

*The frequency code underlies the sound-symbolic use
of voice pitch¹*

JOHN J. OHALA

22.1. Introduction

In this paper I propose that certain global uses of intonation across languages exhibit sound symbolism, i.e. they show a motivated link between the shape of an intonation pattern and its meaning or function. This is not a new claim (Hermann 1942) and, in general, there is a very large literature claiming the existence of sound symbolism in other, usually segmental or lexical, domains. But there are several good reasons for being sceptical of such claims, including those I make here.

First, it runs counter to the dominant Saussurian dictum that "the sign is arbitrary," i.e. that the link between sound and meaning is conventional, not natural. This has been a productive working principle and we should not weaken its application to language without good reason. Actually, some amount of acoustic iconism or onomatopoeia in language has always been acknowledged, but it was usually held to represent a negligible fraction of the entire language.

Second, there has typically been no convincing theory offered as to why sound symbolism should exist in languages, nor for the most part has anyone offered a motivation for linkage between particular phonetic features and semantic features. Notable exceptions to this can be found in the work of Paget (1930) and Fonagy (1983), among others, although none of these can be said to have had widespread influence. None of the phonetically based theories of sound symbolism that do exist has been able to unify the various claims about sound symbolism in consonants, vowels, tones, and intonation; e.g. Paget's theory that speech originated as audible gestures made with the mouth would fail to encompass systematic cross-language use of fundamental frequency (F_0) in intonation.

Third, most of the evidence offered for sound symbolism lacks the rigor found in linguistic argumentation at its best, e.g. as demonstrated in linguistic reconstruction via the comparative method. Though immensely fascinating, most of the literature claiming cross-language sound symbolism is anecdotal (e.g. Jespersen 1933) and, a sceptic might legitimately charge, exhibits selectivity in the gathering

of data. (Even so, there have been numerous fairly well-controlled experimental and statistical studies, to be cited below, which cannot be ignored.)

The most I can hope to do in this paper is to reduce the level of scepticism surrounding sound symbolism by proposing a unifying, ethologically based and phonetically plausible theory of aspects of sound symbolism. I will not only attempt to identify the motivation for a sound-meaning connection but will also suggest a link between sound symbolism in vowels, consonants, tones, and intonation. My basic strategy is to argue that sound symbolism is a manifestation of a much larger ethological phenomenon that is also seen in the vocal communication and certain facial expressions of other species. Furthermore, the theory will also account for sexual dimorphism of the vocal anatomy in humans and other species, since phonatory mechanisms have evolved to exploit a specific sound-meaning correlation. To the extent possible I will cite experimental results in support of my claims although, admittedly, much more empirical work needs to be done.

22.2. F_0 in speech

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.

22.2.1. *Universal tendencies for fundamental frequency (F_0) and sentence type*

It has been frequently noted that languages use high and/or rising F_0 to mark questions – and low and/or falling F_0 to mark statements (Hermann 1942; Bolinger 1964, 1978; Ultan 1969; Cruttenden 1981). Although there are exceptions to this pattern, notably when questions are marked by special words or word order, 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. For example, consider the word *cup* in various languages: English /kʌp/, Spanish /taza/, Hindi /pjala/ – and these are languages that are genetically related! Moreover, the pattern found in intonation 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 convincing (Ohala 1970, 1977, 1978, 1982a, 1983, 1990).

22.2.2. F_0 and affect

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 non-linguistic 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/or falling F_0 (Bolinger 1964, 1978).

In so-called “assertiveness training,” the trainees are told explicitly to use as low a pitch of voice as they are comfortable with in order to enhance their assertiveness or image of self-confidence. Radio and television announcers tend to have a low pitch of voice (in comparison to the general population), so that they may sound authoritative. Actors and actresses are generally “locked into” certain types of roles as a function of their voice pitch, e.g. it would seem ludicrous to have Macbeth or Falstaff played by actors with high-pitched voices.

Some of the exceptions to the generalization that rising F_0 is associated with questions and falling F_0 with statements involve an overlay of emotional or socially dictated attitudes (see Ching 1982).

The experimental literature reveals some apparent conflict on the F_0 correlates of affect, however. Whereas Apple *et al.* (1979) found that a higher F_0 of voice made a speaker sound “less truthful, less emphatic, and less ‘potent’ (smaller) and more nervous” (cf. comparable results, but with different labels, by Fairbanks and Pronovost 1939; Williams and Stevens 1972; Brown *et al.* 1974; Uldall 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 (Ohala 1982b). Short samples (4 sec) of spontaneous speech produced by two male and two female adult speakers of American English were digitally processed in such a

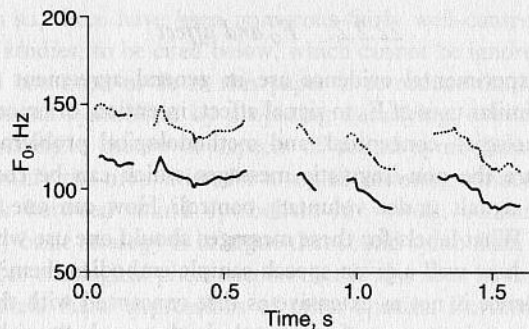


Figure 22.1. The F_0 contours of two samples of "stripped speech" (see text) presented as a pair to listeners to determine which sounded "more dominant, more self-confident." The contour depicted as a dotted line was identical to that depicted by the solid line except that it was upshifted in frequency by a factor of 1.25. The latter, with lower frequency, was judged "more dominant" in 92% of the judgments.

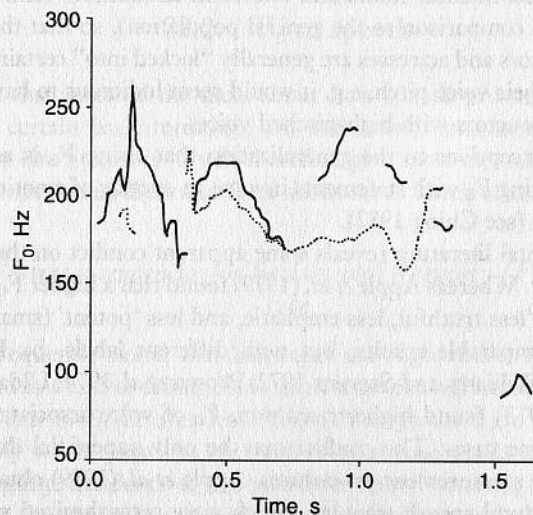


Figure 22.2. The F_0 contours of two samples of "stripped speech" (see text) presented as a pair to listeners to determine which sounded "more dominant, more self-confident." The contour depicted by the solid line was judged "more dominant" in 92% of the judgments.

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, in the judgments for the two samples presented graphically in figure 22.1, 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 as sounding 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 22.2, where the sample shown as a solid line, even though it has a higher peak F_0 , was judged as sounding more dominant (92% of all judgments) 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.

22.2.3. 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; Chao 1947; Whitaker, 1955–1956; Welmers 1973). This observation was made earlier by Bolinger (1978) and Liberman (1978: 92). 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. Some examples are given in table 22.1.

22.3. F_0 in non-human 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,

Table 22.1 Examples of the "sound-symbolic" use of tone

Language	"Small" and high tone	"Large" and low tone
Ewe	[kítsíkítsí] "small"	[gbágbágbà]
Yoruba	[bíri] "be small"	[bíri] "be large"
Cantonese	[to 21] "terrace, stage"	[to 215] "table"

and the frog (Davies and Halliday 1978; Ryan 1980). Morton (1977) provided the following explanation for this sound-function correlation.

Animals in competition for some resource attempt to intimidate their opponent by, among other things, trying to appear as large as possible (because the larger individuals would have an advantage if, as a last resort, the matter had 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 receiver's visual field. There are many familiar examples of this: threatening dogs erect the hair on their backs and raise their ears and tails, cats arch their backs, birds extend their wings and fan out their tail feathers. Some animals have even developed permanent (i.e. non-plastic) 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 F₀ of voice can also indirectly convey an impression of the size of the signaler*, since F₀, 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 F₀ as possible. On the other hand, to seem small and non-threatening a vocalization which is tone-like and high in F₀ is called for. It is also possible in some cases that this latter behavior represents a form of infant mimicry (Ewer 1968: 211, 215, 232ff.; Tembrock 1968). If so, this is a particularly effective way of pacifying a would-be aggressor since, for obvious reasons, 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. I will henceforth call this cross-species F₀-function correlation "the frequency code."

The frequency code in non-human vocalizations suggests an explanation for the three phenomena mentioned above. Its application to the affective use of F₀ of voice to communicate aggression, assertiveness, dominance, etc. and high F₀ to convey social subordinacy, politeness, non-threat, etc. parallels almost exactly the function of F₀ in the non-human cries. In the case of the typical F₀ 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 co-operation 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 F₀ used should be, and *is*, just the opposite to that found in questions. The F₀ 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 F₀ during speech often parallels, both literally and functionally, the rise and fall of some speakers' eyebrows. (See note 2 below.)

The explanation for the systematic use of F₀ in the choice of tones in sound symbolism is somewhat more problematic. The Yoruba speaker who utters the words /bíri/ and /bíri/ is presumably not trying to appear large and small, respectively, or even dominant or submissive. Rather it is the size of the referent of the word which is symbolized by the tone. But there is still this common element: F₀ is used to make the receiver *react* as if something in the environment were large (or small, as the case may be). If the purpose of communication is to effect a change in the receiver – 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 F₀ 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 /kaʊ/ 'cow' and French /bœf/ "boeuf," both of which had a common pronunciation a few millennia ago. In contrast, the sound-meaning correlation found in intonation and in sound-symbolic vocabulary seems to be less subject to deviations.

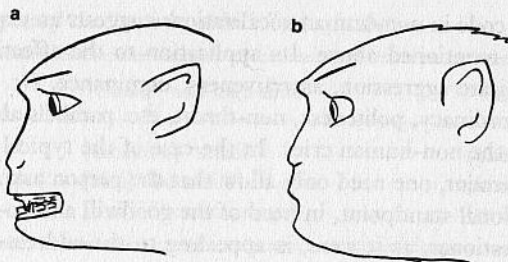


Figure 22.3. Two facial expressions of monkeys. (a) Expression of submission. (b) Expression of aggression (redrawn from van Hooff 1962).

22.4. The frequency code in other communicative domains

In the remaining sections I will attempt to reinforce my argument that an innate frequency code links the above-mentioned use of F_0 in human communication by integrating a few more communicative phenomena into it (although they will not concern prosody *per se*).

22.4.1. 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 non-humans 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 have 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," "content," "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. (figure 22.3).

On the face of it, the shape of the smile and o-face are not well matched to the meaning or functions assigned to them. Why, during a non-threatening display, should the teeth, potential weapons, be exposed (Izard 1971)? And why, during a threatening display (that is, the o-face), should the teeth be partially hidden?

A variety of imaginative accounts has 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: 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. Zajonc (1985) suggests that many facial expressions, the smile included, serve as "ligatures on facial blood vessels and thereby regulate cerebral blood flow, which, in turn, influences subjective feelings." All of these suggestions are worthy of serious consideration, but all, I think, have drawbacks. They either apply only to primates in general or humans in particular (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 above. 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 so that the size of the vocalizer may be conveyed not only by the F_0 of vocalizations but also by 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). 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 22.4. 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 is thus to shift the 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

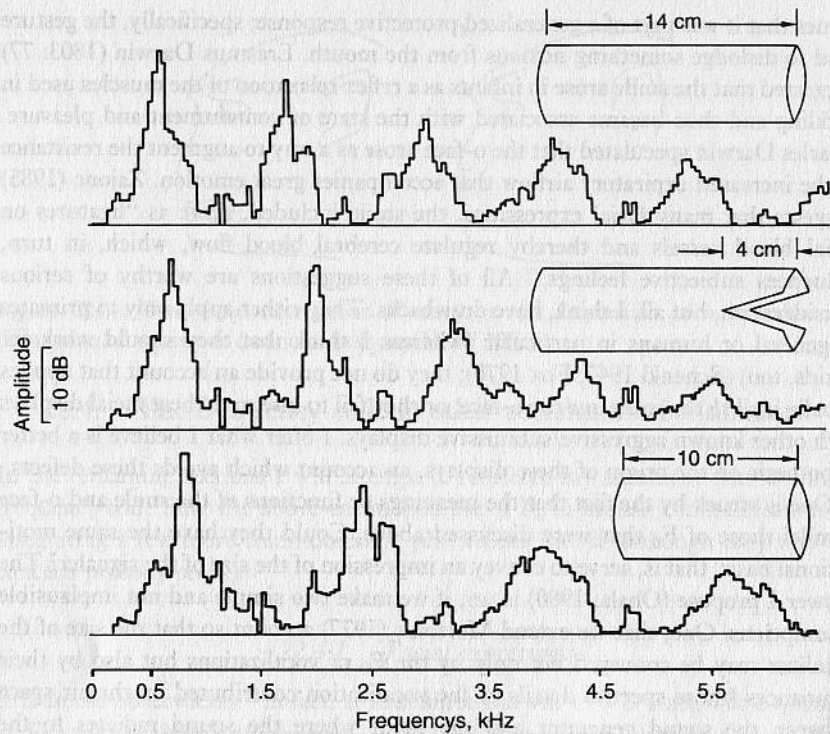


Figure 22.4. The transfer functions of three simulated vocal tracts. Top: a uniform tract 14 cm long. Middle: the same tract with simulated mouth-corner retraction. Bottom: a uniform tract 10 cm long. The effect of the “mouth-corner retraction” is to raise the resonant frequencies towards those characteristic of a shorter tract.

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 as 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 and dubious scenarios whereby these functions could become integrated into agonistic displays.

One can legitimately challenge these theories with this question: when face to face with a competitor, what survival value is there for an animal to give a non-functional enactment of these vegetative behaviors, e.g. regurgitation? The answer is that any animal exhibiting behavior that doesn't deal effectively with a threat probably will not survive long enough to pass on its non-functional behavior to the next generation. It makes much more sense to view the smile and the o-face as signals which attempt to influence the behavior of competitors by either scaring them or inhibiting aggression. The account I give also has the advantage over previous accounts of (1) providing a principled relationship between the smile and o-face, (2) accounting for the presence of these expressions in the many diverse species it has been observed in, and (3) like Morton's (1977) analysis of F_0 , 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.*²

Bauer (1987) has recently provided support for the claim that the frequency code underlies the vocal aspect of agonistic facial displays in chimpanzees by showing a correlation between the mouth dimensions and simultaneous F_0 . There is no plausible physiological motivation for such a correlation; it is reasonable to assume then that both high F_0 and mouth-corner retraction go together to mutually enhance the acoustics of a submissive display.

22.4.2. 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 although not all languages have tones, 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 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 frequencies (of bursts, friction noise and/or formant transitions) than labials and back velars. In the case of vowels, high front vowels have higher F_2 and low back vowels the lowest F_2 (Fischer-Jørgensen 1978 gives evidence that the relevant dimension for vowels is F_2-F_1). Table 22.2 presents a few examples of this type of vocabulary (see also other papers in this volume).

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

Table 22.2 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, 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/

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. (Other tests, both statistical and psychological, of the systematicity of the segment-meaning correlation in languages have been reported by, among others, Usnadze 1924; Newman 1933; Thorndike 1945; Chastaing 1958, 1964a, 1964b, 1965; Fischer-Jørgensen 1967, 1968, 1978; Ultan 1978; Woodworth 1991; 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, i.e. higher frequencies are associated with smallness, lower frequencies with largeness, because these are the frequencies characteristically emitted by respectively small and large things.

22.4.3. Sexual dimorphism of the vocal anatomy

I discuss now the piece of evidence which, more than any other, suggests that the frequency code is innate,³ i.e. part of humans' (and other species') genetic makeup.

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 yet been possible 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 favor of an innate disposition

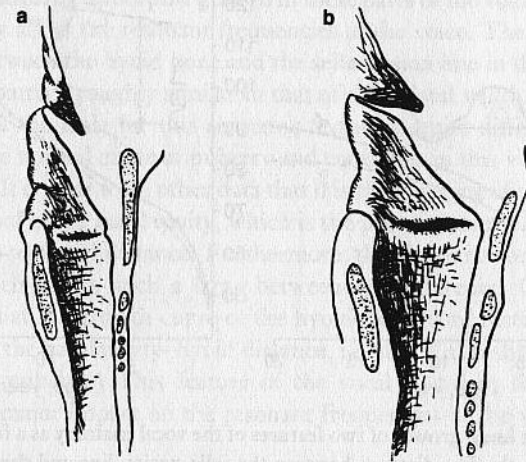


Figure 22.5. Sagittal sections of the larynges of a 15-year-old female (a) and a 19-year-old male (b). The vocal cords of a 19-year-old female (were such available for more precise comparison) would be only approximately 15% longer than those shown here (redrawn from Negus 1949).

for two-legged walking, namely, not only the anatomical structure of our legs and hip joints (in contrast to those of the arms and shoulders and wrists) 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; see figure 22.5). 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

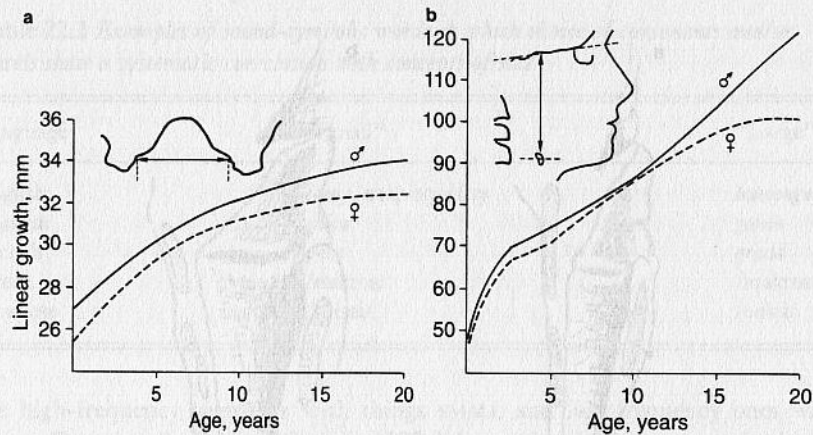


Figure 22.6. The lineal growth of two features of the vocal anatomy as a function of age. (a) The palatal width. (b) The distance between the sella-nasion line and the hyoid bone. The solid line gives the function for males, the dotted line that for females (redrawn from Goldstein 1980).

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 in the human vis-a-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 the 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 22.6, which presents data collated 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. A similar pattern of growth has been reported for the velum (Subtelný 1957).

Differential or sexually dimorphic growth in these parts of the vocal anatomy would not significantly affect the resonant frequencies of the voice. The growth curve of the distance between the hyoid bone and the sella-nasion line in the female, on the right, shows a pattern roughly 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 hyoid-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.) This feature of the vocal anatomy, the length of the pharynx, has a major impact on the resonant frequencies of the voice, namely, to lower them.

Generally such sex and age dimorphism occurs at the time it is 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 a 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 (Schön 1971), many species of ducks, swans, and geese, the whooping (and other) cranes (Roberts 1880), and the elephant seal (Shipley *et al.* 1981). In the case of the elephant seal, the male but not the female has a rather long proboscis which is used in phonation: the trunk-like snout is inserted into the mouth and may function like the fist of the French-horn player to modulate and lower the frequencies of the emitted sound. Equally, the size of the snout itself may influence the acoustic output: there is evidence that the length of the proboscis correlates inversely with the dominant frequency of the phonation and correlates directly with the success of maintaining a harem in the face of competition from other males (Bartholomew and Collias 1962; Shipley, personal communication, 1981). 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

Table 22.3 Summary of the relationship between the phenomena discussed in the text

Meaning		Shape of signal		
primary	secondary	visual	acoustic	
		nonplastic	plastic	plastic I (nonlinguistic) / plastic II (linguistic)
To appear large	threat, intention to prevail in a contest, dominance, self-sufficiency	e.g.: bison'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	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., [ɑ, ɔ, u], consonants with low acoustic frequency, e.g., [grave] (labial and back velar), [flat] (labialized, retroflexed, velarized, or pharyngealized), voiced
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	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, 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, y, e], consonants with high acoustic frequency, e.g., [acute] (apical and palatal), [sharp] (palatalized) voiceless, ejectives
				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’])
		the visual component of the smile	← via ritualization	

sternum and the external skin. This bird, like the cranes, which also have unusually long tracheae, has a very loud call.

In all the cases I have mentioned there is evidence of some sexual dimorphism of the vocal anatomy such that the male has the larger vocal cavity (Winn and Schneider 1977: 822). It goes without saying that these cases cannot be explained as special adaptations to erect posture, lack of a snout, or a 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. As alluded to above, all of these enlarge the angle which the head subtends in the viewers' eyes, thus making the individual appear larger and more awesome. The humps on bisons and gnus, mentioned above, probably also function in a similar way. I think it is also the case that *the enlargement of the vocal apparatus occurs to enhance the acoustic component of aggressive displays*. 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 more often 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 predisposition in the receiver – the listener – to recognize the “meaning” of its acoustic consequences in vocalizations. Ergo, the frequency code must be innate.

Table 22.3 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.

22.5. 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, non-threatening attitude, desire for the goodwill of the receiver, etc., and low F_0 conveys largeness, threat, self-confidence and self-sufficiency. In support of this hypothesis 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.

I do not mean to imply that acoustic frequency is the *only* phonetic feature that can figure in imparting inherent meaning or function to vocal sounds. Other features are also plausible candidates for this, including repetition (reduplication) and such “amplitude envelope” features such as continuancy and rate of onset or decay of sound. These deserve further research as they operate both in human speech and in other species' vocalizations.

NOTES

- 1 This is a revised version of Ohala 1984, and incorporates material from Ohala 1983. These two articles used with permission of S. Karger AG, Basel. The first-cited publication was based on a paper presented at the meeting “Prosody, Normal and Abnormal: An Interdisciplinary Symposium on Suprasegmentals of Speech,” Zurich, April 6–8, 1983, organized by Dr. D. Weniger under the auspices of the Association Européenne de Psycholinguistique and funded by the Schweizerischer Nationalfonds. Measuring the transfer function of the Plasticine vocal-tract models was done with the assistance of S. Pearson. I am grateful to P. Marler, E. Morton, T. Priestly, and J. Wheatley for advice and bibliographic tips. I also gratefully acknowledge the support of the Harry Frank Guggenheim Foundation for the research reported here.
- 2 As was noted above, speakers occasionally raise and lower their eyebrows along with rises and falls in F_0 . There may be a principled reason for this. The raising and lowering of eyebrows in facial expressions may, like the smile and the o-face, help to convey an impression of the size of the signaler (with all the accompanying significance of apparent size). I would speculate that the eyebrows function as a kind of pseudo-boundary to the eyes: raised eyebrows enhance an impression that the eyes are large, lowered brows that they are small. We probably have an innate reaction to relative eye diameter, since the ratio of eye diameter to head diameter is otherwise a good cue to the age and thus the size of the individual: this ratio is far higher in infants than in adults. See Eibl-Eibesfeldt (1971: 21ff.) on the innate appeal of infantile body dimensions and imitations of them.
- 3 I accept and do not wish to fall 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 post-natal stimulation for its full development and implementation."

- 4 There is some evidence, however, that the difference in male and female formants is greater than 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.

REFERENCES

- Andrew, R. J. 1963. The origin and evolution of the calls and facial expressions of the primates. *Behaviour* 20: 1-109.
- Apple, W., L. A. Streeter, and R. M. Krauss. 1979. Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology* 37: 715-727.
- Bartholomew, G. A. and N. E. Collias. 1962. The role of vocalization in the social behavior of the elephant seal. *Animal Behavior* 10: 7-14.
- Bauer, H. R. 1987. The frequency code: oral-facial correlates of fundamental frequency. *Phonetica* 44: 173-191.
- Bolinger, D. L. 1964. Intonation as a universal. In H. G. Lunt (ed.) *Proceedings of the Ninth International Congress of Linguists, 1962*. The Hague: Mouton, 833-844.
- Bolinger, D. L. 1978. Intonation across languages. In J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (eds.) *Universals of Human Language*, vol. 2: *Phonology*. Stanford: University Press, 471-524.
- Bolwig, N. 1964. Facial expressions in primates with remarks on a parallel development in certain carnivores (a preliminary report on work in progress). *Behaviour* 22: 167-192.
- Brown, B. L., W. J. Strong, and A. C. Rencher. 1974. 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. *Journal of the Acoustical Society of America* 55: 313-318.
- Chao, Y. R. 1947. *Cantonese Primer*. Cambridge, MA: Harvard University Press.
- Chastaing, M. 1958. Le symbolisme des voyelles: significations des "i". I and II. *Journal de Psychologie* 55: 403-423; 461-481.
- 1964a. L'opposition des consonnes sourdes aux consonnes sonores et muettes: a-t-elle une valeur symbolique? *Vie et Langage* 147: 367-370.
- 1964b. Nouvelles recherches sur le symbolisme des voyelles. *Journal de Psychologie* 61: 75-88.
1965. Dernières recherches sur le symbolisme vocalique de la petitesse. *Revue Philosophique* 155: 41-56.
- Ching, M. K. L. 1982. The question intonation in assertions. *American Speech* 57: 95-107.
- Clench, M. H. 1978. Tracheal elongation in birds-of-paradise. *Condor* 80: 423-430.
- Cruttenden, A. 1981. Falls and rises: meanings and universals. *Journal of Linguistics* 17: 77-91.
- Darwin, C. 1872. *The Expression of the Emotions in Man and Animals*. London: Murray.
- Darwin, E. 1803. *The Temple of Nature; Or, the Origin of Society: A Poem with Philosophical Notes*. London: Johnson.
- Davies, N. B. and T. R. Halliday. 1978. Deep croaks and fighting assessment in toads *Bufo bufo*. *Nature* 274: 683-685.
- DuBrul, E. L. 1976. Biomechanics of speech sounds. In S. R. Harnad, H. Steklis, and J. Lancaster (eds.) *Origin and Evolution of Language and Speech*. Annals of the New York Academy of Sciences 280: 631-642.
- Eibl-Eibesfeldt, I. 1971. *Love and Hate. The Natural History of Behavior Patterns*. New York: Holt, Rinehart & Winston.
- Ekman, P., R. E. Sorenson, and W. V. Friesen. 1969. Pan-cultural elements in facial displays of emotion. *Science* 164: 86-88.
- Ewer, R. F. 1968. *Ethology of Mammals*. New York: Plenum.
- Fairbanks, G. and W. Pronovost. 1939. An experimental study of the pitch characteristics of the voice during the expression of emotions. *Speech Monographs* 6: 87-104.
- Fant, G. 1975. Non-uniform vowel normalization. *Quarterly Progress and Status Reports, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm*, no. 2/3: 1-19.
- Fischer-Jørgensen, E. 1967. Perceptual dimensions of vowels. In *To Honor Roman Jakobson*, vol. 1. The Hague: Mouton, 667-671.
1968. Perceptual dimensions of vowels. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 21: 94-98.
1978. On the universal character of phonetic symbolism with special reference to vowels. *Studia Linguistica* 32: 80-90.
- Fonagy, I. 1983. *La vive voix. Essais de psycho-phonétique*. Paris: Payot.
- Fox, M. W. 1970. A comparative study of the development of facial expressions in canids: wolf, coyote and foxes. *Behaviour* 36: 49-73.
- Goldstein, U. 1980. An articulatory model for the vocal tracts of growing children. PhD dissertation, MIT.
- Guthrie, R. D. 1970. Evolution of human threat display organs. *Evolutionary Biology* 4: 257-302.
- Hermann, E. 1942. *Probleme der Frage*. Nachrichten von der Academie der Wissenschaften in Göttingen. Philologische-Historische Klasse, nr. 3/4.
- Van Hooff, J. A. R. A. M. 1962. Facial expressions in higher primates. *Symposium of the Zoological Society of London* 8: 97-125.
1967. The facial displays of the Catarrhine monkeys and apes. In D. Morris (ed.) *Primate Ethology*. Chicago: Aldine, 7-68.
1972. A comparative approach to the phylogeny of laughter and smiling. In R. Hinde (ed.) *Non-Verbal Communication*. Cambridge: University Press, 209-238.
- Izard, C. E. 1971. *The Face of Emotion*. New York: Appleton Century Crofts.
- Jakobson, R. and L. R. Waugh. 1979. *The Sound Shape of Language*. Bloomington: Indiana University Press.
- Jespersen, O. 1933. Symbolic value of the vowel *i*. In *Linguistica. Selected Papers in English, French and German*. Copenhagen: Levin & Munksgaard, 283-303.
- Kahane, J. C. 1978. A morphological study of the human prepubertal and pubertal larynx. *American Journal of Anatomy* 151: 11-20.

- Kahn, M. 1975. Arabic emphatics: the evidence for cultural determinants of phonetic sex-typing. *Phonetica* 31: 38–50.
- Lieberman, M. 1978. *The Intonational System of English*. Bloomington: Indiana University Linguistic Club.
- Lieberman, P. 1967. *Intonation, Perception and Language*. Cambridge, MA: MIT Press.
1972. On the evolution of human language. In A. Rigault and R. Charbonneau (eds.) *Proceedings of the Seventh International Congress of Phonetic Science, 1971*. The Hague: Mouton, 258–272.
- MacKay, D. M. 1969. *Information, Mechanism and Meaning*. Cambridge, MA: MIT Press.
- Morton, E. W. 1977. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist* 111: 855–869.
- Negus, V. E. 1949. *The Comparative Anatomy and Physiology of the Larynx*. New York: Hafner.
- Newman, S. S. 1933. Further experiments in phonetic symbolism. *American Journal of Psychiatry* 45: 53–75.
- Ohala, J. J. 1970. *Aspects of the Control and Production of Speech*. UCLA Working Papers in Phonetics 15.
1977. The physiology of stress. In L. M. Hyman (ed.) *Studies in Stress and Accent. University of Southern California Occasional Papers in Linguistics* 4: 145–168.
1978. The production of tone. In V. A. Fromkin (ed.) *Tone: A Linguistic Survey*. New York: Academic Press, 5–39.
1980. The acoustic origin of the smile. *Journal of the Acoustical Society of America* 68: S33.
- 1982a. Physiological mechanisms underlying tone and intonation. In H. Fujisaki and E. Garding (eds.) *Preprints, Working Group on Intonation, Thirteenth International Congress of Linguists, Tokyo, 29 Aug.–4 Sept. 1982*, 1–12.
- 1982b. The voice of dominance. *Journal of the Acoustical Society of America* 72: S66.
1983. Cross-language use of pitch: an ethological view. *Phonetica* 40: 1–18.
1984. An ethological perspective on common cross-language utilization of F_0 in voice. *Phonetica* 41: 1–16.
1990. Respiratory activity in speech. In W. J. Hardcastle and A. Marchal (eds.) *Speech Production and Speech Modelling*. Dordrecht: Kluwer, 23–53.
- Paget, R. 1930. *Human Speech*. London: Routledge & Kegan Paul.
- Roberts, T. S. 1880. The convolution of the trachea in the sandhill and whooping cranes. *American Naturalist* 14: 108ff.
- Roederer, J. G. 1982. The search for a biological survival value of music. *Journal of the Acoustical Society of America* 72: S92.
- Ryan, M. J. 1980. Female mate choice in a neotropical frog. *Science* 209: 523–525.
- Sachs, J. 1975. Cues to the identification of sex in children's speech. In B. Thorne and N. Henley (eds.) *Language and Sex: Difference and Dominance*. Rowley, MA: Newbury House, 152–171.
- Sapir, E. 1929. A study in phonetic symbolism. *Journal of Experimental Psychology* 12: 225–239.
- Schenkl, R. 1947. Ausdrucksstudien an Wölfen. *Behaviour* 1: 81–129.
- Scherer, K. R., H. London, and J. J. Wolf. 1973. The voice of confidence: paralinguistic cues and audience evaluation. *Journal of Research in Personality* 7: 31–44.
- Schön, M. A. 1971. The anatomy of the resonating mechanism in Howling Monkeys. *Folia Primatologica* 15: 117–132.

- Shiple, C. O. 1981. *Development of vocalization on Northern elephant seal bulls*. PhD dissertation, University of California, Los Angeles.
- Shiple, C., M. Hines, and J. S. Buchwald. 1981. Individual differences in threat calls of Northern Elephant Seal bulls. *Animal Behavior* 29: 12–19.
- Subtelny, J. D. 1957. A cephalometric study of the growth of the soft palate. *Plastic Reconstructive Surgery* 19: 49–62.
- Tembrock, C. 1968. Land mammals. In T. A. Sebeok (ed.) *Animal Communication*. Bloomington: Indiana University Press, 338–404.
- Thorndike, E. L. 1945. On Orr's hypothesis concerning the front and back vowels. *British Journal of Psychiatry* 36: 10–14.
- Uldall, E. 1960. Attitudinal meanings conveyed by intonation contours. *Language and Speech* 3: 223–234.
- Dimensions of meaning in intonation. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, and J. L. Trim (eds.) *In Honour of Daniel Jones*. London: Longman, 271–279.
- Ulan, R. 1969. Some general characteristics of interrogative systems. *Working Papers in Language Universals*, Stanford, 1: 39–63a.
- Size-sound symbolism. In J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (eds.) *Universals of Human Language*, vol. 2: *Phonology*. Stanford: University Press, 527–568.
- Usnadze, D. 1924. Ein experimentelle Beitrag zum Problem der psychologischen Grundlagen der Namengebung. *Psychologische Forschung* 5: 24–43.
- Welmers, W. E. 1973. *African Language Structures*. Berkeley: University of California Press.
- Westermann, D. 1927. Laut, Ton und Sinn in westafrikanischen Sudan-Sprachen. *Festschrift Meinhof*. Hamburg: Kommissionsverlag von L. Friedrichsen, 315–328.
- Whitaker, K. P. K. 1955–1956. A study on the modified tones in spoken Cantonese. *Asia Major* 5: 9–36; 184–207.
- Williams, C. E. and K. N. Stevens. 1972. Emotions and speech: some acoustic correlates. *Journal of the Acoustical Society of America* 52: 1238–1250.
- Winn, H. E. and J. Schneider. 1977. Communication in Sirenia, sea otters, and Pinnipeds. In T. A. Sebeok (ed.) *How Animals Communicate*. Bloomington: Indiana University Press, 809–840.
- Woodworth, N. L. 1991. Sound symbolism in proximal and distal forms. *Linguistics* 29: 273–297.
- Zajonc, R. B. 1985. Emotional and facial efference: a theory reclaimed. *Science* 228: 15–21.