

ETHOLOGICAL THEORY AND THE EXPRESSION OF EMOTION IN THE VOICE

John J. Ohala

Department of Linguistics
University of California, Berkeley
Berkeley, CA 94720
ohala@cogsci.berkeley.edu

ABSTRACT

A useful source for unifying theories guiding research on the expression of emotions by the voice as well as by accompanying visual gestures (kinesics) is provided by **ethology**, the science devoted to the comparative study of behavior. Ethology, examining human and non-human behavior, maintains that much of behavior is shaped by phylogenetic adaptations. In this paper I will review and present evidence relevant to some of the fundamental theoretical issues addressed by ethology:

- Does a signal of emotion reflect the inner (psycho)physiological state of the signaler or is it rather primarily designed to induce in receivers behavior that benefits the signaler?
- Are there cross-cultural and cross-species similarities in emotional signals?
- Are the separate components of emotional signals, e.g., voice quality, F0, as well as kinesic elements like eyebrow level, independent of each other or are they correlated in a way to mutually enhance the conveyance of a given message?
- What is the basic vocabulary of emotions? How many are there?

1. INTRODUCTION

It is well recognized that when someone speaks, several 'messages' are conveyed: The acoustic signal contains information not only about the linguistic units in the message (phonemes, syllables, words, phrases, etc.) -- i.e., what could be included in a written transcript of the utterance -- but also information about the speaker's regional accent, sex, approximate age, state of health, personal identity, and something about the speaker's attitude or emotional state. Moreover, in face-to-face communication, some of the same 'messages' are conveyed visually through facial expressions and other movements of the body (kinesics). There is renewed interest in the context of human-machine interaction via spoken language in identifying those features, acoustic and visual, that convey

attitude and emotion. In automatic speech recognition these features may help in the decoding of spoken language by indicating where the speaker is placing emphasis, whether the speaker is cooperative or not or, indeed, perhaps whether the speaker is lying or not. For text-to-speech, including synthesis which incorporates an image of a talking face, they could help to create more naturalistic synthesis incorporating some of the natural human emotions. In this paper I propose that ethology constitutes a valuable resource for data, methods, and theories relevant to the question of how attitude and emotions might be conveyed in speech and its accompanying kinesics. Moreover, ethological accounts have the best chance of providing a unified theory of these behaviors.

Ethology is the comparative study of behavior [7]. In a general way it attempts to discover and explain generalizations covering the behavior of organisms. Today most ethological explanations are couched in evolutionary terms: demonstrating how a given behavior, if it is characteristic of a species and has a fairly stereotypical form, is a phylogenetic adaptation which enhances the survival of the behavior and its kin. Not coincidentally, Darwin [4] was one of the first to speculate in a systematic way about the genetic basis of certain behaviors, especially the expression of the emotions in man and animals (although there were many other earlier, less rigorous, speculations on the evolution of behavior [5]). The comparative study of the expression of emotions has, in fact, reaped a rich harvest. There are remarkable similarities -- both macro- and micro-patterns -- in the expression of emotions in humans and various non-human species. This is particularly true among species able to exploit some of the same signaling modalities as humans, i.e., the vocal-auditory channel and facial expressions. This include especially the mammalian orders and families of primates and canids but for some patterns extends to other groups within the avian and amphibian classes. (We have the impression, probably quite valid, that we can "understand" the facial and acoustic aspects of dogs' and monkeys' displays; at a rather basic level they are morphologically similar to our own.)

2. THE 'FUNCTION' OF EMOTIONAL DISPLAYS: VOICE F0

One of the important theoretical and practical issues surrounding emotional signals is whether they directly stem from or mirror the inner physiological (including the psycho-physiological) state of the signaler or whether to some degree they constitute contrived or even deceitful signals designed not so much to reflect the state of the signaler but rather to influence the behavior of the receiver -- influence it, that is, in a way that would benefit the signaler. There is little doubt that certain vocal and kinesic gestures are closely linked to the signaler's inner state even in situations where it might be disadvantageous to have this information "leak out", e.g., the slightly tremolo evident in the F0 of nervous speakers (as well shaking in the hands and excessive perspiration). In the same category may be the depressed F0 range of depressed individuals -- often accompanied by general lethargy and abnormally slow movements of the rest of the body. But the real issue is whether some of the most familiar emotional signals might better be viewed as designed to produce favorable responses from the receivers of the display. Ethological studies suggest that this is the case.

3. VOICE F0

Morton [11] documented an impressive cross-species similarity in mammals and birds of the sound shape of the acoustic component of agonistic displays (those produced in face-to-face competitive encounters). The confident aggressor emits a vocalization that has a low F0 (within the range permitted by the animal's vocal anatomy) and which may be somewhat rough and aperiodic. A submissive vocalization, on the other hand, typically has high F0 and is tone-like. The dog's aggressive growl and submissive yelp are familiar examples. Remarkably, the same pattern, called 'the frequency code' by Ohala [13] is found in species ranging from the rhinoceros to the shrew. (It is also found in some toads [6].) As Morton argued, this pattern follows from (a) a kind of bluff (deceit) by the animal and (b) the exploitation of physical principles. The size of animal is an important determinant of the outcome of a competition (if, after all posturings and 'persuasions' have been exhausted, an actual fight ensues). Therefore to make itself seem as large as possible to intimidate its adversary and thus forestall such a fight, the animal erects its hair or feathers, elevates its tail or tailfeathers, and, in general, manipulates all the plastic features under its control to convey the signal 'I am large and a threat to you'. Since the F0 of the 'voice' of an animal is inversely related to its bodily size, a low F0 also enhances the overall impression of the vocalizer being large and threatening. On the other hand an animal that submits to an aggressor in order to avoid a fight does the opposite in order to convey an impression that it is small and non-threatening. Some elements of submissive displays clearly employ mimicry of infants and thus probably exploit the very powerful genetically-dictated inhibition against harming infants. Submissive dogs (or, indeed, most canids) that roll onto their back exposing their stomach are said to be imitating a routine common to puppies inviting maternal licking. Ohala [13] [14] presents evidence that the frequency code underlies the shape of humans' vocalizations, too: low F0 to signal threat or self-confidence, high F0 to signal non-threat or deference. The frequency code is also invoked to explain the near-universal

pattern whereby statements are coded by a low or falling F0 and yes-no questions, a high or rising F0 (where this is not otherwise coded by grammatical particles or syntactic means). Several important principles are evident in these ethological accounts of the correlation between signal shape and signal function.

- First, that there is a predictable and recognized physical relation between some parameter, P , of the signal and a feature, F , of the signaler that has some functional import, i.e., $P \propto F$, e.g., the rate of vibration of the laryngeal (or syringeal) folds and the overall size of the signaler (where size is a predictor of success in competition).
- Second, that within a limited range the signaler can vary P .
- Third, that depending on what seems will yield the most favorable outcome from a competitive encounter, the signaler can vary P in order to convey the impression of F' different from its intrinsic F .

(Two important qualifications must be made: First, in addition to parameters that can be varied, one can also often identify anatomically implastic parameters that serve a similar function in signaling, e.g., piloerection in canids is plastic but the large mane of hair around the head of male lions and some monkeys are implastic. Second, though verbs of volition may sometimes be used to describe the signaler's intent, e.g., "The aggressive bird makes itself look bigger by ruffling its feathers, extending its wings and elevating its tail feathers", this is ethological 'shorthand' for "The bird reacts [in such a way] because such behavior provides it (and provided its ancestors) a selectional advantage; it is genetically predisposed to act thus." No conscious strategy of willful deceit is attributed to the signaler -- even to human signalers.)

The above principles probably apply to other vocal and kinesic displays:

4. THE SMILE AND THE 'O-FACE'

Ohala [12] [13] [14] speculates that the smile, a non-threatening display, and what is called the 'o-face' (i.e., with lips constricted and protruded), exhibited in threats, may have originated as part of the acoustic component of submissive/aggressive displays: retracting the corners of the mouth (= 'smile') effectively shortens the vocal tract and thus raises the resonant frequencies; constricting and protruding the lips (= 'o-face') lengthens the vocal tract and thus lowers the resonant frequencies. This would contribute to the impression of the overall size of the signaler since normally resonance vary inversely with other linear bodily dimensions. Through ritualization the smile may have become to a large extent a visual (kinesic) signal but its acoustic origin is revealed in apes monkeys where much the same mouth shape invariably accompanies high-pitched submissive vocalizations [2]. This scenario resolves the long-standing paradox of why the smile, a affiliative display, shows so many of the teeth, normally

regarded as potential weapons and thus seemingly more appropriate for an aggressive display.

(It is widely thought that among humans the smile, at least, is a reflection of inner contentment and well-being. However, Kraut and Johnston [10] have demonstrated convincingly “a robust association of smiling with a social motivation and an erratic association with emotional experience”.)

5. RAISED AND LOWERED EYEBROWS

It is well documented and it is part of common lore that raised eyebrows signal non-threat or friendly greeting whereas lowered eyebrows convey threat [7] [8]. Again, this is true not only with humans but with non-human primates and canids, as well [9]. A functional account of this pattern is often given in terms of the needs of vision: raised eyebrows when a wide field of vision is required and lowered brows for narrow, focused, viewing [1]. But this would predict that in a situation where the signaler is conveying non-threat, i.e., submission or capitulation, in the face of a specific recognized threat, lowered brows would be expected in order to closely monitor source of the threat. But this is not what is found; rather raised eyebrows and widened eyes are found generally for the expression of non-threat, no matter what the circumstances. An alternative, ethologically-based, account is possible [14] which is more in line with the explanations given above for voice F0, voice resonances (and hence for mouth shape), viz., that eyebrows, like voice F0 and resonances, can help to convey an impression of the **size** of the signaler. As is well known, the dimensions of the eyes do not change as much as overall bodily dimensions during maturation. Head size thus changes much more than eye size. Thus the ratio of eye diameter to head diameter, gradually gets smaller, reaching asymptote some time after sexual maturation. As a consequence, a visual estimation of this diameter is a rough indicator of age and size of the signaler and thus its degree of threat. Subtle variations in the portion of the eye externally visible accompany displays conveying threat or non-threat: threat involves a narrowing of this opening -- the so-called ‘squint’. The eyebrows are not actually the outer visible limits of the eye but they convey a **visual impression** of their boundaries. This, I would maintain, is the motivation for the use of eyebrow position in threat and non-threat displays. (This account also helps to explain the popularity of cosmetics such as mascara and eyeliner which, as any cosmetologist will immediately confirm, make the wearer’s eyes ‘look bigger’. The same principle would underlie the recent trend towards larger and larger diameter eyeglasses which, again, convey an impression of eye size and by being large make the wearer appear non-threatening and thus attractive.)

6. CRYING AND TEARS

From a purely physiological point of view, displays of crying and the appearance of tears on the face are extremely puzzling. Certainly the primary function of tears is to cleanse the eyes. Why, then, should tears appear when an individual experiences frustration, imminent threat and the like, even from sources that

are not likely to harm or pollute the eyes? Even more puzzling is why those who witness crying and tears should have such a powerful sympathetic response towards the crier. Roes [15] has proposed an ingenious and novel ethological explanation. Based on the well-established sympathy and helping response exhibited toward infants, he proposes that crying gives the crier some of the appearance of newborns and thus elicits the same sympathy that they do: tears wet the face as does the amniotic fluid in the case of the newborn. The screaming and the wrinkling of the skin around the eyes -- in some case the reddening of the skin -- also evoke the image of a newborn child and thus an instinctive helping response. It is the importance of this response in guaranteeing care to helpless infants and thus their survival which accounts for its origin and maintenance. Any individual past the newborn stage who cries exploits this important phylogenetic adaptation.

7. STARING AND LOOKING AWAY

Another parameter in threatening/non-threatening displays is where the signaler looks. A fixated stare into the eyes of another is taken as a threat whereas non-threat typically involves looking away or avoiding the gaze of the receiver [3]. This is also a cross-species pattern as evidence by the quite valid advice on how to avoid the unwanted attention or attack of a strange dog: *don't look it in the eye*. There is considerable culturally-dictated variability in this pattern, to be sure. In North America an individual who doesn't look his/her interlocuter ‘in the eyes’ is not respected whereas in many other societies looking someone directly in the eye is regarded as disrespectful and threatening. Both patterns are consistent with the generalization that a direct fixated stare is on the ‘aggressive’ end of the aggressive - submissive continuum; the variation comes in whether a given culture encourages or values more or less aggressive attitudes. (Actually, even in North America, a long, unwavering, stare into the eyes of another is unsettling. Anecdotally, staring “contests” occasionally occur where the one who diverts his/her gaze first is regarded as the “loser”.) Friendliness can be signaled not only by not looking directly into the eyes of the receiver for more than a few seconds at a time (or not at all -- in some cultures a deferential person not only averts their gaze but also keeps their head down) but also by *tilting the head* (about the anterior-posterior axis) so that the signaler’s eyes do not “lock onto” the eyes of the receiver.

I do not know how this particular signaling device arose, i.e., how it has survival value, but the important point is that it is widespread among humans and certain non-humans species alike. It is therefore a prime candidate for a phylogenetically-determined behavior used in the expression of emotion and attitude.

8. HOW BIG IS THE VOCABULARY OF EMOTION?

In the experimental and descriptive literature on the expression of emotions one finds considerable variation in the emotions or

attitudes thought to be conveyed by voice and kinesics. These include, happiness, sadness, boredom, surprise, disgust, fear, anger, frustration, irony, sarcasm, shame, scorn, confidence, etc. Ethological theory may be able to help in constructing a principled taxonomy of such messages.

At a fundamental level one should recognize signals that convey messages where the transmission of the message itself has survival value. Such signals should be found cross culturally and, perhaps cross species. Signals that convey threat and non-threat, including the smile, would fit into this category. As Kraut and Johnston, quoted above, have stated, such signals would be motivated more by the effect they would have on receivers than by the inner (psycho)physiological state of the signaler. Signals used in courtship (which overlap considerably with threat and non-threat signals) would probably also belong in this category.

At another level would be signals which do reflect the (psycho)physiological state of the signaler. In most cases, arguably, displaying these conditions does not serve the best interests of the signaler but they presumably arise from physiological states which are themselves beneficial, i.e., adaptive, responses to external conditions. They are adaptive in that they prepare the organism to effectively deal with some situation but they inadvertently "leak out" as a signal. This would include excessive perspiration and the tremor in the voice and hands in someone who is nervous or fearful.

Finally, at another level one can recognize ways that humans convey their attitudes about the receiver, about the content or referent of their utterance, or about themselves. This would include irony, sarcasm, indifference, and the like. Such messages do not confer obvious survival benefit to the signaler and are probably acquired, i.e., learned. Thus they are likely to vary considerably from culture to culture and perhaps even from one individual to another. Unlike the signals mentioned above, they are also likely to require much pragmatic and higher-order linguistic context in order to be appropriately communicated.

9. CONCLUSION

As speech research moves beyond the purely linguistic content of utterances and seeks to discover the emotions or attitudes of the speaker from an analysis of voice qualities and kinesics, ethology, the comparative study of behavior -- human and non-human -- can provide it useful data, methods, and theories. In particular, it offers some theories which can unify both vocal and kinesic aspects of displays, e.g., that the common motivation in elevated F0, the smile, and raised eyebrows in friendly displays is to convey indirectly the relatively small size and thus the non-threatening status of the signaler.

10. REFERENCES

1. Andrew, R. J., "Information potentially available in mammal displays," in R. A. Hinde (ed.), *Non-verbal communication*. Cambridge University Press, Cambridge, 179-206, 1972.
2. Bauer, H. R., "The frequency code: oral-facial correlates of fundamental frequency," *Phonetica* 44.173-191, 1987.
3. Blurton-Jones, N. G., "Non-verbal communication in children," In R. A. Hinde (ed.), *Non-verbal communication*. Cambridge University Press, Cambridge, 271-296, 1972.
4. Darwin, Ch . *The expression of the emotions in man and animals*, John Murray, London, 1872.
5. Darwin, E., *The temple of nature; or the origin of society: a poem with philosophical notes*, London, 1803.
6. Davis, N. B. & Halliday, T. R., "Deep croaks and fighting assessment in toads *Bufo bufo*," *Nature* 274.683-685, 1978.
7. Eibl-Eibesfeldt, I., *Ethology. The biology of behavior*, 2nd ed. Holt, Rinehart, and Winston, New York, 1975.
8. Ekman, P. "Pan-cultural elements in facial displays of emotion," *Science* 164.86-88, 1969
9. van Hoof, J. A. R. A. M., "Facial expressions in higher primates," *Symposium of the Zoological Soc. of London, No. 8: Evolutionary aspects of animal communication*, 97-125, 1962.
10. Kraut, R. E. & Johnston, R. E., "Social and emotional messages of smiling: An ethological approach," *J. Personality & Social Psych.* 37.1539-1553, 1979.
11. Morton, E. W., "On the occurrence and significance of motivation-structural rules in some bird and mammal sounds," *American Naturalist* 111.855-869, 1977.
12. Ohala, J. J., "The acoustic origin of the smile [Abstract]," *J. Acoust. Soc. Am.* 68.S33, 1980.
13. Ohala, J. J. "An ethological perspective on common cross - language utilization of F0 of voice," *Phonetica* 41.1 - 16, 1984.
14. Ohala, J. J. "The frequency codes underlies the sound symbolic use of voice pitch," In L. Hinton, J. Nichols, & J. J. Ohala (eds.), *Sound symbolism*. Cambridge: Cambridge University Press. 325-347, 1994.
15. Roes, F. L., "On the origin of crying and tears," *Human Ethology Newsletter* 5.5-6, 1989.