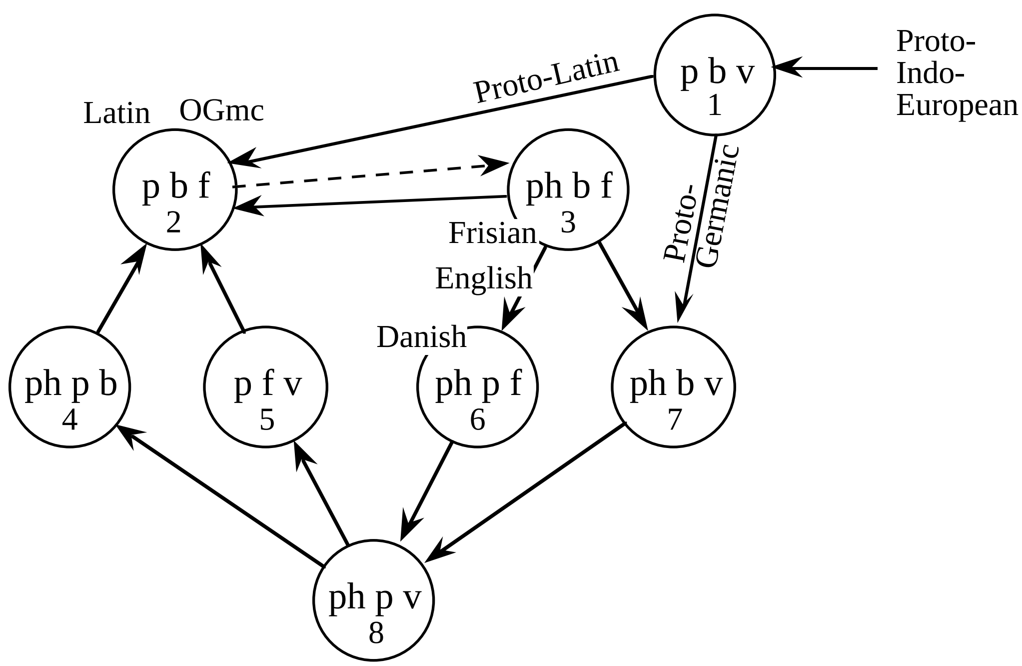

Figure 1. The directions of the first and second Germanic consonant shifts in word-initial position. The drawn arrows denote the changes that are favoured in our model for the labials. The two possible systems that are not shown are {b, v, f} and {ph, f, v}. They have only arrows pointing away from them. The velars feature the same favoured directions of sound change, except for one, which is shown by the dashed arrow. Explanations are found in the text.

5. Hierarchy of articulatory effort

The first hierarchy that we consider, is that of articulatory effort. Compare [pá] with [bá]. The strong word-initial accent brings about a high tone on [á]. The articulatory feature corresponding to this high tone is [stiff vocal cords]. You have to make your vocal cord tension high in order to produce the high tone. However, for the vocal cords to vibrate during [b], they have to be slack. That is because the condition for vocal cord vibration is something like:

transglottal pressure > glottal area * vocal cord tension

(For an elaborate treatment, see Halle & Stevens 1971). So in case of supralaryngeal closure as in [b], the transglottal pressure is low and the vocal cords will vibrate if the orifice is small enough or the vocal cords are slack enough. As a matter of fact, several tone languages require consonants to be transparent for tone. They accordingly implement a voiced plosive on a high tone in either of two ways: first, they could improve transglottal pressure by prenasalization [ᵐb], thus shortening the duration of supralaryngeal closure, without necessarily affecting the labial closure interval. Transglottal pressure can also be improved by lowering the larynx during closure, which increases the pressure beneath the glottis and decreases the pressure above. Alternatively, the vocal cords could be brought close together, that is, the consonant is glottalized. Implosives like [ɓ] combine the two, simultaneously lowering the larynx and strongly adducting the vocal cords. If a tone language has the plain [b] and nevertheless shows tone spreading, there is a good chance that the spreading of high tones is blocked by plain voiced obstruents, like the spreading of low tones is blocked by voiceless obstruents (Hyman & Schuh 1974).

So we will have to adjust the vocal cord tension to make a plain [b] voiced. This is the case in the Germanic languages, where consonants do not have to be transparent for tone. In Germanic languages that use tone at all, voiceless obstruents, for instance, are opaque for low tone spreading. So there is a vocal cord tension contour in [bá] and not in [pá]. That constitutes the first part of the articulatory effort hierarchy: [pá] is easier to pronounce than [bá].

The remaining parts of the articulatory effort hierarchy are the following. Plosives are easier to pronounce than fricatives. It is easier to run into a wall than to suddenly halt one inch in front of it (this is not my metaphor). So [pá] and [bá] are easier to pronounce than [fá] and [vá]. Furthermore, we will consider [v] to be an approximant, which is spontaneously voiced and therefore does not involve a vocal cord tension contour. The fact that [v] is an approximant further means that it requires less precision than [f] does, say, it could halt anywhere between one and five inches from the wall. The last contributor to our hierarchy of articulatory effort is [pʰ]. It involves an abduction of the vocal cords almost to the position during respiration and is therefore more difficult before a vowel than [p], which has no glottal contours at all.

Summary (the arrows point in the direction of the favoured sound change):

Articulatory features	p á	←	b á	←	v á	←	f á	←	pʰ á
slack vocal cords	–		+–		–		–		–
precision					*		**		
spread glottis	–		–		–		–		+–

6. Hierarchy of perceptual salience

This boils down to a hierarchy of perceptual contours (the arrows again point in the direction of the favoured sound change):

Perceptual features	p^h á	←	p á	=	f á	←	b á	←	v á
voiced	–+		–+		–+		+		+
continuant	–+		–+		+		–+		+
noise	+–		–		+–		–		–
number of contours	3		2		2		1		0

As for perceptual salience, [pá] is also preferred over [bá], as there is a voicing contour in [pá] and no voicing contour in [bá]. This means that the sound change [bá] → [pá] satisfies at least two of our three criteria, whether the perceptual distinctions between words improve or not. This again means that if we prohibit any actual merger, the sound change [bá] → [pá] is allowed to take place in all situations where the word [pá] is not already present. This is shown in figure 1 by the arrows pointing from the system numbered '3' (Frisian) to the system numbered '6' (Danish) and from system '7' to system '8'.

7. Hierarchy of perceptual distinctiveness

And now for our third criterion, auditory distinctions between words. Voiced sounds are acoustically marked in the sense that the presence of voicing is visible in the acoustic signal, whereas the absence of voicing entails the absence of something in the signal. The same goes for continuancy: continuant sounds show something (more energy) in the acoustic signal, and non-continuants do not. Accordingly, these two features can mask one another and it is this property that led us in the first place to build up our so-called universal inventory with only two voiced consonants and only two continuants, in order to reflect the larger contrasts that are possible within the voiceless consonants and within the stops.

So now we have three voiceless consonants and only two voiced ones. What we could expect, then, from minimal information considerations, is that the three voiceless consonants are perceptually more alike than the two voiced ones. Figure 2 shows that the unvoiced [p], [p^h] and [f] are closer together than the voiced [b] and [v]. It is still true, of course, that the distinction between [p] and [f] is greater than that between [b] and [v]. The figure also shows that the stops [b], [p] and [p^h] are closer together than the continuants [f] and [v], though, again, the contrast between [b] and [p^h] is larger than that between [f] and [v].

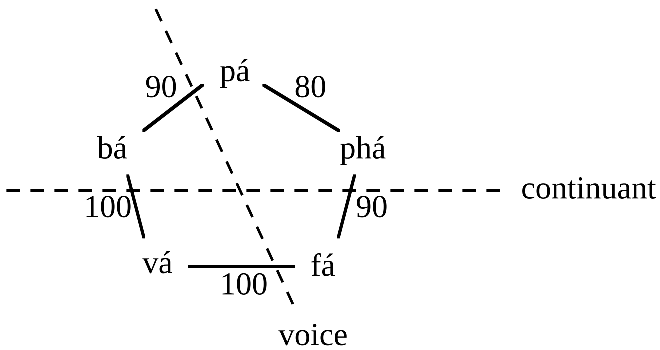

Figure 2. Perceptual distances between five labial consonants. The numbers represent nothing more than the ordering.

Let us assume that for sound inventories to be adjacent to one another, it is required that they differ in only one element and that the differing elements are adjacent in figure 2. Counted in this way, then, we have 15 pairs of adjacent systems. Figure 1 shows 11 of them: the number of drawn arrows in figure 1. As an example, consider the system that is numbered '8' in figure 1. It can change its [vá] into [bá] because articulatory effort and perceptual salience improve in that case. Alternatively, it can change its [p^há] to [fá] because that would improve articulatory effort and perceptual distinctiveness within the system: the smallest distance in the system increases from '80' (the distance between [p] and [p^h]) to '100' (the distance between [f] and [v]).

We can now explain the first Germanic consonant shift. If we assume after Proto-Indo-European a system like {*pa*, *ba*, *va*} (numbered '1' in figure 1), we see that there are two different routes to arrive at the favoured sound system {*pa*, *ba*, *fa*}. One route is quite direct, devoicing [v], which is what Latin may have done in prehistoric ages (nobody knows). The second route starts with aspirating the [p], which is what Proto-Germanic may have done. This language then necessarily obtained system '8', then '4' or '5', and then number '2'. This is why Latin and Old-Germanic both feature the favoured {p, b, f} system, though these systems are shifted relatively to one another.

There remains one problem. The two sound systems {*bá*, *fá*, *pá*} and {*bá*, *fá*, *p^há*} are adjacent to one another and seem equally dispersed in that in both cases the minimal distance within the system is '90' (refer to figure 2). In other words, [p] is equally distant from [b] as its proposed variant [p^h] is from [f]. Therefore, the criterion of perceptual distinctiveness cannot choose between these two sound systems. Furthermore, the system {*bá*, *fá*, *pá*} is the 'easier' one, and the system {*bá*, *fá*, *p^há*} is the more 'salient' one. This means that with the present amount of information, we cannot decide which of the two is the better sound system.

So we need one more bit of information to decide this issue. This bit is found in an asymmetry of the speech organs. The velar [g] is less voiced than the labial [b]. This is again due to the condition for vocal cord vibration that we invoked earlier. For the back closure of [g], the cavities above the glottis are filled earlier with air than in [b] and so voicing will stop earlier in [g] than in [b] due to the more rapid drop in transglottal pressure. This means that the voicing contrast between [g] and [k] is smaller than the voicing contrast between [b] and [p] or between [d] and [t]. This is why there are so many languages that do feature a [d]-[t] contrast without having a voiced counterpart to [k]. To mention a few of these: Dutch changed its [g] to the fricative [ɣ], Czech changed it to [ɦ], Japanese (McCawley 1968) to [ŋ] (in intervocalic position), and Arabic changed an older [g] to the palato-alveolar affricate [dʒ] (Moscati, Spitaler, Ullendorff and Von Soden 1964).

The difference that we found between anterior consonants and the velars, is shown in figure 3. Assuming that the mean distance between voiced and voiceless stops is '90' (what we must do in the absence of more information), the voicing contrast between anterior consonants must be thought of as being greater than '90', and that between the velars as being smaller than '90'. For the labials, then, system '3' in figure 1 will change into system '2', which is shown by the drawn line there. For the velars, however, the reverse is the case. This is shown by the dashed line in figure 1. We see here that the velars make a case for a circular optimization, that might never stop.

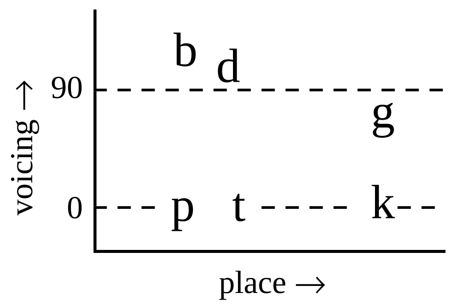

Figure 3. Different voicing contrasts due to an asymmetry in the speech organs.

We can now also explain the second Germanic consonant shift. It starts out from the system numbered '2' in figure 1 (an example is Western Dutch). First, the voiceless stops are aspirated, as has actually happened in almost all Germanic languages. Frisian is an example of a language in this phase (number '3'). This change must be caused, however, by the [g]-[k] confusion. If there is no [g], there is no aspiration, which is the case in Dutch. Secondly, the voiced stop is devoiced, as in Danish and Icelandic, or the voiceless fricative is voiced, as in some English dialects. The High German dialects affricated, instead of aspirated, their plosives, and went some steps further (as far as the system numbered '5' for intervocalic consonants). All these changes are blurred, of course, by such events as the replacement of [x] by [h], the presence of [w] and the nasals, and the particular problems with the dental fricatives.

8. Language universals.

Why did the other Indo-European languages not undergo the second consonant shift? First, the condition for strong favouring of voiceless plosives is only valid in strongly accented syllables. This withholds languages like Czech and French from these changes. The strong accent of Russian, on the other hand, is not always fixed on the same syllable, so a lot of paradigmatic pressure must be overcome before such a language will assume this kind of phonetic changes (though Russian appears to have no problem with paradigmatic vowel quality alternations).

A hypothesis that can be derived from our theory is the following:

> In languages that have a [g] and a [k], the [k] may become aspirated (of course, if there is not already a [kʰ]). The [p] and [t] may then, and only then, also become aspirated. Therefore, we do not expect many aspirates in languages with a gap at [g].

This hypothesis is corroborated by the data of the languages of the world, as they have been compiled in the UPSID database (Maddieson 1984), that contains information on the sounds of 317 languages from all over the world. 88 of these languages have a coronal and velar stop system of {t, d, k, g} and 13 have {t^h, d, k^h, g}. However, the languages with a gap at [g] show a slightly different pattern: 10 languages have a coronal and velar stop system of {t, d, k} whereas no language has {t^h, d, k^h}. So, of the languages with [g], 13% has aspirates, and of the languages with a gap at [g], 0% has aspirates. A χ^2-test gives an accident probability of 0.11. This is not to say, of course, that languages without [g] *cannot* have aspirates. Many Low German dialects are relevant examples (apparently not in UPSID). The theory suggests, however, that the change [g] → [ɣ] in these languages must have occurred *after* the [g] caused the [k] to become aspirated.

9. Conclusion

We found in this paper an explanation of the first Germanic consonant shift as taking a different route than Latin to the same favoured sound system, and an explanation of the second Germanic consonant shift as triggered by the aspiration that was caused by the confusion of [g] and [k]. Moreover, we found an implicational universal for consonant systems: if a language has [t^h], [k^h] and [d] but no [t] or [k], it has [g].

References

Brugmann, K. & Delbrück, B. (1897): *Grundriß der vergleichenden Grammatik der indogermanischen Sprachen*, vol. I, Einleitung und Lautlehre. Trübner, Strasbourg.

Halle, M. & Stevens, K.N. (1971): A note on laryngeal features, *MIT Research Laboratory of Electronics Quarterly Progress Report 101*, 198-213.

Hyman, L.M. and Schuh, R.G. (1974): Universals of tone rules: evidence from West-Africa, *Linguistic Inquiry* 5, 81-115.

Kawasaki, H. (1982): *An Acoustical Basis for Universal Constraints on Sound Sequences*. Dissertation, University of California, Berkeley.

Maddieson, I. (1984): *Patterns of Sounds*. Cambridge University Press.

McCawley, J.D. (1968): *The Phonological Component of a Grammar of Japanese*. Mouton, The Hague.

Meillet , A. (1922): *Introduction à l'étude comparative des langues indo-européennes*. Hachette, Paris.

Moscati, S.; Spitaler, A.; Ullendorff, E. and Soden, W. von (1964): *An Introduction to the Comparative Grammar of the Semitic Languages*. Otto Harrassowitz, Wiesbaden. 2nd printing, 1969.

Passy, P. (1890): *Etude sur les changements phonétiques et leurs caractères généraux*. Librairie Firmin - Didot, Paris.