

Why Make Life Hard? Resolutions to Problems of Rare and Difficult Sound Types

Author(s): Ian Maddieson

Proceedings of the Twenty-Fourth Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Phonetics and Phonological Universals (1998), pp. 367-380

Please see “How to cite” in the online sidebar for full citation information.

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via <http://linguistics.berkeley.edu/bls/>.

The Annual Proceedings of the Berkeley Linguistics Society is published online via [eLanguage](#), the Linguistic Society of America's digital publishing platform.

Why make life hard? Resolutions to problems of rare and difficult sound types

Ian Maddieson
University of California, Los Angeles

Introduction

Ideas inspired by the models of evolutionary biology have provided valuable insights into the shaping of the phonetic systems of human languages. One fruitful perspective is to view language as a field within which a variety of phonetic entities — the possible sounds that humans can make — are in competition for survival according to how well-adapted they are to linguistic function. The usual terms of this discussion are to view sounds, or systems of sounds, as being subject to twin pressures for articulatory economy and auditory distinctiveness. The 'fittest' are those which achieve a successful balance between these demands. Obviously the biological analogy is imperfect — there is nothing equivalent in sound systems to the genetic code — but nonetheless the notion that phonetic elements are in an ecological competition for survival is fruitful, and goes a long way toward providing an account of such things as why some sounds are more common and others are more rarely found in the sound inventories of the world's languages (see, for example, Lindblom 1980, Lindblom and Maddieson 1988).

One challenge which this approach presents, however, is to account for the stable occurrence — as part of the established system of a language, not as occasional variants, or transient 'experiments' during acquisition — of sounds that seem to fail to satisfy the ecological demands. In the biological world mal-adapted life forms die out quickly. How come phonetic elements which appear to be mal-adapted are able to have long life? This is the problem that will be the focus of this paper. It will start with an attempt to sketch a more satisfactory picture of the ecological demands that the phonetic patterns of language must satisfy. It will then discuss three classes of sounds found in traditional typologies that appear to be ecologically unfit for survival in light of these demands. An examination of three language-specific instances of sounds that putatively fall into these classes shows that in each of these cases the ecological damage is mitigated by arrangements of details of articulatory positions or timing.

Contrastivity and Connectedness

I have suggested elsewhere (Maddieson 1997) that the two fundamental 'ecological' conditions the phonetic patterns of language must satisfy are ones that may be conveniently subsumed under the labels Contrastivity and Connectedness.

Contrastivity is above all the requirement that a language must show differentiation in sound, rather than being an undifferentiated noise. Without differentiation, essentially only the message "I'm over here" can be transmitted and received (roughly the meaning of certain monotonous insect calls, or frog croaks). A language cannot be built unless messages can begin in different ways and continue in different ways, making possible the construction of a lexicon and the

other elements required. But mere difference is not enough. Both speakers and listeners need to be able to reliably identify the same message as being the same, and different messages as being different. This favors the selection of sound elements with stable characteristic 'signatures' in both their motor and auditory patterns so that they may be recognized and memorized. Note the stress placed here on the importance of distinctiveness among the motor patterns, not just on auditory distinctiveness. Production patterns must be as distinctive as auditory/perceptual patterns in order for a speaker to be able to encode the differences between one word (morpheme, etc) and another.

These requirements create the property that is usually labeled phonological contrast: utterances must contain parts that are differentiated from each other and recognizable when they recur. Contrastivity is also behind the alternation of louder and quieter sounds that is the basis of syllabification.

Equally, a language needs to be produced as a continuous stream, its parts connected to each other just as essentially as they must be differentiated from each other. This property is given the label Connectedness. Note that one of the implications of a connectedness requirement, since the position in which a given word or other element will occur is variable, is that the form of any item must be adapted for variable environments. The suggestion here is that it is the need to achieve connectedness — rather than an explicit minimization of articulatory effort — that favors limits on the range of articulatory gestures, especially in adjacent parts of an utterance. Making an extreme gesture, for example sticking the tongue out as far as possible between the teeth, makes it harder to connect to the next position (unless it happens to be the same, and variability of context ensures that it will more frequently not be the same). Connectedness therefore favors articulatory displacements close to the mean position rather than extreme ones, as well as moderate efforts to raise subglottal pressure, limited modulation of fundamental frequency, etc.

Variation in the articulatory instantiation, i.e. coarticulation, of a given item (word, etc.) facilitates its concatenation with other items in contexts that themselves vary. In the perceptual domain, coarticulatory variation spreads information over longer durations in the signal, and hence helps to reduce the likelihood of transmission error. Sound elements that resist coarticulation are therefore disfavored. However, beyond a certain degree, variation begins to conflict with the need to be able to identify repetitions of the same item, coming into conflict with the requirement of contrastivity. Variation will therefore tend to remain within limited bounds. (Of course, certain types of variation become conventionalized over a period of time and take on new roles in word identification, boundary marking, etc. In other words, variability is also a productive source of new contrasts.)

The two principles proposed here are characterized in very informal terms (though hardly less so than those usually given to the older formulations of articulatory ease and auditory distinctiveness). This is inevitable, given how little we know at this stage about such matters as how speech motor patterns are

represented in memory and how signals as complex as those in speech are perceptually processed. However, it is nonetheless possible to use these principles as the basis for making some predictions about expected and unexpected patterns. From this point on, the paper will focus on certain predictions concerning disfavored segments.

Disfavored segments

The requirement of contrastivity leads first to the expectation that languages will prefer sequences of sounds that contain sufficiently differentiated elements. However, because identifiability is also crucial, individual elements that are hard to identify as distinct in motor or auditory terms because they are inherently bland (non-salient), confusable with competing sounds, or produce unstable acoustic outputs from small variability in production details are among the kinds of segments that are predicted to be disfavored. For lack of connectedness it is elements which are resistant to contextual variation — for example, because they make use extreme gestures or require precise positioning of several articulators — which would be expected to be disfavored. Three classes of segments which, given their 'text-book' definitions, might be expected to be disfavored for one or both of these reasons are doubly-articulated plosives, doubly-articulated fricatives, and ejective fricatives. These classes are certainly comparatively rare, though all have been mentioned as occurring in at least some languages.

We might expect that doubly-articulated plosives are not well-connected since they seem poorly adaptable to context, given that two articulators are pre-empted. This is to extend to this case an argument that was first made in connection with clicks, another class of segments in which two articulators are employed. (Compare the 'back vowel constraint' limiting click-vowel co-occurrence in !Xóǀ (Traill 1985), and Sands' (1991) conclusion, based on acoustic measurements, that clicks in Xhosa do not coarticulate with surrounding vowel environments.) Perceptually, stops are primarily distinguished one from another by the transitions preceding and following them, and when a burst is present, by the spectrum of the burst. Given the text-book definition of doubly-articulated plosives as involving two simultaneous articulations, we might also expect that they are not well-contrasted since simultaneous articulations would make it hard to recover cues to the two separate motor gestures involved, impeding their identifiability. Moreover, given a target of simultaneity of articulation, small variations in inter-articulator timing would produce great variation in the acoustic signal, according to which gesture leads or lags the other and hence dominates the transitions and bursts.

In the case of doubly-articulated fricatives, there are reasons to think that the problems are more severe than for stops. Doubly-articulated fricatives might be thought to be not well-connected for the same reason as doubly-articulated stops: two articulators are involved and this impedes coarticulation. For fricatives, perceptual identification is usually taken to be primarily based on the spectral characteristics of the frication noise itself. The impedance of a fricative constriction creates an elevated air pressure level in the mouth behind the constriction as air flows from the lungs. The driving pressure of the air supply coming from the

lungs must be higher than this in order to maintain a sufficient flow of air past the constriction to generate the frication noise. Now, if there are two constrictions, as illustrated in Figure 1, then air flow past the further back of the two has to overcome the impedance created by the more forward of the two, demanding a higher driving pressure from the lungs. This higher subglottal pressure would only be required while this particular articulatory configuration was maintained and would be inappropriate for flanking sounds. Hence, there is an additional reason for thinking doubly-articulated fricatives are poorly-connected in speech.

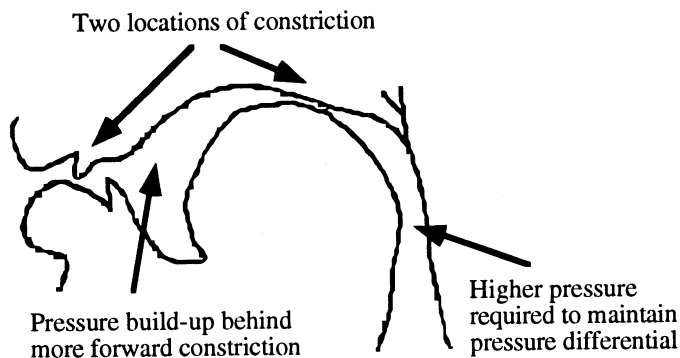


Figure 1. Schematic representation of production of a doubly-articulated fricative.

Moreover, doubly-articulated fricatives are not well-distinguished since the more forward of the constriction locations will act as an acoustic filter on the noise spectrum generated at the rear one, masking its place-specific characteristics, and making it hard to recover the articulatory gestures involved in production of the sound. We would expect doubly-articulated fricatives also to be acoustically unstable due to the very critical balance of various aerodynamic factors required to produce them; small variations around a mean would have greater consequences than would be the case for fricatives with only a single constriction.

Ejective fricatives, like other ejective segments, are judged to be poorly connected in speech because of the impossibility of producing a flow of speech using the ejective mechanism. By its nature, this can only be used for one segment-length interval before the laryngeal parameters of height and constriction must be re-set. However, ejective fricatives suffer from a further problem in that, in their textbook form, they would be of necessarily short duration and, for this reason, poorly distinctive. This is because raising the larynx with the glottis closed — the ejective mechanism— produces only a small change in oral cavity volume and this volume change is sufficient for brief frication only, as illustrated in Figure 2.

Based on data in the literature it is possible to make an approximate estimate of the anticipated duration of an ejective fricative. The average volume velocity estimates for the air flow in a pulmonic voiceless oral fricative are about $300 \text{ cm}^3/\text{s}$

(200-400 cm³/s in Shadle (1997); 330 cm³/s in Catford (1977)). An average fricative of 100 ms duration would therefore require a flow volume on the order of 30 cm³. We estimate the approximate magnitude of ejective larynx raising to be on the order of 1-2 cm and, using the volumes of the lower pharyngeal area shown in Baer et al's (1991) MRI study as a guide, the consequent reduction in the volume in the oral cavity to be 10 cm³ or less. This would produce enough outward air flow to sustain an ejective fricative with a duration on the order of 30 ms only, and quite likely less, as the estimates probably err on the generous side.

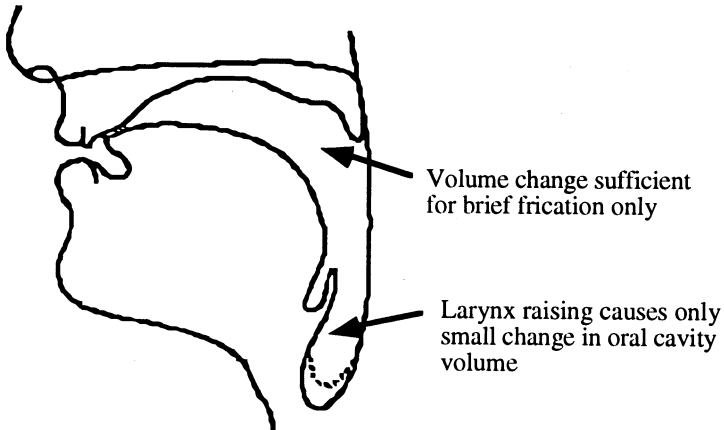


Figure 2. Schematic representation of production of an ejective fricative.

Thus, all three sound types appear disfavored for lack of contrastivity and connectedness. The severity of the problems mentioned also seem to correlate quite well with reported frequency. Doubly-articulated plosives are much rarer than singly-articulated ones; and the more seriously problematic doubly-articulated fricatives have been claimed to exist in only a handful of languages. Ejective fricatives are reported in far fewer languages than are ejective stops, and both are very much rarer than their pulmonic counterparts.

However, text-book descriptions and rough theoretical calculations do not always tell the whole story. As will be shown below, in individual cases the phonetic details concerning the way in which a 'text-book' category is realized show important deviations from standard descriptions. One might simply conclude that there are fewer cases of the textbook type than expected (and doubly-articulated fricatives may not exist at all, as suggested by Ladefoged and Maddieson (1996)), and leave it at that.

The perspective suggested here is to view the three types of 'ecologically unfit' segments discussed above in terms of the combinations of properties they represent, respectively:

— transitional cues to two places of articulation (doubly-articulated plosives)

— inherent cues to two places of articulation (doubly-articulated fricatives)

— laryngeal constriction + frication (ejective fricatives)

and to show how these combinations can get off the endangered list if they are 'partitioned', e.g. in the temporal or strictural domains, while retaining the organization typical of a single segment.

Doubly-articulated plosives

Segments written /kp/, /gb/ occur in many West African languages (and in some languages of New Guinea). These are typically described as doubly-articulated plosives with simultaneous bilabial and velar closures. A classic description of a /kp/ segment is that offered by Westermann and Ward (1933: 58):

"The two articulations must be simultaneous, i.e. when the sound occurs between two vowels there must be no onglide to the k heard before the lips come together for the p position."

This description is closely echoed by Ladefoged (1968: 9) who talks of labial-velar stops produced by "the simultaneous articulation of k and p or g and b".

We pointed out above the problems that strict simultaneity of articulations would engender. However, if the two articulatory gestures are slightly offset in time the presence of two articulations can be signaled much more strongly, since distinct onset and offset transitions can be heard. The risk of excess variability is also reduced by fixing the relative timing of the articulations. Maddieson (1993) showed that in Ewe the two articulatory gestures involved are indeed slightly offset in time in just this way, with the velar leading the labial one by some 20 ms.

Data on lip and tongue movement was collected using electromagnetic articulography from two Ewe speakers from Kpando, Ghana, reading short phrases including, among others, the following words:

ákpá	<i>too much</i>	ekpé	<i>stone</i>
àgbà	<i>load, trouble</i>	ekpo	<i>log</i>

Data on only one speaker will be shown here, but the patterns are the same for both. The onset and release of closure for the stops /kp/ and /gb/ were determined for each word from the acoustic records, and the movement tracks averaged for ten repetitions, aligned at the instant of acoustic release. Derivative measures, such as the distance between the upper and lower lip receivers (inter-lip distance) were also calculated. Figures 3 and 4 show results for ákpá and àgbà. Averaged vertical movement tracks over time of the tongue back and the lower lip, as well as the derived inter-lip distance measure (which effectively adds the much smaller movement of the upper lip to that of the lower lip) are plotted against time. The time interval shown is 400 ms, with the acoustically-determined release instant at 300 ms from the start of the window. In order to show the movements in a normalized space, the actual articulator heights were converted to standard scores. In both cases, the velar and labial gestures are highly overlapped, but are not simultaneous. Auditorily the onset sounds velar and the release labial. The production of these segments must explicitly be made contrary to the specification given by Westermann and Ward.

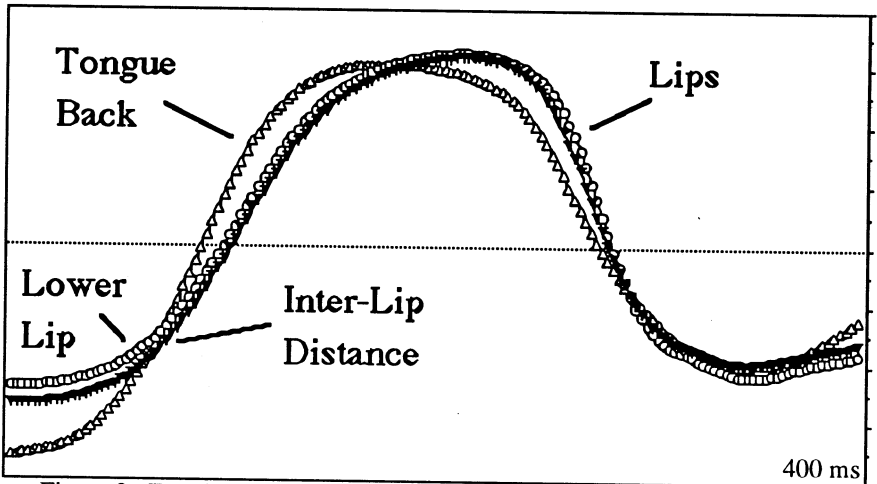


Figure 3. Ewe /akpa/ mean of ten repetitions. (Duration shown is 400 ms).

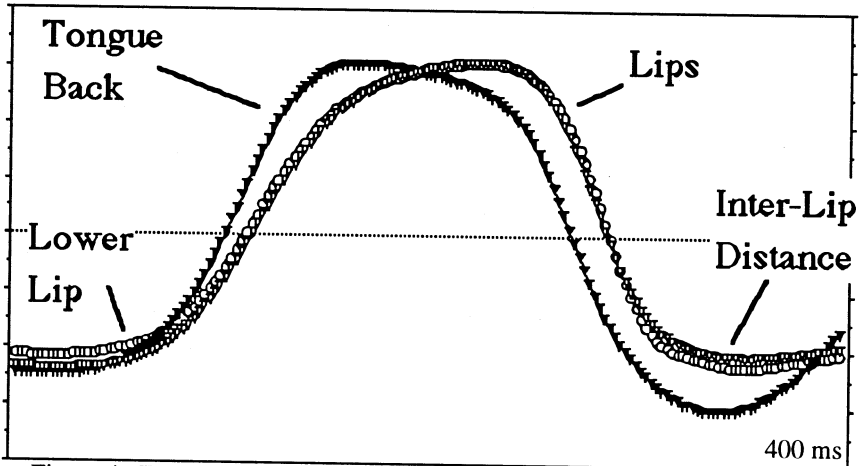


Figure 4. Ewe /agba/ mean of ten repetitions. (Duration shown is 400 ms).

The timing offset of the two articulations leads to a slightly longer total closure phase for labial-velar stops than for singly-articulated ones, as shown in Table 1.

Table 1. Mean consonant closure durations (in ms)

Velar:	/k/ 142	/g/ 133
Bilabial:	/p/ 158	/b/ 150
Labial-velar	/kp/ 174	/gb/ 179

Labial-velars are not, however, a sequence of separate segmental articulations, but a coordinated single entity, with very little timing variation. After averaging the

standard deviation of articulator positions for any time point is on the order of .1 to .2 mm, which is close to the accuracy limits of the measurement technique used. Contrast the English sequence /k/ + /p/, shown in Figure 5, which has generally less overlap of the two articulations and more variation in their timing.

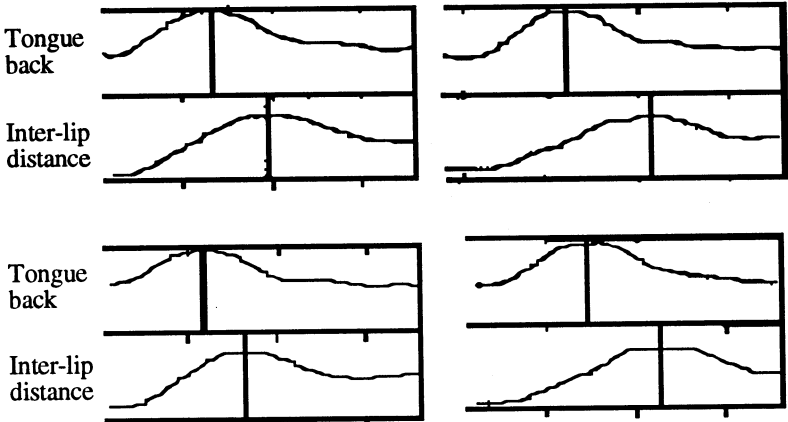


Figure 5. Movement traces (from x-ray microbeam) of 4 repetitions of the /k + p/ sequence in the English phrase 'cock puddles'. Vertical lines show the consonant centers for the velar and labial gestures involved. (Data courtesy of Louis Goldstein, Haskins Laboratories.)

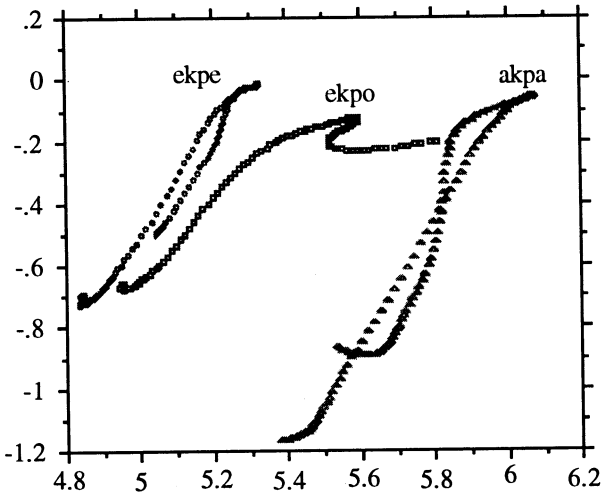


Figure 6. Coarticulation of Ewe labial-velars. Movement of the tongue back in two dimensions as a function of vowel environments. Distances on vertical and horizontal scales are in cm. The front is to the left of the figure.

Not only are the labial and velar gestures non-simultaneous, but contrary to a priori expectations, the velar component of labial-velars co-articulates well with the tongue body position of different vowel environments. This is shown in Figure 6. The same point on the tongue makes contact with the palate at three different locations over a span of almost a centimeter depending on the vowels in the three words shown.

Doubly-articulated fricatives

Avatime, another language of Ghana, according to Ford (1988), has a consonant inventory which seems to show a regular pattern of combining velar and bilabial articulations, so that a partial chart reads as follows:

p	k	kp
b	g	gb
ɸ	x	xɸ
β	ɣ	ɣβ

In view of the severe problems noted in producing doubly-articulated fricatives, it would be surprising if the combination of labial and velar articulation in fricatives was produced as indicated (Ladefoged (1964) wrote ɸ and ɣβ for xɸ and ɣβ but also describes these segments as fricatives).

As is shown in more detail in Maddieson (1998) the labial constriction in the fricatives which are [labial] + [dorsal] is not narrow enough to constrict air flow, and hence does not impede frication at the back closure. In fact, the nature of this articulation is quite different from that for a simple bilabial fricative, involving rounding and forward projection of the lips rather than a narrow approximation along their length. In other words, the putative xɸ , ɣβ are actually x^w , ɣ^w .

The method employed to study these productions was to video-tape the lip articulations of 4 speakers as they produced a variety of words containing labial articulations. Dots were placed on the outer lip surfaces to facilitate measurement. A frame from the videotape showing one speaker's most constricted lip position in the word /ax^wa/ "charcoal" is shown in Figure 7. (The video image has been dithered to improve printability.)

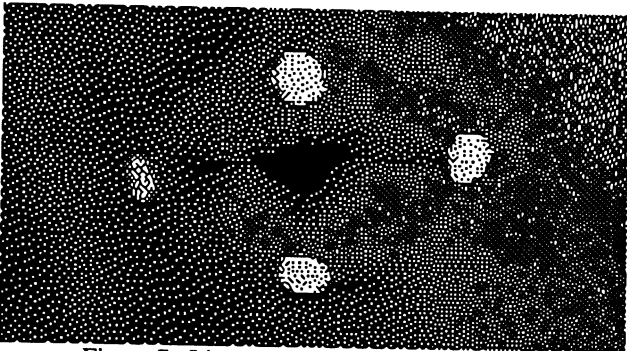


Figure 7. Lip-rounding in Avatime /x^w/.

We thus see that although frication at two locations is absent, a more general feature of labiality is easily combined with frication at another location by making it a vowel-like articulation with long-range transitional effects on adjoining vowels.

Ejective fricatives

For ejective fricatives, the problem is combining glottal constriction with sufficient duration of frication to cue place. In Yapese, an Austronesian language spoken in the Micronesian state of Yap, Hsu (1969) and Jensen (1977) specifically describe both the glottalized stops and fricatives as ejective. In his grammar, intended to be reasonably intelligible to a lay reader, Jensen puts it as follows:

“[...] for p', in contrast with p, while the lips are closed the glottis is also closed. With the lips closed, the closed glottis is raised very rapidly and suddenly. This motion compresses the air which is trapped in the mouth between the closed glottis and the closed lips, so that when the lips are suddenly opened, a popping sound is made. [...] Only then is the closed glottis opened. [...] The other glottalized stops are made in the same way, except [for place of articulation].” (Jensen 1977: 28)

“Glottalized fricatives are pronounced in a similar way to glottalized stops. [...] In pronouncing f', the lips are not closed completely. Rather, the lower lip is pressed against the upper teeth in such a way as to restrict the flow of air. So when the closed glottis is raised (as for p'), the air is forced out [...] making a hissing (friction) noise. Then the glottis is opened. th' is pronounced in the same way, except that the restriction [...] is produced at the teeth.” (Jensen 1977: 31)

The production of the glottalized fricatives was examined using both audio recordings and aerodynamic data provided by a number of speakers of Yapese. The principal words used as examples of the sounds being examined are the following (maintaining the use of ['] to mark the glottalized category):

f	faaŋ	<i>platform</i>	f'	f'aaŋ	<i>eel</i>
f	faak	<i>his child</i>	f'	f'œθ	<i>apportion</i>
θ	θaam	<i>outrigger</i>	θ'	θ'aab	<i>cut</i>
θ	θiin	<i>language</i>	θ'	θ'iib	<i>pot</i>
θ	maaθ	<i>touch</i>	θ'	maaθ'	<i>severed</i>

Measurements of the frication duration showed that plain and glottalized fricatives had similar noise durations, contrary to the theoretical expectations for 'text-book' ejective fricatives. Mean frication durations for the four fricatives compared are shown in Figure 8. There is a salient difference in duration according to place, with the dentals longer than their labio-dental counterparts, but no significant difference between the pulmonic and glottalized categories.

Analysis of the production of the glottalized category shows that in fact these segments are not ejectives. Audio waveforms, and oral flow and intra-oral air

pressure traces are shown in Figures 9 and 10 for representative tokens of /f/ and /f'/ in word-initial position produced by a female speaker. Comparison of the oral pressure traces shows patterns corresponding to the fricative portions that are quite similar in amplitude and duration, giving them similar overall shapes. Differences lie in the fact that the pressure rise occurs considerably earlier with respect to vowel onset in Figure 10, and is then followed by a flat (silent) interval and a glottalized onset to the vowel. In other words, /f'/ is a sequence of an ordinary pulmonic fricative and a glottal stop.

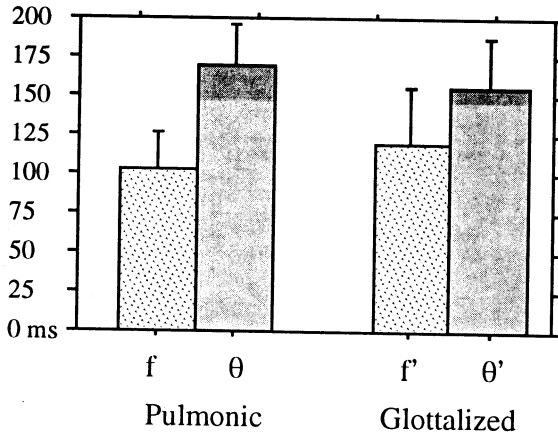


Figure 8. Friction noise durations of four Yapese fricatives. Mean durations of 2-4 tokens per category from each of three speakers.

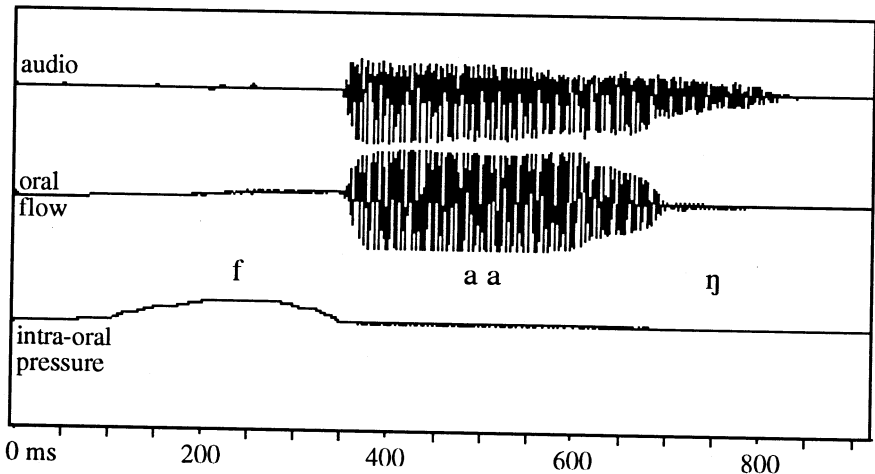


Figure 9. Aerodynamic records illustrating the Yapese word /faan/ 'platform'.

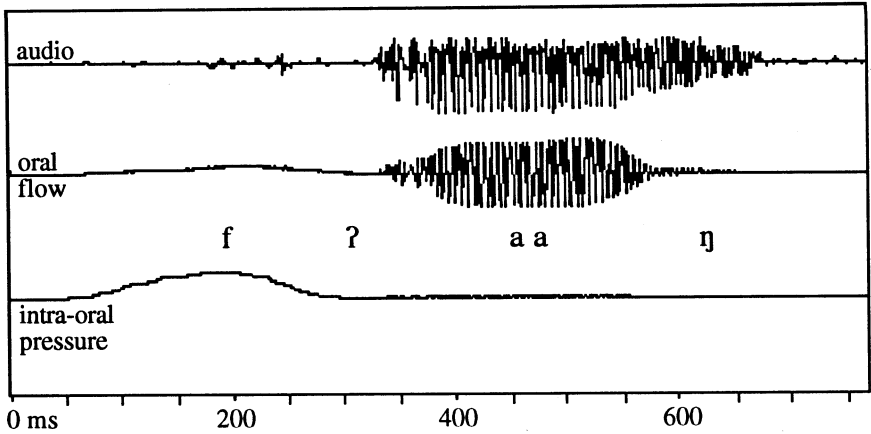


Figure 10. Aerodynamic records illustrating the Yapese word /f'aary/ 'eel'

Yapese does indeed use the ejective mechanism in its stops. Figure 11 shows records of a word beginning with /p'/. With the oral passage closed the ejective mechanism generates quite high intra-oral pressures (MacEachern 1996), higher than those typically generated by air pressure from the lungs during speech. Measurements of the peak intra-oral pressure during selected fricatives and stops in Yapese are shown in Figure 12 separated by speaker. Note that the peak pressure is very similar for all the segments in which this peak value is attributed to pulmonic effort (all fricatives and the pulmonic stop /p/), but is much higher for /p'/. (These means are in fact an underestimate because some ejective stops showed clipping of the peak value, as in Figure 11.)

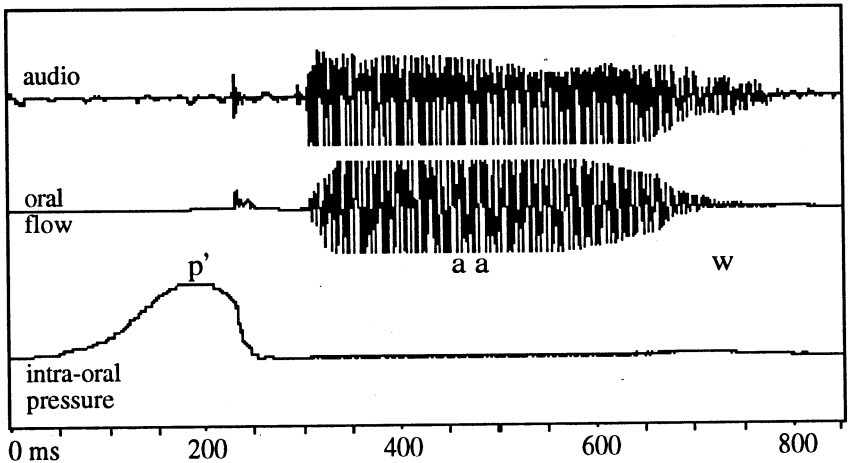


Figure 11. Aerodynamic records illustrating the Yapese word /p'aaw/ 'banana'

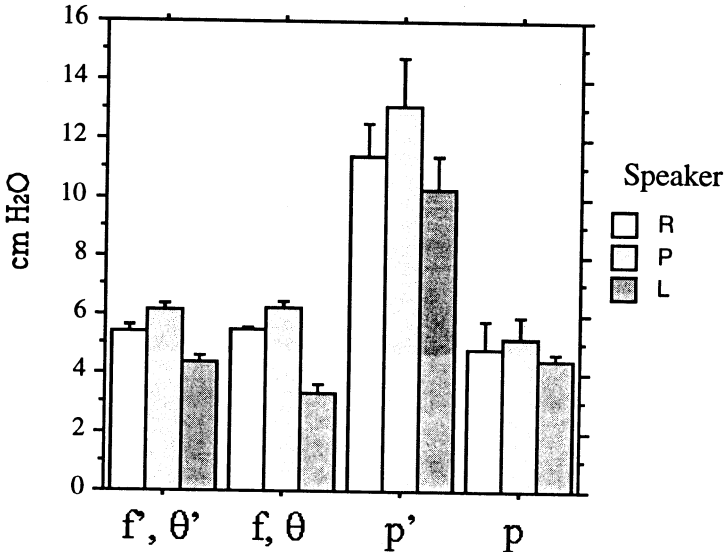


Figure 12. Peak intra-oral pressure: ejective stops vs fricatives and pulmonic stops (means of 3-12 tokens per category per speaker).

Although the Yapese solution is not the only way to combine glottal constriction and frication (see Maddieson, Bessell, and Smith 1996 for a different strategy employed in Tlingit), it is one way to be able to generate sufficient frication duration for place identification while maintaining distinctiveness from plain fricatives.

Conclusion

The three sound types which appeared disfavored for lack of contrastivity and connectedness if we took at face value the descriptions offered, are shown in specific cases to be organized so that the expected problems do not arise. In general the solution is to have a 'partition' of the components so the constituents are not so hard to identify or to produce:

- transitional cues to two places of articulation can be provided by having overlapping but temporally offset articulations
- the problem of providing inherent cues to two places of articulation in a fricative can be avoided by using different degrees of stricture
- laryngeal constriction + frication can be easily combined by having non-overlapping articulations.

In short, the difficulty of segments and their fitness for survival cannot always be appropriately determined from standard descriptions. We suggest that understanding of what conditions are truly imposed on the sound structure of language can be enhanced by detailed examinations of what happens in individual languages. It turns out life may not always be so hard as we make it appear.

References

- Baer, T., Gore, J. C., Gracco, L. C. and Nye, P. W. 1991. Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels. *Journal of the Acoustical Society of America* 90: 799-828.
- Catford, J. C. 1977. *Fundamental Problems of Phonetics*. Indiana University Press, Bloomington, IN.
- Hsu, R. W. 1969. *Phonology and Morphophonemics of Yapese*. Ph. D. dissertation. University of California, Berkeley. Pp 183.
- Jensen, J. Thayer (with L. D. Pugram, J. B. Iou, and R. Defeg). 1977. *Yapese Reference Grammar* (PALI Language Texts: Micronesia). University of Hawaii Press, Honolulu.
- Ladefoged, P. and Maddieson, I. 1996. *The Sounds of the World's Languages*. Blackwell, Oxford.
- Lindblom, B. 1980. Models of phonetic variation and selection. *PERILUS* 11: 65-100.
- Lindblom, B. and Maddieson, I. 1988. Phonetic universals in consonant systems. In *Language, Speech and Mind: Studies in Honor of Victoria A. Fromkin*, ed. L. M. Hyman and C. N. Li. Routledge, London and New York: 62-80.
- MacEachern, M. R. 1996. Computer modeling of the production of ejectives using aerodynamic and acoustic data (Abstract). *JASA* 100: 2661.
- Maddieson, I. 1997. Phonetic universals. In *Handbook of Phonetic Sciences*, ed. W. J. Hardcastle and J. Laver. Blackwell, Oxford:
- Maddieson, I. 1998. Disappearing vowel harmony and doubly-articulated fricatives: two myths about Avatime phonology. In *Language History and Linguistic Description in Africa*, ed. I. Maddieson and T. Hinnebusch. Africa World Press, Trenton, NJ.
- Maddieson, I., Bessell, N. and Smith, C. L. 1996. A Preliminary Report on the Phonetics of Tlingit. *UCLA Working Papers in Phonetics* 93: 125-148.
- Sands, B. 1991. Evidence for click features: Acoustic characteristics of Xhosa clicks. *UCLA Working Papers in Phonetics* 80: 6-37.
- Shadle, C. H. 1997. The aerodynamics of speech. In *Handbook of Phonetic Sciences*, ed. W. J. Hardcastle and J. Laver. Blackwell, Oxford: 33-64.
- Trail, A. A. 1985. *Phonetic and Phonological Studies of !Xóõ Bushman*. Helmut Buske, Köln.