

An Ethological Perspective on Common Cross-Language Utilization of F_0 of Voice¹

John J. Ohala²

Phonology Laboratory, Department of Linguistics, University of California, Berkeley, Calif., USA

Abstract. The author suggests that the following seemingly disparate phenomena have an underlying relationship: (a) cross-language similarities in the intonation contours for statements versus questions, (b) cross-cultural similarities in the vocal expression via intonation of attitude and affect, (c) cross-language patterns in the use of tone, vowels, and consonants in 'sound symbolic' vocabulary, (d) cross-species use of F_0 in threatening or non threatening vocalizations, (e) cross-cultural and cross-species use of certain facial expressions (involving distinct mouth shape), and (f) the existence of sexual dimorphism in the vocal anatomy of humans (and certain non humans). He argues that all arise due to an innately specified 'frequency code', which associates high acoustic frequency with the primary meaning of 'small vocalizer' and thus such secondary meanings as 'subordinate, submissive, non threatening, desirous of the receiver's goodwill, etc.' and associates with low acoustic frequency the primary meaning of 'large vocalizer' and such secondary meanings as 'dominant, aggressive, threatening, etc.'

Introduction

In this paper I propose that some of the global uses of speech prosody which are similar cross-linguistically and cross-cultu-

rally can best be understood by viewing them from a comparative ethological perspective, specifically, by seeing how they conform to general principles governing the form of agonistic displays in a wide variety of species. To support my main hypothesis about the origins and functions of speech prosody I offer further hypotheses that other types of signals, verbal and nonverbal, visual and auditory, in humans and nonhumans, have similar histories and functions. I would liken what is presented here to a piece of *patola* fabric – a style of weaving done in Gujarat, India. In this

¹ Revised version of a paper read at a meeting in Zurich, April 6-8, 1983, on 'Prosody, Normal and Abnormal. An Interdisciplinary Symposium on Suprasegmentals of Speech', which was organized by Dr. D. Weniger under the auspices of the 'Association Européenne de Psycholinguistique' and funded by the 'Schweizerischer Nationalfonds'.

² I gratefully acknowledge the support of the Harry Frank Guggenheim Foundation for the research reported here.

method of weaving the threads first receive their coloring by being tied and dyed and subsequently woven together to form a pre-planned pattern. Each of my hypotheses is like one of these threads – by itself not always strong and showing a not very remarkable pattern. However, when the hypotheses, like the threads, are woven together they reinforce each other both structurally and in a way that reveals a striking, perhaps even a pleasing, pattern. I realize, of course, that the strength of the fabric needs to be tested and evaluated on more than esthetic grounds; this remains as a task for the future.

I will assume that most of the facts about how speech prosody is used for the expression of certain basic meanings are well known and have been adequately documented in the literature. I will therefore give most attention to establishing connections between these facts and facts about other forms of communication [Ohala, 1980, 1982a, b, c, 1983].

Thread 1: Universal Tendencies for Fundamental Frequency (F_0) and Sentence Type

The first thread on the loom is the often-noted tendency for languages to use high and/or rising F_0 to mark questions – especially yes-no questions – and low and/or falling F_0 to mark statements [Hermann, 1942; Bolinger, 1964, 1978; Ultan, 1969]. Although there are exceptions to this pattern, the high cross-language incidence of this particular sound-meaning correlation makes it quite unlike the typically arbitrary sound-meaning correlation that exists for most lexical and grammatical entities.

Moreover, the pattern is too widespread to be explained by borrowing, descent from a common linguistic source, or chance. It follows that there is something common to all human speakers, at all stages in history, which creates this phenomenon. Nevertheless, attempts to explain it by reference to universal physiological constraints [Lieberman, 1967] have so far not been successful [Ohala, 1970, 1977, 1978, 1982a].

Thread 2: The ‘Affective’ Use of F_0

Anecdotal and experimental evidence are in general agreement that there are cross-culturally similar uses of F_0 to signal affect, intention, or emotion. There are numerous terminological, conceptual, and methodological problems in this area, however. What are the nonlinguistic messages which can be conveyed by the voice? Are these signals under voluntary control? How can one obtain natural samples of them? What labels for these messages should one use when instructing listeners to judge how well a given speech sample embodies them? Nevertheless, although the evidence is not as extensive as that concerned with the use of F_0 to mark sentence types, it seems safe to conclude that such ‘social’ messages as deference, politeness, submission, lack of confidence, are signaled by high and/or rising F_0 whereas assertiveness, authority, aggression, confidence, threat, are conveyed by low and/or falling F_0 [Bolinger, 1964].

The experimental literature reveals some apparent conflict on this point, however. Whereas Apple et al. [1979] found that a higher F_0 of voice made a speaker sound ‘less truthful, less emphatic, and less “po-tent” (smaller) and more nervous’ [cf. com-

parable results, but with different labels, by Fairbanks and Pronovost, 1939; Williams and Stevens, 1972; Brown et al., 1974; Udall, 1960, 1964], Scherer et al. [1973] found higher maximum F_0 of voice associated with greater confidence in some cases. The conflict may be only superficial due to different experimental and measurement procedures. Apple et al. [1979] obtained listeners' evaluations of natural speech samples which were resynthesized with an overall upshifted, downshifted, or unaltered F_0 , with all other parameters left unchanged (except duration, in one condition). Scherer et al. [1973] presented listeners with unaltered samples of speech which were allowed to vary naturally in a variety of acoustic parameters from one token to the next. Under these circumstances they found peak F_0 to show an occasional correlation with listeners' perception of greater confidence.

To attempt to resolve this conflict I conducted the following study. Short samples (≤ 4 s) of spontaneous speech produced by 2 male and 2 female adult speakers of American English were digitally processed in such a way as to remove all spectral details but to retain the original amplitude and F_0 contour, the latter of which was either linearly upshifted or downshifted by varying amounts or left unchanged. In this way the sex of the speakers and the actual linguistic content of the sentences were completely masked. These samples of 'stripped speech' were presented in pairs to American English-speaking listeners who were asked to judge which voice of each pair sounded more dominant or self-confident. The results indicate that, other things being equal, lower F_0 does make a voice sound more dominant. This agrees with the results of Apple et al. [1979]. This is evident, for example,

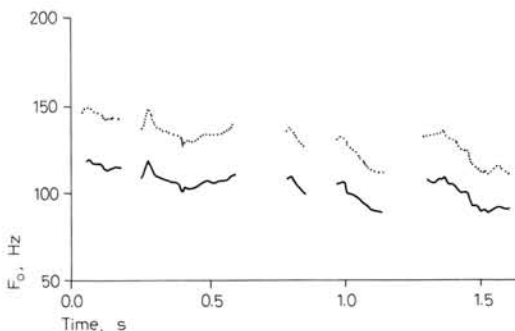


Fig. 1. The F_0 contours of two samples of 'stripped speech' (see text) presented as a pair to listeners to determine the one which sounded 'more dominant, more self-confident'. The contour depicted as a dotted line was identical to that depicted by the solid line except that it was upshifted in frequency by a factor of 1.25. The latter was judged 'more dominant' in 92% of the judgements.

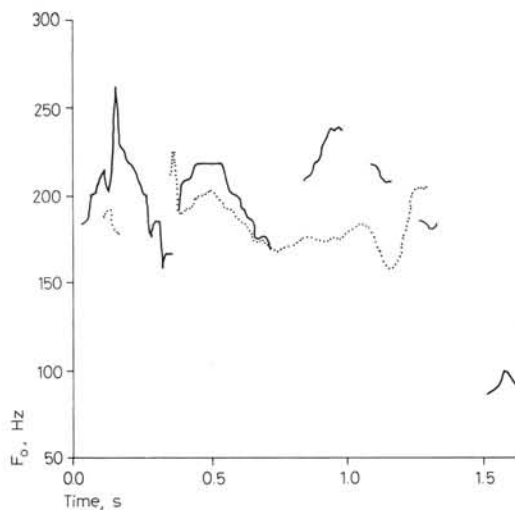


Fig. 2. The F_0 contours of two samples of 'stripped speech' (see text) presented as a pair to listeners to determine the one which sounded 'more dominant, more self-confident'. The contour depicted by the solid line was judged 'more dominant' than that depicted by the dotted line in 92% of the judgements.

in the judgements for the two samples presented graphically in figure 1, which are derived from the same speech sample but with one of them upshifted from the original by a factor of 1.25 (when F_0 is expressed in Hertz). The sample with the lower F_0 was judged more dominant than the sample with the higher F_0 by 92% of the listeners. However, when 'other things' were not equal, the one feature which contributed most to making a voice dominant was a steep terminal fall in F_0 . This is shown in figure 2 where the sample shown as a solid line, even though it has higher peak F_0 , was judged as sounding more dominant (92% of all judgements) than the sample shown as a dotted line, even though the latter is lower in F_0 during most of its duration. The sharp F_0 terminal fall, lacking in the other sample, seemed to be the determining factor in listeners' evaluations. This result is compatible with those of Scherer et al. [1973], but it suggests that the occasionally higher peak F_0 in the voices exhibiting greater confidence is there in order to make the terminal fall seem to be even steeper, i.e., by virtue of having fallen from a greater height.

Thread 3: Tone in Sound Symbolism

The documentation is not extensive, but there is an apparent cross-language tendency in certain tone languages to use tone systematically in a 'sound symbolic' way [Westermann, 1927]. Specifically, high tone tends to be associated with words denoting or connoting SMALL (and related concepts such as DIMINUTIVE, FAMILIAR, NEAR, or NARROW), whereas low tone is associated with the notion LARGE, etc. [Ohala, 1982a, 1983]. Some examples are given in table I.

Table I. Examples of the 'sound symbolic' use of tone

Language	'Small' and high tone	'Large' and low tone
Ewe	[kitsikítsi] 'small'	[gbàgbàgbà] 'be large'
Yoruba	[biri] 'be small'	[biri] 'be large'
Cantonese	[to 21] 'terrace, stage'	[to 215] 'table'

Thread 4: F_0 in Nonhuman Vocalizations

A systematic F_0 -meaning correlation is also found in the vocal signals of other species. Morton [1977] documented the existence of a remarkable cross-species similarity in the form-function relationship of the acoustic component of 'close-contact agonistic displays' (i.e., the signals given during face-to-face competitive encounters). The sounds made by a confident aggressor (or one who wanted to appear so) are typically rough and have a low F_0 ; submissive or non-threatening individuals' cries are typically tone-like and have a high F_0 . The dog's threatening growl and submissive whine or yelp are familiar examples of this. The same pattern is found in vocalizing species as diverse as the chickadee, the Indian rhinoceros, and the frog. Morton [1977] provided the following explanation for this sound-function correlation:

Animals in competition for some resource try to intimidate their opponent by, among other things, trying to appear as large as possible (because the larger individual would have an advantage if, as a last resort, the matter has to be settled by actual combat). Size (or apparent size) is primarily

conveyed by visual means, e.g., erecting the hair or feathers and other appendages (ears, tail feathers, wings), so that the signaler subtends a larger angle in the vision of the receiver. Some animals have even developed permanent (i.e., nonplastic) size markers, e.g., the bison's and gnu's hump, the mane of the male lion, and the growth of hair around the perimeter of the face in so many primate species, including male humans [Guthrie, 1970]. As Morton [1977] points out, however, the F_0 of voice can also indirectly convey an impression of the size of the signaler since F_0 , other things being equal, is inversely related to the mass of the vibrating membrane (vocal cords in mammals, syrinx in birds) which, in turn, is correlated with overall body mass. Also, the more massive the vibrating membrane, the more likely it is that secondary vibrations could arise thus giving rise to an irregular or 'rough' voice quality. To give the impression of being large and dangerous, then, an antagonist should produce a vocalization as rough and as low in F_0 as possible. On the other hand, to seem small and nonthreatening a vocalization which is tone-like and high in F_0 is called for. It is also possible in some cases that this latter behavior represents a form of infant mimicry. If so, this is a particularly effective way of pacifying a would-be aggressor since, for obvious reason, natural selection has left most species with a very strong inhibition against harming conspecific infants.

Morton's [1977] analysis, then, has the advantage that it provides the same motivational basis for the form of these vocalizations as had previously been given to elements of visual displays, i.e., that they convey an impression of the *size* of the signaler.

This analysis of the sound-meaning correlations in nonhuman vocalizations suggests an explanation for the three phenomena mentioned above. Its application to the affective use of F_0 is obvious: the use of low F_0 of voice to communicate aggression, assertiveness, dominance, etc. and high F_0 to convey social subordination, politeness, nonthreat, etc. parallels almost exactly the function of F_0 in the nonhuman cries. In the case of the typical F_0 contours for question and statement, one need only allow that the person asking a question is, from an informational standpoint, in need of the goodwill and cooperation of the receiver. The questioner, as it were, is appealing to the addressee for help. The high-pitched whine of the loser (or anticipated loser) of a battle has much the same meaning. The person making a statement is self-sufficient – again, from an information standpoint. Thus the F_0 used should be, and *is*, just the opposite to that found in questions. The F_0 of voice is used, as it were, as a *gesture* which *accompanies* or *is superimposed* on the linguistic message in order to enhance, elaborate, or even, in some cases, to contradict its meaning. In much the same way we use kinesic signals ('body language') to modify the meaning of our verbal messages. In fact, the rise and fall of F_0 during speech often parallels, both literally and functionally, the rise and fall of some speakers' eyebrows.

The explanation for the systematic use of F_0 in the choice of tones in sound symbolism is somewhat more problematic. The Yoruba speaker who utters the word /biri/ and /biri/ is presumably not trying to appear small and large, respectively, or even submissive or dominant. Rather it is the size of the referent of the word which is symbol-

ized by the tone. But there is still this common element: F_0 is used to make the receiver *react* as if something in the environment is small (or large, as the case may be). If the purpose of communication is to effect a change in the receiver – in higher animals one might say a change in the ‘cognitive map’ of the receiver [MacKay, 1969] – then the use of different extremes of frequency in the signal is quite an effective way to accomplish this, whether with an emotive or denotative intent.

I think the amazing cross-language and cross-cultural similarity of these uses of F_0 represent by *themselves* a strong argument for their being innately determined. I do not think that the consistency we find in the shape and meaning of these signals could result from a culturally maintained template. To see why this is so, consider that the phonetic shape of the bulk of any language’s vocabulary *is* maintained by a cultural template, but since the sound-meaning correlation is arbitrary, this template is subject to gradual distortion with the passage of time; thus, sound change gives rise to such radically different pronunciations as English /kao/ ‘cow’ and French /bœf/ ‘boeuf’, both of which had a common pronunciation a few millenia ago. In contrast, the sound-meaning correlation found in intonation and in sound symbolic vocabulary seem to be less subject to distortions.

In the remaining sections I will weave a few more threads into the fabric of my argument (although they will not concern prosody, *per se*), first, because these additional threads reinforce the patterns already described and, second, because they help to support my contention that this code is innate.

Thread 5: Facial Expressions

It is an old observation – in fact, at least since Darwin’s 1872 work, *The Expression of Emotions in Man and Animals* – that humans and nonhumans show certain similarities in their facial expressions. Although Darwin’s evidence was extensive it was largely of an anecdotal sort. Now, however, careful ethological surveys and, in some cases, experimental work has verified Darwin’s claims [Ekman et al., 1969; Andrew, 1963; van Hooff, 1962, 1967, 1972]. The *smile* or lip corner retraction, one of these cross-cultural and cross-species facial displays, is used to express attitudes or emotions variously characterized as ‘submissive’, ‘contentment’, ‘desirous of the goodwill of the viewer’, etc. Another one is the facial expression that is the opposite of the smile but which does not have a convenient name: it involves drawing the corners of the mouth forward, even to a protrusion of the lips. For the sake of convenient reference, I call this the ‘o-face’. It is used to express aggression, disapproval, the desire for the viewer to leave the signaler’s presence, etc. (fig. 3).

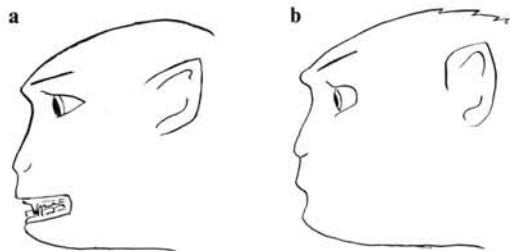


Fig. 3. Two facial expressions of monkeys. Expression of submission. **b** Expression of aggression [redrawn from van Hooff, 1962].

On the face of it, the shape of the smile and o-face are not well matched to the meanings or functions assigned to them. Why, during a nonthreatening display should the teeth, potential weapons, be exposed [Izard, 1971]? And why, during a threatening display, the o-face, should the teeth be partially hidden?

A variety of imaginative accounts have been given for the origin of these two facial expressions. Some have suggested that the smile arose in primates as a play bite or an invitation to grooming [Bolwig, 1964; Eibl-Eibesfeldt, 1971]. Andrew [1963] argues that it was part of a generalized protective response; specifically, the gesture used to dislodge something noxious from the mouth. Erasmus Darwin [1803, p. 77] suggested that the smile arose in infants as a reflex relaxation of the muscles used in suckling and thus became associated with the state of contentment and pleasure. Charles Darwin speculated that the o-face arose as a way to augment the resistance to the increased expiratory airflow that accompanies great emotion. All of these suggestions are worthy of serious consideration but all, I think, have drawbacks. They either apply only to primates (whereas I think that they should work for canids, too) [Schenkl, 1947; Fox, 1970]; they do not provide an account that applies equally well to the smile *and* the o-face, or they fail to integrate these facial displays with other known aggressive/submissive displays. I offer what I believe is a better hypothesis on the origin of these displays, an account which avoids these defects.

One is struck by the fact that the meanings or functions of the smile and o-face parallel those of F_0 that were discussed under Threads 2 and 4. Could they have the

same motivational basis, that is, serve to convey an impression of the size of the signaler? The answer, I propose [Ohala, 1980], is yes, if we make two simple and not implausible assumptions. One, that we extend Morton's [1977] account such that not only the F_0 of vocalizations but also their resonances (those spectral details of the vocalization contributed by the air space between the sound generator and the point where the sound radiates to the atmosphere) may convey the size of the vocalizer. Second, we must assume that the smile and the o-face originally served to modify the resonances of accompanying vocalizations. High resonances are typical of short vocal tracts which, in turn, are indicative of a small vocalizer and, conversely, low resonances of a larger vocalizer. Retracting the mouth corners in effect shortens the vocal tract and raises its resonances (this is particularly true in species with a snout where fully retracting the mouth corners can reduce the effective length of the resonator by some 40% or more). This resonance shift can be demonstrated by the use of Plasticine models with and without a simulated mouth corner retraction as shown in figure 4. Cylindrical models of vocal tracts with the dimensions indicated were coupled to horn drivers and excited by low-frequency (50 Hz) pulse trains. The resulting sound was sampled by a high-quality microphone placed 10 cm from the opposite end and then fed to a spectrum analyzer. As can be seen the resonance peaks shift upwards in the model with simulated mouth corner retraction, e.g., the second resonance increases from 1,700 to 1,970 Hz. (The spectrum of a shorter resonator is also shown for comparison; the effect of the simulated mouth corner retraction thus is to shift the

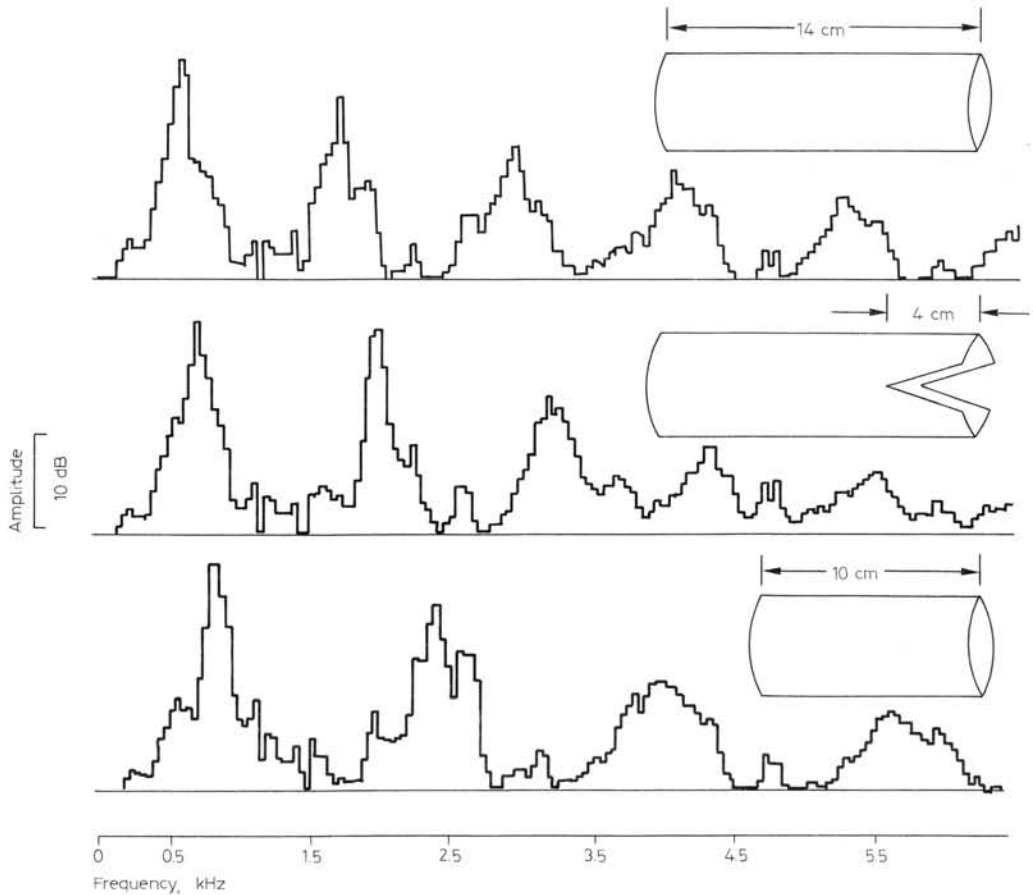


Fig. 4. The transfer functions of three simulated vocal tracts. Top: A uniform tract 14 cm long. Middle: The same tract with simulated mouth corner retraction. Bottom: A uniform tract 10 cm long. The effect of the 'mouth corner retraction' is to raise the resonant frequencies towards those characteristic of shorter tract.

resonances towards those of the shorter resonator.)

It is true that in humans and some primates the smile is often done soundlessly and even with the mouth closed, so that in these cases it could not serve to modify the acoustic shape of a vocalization. However, it has been noted that the high- F_0 submissive screams of many primates are almost invariably accompanied by mouth corner

retraction (if not the other way around), and the low- F_0 cries are typically accompanied by lip corners brought forward [Andrew, 1963]. It is plausible to assume that through what ethologists call ritualization this peculiar mouth shape (which I claim originally served an acoustic purpose) became reinterpreted as an independent visual display having the same meaning or function of the original vocalization.

This account avoids what I believe has been the mistake of some of the earlier speculations on the origin of the smile and o-face, namely to posit that their shapes were originally functional for certain vegetative activities, e.g., eating, biting, regurgitating, respiration, and then to construct quite complex scenarios whereby these functions could become integrated into agonistic displays. It also has the advantage over previous accounts of (1) providing a principled relationship between the smile and o-face, (2) in accounting for the presence of these expressions in the many species it has been observed in, and (3) like Morton's [1977] analysis of F₀, it brings these displays under the same explanatory umbrella as has previously been provided for the visual components of agonistic displays, i.e., that they convey (or originally conveyed) an impression of the size and therefore the degree of threat posed by the signaler.

Thread 6: Consonants and Vowels in Sound Symbolism

There is extensive documentation of a cross-language similarity in the use of certain consonants and vowels in sound symbolism. The evidence is stronger than in the case of tone because not all languages have tones but all languages have consonants and vowels. Words denoting or connoting SMALL or SMALLNESS (and related notions) tend to exhibit a disproportionate incidence of vowels and/or consonants characterized by high acoustic frequency. Words denoting or connoting LARGE use segments with low acoustic frequency. In consonants, voiceless obstruents have

Table II. Examples of sound symbolic words in which choice of consonants and/or vowels show a systematic correlation with concepts of size

Language	'Small'	'Large'
English	<i>teeny, wee</i> <i>itsy-bitsy</i>	<i>humongous</i>
Spanish	<i>chico</i>	<i>gordo</i>
French	<i>petit</i>	<i>grand</i>
Greek	/mikros/	/makros/
Japanese	/tʃiisai/	/ookii/

See also the Ewe example in table I.

higher frequency than voiced because of the higher velocity of the airflow, ejectives higher than plain stops (for the same reason) and dental, alveolar, palatal and front velars higher than labials and back velars. In the case of vowels, high front vowels have higher F₂ and low back vowels the lowest F₂. Table II presents a few examples of this type of vocabulary.

To be sure, there are exceptions to this pattern. The English words *small* and *big* are examples. In spite of such exceptions, subjects of various language backgrounds have, in numerous psycholinguistic tests, shown a clear preference for associating the high frequency segments with things SMALL and low frequency ones with LARGE. For example, *Edward Sapir* in 1929 did a test in which he required subjects to assign nonsense words like [gil] and [gɔl] as names for smaller or larger versions of objects. There was a significant tendency for forms like [gil] to be assigned to the smaller object and [gɔl] to the larger [see also the review of this literature by *Jakobson and Waugh*, 1979].

If we assume that the resonances (spectral shapes) of vocalizations can carry an

impression of size as discussed above, then the pattern of segment utilization in this way is explained in the same way as was the use of tone in sound symbolism.

Thread 7: Sexual Dimorphism of the Vocal Anatomy

I discuss now the piece of evidence which, more than any other, suggests that the frequency code is innate³, i.e., part of human's (and other species') genetic make-up.

Establishing the innate character of a given form of behavior is very difficult. One might think that sensitivity to the special qualities of music is innate since it is a behavior so widely distributed among humans of all cultures. Nevertheless, it has not been possible yet to prove the innate character of such behavior [Roederer, 1982]. Two-legged walking is another interesting case. One might want to argue that we walk on our two lower limbs because we learn to do so. There are even anecdotes that feral children – so-called ‘wolf children’ – left on their own or ‘adopted’ by wild animals, walk on ‘all fours’ not as ‘civilized’ humans do. There is, however, conclusive anatomical evidence in

³ I accept and do not wish to run afoul of the legitimate criticisms leveled at the simplistic labeling of behaviors as ‘innate’ or ‘learned’. Innate behavior usually has some ‘learned’ or postnatal component, e.g., some amount of practice or ‘triggering’ by appropriate environmental stimuli, and every learned or acquired behavior must have an innate component, e.g., the anatomical organs or the sensory mechanisms needed in the execution of the behavior. I use ‘innate’ in the sense ‘having a genetic predisposition which, however, may require extensive postnatal stimulation for its full development and implementation’.

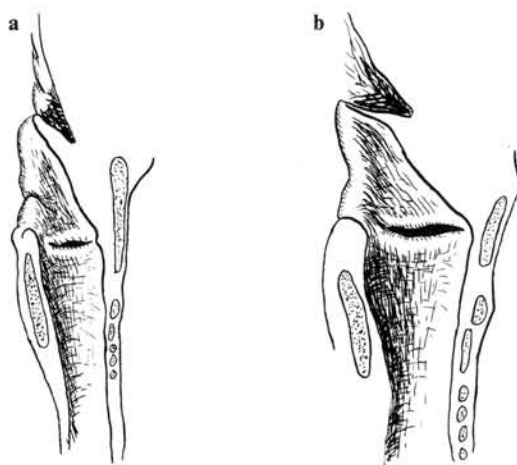


Fig. 5. Sagittal sections of the larynges of a 15-year-old female (a) and a 19-year-old male (b). The vocal cords of a 19-year-old female (were such available for more precise comparison) would only be approximately 15% longer than those shown here [redrawn from *Negus*, 1949].

favor of an innate disposition for two-legged walking, namely, not only the anatomical structure of our legs and hip joints but also the fact that the skin on the soles of the feet is thicker than the skin on the palms of the hands – even several weeks before birth. Is there comparable anatomical evidence in favor of the frequency code being innate? I suggest that there is. As such, it does not manifest itself in the womb but it is still quite clearly genetically determined. This is the evidence of dimorphism in the vocal anatomy of adult males and females.

The facts are well-known but I do not think their significance has been fully appreciated. The adult male larynx is approximately 50% larger than the adult female's in the anterior-posterior dimension [Negus, 1949; Kahane, 1978] (fig. 5). The difference is less marked in the lateral dimension. In other words, it is larger precisely in a way

that would give the male longer vocal cords and thus a lower F_0 . The male larynx is also lower in the throat than the female's, thus making the vocal tract about 15 – 20% longer. This gives the male voice lower resonances.⁴ Now, what is the significance of these facts? They have been widely noted, but I know of only one attempt to give a functional interpretation to them: *Negus* [1949] speculated that the larger larynx of the male is necessitated by his having to engage in more vigorous physical activity than the female and therefore needing, as it were, a larger intake valve to his lungs. However, this would not explain why the male larynx is disproportionately larger only in the anterior-posterior dimension and not the lateral dimension and it would not explain why the male larynx is lower in the throat.

Although the reasons for the sexual dimorphism have not been adequately studied, the low larynx position of the human vis-à-vis other primate species *has* been the subject of much discussion and speculation. It will be useful to discuss these two phenomena together. It has been claimed that this low larynx in humans is a special adaptation to (a) erect posture, (b) lack of a snout, and/or (c) the ability to speak [*Negus*, 1949; *Lieberman*, 1972; *DuBrul*, 1976]. All of these characteristics are indeed found

only in or predominantly in humans, not in apes or monkeys. Nevertheless I believe these hypotheses lack plausibility.

First of all, as mentioned, the larynx is *not* remarkably low in adult females. To maintain these previously mentioned hypotheses one would also have to assert, implausibly, that women were less well adapted to erect posture, lack of a snout, or the ability to speak.

Second, it is important to note *when* this sexual dimorphism occurs developmentally. It occurs at puberty. Prior to that both sexes have virtually identical vocal anatomy. This can be illustrated in figure 6, which presents data organized by *Goldstein* [1980]. This figure shows average growth curves for, on the left, palatal width and, on the right, the distance between the sella-nasion line and the hyoid bone. Age is the horizontal axis. The palatal width data exhibit rather typical growth curves: rapid initial growth which starts to taper off during the teens. There is a slight difference between the sexes, but the difference is small and fairly constant. The growth curve of the distance between the hyoid bone and the sella-nasion line in the female, on the right, shows a pattern similar to that of the palatal width. In contrast, the growth curve of the male for this anatomic feature is quite different. It starts to deviate from the normal curve at puberty and continues in this way until approximately age 20. (It is clear from other data that it is the lowering of the hyoid, not the raising of the roof of the nasal cavity, which is the primary source of the increase in the nasal cavity-to-hyoid distance. Furthermore, this is not the only section of the vocal tract which shows such a large between-sex difference; *Goldstein* [1980] demonstrated that the growth curve of the hy-

⁴ There is some evidence, however, that the difference in male and female formants is greater than what could be explained solely by anatomical differences and that some of the observed differences may be learned [*Fant*, 1975; *Sachs*, 1975; *Kahn*, 1975]. This suggests that speakers are aware of the sex-determined differences in speech and that they may choose to emphasize their masculinity or femininity by producing speech which exaggerates these differences.

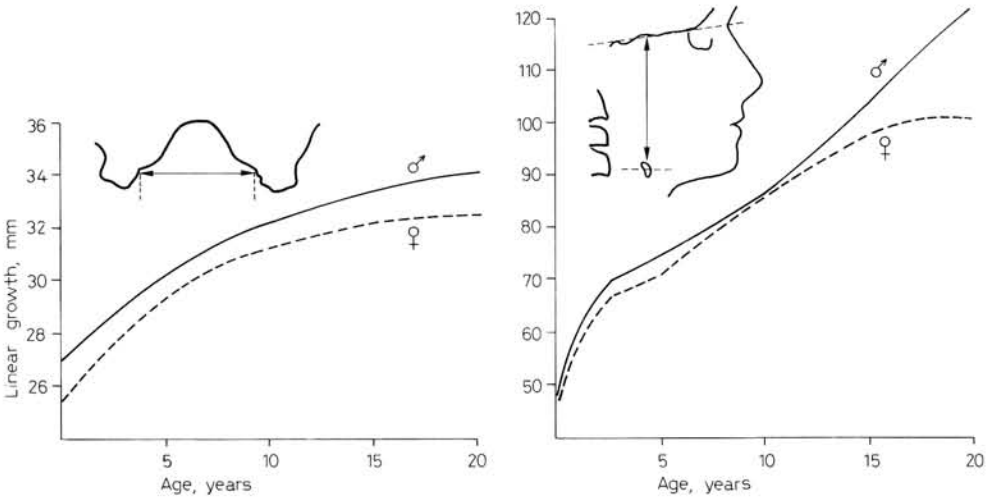


Fig. 6. The linear growth of two features of the vocal anatomy as a function of age. **a** The palatal width. **b** The distance between the sella-nasion line and the hyoid bone. The solid line gives the function for males, the dotted line for females [redrawn from Goldstein, 1980].

Table III. Summary of the relationship between the phenomena discussed in the text.

Meaning		Shape of signal	
primary	secondary	visual	
		nonplastic	plastic
To appear large	threat, intention to prevail in a contest, dominance, self-sufficiency	e.g.: bison's, gnu's hump; male lion's mane; growth of hair on perimeter of face of many primates, including human male	e.g.: piloerection, extension of tail, ears in mammals; extension of wings and tail feathers in birds; arching of back in cats; wearing top hats, epaulets, elevator shoes in humans
To appear small	nonthreat, submission, appeasement, desirous of goodwill and cooperation of receiver	e.g.: opposite of above, including retraction of ears, tail; infant mimicry; cowering	
		the visual component of the smile	

oid-vocal cord distance is similar in shape to that of the nasal cavity-hyoid distance, i.e., that it also shows rapid growth in the male at puberty.) Generally such sex and age dimorphism occurs at the time they are needed. The male deer, for example, grows full antlers only by the time he is ready to compete for a mate. It should be obvious that the conditions of erect posture, lack of snout, and onset of speech come long before puberty. We can therefore rule out these factors as having anything to do with the low larynx of human males. On the other hand, a number of other secondary sexual characteristics show up in the male at puberty, e.g., the growth of facial hair. We might usefully entertain the idea, then, that whatever the reason is for the growth

of facial hair, the same reason may apply to the enlargement of the vocal anatomy. I will elaborate on this below.

Third, many other species besides humans show an anatomical enlargement of the vocal anatomy – and often not in a way that could be explained as an adaptation to erect posture, lack of snout, speech, special respiratory requirements, or, for that matter, any other purely vegetative needs. Among the many species which have this trait are the gorilla, the howler monkey, many species of ducks, swans and geese, the whooping (and other) cranes, and the elephant seal. In the case of the elephant seal, the male but not the female has a rather long proboscis which is used in phonation and which serves to make his

acoustic		
nonplastic	plastic I (nonlinguistic)	plastic II (linguistic)
e.g.: Longer tracheae of geese, cranes, <i>Phonygammus</i> ; bulla in male of some water fowl, e.g., mergansers, wood duck; proboscis in male elephant seal; human male's longer vocal cords, longer vocal tract	vocalizations with low F ₀ and low resonances (implemented by reducing tension on vibrating membrane – vocal cords or syringeal membrane – and by lengthening vocal tract, including protrusion of lips [= 'o-face'])	in intonation, low and/or falling F ₀ for statements; in sound symbolism, concept LARGE conveyed by use of low tone, vowels with low F ₂ , e.g., [a, ɔ, u], consonants with high acoustic frequency, e.g., [grave] (labial and back velar), [flat] (labialized, retroflexed, velarized, or pharyngealized), voiced
	vocalizations with high F ₀ and high resonances (implemented by increasing tension on vibrating membrane and by shortening vocal tract, including retracting mouth corners [= 'smile'])	in intonation, high and/or rising F ₀ for questions; in sound symbolism, concept SMALL conveyed by use of high tone, vowels with high F ₂ , e.g., [i, ɪ, y, e], consonants with low acoustic frequency, e.g., [acute] (apical and palatal), [sharp] (palatalized), voiceless, ejectives
via ritualization		

calls louder and more resonant. One of the most extreme cases of enhancement of the vocal apparatus is the bird of paradise *Phonygammus* which although only about 25 cm long itself, has a trachea over 80 cm long [Clench, 1978]. The extra length is coiled up between the sternum and the external skin.

In all the cases I have mentioned there is evidence of sexual dimorphism of the vocal anatomy such that the male has the larger vocal cavities. It goes without saying that these cases cannot be explained as special adaptations to erect posture, lack of a snout, or to speech. Therefore I see little reason to invoke these factors in explaining the same phenomenon in humans and human males in particular.

But now we come to the question of why these vocal enlargements do occur, why there is sexual dimorphism evident, why this crops up only at puberty, and what this has to do with males' beards.

As for the beards, a very plausible case has been made by Guthrie [1970] that facial hair is present to enhance the visual aspect of aggressive displays. Other primate species – and male lions, of course – also exhibit peculiar hair growth around the perimeter of the face. All of these enlarge the angle which the face subtends in the viewers' eyes thus making the individual appear larger and more awesome. The humps on bison and gnus, mentioned above, probably also function in a similar way. I think the enlargement of the vocal apparatus also occurs to enhance aggressive displays – the acoustic component. Males, by their role in the family unit and the fact that they compete for the favors of the female – i.e., they are subject to what Darwin called sexual selection – would be the ones to develop such

deviations from the 'norm'. However, they would only need these aggressive decorations when they are ready to compete for and retain the favors of a female, that is, at the time of sexual maturity.

There would obviously have to be an innate predisposition for these anatomical developments even if the actual triggering of the growth is regulated by hormonal secretions, the intensity of which might be influenced by environmental factors. There would be no 'payoff' for the evolution of such an elaborate anatomical pattern if there was not an innate recognition of the 'meaning' of the acoustic consequences in vocalizations. Ergo, the frequency code must be innate.

Table III summarizes many of the points I have argued for in this paper and depicts more clearly the connections I have tried to establish between different phenomena.

Conclusion

I have argued that uses of voice F_0 in speech where the sound-meaning correlation shows cross-language consistency, e.g., in intonation, the communication of 'affect', and in sound symbolic vocabulary, can be explained by reference to the factors which have influenced the shape of the acoustic component of agonistic displays in virtually all vocalizing species. The sound-meaning correlations found in these cases adhere to the 'frequency code', which also governs the vocalizations of other species, namely, where high F_0 signifies (broadly) smallness, nonthreatening attitude, desirous of the goodwill of the receiver, etc., and low F_0 conveys largeness, threat, self-confidence and self-sufficiency. In support of this hy-

pothesis I have reviewed data from other domains which, I claim, can also be explained by the frequency code: (a) the shape of certain facial expressions involving specific mouth shapes, e.g., the smile, (b) the cross-linguistic similarities in choice of consonants and vowels in sound symbolism, and (c) the existence of sexual dimorphism in the vocal anatomy of humans and other species.

References

- Andrew, R. J.: The origin and evolution of the calls and facial expressions of the primates. *Behaviour* 20: 1–109 (1963).
- Apple, W.; Streeter, L. A.; Krauss, R. M.: Effects of pitch and speech rate on personal attributions. *J. Pers. soc. Psychol.* 37: 715–727 (1979).
- Bolinger, D.: Intonation as a universal; in Lunt, Proc. 9th Int. Congr. Linguists, 1962, pp. 833–844 (Mouton, The Hague 1964).
- Bolinger, D.: Intonation across languages; in Greenberg, *Universals of human language*, pp. 471–524 (Stanford University Press, Stanford 1978).
- Bolwig, N.: Facial expressions in primates with remarks on a parallel development in certain carnivores (a preliminary report on work in progress). *Behaviour* 22: 167–192 (1964).
- Brown, B. L.; Strong, W. J.; Rencher, A. C.: Fifty-four voices from two: the effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. *J. acoust. Soc. Am.* 55: 313–318 (1974).
- Clench, M. H.: Tracheal elongation in birds-of-paradise. *Condor* 80: 423–430 (1978).
- Darwin, C.: The expression of the emotions in man and animals (Murray, London 1872).
- Darwin, E.: The temple of nature; or, the origin of society: a poem, with philosophical notes (Johnson, London 1803).
- DuBrul, E. L.: Biomechanics of speech sounds; in Harnad, Steklis, Lancaster, *Origin and evolution of language and speech*. Ann. N. Y. Acad. Sci. 280: 631–642 (1976).
- Eibl-Eibesfeldt, I.: Love and hate. The natural history of behavior patterns (Holt, Rinehart & Winston, New York 1971).
- Ekman, P.; Sorenson, R. E.; Friesen, W. V.: Pan-cultural elements in facial displays of emotion. *Science* 164: 86–88 (1969).
- Fairbanks, G.; Pronovost, W.: An experimental study of the pitch characteristics of the voice during the expression of emotion. *Speech Monogr.* 6: 87–104 (1939).
- Fant, G.: Non-uniform vowel normalization. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., No. 2/3, pp. 1–19 (1975).
- Fox, M. W.: A comparative study of the development of facial expressions in canids; wolf, coyote and foxes. *Behaviour* 36: 49–73 (1970).
- Goldstein, U.: An articulatory model for the vocal tracts of growing children; doct. diss. MIT (1980).
- Guthrie, R. D.: Evolution of human threat display organs. *Evol. Biol.* 4: 257–302 (1970).
- Hermann, E.: Probleme der Frage. *Nachr. Akad. Wiss. Göttingen, Philologisch-Historische Klasse*, Nr. 3/4 (1942).
- Hooff, J. A. R. A. M. van: Facial expressions in higher primates. *Symp. zool. Soc. Lond.* 8: 97–125 (1962).
- Hooff, J. A. R. A. M. van: The facial displays of the Catarrhine monkeys and apes; in Morris, *Primate ethology*, pp. 7–68 (Aldine, Chicago 1967).
- Hooff, J. A. R. A. M. van: A comparative approach to the phylogeny of laughter and smiling; in Hinde, *Non-verbal communication*, pp. 209–238 (Cambridge University Press, Cambridge 1972).
- Izard, C. E.: The face of emotion (Appleton Century Crofts, New York 1971).
- Jakobson, R.; Waugh, L. R.: The sound shape of language (Indiana University Press, Bloomington 1979).
- Kahane, J. C.: A morphological study of the human prepubertal and pubertal larynx. *Am. J. Anat.* 151: 11–20 (1978).
- Kahn, M.: Arabic emphatics: the evidence for cultural determinants of phonetic sex-typing. *Phonetica* 31: 38–50 (1975).
- Lieberman, P.: Intonation, perception and language (MIT Press, Cambridge 1967).
- Lieberman, P.: On the evolution of human language; in Rigault, Charbonneau, Proc. 7th Int. Congr. Phonetic Sci. 1971, pp. 258–272 (Mouton, The Hague 1972).

- MacKay, D. M.: Information, mechanism and meaning (MIT Press, Cambridge 1969).
- Morton, E. W.: On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am. Nat.* 111: 855–869 (1977).
- Negus, V. E.: The comparative anatomy and physiology of the larynx (Hafner, New York 1949).
- Ohala, J. J.: Aspects of the control and production of speech. Working Papers in Phonetics, UCLA, No. 15 (1970).
- Ohala, J. J.: The physiology of stress; in Hyman, Studies in stress and accent. S. Calif. Occ. Papers Ling., No. 4, pp. 145–168 (1977).
- Ohala, J. J.: Production of tone; in Fromkin, Tone: a linguistic survey, pp. 5–39 (1978).
- Ohala, J. J.: The acoustic origin of the smile. *J. acoust. Soc. Am.* 68: 33 (1980).
- Ohala, J. J.: Physiological mechanisms underlying tone and intonation; in Fujisaki, Garding, Preprints working group on intonation. 13th Int. Congr. Linguists, Tokyo 1982, pp. 1–12 (1982a).
- Ohala, J. J.: The phonological end justifies any means. Preprints of the plenary session papers. 13th Int. Congr. Linguists, Tokyo 1982, pp. 199–208 (ICL Editorial Committee, Tokyo 1982b).
- Ohala, J. J.: The frequency code and its effect on certain forms of speech and facial expressions; in House, Proc. Symp. on Acoust. Phonetics Speech Modeling, paper 23/81, pp. 1–31 (Institute for Defense Analyses, Princeton 1982c).
- Ohala, J. J.: Cross-language use of pitch: an ethological view. *Phonetica* 40: 1–18 (1983).
- Roederer, J. G.: The search for a biological survival value of music. *J. acoust. Soc. Am.* 72: S92 (1982).
- Sachs, J.: Cues to the identification of sex in children's speech; in Thorne, Henley, Language and sex: difference and dominance, pp. 152–171 (Newbury House, Rowley 1975).
- Sapir, E.: A study of phonetic symbolism. *J. exp. Psychol.* 12: 225–239 (1929).
- Schenkl, R.: Ausdrucksstudien an Wölfen. *Behaviour* 1: 81–129 (1947).
- Scherer, K. R.; London, H.; Wolf, J. J.: The voice of confidence: paralinguistic cues and audience evaluation. *J. Res. Personality* 7: 31–44 (1973).
- Uldall, E.: Attitudinal meanings conveyed by intonation contours. *Lang. Speech* 3: 223–234 (1960).
- Uldall, E.: Dimensions of meaning in intonation; in Abercrombie, Fry, MacCarthy, Scott, Trim, In honour of Daniel Jones, pp. 271–279 (Longmans, London 1964).
- Ullman, R.: Some general characteristics of interrogative systems. Working Papers Lang. Universals, Stanford 1: 39–63a (1969).
- Westermann, D.: Laut, Ton und Sinn in westafrikanischer Sudan-Sprache. *Festschrift Meinhof*, pp. 315–328 (Hamburg 1927).
- Williams, C. E.; Stevens, K. N.: Emotions and speech: some acoustic correlates. *J. acoust. Soc. Am.* 52: 1238–1250 (1972).

Received: December 6, 1983

Accepted: December 14, 1983

Prof. John J. Ohala,
Phonology Laboratory,
Department of Linguistics,
University of California,
Berkeley, CA 94720 (USA)