Facial Expression as a Behavioral Indicator of Emotional States

H. Ellgring
Max-Planck-Institute for Psychiatry, Munich, FRG

Summary

This article gives an overview of possibilities for the assessment of facial behavior. With regard to validity, results from a longitudinal study of 36 depressed patients and nine controls as well as of ten schizophrenic patients and their relatives will be referred to. These results are used to illustrate the following principles which have to be taken into account when studying facial behavior: a) communication strongly facilitates facial expression, b) activation of facial behavior follows the "principle of least effort", and c) the principle of individual specificity applies to the association of nonverbal behavior and mood states. Making allowance for these principles has, among others, consequences a) for situations or conditions under which to assess behavior (specifically conditions of communication), b) for data analysis (e.g., dealing with frequent and rare events), and c) for empirical or experimental strategies (e.g., aggregation of single-case longitudinal comparisons). From the results on facial behavior during depression it can be concluded that the nonverbal reaction tendencies of endogenous and neurotic depressed patients differ. Moreover, the differential behavioral patterns observed cast doubt on the assumption of a homogeneity of affects in depression. Taking into account the conditions which govern it, facial behavior has proved to be a valid and, especially, a differential indicator for pathologic affective states and their changes. Given the fact that a psychiatric illness generally incorporates emotional problems it is more than surprising that little attention has been paid to the systematic study of emotional behavior. Some of the reasons for this will be clarified in the following.

Function and Use of Facial Behavior

Facial Expression and the Limbic System

The links between the nucleus facialis and limbic structures show that facial behavior is closely connected to those processes of the brain involved in emotions (Rinn, 1984; Ploog, 1986). The complex muscular structure of the face, which has evolved during phylogenesis, seems to reflect a variety of discrete emotions such as anger, fear, happiness, surprise, etc. (Tomkins, 1982; Izard, 1971) in a most immediate way.

In this respect, facial behavior seems to be an even more differentiated indicator than psychophysiological reactions. A psychophysiological reaction, such as an increase in heart rate, may be due to either positive or negative arousal, while different emotions are assumed to be expressed by different behavior types.

Mimik als Verhaltensindikator emotionaler Zustände

Communicative function of facial behavior

Underlying the facial surface there is a complex structure of densely innervated muscles. Only three of these muscles serve a biological function other than communication: the orbicularis oculi muscle, the orbicularis oris muscle, and the masseter muscle close and open eyes, mouth, and jaw. The main function of all the other muscles, which emerged comparatively late during phylogensis, appears to be the communication of emotional information (Ellgring, 1987).

Since facial behavior came increasingly under voluntary control during ontogenesis, both, its communicative function and its expressive function have to be taken into account. Thus, facial behavior originates from the activity of the limbic system as well as that of cortical areas. This leads to some specific problems when interpreting behavior as either immediate expression or voluntarily controlled communication.

This combination of the connection of facial muscle innervation with the limbic system and the complex nature of a muscular structure with the main function of communicating emotional information makes it quite logical to assess this behavior in order to differentiate between pathologic emotional states.

Neglect of Behavioral Information in Current Diagnostic Systems

In formal psychiatric diagnostic systems no explicit reference is made to facial behavior, either in ICD (World Health Organisation, 1978), in AMDP (Angst et al., 1986), or in DSM III-R (American Psychiatric Association, 1987). The only descriptive term for nonverbal aspects is “slowed and reduced speech” as a criterion for a major depressive episode. This is in contrast to nearly all the descriptions given in clinical textbooks which point to altered expressive behavior as an important pathologic sign.

The omission of facial behavioral terms by no means implies that this information does not contribute to clinical impression and diagnostic decision. In fact, whenever pathologic mood or affects are recognized, it can safely be assumed that expressive behavior plays a major role in these judgments: “...the presence of depressed mood can be inferred from others observing that the person looks sad or depressed” (DSM-III-R, p. 219).

One of the main problems with respect to the notation of facial information appears to be its complex nature. It would seem that our verbal repertoire is better suited to describing an inferred mood state than the various types of behavior used as a basis for that inference.

The methodological pitfalls of these subjective inferences are well known, especially as regards the reliability of judgments based on implicit information. The adequate measurement of behavior in order to obtain objective indicators of mood and affects therefore remains a crucial point.

Measurement of Facial Behavior

Since there is still disagreement about the feasibility and the specific methods of assessing facial behavior reliably, the existing methods will be reviewed briefly. Results obtained with one of these methods will be given to show how the observational approach works. Generally speaking, four different approaches are used: a) rating scales; b) systematic observation and coding of behavior; c) automated measurements of facial movements; d) electromyogram of facial muscle activity.

a) Rating scales for facial behavior. Although ratings are easy to use, their necessarily subjective character is a disadvantage. The observation of events as short as most facial actions are has to be integrated into one rating. Using this method, it is difficult for the observer simultaneously to describe a general behavior tendency, such as a general sad expression, together with such short-term events as expressions of anger or disgust.

If ratings are compared with systematic observation of behavioral events, it becomes clear that a rating of “facial variability” relies mainly on the frequencies of smiles. In this case, raters obviously disregard for the most part, the variable occurrence of other facial behavior indicating differential negative affects (Ellgring, in press). It is still unclear whether a specific rating for defined aspects of facial behavior can be reliably obtained and whether its validity is similar to a general rating of mood.

b) Facial measurement by systematic observation. Given the technical possibility of variable-speed video analysis, facial activity can be described by means of coding single behavioral events. The Facial Action Coding System (Ekman and Friesen, 1978), for example, allows a description based on a functional anatomical basis. Visible changes on the surface of the face are coded according to the underlying muscular action. Thus, raising the eyebrows using the frontal muscles is coded as “Action Unit 1 + 2”, pulling the eyebrows together using the corrugator muscle is coded as “Action Unit 4”.

Coding by trained observers has to be done using variable motion, mainly slow motion, since most of the facial events have a duration shorter than two seconds. The short duration of these events has to be kept in mind when considering electromyographic measurement (see below).

One of the main problems with this kind of analysis is that observations and codings are highly time-consuming. Its advantages, among other things, are a) that it assesses and describes complex behavioral patterns; b) that it is more descriptive than interpretative; c) that it guarantees adequate treatment of the onset, duration, and intensity of a movement; d) that the validity of the assessment is ensured by direct observation of visible behavior; e) that the visible, and thus potentially socially active, part of expressive behavior is described.

c) Automated measurement of facial movements. Recent developments in automated facial measurement make use of video pattern recognition. A limited number
of variable and fixed points are marked on the surface of the face. Movements are assessed by the changing distance between variable and fixed points. Such an approach was proposed by Heimann (1958).

The application of an automated method to facial behavior in Parkinson patients revealed that their facial impairments could be validly distinguished from those of normals (Katsikis and Pilowsky, 1988). Although there are still technical and computational problems, e.g., when head movements occur, this approach has three advantages: a) no observers are necessary; b) it measures asymmetries; and c) an exact recording of temporal characteristics is available.

Some limitations arise from its two-dimensional representation and from the fact that only a finite number of points can be allocated on the face. Since the technical facilities for this approach have only recently become available, some methodological studies are still necessary to evaluate the range of applicability of this method.

d) EMG of facial muscle activity. Electromyographic (EMG) measurement of facial muscular activity is normally done using surface electrodes. In various studies it was possible to show that EMG activity in the corrugator muscle area is most relevant for the assessment of depressed mood (for a review, see Fridlund and Izard, 1983) and the prediction of drug treatment effectiveness (Greden et al., 1984). Less clear is the function of zygomatic muscle activity, the muscle involved in the smile. Since greater activity of this muscle has also been found in depressed patients this might possibly point to high muscular tension generally, not only in the face.

Together with the ambiguities of treating short facial actions within a longer period of recording, "cross-talk" of muscular activity, especially in the mouth area, seems to be an as yet unsolved problem. Ethical considerations forbid the routine use of wire electrodes for direct recording from the muscle. Surface electrodes are easier to use. Their disadvantage, however, is that they record rather unspecified electric activity. Thus, a "smile" can be recorded when in fact the corners of the mouth are pulled down. Pattern recognition approaches (Fridlund, Schwartz, and Fowler, 1984) may help to solve some of these problems.

Advantages of EMG measurement are a) its comparatively easy applicability; b) its assessment of temporal characteristics of facial activity; c) its presumed identification of muscular activity and tension below the level of visibility. However, the unclear relationship between muscular activity which is electromyographically recorded on the surface and visible muscular activity still points to some problems as regards the validity of the measurements.

Conclusions with Regard to the Methods Available

Comparing the methods of measuring facial behavior, one should note that each covers specific aspects of facial activity in a better way than the others. Therefore, no one method can be regarded as definitely superior: rather, they complement each other. Which is most suitable depends on the kind of information one wants to obtain: behavioral observation provides information on complex emotional or affective events; automated video analysis gives insight into temporal aspects of the behavioral flow and into asymmetric activation in the face; electromyographical measurement informs about tension and as yet in visible activation.

Clearly, therefore, it would be premature to recommend one method or the other. The authors' interest in the communicative function led to a decision for the observational method. Supplementing this with the other methods described above would clearly be desirable.

Empirical Findings on Depression

Although there is evidence that depression can be rated from nonverbal behavior (Waxer, 1974; Renfordt and Busch, 1978; Pansa-Henderson, de L'Horne, and Jones, 1982) surprisingly few studies have actually assessed facial behavior in psychiatric patients; either by systematic observation or by EMG. Ekman and Fridlund (1987) feel that this sparseness is due to a lack of tools for measuring facial behavior and a lack of adequate models for relating facial behavior to affective disorders.

The empirical findings reported in the present article are not yet intended as figures of reference or normative data. They rather serve to support the argument that data on facial behavior are valid when taking into account the proper preconditions and the principles that govern this behavior.

As mentioned above, facial EMG has repeatedly been shown to predict treatment outcome in depression (Carney et al., 1981; Greden et al., 1984). A higher initial level of EMG activity in the corrugator and zygomatic muscles during imaging tasks predicted a better clinical outcome in depressed patients after treatment with antidepressants than lower levels of EMG activity. These findings suggest that a differentiation of mood disorders according to their emotional responsiveness can help when predicting the effects of antidepressant drug treatment. Since both corrugator activity (pulling brows together) and zygomatic activity (lifting mouth corners in a smile) were higher in good responders, general emotional responsiveness, rather than the experience of specific negative or positive emotions, has to be assumed. However, this activity was assessed during an imaging task and differential emotional expression might occur at other instances.

Facial Behavioral Patterns

Very few data are available from systematic observations of depressed patients. Ekman and Fridlund (1987) provide data on a few patients with major or minor depression. They describe the four major depressives showing more sadness and disgust and fewer "unfelt-happy" expressions than the three minor depressive patients. In the present longitudinal study of 36 depressed patients and nine controls, distinct patterns of facial behavior emerged when microanalyzing video tapes of standardized interviews (Ellgring and Ploog, 1985; Ellgring, in press). Not surprisingly, the frequency of smiles was found to be lower during depression than during remission. Even during remission, however, for endogenous as
well as for neurotic depressed patients, this frequency remained only one-third of the frequency shown by normals. For general facial activity, a reverse trend was found on comparison of the two subgroups of patients. Whereas general facial activity increased with improvement in endogenous depressed patients it decreased in neurotic depressed patients.

A cluster analysis showed that depressed patients displayed a variety of facial action patterns as opposed to a homogeneous "conversational pattern" among normals. Moreover, predominant facial actions in patients pointed to affects of anger, fear, and disgust; i.e., to negative affects other than sadness. From these results it seems reasonable to assume differential emotional patterns underlying a state which, in a general classification, appears as a homogeneous state of depression. This supports the concept of heterogeneity among endogenous and neurotic depressives, which has also been suggested by Greden et al. (1984).

Three Principles for Expressive Behavior

An adequate empirical or experimental treatment of nonverbal behavior requires that some of their principles be considered. There are certainly more than the three described below. However, ignoring them may have caused some of the problems which have hindered the use of nonverbal information in the past. These three principles are a) that communication is a facilitating factor for facial expression; b) the "principle of least effort" in the usage of facial elements; c) the principle of individual specificity in nonverbal behavior. These principles can only be described briefly here. (For a more detailed discussion of relevant data see Ellgring, in press.)

a) Communication as a facilitating factor for facial expression. Under normal conditions most facial expressions occur during active communication. Facial behavior is coordinated with speech. About 80% to 90% of facial expressions occur when an individual is speaking (see Fig. 1).

As can be seen from Figure 1, there is a clear association of facial behavior with the speaker role in nearly all of the six groups studied. This is indicated by the proportions whereby > > 50% of facial actions are shown when speaking and far fewer when listening. The only exceptions can be found in some of the endogenous patients and in the schizophrenic groups. This points to a dissociation of facial behavior and communication in psychotic patients. Generally, however, the probability of visible facial behavior is increased during communication. This implies that experimental situations should permit verbal communication when facial behavior is of interest.

b) The "principle of least effort" in facial behavior. The "principle of least effort", as applied to facial behavior states that only part of the available repertoire is actually activated during communication. During a five-minute segment of an interview, individuals tended to activate between three and four facial actions out of 13 selected for coding, with no significant difference between depressed patients and normal controls (Ellgring, in press). This repertoire of 13 actions had been selected out of 35 possible actions on the basis of 40 clinical interviews with 20 endogenous depressed patients.

As is shown in Figure 2, the frequency distribution of these action units corresponds to those described by Zipf (1949) for the use of words.

Facial Action Units in Figure 2 are arranged according to their frequency. It can clearly be seen that, in this sample, a few elements of the repertoire were used quite often, while most of the elements occurred rarely or not at all.

This kind of distribution represents the "principle of least effort" stated by Zipf (1949) in his study of the use of words by various writers. If the results with facial behavioral elements are interpreted with reference to this principle it can be concluded that individuals use facial expressions in a quite economical way: only a few of the possible elements are activated in affective communication. This poses some methodological problems, especially as regards the statistical treatment of data. One question is whether rare events may indicate an important content, even though their probability is too low to appear in a statistical average.

c) The principle of individual specificity. This principle, which was established for physiological reactions to stress (Lacey and Lacey, 1958; Engel, 1960), appears to apply equally to nonverbal behavior in depression. In the abovementioned longitudinal study on depressed patients and controls, six nonverbal parameters were assessed from the videotaped clinical interviews: general facial activity, specific facial reaction of smiling, facial repertoire, number of gestures, amount of eye-contact, and amount of speech activity independent of content.

These types of behavior change substantially when individuals are compared over the course of depression. In only one of the cases, however, did all the parameters show the expected increase with improvement of the pathologic state. In all of the other cases, only some of the nonverbal behavioral elements showed such a change (see Fig. 3).
Facial Expression as a Behavioral Indicator of Emotional States

As can be seen from Figure 3, nearly all patients (except three out of the neurotic depressed group) showed a substantial change in at least one of the parameters. In endogenous patients, there is a tendency for this to happen in more parameters than in neurotic depressed patients.

Although eye-contact behavior and smiling changed most often, it was not possible to detect a hierarchical order among the different behavioral elements. Thus, behavioral elements seem to be related in a logical "or" connection, and they cannot reasonably be combined in an additive way. The various patterns of nonverbal reactions to changes in mood state thus represent an individual response-specificity (Engel, 1960). This again would be in accordance with the "principle of least effort", ensuring economical communication of one's psychological state.

From a methodological point of view, however, this implies that single-case longitudinal analyses are necessary to individually assess behavioral changes. Whether different patterns of nonverbal reactions correspond to different functional impairments is still an open question. Differences observed between endogenous and neurotic depressed patients support this view.

Future Developments

Considering the functions of nonverbal behavior and the principles that apply to it, the usage of facial expression as an objective indicator for emotional states should incorporate some of the aspects listed below.

a) Situations. Analysis of facial behavior requires conversation or test situations which facilitate the activation of this behavioral system. Its social and communicative function and close association with verbal behavior means that tasks such as imaging are less suited for observing visible expression. They may be more suitable for assessing general muscular tension. Dialogs, interviews, or tasks requiring vocal-verbal output would be more adequate.

b) Precision of measurement. With regard to measurement of facial EMG, methodological studies are required which compare electrophysiological measurements with both observational data and automated analyses. This would help to clarify ambiguities with regard to the possible causes of increased facial EMG activities found in some depressed patients. Whether this is due to unspecific increased muscular tension or to the activations of specific facial muscles could be cross-validated using observational data.
c) Differential affect assessment. Observational as well as EMG studies suggest that, within the general class of depression, subgroups differ with regard to their facial behavioral tendencies. Moreover, observational data on facial behavior indicate differential affects which are predominantly expressed by individuals. Further studies should make use of this information potential: aside from assessing unspecified arousal, facial expression can provide access to differential emotional functions and thus give rise to different treatment.

In sum, the analysis of facial behavior is a useful tool for assessing emotional states. Its usefulness depends, however, on the proper consideration of the conditions and principles that govern this behavior. What is implicitly done in clinical judgments — i.e., learning the individual specific usage of behavioral elements when interacting with the patient and observing the individual case over time — has also to be incorporated into the systematic assessments and comparisons of facial expression.

References


World Health Organization: Mental disorders: Glossary and guide to their classification in accordance with the Ninth Revision of the International Classification of Diseases. World Health Organization, Geneva 1978

Prof. Dr. H. Ellgring
Institut für Psychologie
Freie Universität Berlin
Habelschwerder Alle 45
D-1000 Berlin 33 (West)
Germany