To appear in J.A. Harrigan, R. Rosenthal, & K. Scherer (Eds.), *Handbook of nonverbal behavior research methods in the affective sciences*. NY: Oxford.

DRAFT: August 20, 2004

RUNNING HEAD: Measuring Facial Action

Measuring Facial Action by Manual Coding, Facial EMG, and Automatic Facial Image Analysis

Jeffrey F. Cohn

University of Pittsburgh

Paul Ekman

University of California San Francisco

Introduction

Of all the nonverbal behaviors ─ body movements, posture, gaze, proxemics, voice ─ the face is probably the most commanding and complicated, and perhaps the most confusing. In part, the face is commanding because it is always visible, always providing some information. There is no facial equivalent to the concealment maneuver of putting one's hands in one's pockets. Whereas sounds and the body movements that illustrate speech are intermittent, the face even in repose may provide information about some emotion or mood state. Many nonverbal behaviors simply do not occur when a person is alone, or at least do so very rarely. For example, it would be unusual for someone to shrug or gesture hello when totally alone. Yet facial expressions of emotion may be quite intense even when a person is alone. They are not occasioned only by the presence of others. In fact, social situations can dampen facial expression of emotion (Ekman & Friesen, 2003).

The face is commanding also because it is the location for the senses of smell, taste, sight, and hearing. It is the site of the intake organs for inputs of air, water, and food necessary to life. It is the output source for speech, and what we hear in part is determined by the lip movements we see with the speech (McGurk & MacDonald, 1976). It commands attention because it is the symbol of the self. The faces of those we care about are hung on walls, displayed on desks, carried in wallets.

Multimessage-multisignal system

This commanding focus of attention is quite complex. The face can be considered as a multimessage, multisignal semiotic system(Ekman & Friesen, 1978). The face conveys not only the message of individual identity, but also messages about gender and race. Certain changes in the face reveal, more or less truthfully, age. There are standards for beautiful and ugly, smart and stupid, strong and weak faces. And apart from stereotypes, there have been claims for accurate information about personality traits, psychopathology, and intelligence from facial behavior (Bruce & Young, 1998).

These different messages (identity, gender, beauty, traits, etc.) have as their source one of four types of facial signal systems: static, slow, artificial, and rapid. *Static* signs include the size, shape, and relative locations of the features and the contours produced by the underlying bony structure. These static signs are the likely vehicles for transmitting information about identity and beauty. Examples of *slow* sign vehicles would be the accumulation of wrinkles, pouches, and bags, which occur with and convey information about age. *Artificial* signs, such as cosmetics and plastic surgery, attempt to disguise these slow age signs. The *rapid* signs include the actions produced by the muscles (typically called expressions or displays), as well as changes in muscle tonus, blood flow, skin temperature, and coloring.

Most research on the face has focused just upon these rapid signs, in particular, the momentary movements of the face and the muscle tonus changes as sign vehicles for information about emotion and mood. Rapid signs may also be relevant sources for other messages, for correct or incorrect information about traits, attitudes, personality, and so on. Our focus in this chapter is upon methods for measuring momentary facial movement (expressions). We first distinguish between *sign-vehicle* and *judgment* based measurement and then focus on three approaches to measuring *sign vehicles* of facial action: human-observer based coding systems, facial electromyography, and automated measurement by computer vision, an emerging approach that shows promising concurrent validity with manual coding, increased efficiency, and powerful capabilities for analyzing the timing of facial action.

Sign- versus judgment-based approaches

Ekman and Friesen (Ekman, 1964, 1965; Ekman & Friesen, 1969) distinguished two conceptual approaches for studying nonverbal behavior, namely, measuring judgments about one or another message and measuring the sign vehicles that convey the message.[1] Often either approach can be used to answer a question. Take, for example, the question whether facial expressions vary with psychopathology. Suppose a sample was available of facial behavior during interviews with patients who had a diagnosis of schizophrenia or depression, and with a control group who had no psychiatric problems. To utilize the *message judgment* approach, the facial movements in these interviews would be shown to a group of expert clinicians, who would be asked whether each person they viewed was normal, schizophrenic, or depressive. If the judgments were accurate, this would answer the question, showing that facial expressions do convey messages about psychopathology. To utilize the *measurement of sign vehicles* approach, some or all of the facial movements would be classified or counted in some fashion. If the findings showed, for example, that depressives raised the inner corners of their eyebrows more than the other two groups, whereas schizophrenics showed facial movements that very slowly faded off the face, this would also answer the question affirmatively.

Although both approaches can answer the same or related questions, they yield different information. The message judgment approach would show that expert clinicians can tell from viewing a face whether a person is schizophrenic, depressive, or normal. That cannot be learned from the other approach, which does not determine whether observers can accurately judge this message. But by measuring the sign vehicles it is possible to find out exactly what differs in the faces of the diagnostic groups: Is it the timing or the particular movements, or both, that show whether a person is depressive or schizophrenic? That cannot be learned from the first approach, which never determines exactly what the observers respond to when making their judgments.[2] Let us turn now to some of the other relationships between the outcomes of these two approaches. Consider these cases:

1. Negative findings with message judgment and positive findings with sign vehicle measurement. This suggests that people (at least those used in the study) do not know what to look for or cannot see the differences in facial behavior. Careful measurement of the facial sign vehicles might have revealed hitherto unknown differences. Once known, these clues to psychopathology might make it possible for observers to make judgments accurately. Or perhaps the clues are such that people will never be able to make this judgment accurately when viewing the behavior at real time. The differences in facial behavior might be too subtle to be seen without repeated or slowed viewing and precise measurement.
2. Positive findings with message judgment and negative findings with sign vehicle measurement. The positive results show that there must be some difference in the facial sign vehicles, for how else would the observers achieve accuracy in their judgment? This outcome shows that something must be faulty in the measurement of the sign vehicles. Either the measurement was not reliable or it was selective rather than comprehensive. The sign vehicles may have omitted movements or related cues, such as blushing, that may have differed between diagnostic groups and there was bad luck in selecting just those sign vehicles that did not differ.
3. Negative findings with message judgment and negative findings with sign vehicle measurement. This all-too-frequent outcome may occur because the face simply does not provide information about the topic being studied. Or something may have been faulty in the sampling. For example, there may not have been sufficient care in obtaining high agreement

among experts about the diagnosis of the patients. Or perhaps the patients were receiving medications that suppressed some behavioral differences. Also, this outcome does not eliminate the possibility that there were differences in facial movement related to psychopathology that the observers did not know about or could not see (thus the message judgment approach failed), and that were missed by a faulty technique for measuring the facial sign vehicle. Was the measurement of sign vehicles comprehensive rather than selective? If it was selective, the possibility always remains that movements unrelated to psychopathology were measured.

The difference between these two approaches ─ message judgment and the measurement of sign vehicle ─ has sometimes been confusing, because both may involve observers and many of the methodological issues, such as inter-observer agreement, are similar (See Rosenthal, , this volume). It is what the observers do that matters. In message judgment they make *inferences* about something underlying the behavior ─ emotion, mood, traits, attitudes, personality, and the like; for this reason they typically are referred to as "judges" or "raters." In measuring sign vehicles the observers *describe* the surface of behavior; they count how many times the face moves, or how long a movement lasts, or whether it was a movement of the *frontalis* or *corrugator* muscle. As an example, upon seeing a smiling face, an observer with a judgment-based approach would make judgments such as "happy," whereas an observer with a sign-based approach would code the face as having an upward, oblique movement of the lip corners.

Observers with a sign-based approach are supposed to function like machines, and typically are referred to as "coders." In the final section of this chapter, we consider progress toward actually replacing human coders with machines. Considerable progress toward this goal has been made through research in computer vision, and the final section of this chapter reviews this work and the prospects for automatic coding by computer facial image analysis.

Though message- and sign-based approaches can sometimes answer the same questions, they can also answer different questions, for they focus on different phenomena. Message judgment research is not typically focused on the face. The face is but an input, although there may be study of different types of faces, as in the psychopathology example. In message judgment studies the focus is instead on the person observing the face and/or on the message obtained. Questions have to do with whether a difference is detectable or accurate; there are individual differences among observers, reflecting skill, gender, personality, and the like; messages obtained are best represented as dimensions or categories.

Facial sign vehicles are measured when the focus is upon unearthing something fairly specific about facial behavior itself, not about the perception of the face. It is the only method that can be used to answer such questions as:

1. To what extent is the facial activity shown by newborns and infants systematic, not random, and which particular actions first show such systematic organization? To answer this question, facial behavior shown during samples taken at different developmental points or in different situational contexts can be measured. Then the probabilities of particular co-occurrences and sequential patterns of facial actions can be evaluated (Cohn & Tronick, 1983; Oster & Ekman, 1978). Which particular facial actions are employed to signal emphasis in conversation? Facial actions that co-occur with verbal or vocal emphasis must be measured to determine whether there are any actions that consistently accompany any emphasis (Ekman, 1980).

3. Is there a difference in the smile during enjoyment as compared to a discomfort smile? The particular facial actions evident in smiling movements must be measured when persons are known, by means other than the face, to be experiencing positive and negative affect (Ekman, Friesen, & Ancoli, 1980; Frank, Ekman, & Friesen, 1993).

4. Are there differences in heart rate that accompany nose wrinkling and upper lip raising versus opening the eyes and raising the brows? Facial behavior must be measured to identify the moments when these particular facial configurations occur in order to examine coincident heart rate activity (Levenson, Ekman, & Friesen, 1990).

These examples are not intended to convey the full range of issues that can be addressed only by measuring facial sign vehicles. They should, however, serve to illustrate the variety of questions requiring this approach. One might expect the measurement of sign vehicles approach to have been followed often, as it is required for study of many different problems. But there have been only a few such studies compared to the many that have measured the messages judged when viewing the face. It is much easier to perform the latter sort of study. The investigator need not tamper with the face itself, other than by picking some sample to show. Data are obtained quickly: One can measure observers' judgments much more quickly than one can describe reliably the flow and variety of facial movement.

Until recently, an important obstacle to research measuring sign vehicles has been the lack of any accepted, standard, ready-for-use technique for measuring facial movement. Each investigator who has measured facial movement has invented their technique in large part *de novo*, rarely making use of the work of their predecessors. Some have seemed to be uninformed by the previous literature. Even the more scholarly have found it difficult to build upon the methods previously reported, because descriptions of facial activity are often less clear than they appear upon first reading. A facial action may seem to be described in sufficient detail and exactness until an attempt is made to apply that description to the flow of facial behavior. For instance, descriptions of brow motion that omit specific appearance changes in facial lines and furrows and in the appearance of the upper eyelid omit information that may be needed to discriminate among related but different facial actions.

Three types of methods for measuring facial sign vehicles

Three types of methods for measuring facial sign vehicles are manual coding, facial electromyography (EMG), and automatic facial image analysis. Manual coding has been used the longest and is the most frequent approach for theoretical and applied research in facial expression; it has been especially informative to the development of automatic facial image analysis by computer vision (Cohn, Campbell, Matias, & Hopkins, 1990). Manual coding is unobtrusive and can be used both for live observation and for analysis of pre-recorded analogue or digital images. Facial EMG requires the use of surface or needle electrodes attached to the face and is typically the method of choice in laboratory studies of psychophysiology. Automatic facial image analysis by computer vision is an emerging methodology. Computer vision has been an active area of research for some 30 years (Duda & Hart, 1973). Early work included attempts at automatic recognition of faces (Kanade, 1973). Within the past decade, there has been increasing effort in automatic recognition of facial expression. We review techniques for measurement of facial sign vehicles by each of these approaches, as well some of the initial applications of these techniques to theory and research in facial expression.

Manual Coding Techniques

The 14 techniques for measuring facial actions reviewed in this chapter cover a span of 78 years, from the 1924 report by Landis to the work of Ekman, Friesen, and Hager in 2002. Five were not presented by the authors as methods that could be used by others, but were reported in the course of describing substantive results. They have been included for various reasons. Landis is included because he was among the first to build a measurement system based on the anatomy of muscle action, and his negative findings were influential for the next forty years. Frois-Wittmann (1930) and Fulcher (1942) were both innovative for their times, but their methods and findings have been largely forgotten by the current generation of researchers. McGrew's (1972) behavioral checklist has influenced those studying children from an ethological viewpoint. Nystrom (1974) has been included because there is much interest today in measuring facial action in infants. The other nine techniques reviewed represent all of the systems for measuring facial movement that have been proposed, some of which have attracted considerable interest and research activity.

A few reports describing facial actions in detail have been omitted. Discussions of facial behavior that did not report a procedure for measurement — such as Hjorstjo (1970) and Lightoller (1925), both of which provided enlightening discussions of the anatomical basis of facial movement — are not included. Depictions of facial expressions primarily designed to train observers to recognize emotion, rather than measure facial movement (Ekman & Friesen, 2003) are excluded even though some investigators have used them to measure facial expression. Izard’s Affex (1983), previously called FESM (1979a), has also been excluded because observers are required to judge emotion rather than describe the appearance of facial movement, which would fall under the judgment-based approach. Unlike most message judgment approaches to the measurement of the face, Izard's Affex provides the observers with training about the various clues believed to signal each emotion. There is no way to know, of course, what clues the observers actually rely upon when they make their emotion judgments, because all the investigator obtains is the end point in the observers' inferences. Though the aim of Affex is to provide quick data about emotions, it cannot allow investigation of what indeed are the facial clues to each emotion. Other techniques designed to provide economical measures of emotion, EMFACS (Ekman & Friesen, 1982)and MAX (C. E. Izard, 1983) are considered in this chapter because they involve describing facial appearance rather than making direct inferences about underlying states. Reports that used but did not add new methodological features to one of the techniques reviewed here are excluded.

The measurement techniques that are reviewed share the features of being unobtrusive; requiring a permanent visual record (still image or video) that allows slowed or multiple viewing, rather than being applicable to behavior as it occurs; and relying upon an observer who scores or codes behavior according to a set of predetermined categories or items.

This chapter cannot teach the reader how to measure facial actions. Nor does it fully describe most of the measurement techniques, many of which would require a whole chapter, and some an entire book. Exceptions are the techniques of Birdwhistell (1952), Landis (1924), and Nystrom (1974), each of whom provided a little more detail than what is reported here. Instead, the emphasis of this chapter is upon the criteria to be considered in evaluating any measurement technique, either one of those available or one that the reader might devise. These criteria are: 1) the basis for deriving facial behavior units, 2) comprehensiveness, 3) separation of inference from description, 4) types of image records and persons with which the technique has been or may be used, 5) reliability, 6) validity, 7) individual differences and 8) cost. The strengths and weaknesses of each technique will be made evident, so that the reader is better able to choose which might be best for a particular research problem. Tables 1, 2, and 3 and the chapter Appendix summarize the comparisons and provide examples. The techniques are organized in terms of their basis for deriving units of facial behavior: linguistic, ethological, theoretical, and anatomic.

The basis for deriving units

Each of the 14 human-observer based measurement techniques contains a list of facial actions such as a brow raise, nose wrinkle, lip corners down, and so on. Measurement includes noting whether any action (or, with some techniques, combination of actions) is present. Later we will consider how each technique describes actions and differentiates one action from another, but here we are concerned with the question how the author decided upon his or her particular list. The lists vary in number of items from a low of 22 to a high of 77. Some actions appear in all techniques, other actions in only some techniques, and still others in just one technique. Sometimes behavior that is treated as a single action by one technique appears subdivided as two distinct actions by others. For example, raising the eyebrows is treated as one behavioral unit by some techniques, but appears as three separate units ─ inner brow raise, outer brow raise, and the combination of inner and outer brow raise ─ in other techniques. Most authors did not explain what they considered when they included or excluded a facial action, what basis they had for subdividing a unit another researcher had treated as a single action, or why they found it wise to collapse a distinction drawn by another investigator. In fact, most did not acknowledge the work of their predecessors, but instead acted as if they had invented their system and had no knowledge of differences between it and the systems of their earlier or contemporary colleagues.[3]

Investigators — often failing to specify the sample, setting, or persons viewed — usually said only that they looked at behavior and that their list of facial actions was simply the product of what they saw. Something more is needed, however, to account for the differences among these techniques, even allowing for the fact that each investigator observed a different behavior sample. What stood out, which attributes were noticed when an action occurred, how the flow of behavior was segmented by the investigator probably depended upon theoretical commitments. Only a few were explicit.

Birdwhistell (1952) tried to organize units and select behavior to construct a system to parallel linguistic units. Grant (1969) advocated the selection and organization of measurement units according to function. Brow raising, for instance, was chosen by Grant because it was said to serve an attention-getting function. This puts the cart before the horse, because the measurement technique so constructed was to be used to discover the function of those very behaviors. Among ethologists, Blurton Jones (1971) was most explicit in considering the anatomical basis for facial actions. In the case of brow raising, contraction of the *frontalis* was believed responsible. Blurton Jones did not say that anatomic basis of facial actions was the final or even the major basis for his decisions about what to include, and he did not specify how he arrived at his list of minimal units of behavior.

Ekman, Friesen, and Tomkins (1971), in contrast to the aforementioned investigators, derived their list of facial actions from explicit theory about the facial actions relevant to emotion, rather than from observation of some sample of behavior. The cart-before-horse criticism applies to them also. Although they could learn whether the actions proposed for an emotion accurately reflect that emotion, they could not discover signals for the emotion that they did not know about in advance. Izard, eight years later, also used theory about emotion signals as the basis for selecting actions to score in his measurement technique MAX. His decisions were based on inspection of still photographs of posed emotions that had yielded high agreement among observers who made global judgments about emotion.

The anatomical basis of facial action provided another basis for deriving units of behavior. The measurement units were presumably based on what the muscles allow the face to do. Because we all have the same muscles (for all practical purposes), this approach might be expected to have led the investigators who followed it to arrive at the same listings of facial actions. This is not the case. For example, Landis (1924) had 22 actions and Frois-Wittmann (1930) 28, and yet they both claimed to have based their measurement units on the anatomy of facial action. In part, the discrepancies occurred because of explicit decisions to select only certain actions. Most standard anatomy texts list many, usually not all, facial muscles with rather simple, only partially correct, and usually quite incomplete accounts of how each muscle changes appearance. Most investigators who based their technique on anatomy selected only some muscles, and usually did not explain the basis for their selection. Ekman and Friesen (1978; Ekman, Friesen, & Hager, 2002) and Ermiane and Gergerian (1978) were exceptions, each attempting to determine all the actions the anatomy allows. Both studies attempted to determine this by systematically exploring the activity of each single muscle; Ekman and Friesen also resurrected Duchenne's (1862) technique of determining how muscles change appearance by inserting a needle into and electrically stimulating muscles.

The discrepancies among the techniques of Ekman and Friesen (1978; Ekman et al., 2002), Ermiane and Gergerian (1978), and Izard (1983) are due to differences in purpose and in procedure for obtaining reliability. Both Ekman and Friesen and Ermiane and Gergerian attempted to include in their lists changes in appearance that are independent of each other. If a muscle contraction would produce two or three changes in appearance, these were gathered together as multiple indexes of the activity of one unit or muscle. For example, when the entire fontalis muscle acts, it will (1) raise the eyebrows; (2) produce horizontal furrows running across the forehead (except in infants, who have a fatty pad in the forehead blocking such wrinkles); and (3) expose more of the eye cover fold (the skin between the upper eyelid and the eyebrow). Both Ekman and Friesen and Ermiane and Gergerian listed these multiple signs together as different ways of recognizing that this one action had occurred. Izard, however, treated signs (1) and (2) of *frontalis* muscle activity as separate measurement units, giving each equal, independent, separate status, failing to recognize that they are signs of the same action. He ignored sign (3). Alternatively, Izard failed to distinguish among facial actions that have different anatomic bases. As an example, pulling the lip corners down and raising or pulling up of the lower lip are assigned the same MAX code even though they are produced by contraction of different facial muscles (Oster, Hegley, & Nagel, 1992). These actions are coded separately in FACS (AU 15 and AU 17, respectively).

Izard (C. E. Izard, 1983) also differed from the others in selecting only movements that he judged relevant to emotion. Any movements that did not figure in MAX formulas for prototypic emotions were excluded (Oster et al., 1992). Ekman and Friesen (1978; Ekman et al., 2002) and Ermiane and Gergerian (1978) intended to include all the possible appearance changes that the muscles can produce. This sometimes meant creating more than one measurement unit, if use of different strands of a single muscle or different portions of that muscle was found to produce visible different changes in appearance. For example, Ekman and Friesen and Ermiane and

Gergerian distinguished a number of different facial action units that are based on various uses of what anatomists have termed one muscle ─ the *orbicularis oris*, which circles the lips. Izard included only some of these separate appearance changes.[4] The Ekman and Friesen technique differed from the others in another important respect. Anatomy was only part of their basis for the derivation of measurable units. They also determined whether observers could reliably distinguish all of the appearance changes resulting from the various muscles. If two appearance changes could not be reliably distinguished, they were combined, even if different muscles were involved. If Ekman and Friesen erred, it was on the side of caution, by excluding distinctions that observers with considerable training might perhaps be unable to distinguish. The opposite error may have been made by Ermiane and Gergerian and by Izard (1983). They included distinctions in absence of evidence that each and every distinction could reliably be made by those who learn their system (see section below on reliability).

Comprehensiveness or selectivity

Three aspects of facial movement can be measured either selectively or comprehensively. *Type* refers to whether the facial action was a brow raise, inner brow raise, brow lower, or some other action. *Intensity* refers to the magnitude of the appearance change resulting from any single facial action. *Timing* refers to the duration of the movement, whether it was abrupt or gradual in onset, and so on. Most investigators have considered how to measure only the type of action, not its intensity or its timing. Type of action, intensity, and timing are discussed here and summarized in Table 1.

Type of action

A technique for measuring the type of facial action can be selective, measuring only some of the actions that can occur, or it may claim to be comprehensive, providing a means of measuring all visible facial action. There are advantages and disadvantages in each case. If the technique is selective, it is important to know what has been excluded; and if it claims to be comprehensive, there must be some evidence to establish that this is indeed the case.

The great advantage of a selective technique is economy. Because only some of the mass of facial actions must be attended to, the work can be done more quickly. Suppose an investigator wants to measure whether fear is reduced by exposure to one set of instructions versus another. A measurement technique that allows measurement of just the occurrence of three or four signals of fear would be ideal, because it will not matter if the occurrence of anger, disgust, distress, or some other emotion signal is missed. Even if the technique does not include *all* of the fear facial expressions (and at this time there is no conclusive or even definitive evidence about all the facial actions for *any* emotion), a selective technique could be useful. It might not matter that some or even most fear expressions were not scored, nor that blends of fear with other emotions were not scored; enough might be measured to show the effect. If the findings were negative, however, the investigator would not know whether the cause was an inadequate experimental treatment (in this example, the instructions might not have differed sufficiently) or failure to measure all of the fear expressions. In such an instance the investigator might want to turn to a comprehensive technique.

Insert Table 1 About Here

Some questions require a comprehensive technique and cannot be answered with a selective one. Suppose the investigator wishes to discover which facial actions signal fear, anger, sadness, and so on. Or perhaps they wish to discover whether different actions are employed to serve a linguistic rather than an emotive function, or to learn what people show on their faces when their heart rates show a sharp acceleration, or whether there are cultural or social class differences in facial actions during a greeting. A comprehensive technique would have to be employed. Once there was reasonably conclusive evidence on any of these issues, then such evidence could provide the basis for selective use of portions of a comprehensive system. For example, Ekman, Friesen(1978; Ekman et al., 2002), and Simons (1985), building upon the earlier research of Landis and Hunt (1939) have strong evidence about the particular combination of facial actions and the timing of those actions that index the startle reaction.[5] Once that has been replicated by other laboratories, those interested in the startle in particular could utilize just that portion of Ekman and Friesen's comprehensive scoring technique.

Only a comprehensive technique allows for discovery of actions that the investigator did not know about in advance and permits a complete test of an a priori theory about facial sign vehicles. Another advantage of a comprehensive technique is that it provides a common nomenclature for descriptions of facial behavior. If many investigators were to use the same comprehensive technique, comparison of findings would be facilitated because investigators, even those who used it selectively, would key their units to a single list of facial actions. Investigators considering selective scoring might well want first to study a comprehensive technique, in order to become acquainted with the entire array of facial actions, so that they could be explicit about what it is they are choosing not to measure.

Wedded to these advantages of comprehensive facial scoring is the disadvantage of cost. It takes more time to learn a comprehensive technique, and it takes more time to apply it, for nothing (presumably) is left out.

It is no accident that the only techniques that claim to be comprehensive — Ekman and Friesen (1978) and Ermiane and Gergerian (1978) — were anatomically based. An inductive approach would be too costly if comprehensiveness was the goal. Too large a sample of diversified behavior would have to be observed to have a reasonable likelihood of achieving completeness. By contrast, it should be possible to achieve comprehensiveness by exploring how each muscle works, because the muscles produce the actions observed. This is not as simple as it might first seem, because muscles can act in concert, not just singly. Facial expressions are rarely the consequence of the activity of a single muscle. Even the smile, which is principally the work of the single *zygomatic major* muscle, typically involves two or three other muscles as well, and not every smile involves the same other muscles. Moreover, what happens to appearance when muscles act in concert is not always the sum of the changes associated with each of the components. Analogous to co-articulation effects in speech, contraction of one muscle can modify the appearance change of another. The activity of one muscle also may obscure the presence of another. It is important, therefore, that a comprehensive technique list not simply the ways of recognizing how each single facial action appears, but also the ways of scoring the occurrence of these units of facial action when they combine in simultaneous or overlapping time. Only the Ekman and Friesen technique has done so.

A last issue regarding how comprehensively a technique measures the *type* of facial action is what evidence is provided to demonstrate that the system is what it claims to be. One wants to know whether the universe of facial movement can be described by the technique, or at least what part of the universe has been omitted. If there is uncertainty about comprehensiveness it should be clear whether it is about just some or all actions. An empirical answer would be possible if either of the techniques claiming comprehensiveness (Ekman and Friesen and

Ermiane and Gergerian) had scored large samples of facial actions of males and females of diverse ages, from various cultural, ethnic, and class backgrounds, in a wide variety of social and individual settings. The system of Ekman and Friesen has been used extensively in cross-cultural, developmental, and medical populations, and evidence for comprehensiveness so far is strong. A sample of this literature can be found in (Ekman, 1997).

Alternatively, comprehensiveness could be determined by experimentally generating all possible permutations of facial actions. Ekman and Friesen explored the comprehensiveness of their technique by producing voluntarily on their own faces more than 7,000 different combinations of facial muscular actions. These included all permutations of the actions in the forehead area, and for the lower face all of the possible combinations of two muscles and of three muscles. Although they believe their system is relatively comprehensive[6] only time and application to diverse samples of facial behavior will establish it to be so. Ermiane and Gergerian provided no evidence of comprehensiveness. They determined only that their system would describe the actions of single muscles, and a few of the combined actions of two or three muscles.

Intensity of action

Actions vary not only in type (inner corner brow raise versus raise of the entire brow) but also in intensity. A brow raise may be weak or strong; the lift of the brow, the extent of exposure of the eye cover fold and gathering of skin on the forehead, may be very slight or great. The intensity of a facial action may be of interest for a variety of reasons. For example, Ekman et al. (1980) found that the intensity of *zygomatic major* muscle action was correlated with retrospective self-reports about the intensity of happiness experienced.

Ermiane and Gergerian was the only one of the 13 other techniques to provide for comprehensive measurement of intensity. Nine of the techniques treated facial action as an all-or-nothing phenomenon, or as if there were evidence that variations in intensity are without significance. One (Grant) even confused intensity with type of action, listing as different action types appearance changes that are due only to variations in intensity. A few made provision for scoring the intensity of four or five actions (see Table 1). Good reliability and precision have been found for intensity scoring using FACS (Sayette, Cohn, Wertz, Perrott, & Parrott, 2001). Ekman, Friesen, and Hager (2002) found that the logic provided in the original version of FACS for measuring the intensity of four actions could be extended to the other facial actions, but evidence has not yet been provided that such extensions can be made reliably for all the actions in their technique.

Timing of action

A facial action has a starting and a stopping point. It is often more difficult to ascertain the exact determination of these points than to decide which action occurred. From start to stop, other aspects of timing may be distinguished:

1. *Onset time*: the length of time from the start until the movement reaches a plateau where no further increase in muscular action can be observed.
2. *Apex time*: the duration of that plateau.
3. *Offset time*: the length of time from the end of the apex to the point where the muscle is no longer acting.

Onsets and offsets may vary not only in duration but in smoothness; for example, an onset may increase at a steady rate, or steps may be apparent (K. L. Schmidt, Cohn, & Tian, 2003).

Similarly, an apex may be steady or there may be noticeable fluctuations in intensity before the offset begins. When examined closely the separate actions that compose a facial expression do not start, reach an apex, and stop simultaneously. In even a common expression, such as surprise, the raising of the eyebrows may reach an apex while the dropping of the jaw is still in onset.

For some questions it is possible that simple counts of the occurrence of particular actions may be sufficient, without measurements of onset, apex, and offset. The investigator may want to know only how often or for how long a person raised the brow, wrinkled the nose, or depressed the lip corners. Even when interest is limited to simple summary measures of the occurrence of single actions, there is no rationale for using frequency rather than duration measures (which require stop- start determination) other than economy. A frequency count will under represent those actions that go on for long periods of time and over represent frequent brief actions.

Limiting measurement to single actions is hazardous regardless of whether frequency or duration is measured. Nose wrinkling, for example, may signify one thing when it occurs in overlapping time with a lower lip depression (disgust) and something quite different when it flashes momentarily while the lip corners are pulled upwards (an action that Ekman and Friesen suggest functions like a wink to accentuate a smile). A pulling down of the lip corners may signify sadness when it accompanies raised inner corners of the brows with drooping upper eyelids. When this same action occurs with the entire brow raised and the lower lip pushed up it may be a disbelief gesture. These interpretations, which have not all been tested, cannot be tested unless the timing of actions is measured. What evidence does exist (Ekman & Friesen, 1978) suggests that it is unwise to measure the face as if each action can be counted separately, as if each action has an invariant meaning apart from other actions that overlap in time.

Measurement of combinations of facial actions (what is usually meant by an expression) requires at least a determination that actions overlap, if not precise determination of the stopping and starting points of each action. Ekman and Friesen (1978) further suggest that it is overlap in the apex that is crucial to determining whether actions that co-occur are organized as part of the same event, signal, or expression. Their reasoning is that when one action begins (onset) while another action is fading (offset), it is not likely that they have been centrally directed as part of the same signal. Suppose, for example, that there has been an overlap in the apex of brow lowering, tightening and pressing together of the red parts of the lips, and raising the upper eyelid. Ekman and Friesen have hypothesized that these elements compose one of the anger expressions. Overlap in the apex of these actions would support their notion that an anger signal had occurred and that these actions should be so counted, and not tallied separately. Let us suppose that there was also a nose wrinkle, with an apex overlapping these anger actions. Ekman and Friesen suggest that this would be a blend of disgust with anger. If the nose wrinkling reached its apex as these anger actions were in offset, they suggest that it be characterized as a sequence of anger followed by disgust. Test of these hypotheses requires precise measurement of onset, apex, and offset.

A number of other research questions also require comprehensive measurement of the timing of facial actions. For example, does a brow raise and upper eyelid raise occur before or during an increase in loudness in speech or a deceleration in heart rate? Ekman, Friesen, and Simons (1985) found that onset time is crucial in isolating from idiosyncratic facial actions those muscular actions that always occur in unanticipated startle reactions. Only actions that began within 0.1 second were evident in all unanticipated startles; offset time did not distinguish the idiosyncratic from uniform facial actions. In another situation offset time, rather than onset, may be crucial; for example, Ekman and Friesen (2003) hypothesized that stepped offsets occur more often in deceptive than in felt emotional expressions.

Most of the 14 techniques do not describe procedures for measuring starting and stopping points and ignore onset, offset and apex measurement. The data reported usually consists only of frequency counts. While other features could be coded, no criteria are provided for how to do so. Ekman and Friesen's technique is the only one to describe how to measure these different aspects of timing.

Depicting facial measurement units

It is not as easy as it may at first seem to depict clearly what is referred to by a facial measurement unit. Some authors did not bother, because they did not expect others to try to use their methods. Regrettably, this lack of clarity also has caused some uncertainty about their substantive results. Take the example "down corners mouth," which is found in the measurement techniques of Birdwhistell (1952), Brannigan and Humphries (Brannigan & Humphries, 1972), Grant (1969), and Nystrom (1974). Does this phrase describe instances in which the mouth corners have been pulled down? Or those in which the mouth corners are down because the chin and lower lip have been pushed up in the middle? Or does it refer just to expressions in which the mouth corners are down because the center of the upper lip has been raised? Or is it all of them?

The first column in Table 2 describes how measurements were depicted in each of the 14 techniques. The chapter appendix lists how a particular facial action (brow raise) was depicted by each technique.

Most techniques used but a few words to describe each measurement unit. Some supplemented this description with a few still photographs. Only three techniques went beyond this step to provide more thorough illustration of each unit. Ekman and Friesen, Ermiane and Gergerian, and Izard's MAX technique all provided visual illustrations of every measurement unit. All provided some explanations of the anatomical basis of each action, Ekman and Friesen and Ermiane and Gergerian more thoroughly than Izard. Ermiane and Gergerian provided still photographs of each action and combination considered; Izard provided video, photographs, and drawings; and Ekman and Friesen provided still photographs and video illustrations.

Separating inference from description

Although many investigators have been interested in inferring something about the signal value or function of facial actions, not all have recognized that such inferences should not be intermixed with descriptions in their measurement techniques. The measurement must be made in non-inferential terms that describe the behavior, so that inferences about underlying states, antecedent events, or consequent actions can be tested by empirical evidence.

Mixing inference with description may also make the measurements quite misleading. Few single-muscle actions have an invariant meaning. Take the example of the so-called frown (lowering and drawing the brows together). This action is not always a sign of negative affect; depending upon the timing of the action, what other actions co-occur with it, and the situational context, it may signify quite different matters (Klaus R Scherer, 1992). It would be misleading to be identifying the occurrence of a frown when the brow lowering is signaling concentration, or conversational emphasis.

Insert Table 2 About Here

Because humans make the measurement, inferences cannot be eliminated, but they need not be encouraged or required. If the person scoring a face identifies the brows being lowered and/or drawn together, the scorer may still make the inference that he or she is describing a frown. But Ekman and Friesen (1978) reported that when people use a measurement technique that is solely descriptive, as time passes the scorer increasingly focuses on the behavioral discriminations and is rarely aware of the possible meaning of the behavior. Although there can be no guarantee that inferences are not being drawn, a measurement technique should neither encourage nor require inferences about meaning by the terminology or descriptions it employs.

Both Ekman and Friesen and Izard separated their hypotheses about the signal value of facial actions from the descriptive materials to be used in training a person to measure facial behavior. Ermiane and Gergerian intermixed inferences about the meaning of behavior with the information necessary to learn their descriptive system. Theirs is the only technique to contain inferences about how given facial actions are indicative of specific personality processes and types of psychopathology. Birdwhistell (1952), Blurton Jones (1971), Brannigan and Humphries, Grant (1969), McGrew (1972), Young and Decarie (1977), and Frois-Wittmann (1930) all used some inferential or emotional terms (e.g., *frown, smile, sneer, angry frown*) mixed in with descriptive terms. (This is not always evident from the chapter appendix, because not all who mixed inference with description did so for the brow raise.)

Both Ekman and Friesen and Izard listed hypotheses about the emotion signaled by particular facial actions. Ekman and Friesen were explicit about the particular combinations of units they considered as emotion signals. Izard's MAX contains only those facial actions which, he claims, distinguish among the emotions. Ekman and Friesen have evidence that Izard is wrong, that he has excluded a number of actions relevant to emotions. For example, Izard does not include *levator labii superioris caput infraorbitalis*, which is relevant to both disgust and anger, except when this muscle acts unilaterally. Ekman, Friesen, and Ancoli (1980) found that bilateral evidence of this muscle correlated with the subjective report of disgust. Ekman and Friesen also found that when this action is accompanied by the narrowing of the red margins of the lips (another action ignored by Izard), the signal changes from disgust to anger.[7] As another example, MAX omits action of the *buccinator,* for which unilateral action is associated with contempt (Darwin, 1872/1998; Ekman & Heider, 1988).

Types of records and persons to which the measurement has been applied

Still or motion records

Although a number of techniques claim that they can be used with motion records, most have not dealt with the complexities in the timing of facial action that a motion record reveals. These investigators may never have been confronted with the complexity of the temporal organization of facial actions because of either the type of behavior or the type of record they examined. If only posed expressions were measured (as in the case of Ermiane & Gergerian), variations in timing might not be apparent. Posers generally try to perform all the required movements at once, in overlapping time, with similar very short onsets, long-held apexes, and abrupt short offsets. Preliminary data suggest that the relationship between intensity and duration of smile onsets varies as well between posed and spontaneous smiles. In the former, there parameters are uncorrelated, whereas in the latter they are highly correlated, consistent with automatic movement (Cohn & Schmidt, 2004). An investigator who used his or her method only to score still photographs might not know of these complexities in timing, because the camera shutter freezes all action. Though Izard has scored some motion records, he pre-selected only certain brief segments of videotape to score, segments in which the infants seemed to be emitting expressions that looked like those in posed photographs of adults. Thus he has not dealt with the complexities that a motion record reveals. Other investigators may have failed to consider the timing of facial movement because they tried to apply their systems in real time, as the behavior occurred, and even if they had videotape or film, they may not have examined the records in slowed or repeated replay.

It will be most important for investigators to make use of motion measuring the timing of facial actions whenever they want to study spontaneous behavior, taking a strictly descriptive approach; or interrelate facial activity and some other simultaneous behavior (speech, respiration, body movement, etc.); or distinguish configurations in which the temporary organization of multiple facial actions suggests that they be considered parts of the same signal or expression. (See the discussion below of the research questions that require measurement of timing.)

Modifications for varying age levels

Ideally, a facial measurement system should be applicable to the study of individuals of any age, by making provision for any modifications needed to measure infants or the aged. The appearance of certain facial actions is quite different in neonates and infants from what it is in young children and adults. Oster (1978), who worked with Ekman and Friesen during the final stages in the development of their measurement system, has studied the neuro-anatomical basis for these differences. She has provided (Oster & Rosenstein, Undated) a set of transformations for utilizing the Ekman and Friesen system with neonates and infants. Izard's MAX technique was specifically designed to measure infant facial expression. He provides only a few overly general descriptions of potentially confusing infant-adult differences. For investigators wishing to use MAX to code facial actions in adults (e.g. Sayette, Smith, Breiner, & Wilson, 1992), it becomes important to know about how criteria may change with development. No other investigator has attended to the problem of how coding criteria may change with development.

Parallel problems may occur in measuring facial activity in quite elderly people, because age signs may necessitate some modifications in scoring rules to avoid mistakes in identifying certain actions. No one has considered this.

Reliability

The need for reliability is obvious to psychologists. To some anthropologists and sociologists the quest for reliability has seemed a peculiar madness that deflects psychologists from the real problem at hand. For example, Margaret Mead, in the last years of her life, wrote "Psychologists ... are more interested in validity and reliability than in what they are actually studying" (Mead, 1973). Yet if a measurement system cannot be shown to be reliable, there is no way of knowing whether even the investigator who invented the system recognizes the same facial action when it twice occurs. The need to demonstrate reliability seems especially important with facial behavior. For here there is an enormous variety of behaviors that can occur, with no names for most. And those who have observed facial actions have produced very different catalogs.

Some ethologists (Young & Decarie, 1977) have argued that if the same finding is obtained in two independent studies, there is no need to demonstrate that the measurement technique was reliable. This reasoning should not be applied to the area of facial measurement, where there have been completely contradictory reports by different investigators (e.g., the argument about universality between Birdwhistell and Ekman). If we knew that Birdwhistell and Ekman had each used a reliable measurement technique (preferably the same one), at least we could be certain about what was seen, and search differences in sampling, situation, or interpretation as sources of their disagreement. When a measurement technique is intended to be usable by other investigators, it is especially important for its originator to demonstrate that he or she as well as others can use it reliably. (See also the first section of the Introduction in which reliability was discussed in the context of the relationship between the outcomes of message judgment studies and measurement of sign vehicle studies.)

Let us consider now various aspects of reliability, for it is not a simple matter to establish. A number of requirements can be enumerated:

1. The researcher, rather than just giving an overall index of agreement, should provide data to show that high agreement can be reached about the scoring of specific facial actions. Typically, some actions are easier to recognize than others. Unless reliability data are reported for the scoring of each facial unit, it is not possible to evaluate which discriminations may be less reliable.

2. Data on reliability should be reported from the measurement of spontaneous, not just posed, behavior, and from the flow of behavior as revealed in a motion record, not just from still photographs or slices abstracted from video, which may yield higher agreement.

3. Reliability data should be provided for (*a*) infants, (*b*) children, (*c*) adults, and (*d*) aged populations, because reliability on just one group does not guarantee reliability on the others.

4. The most common source of unreliability in behavioral measurement, whether it be of face or of body, is the failure of one person to see what another scores. Usually this occurs when an action is small in size. This source of disagreement can be attenuated if the technique specifies a threshold that must be surpassed for the action to be scored. Specifying minimum

thresholds alerts the persons doing the scoring to subtle signs and provides explicit bases for decisions about when a change in appearance is likely to be ambiguous. A technique that provides such threshold definitions should therefore yield higher agreement.

5. Reliability should be reported not only for the person(s) who developed the technique, but also for learners who did not previously have experience with facial measurement. Data about the range of reliabilities achieved by new learners should be provided and compared to those for experienced or expert scorers. A technique will be more generally useful if it can be learned independently, without direct instruction from the developer. This usually requires a self-instructional set of materials, practice materials with correct answers, and a final test for the learner to take.

6. Reliability should be reported for the scoring of not just (*a*) the type of action, but also (*b*) the intensity of actions and (*c*) the timing of actions.

Of the 14 measurement techniques, 5 did not report data on any aspect of reliability. Others provided fairly sparse data on reliability, with the exception of Ekman and Friesen and Izard. Even these techniques did not meet all the requirements just listed. Table 3 lists the specific reliability requirements met by each technique.

Insert Table 3 About Here

Validity

Descriptive validity

The validity of a technique designed to measure facial movement entails questions on a number of levels. Most specifically (and concretely), validity requires evidence that the technique actually measures the behavior it claims to measure. When a technique claims to measure brow raise, are the brows actually raised, or is it just the inner corners that are raised? If the technique claims to measure the intensity of an action, such as whether the brow raise is slight, moderate, or extreme, do such measurements correspond to known differences in the intensity of such an action? The problem, of course, is how to know what facial action occurs, what criterion to utilize independently of the facial measurement technique itself. Two approaches have been taken:

1. Performed action criterion: Ekman and Friesen trained people to be able to perform various actions on request. Records of such performances were scored without knowledge of the performances requested. Ekman and Friesen's Facial Action Coding System (FACS) accurately distinguished the actions the performers had been instructed to make.
2. Electrical activity criterion: Ekman and Friesen, in collaboration with Schwartz (Ekman, Schwartz, & Friesen, 1978) placed surface EMG leads on the faces of performers while the performers produced actions on request. Utilizing the extent of electrical activity observed from the EMG placements as the validity criterion, they found that FACS scoring of facial movement accurately distinguished the type and the intensity of the action. (This study is described in more detail in the section on EMG below.)

Utility or validity

Some measurement techniques contain hypotheses about the particular facial actions that signal particular emotions (Ekman & Friesen; Ekman, Friesen, & Tomkins; Ermiane & Gergerian; Izard). For these techniques it is appropriate to ask whether the hypotheses are correct, but the answer does not pertain to the validity of the techniques, only to that of the hypotheses. Suppose the facial behaviors found to signal emotion were exactly the opposite of what had been hypothesized by the developer of the technique. Such evidence would not show that the technique was invalid, only that the hypotheses were wrong. In fact, the discovery that the hypotheses were wrong would itself require that the technique measure facial movement accurately. Suppose a study not only failed to support the investigator's hypotheses about the actions that signal emotions but found that there were no facial actions related to emotion. If one could discount the possibility that the sample did not include emotional behavior, this might suggest that the facial measurement technique was not relevant to emotion. It might have measured just those facial behaviors that are unrelated to emotion. Another technique applied to the same sample of facial behavior might uncover the actions related to emotion.

Two techniques (Ekman and Friesen and Ermiane and Gergerian) claim not to be specific to the measurement of any one type of message such as emotion, but to be of general utility, suitable for the study of any question for which facial movement must be measured. Such a claim can be evaluated by evidence that the technique has obtained results when studying a number of different matters.

Posed expressions. Many techniques have been shown to be able to differentiate poses of emotion or judgments of emotion poses: Ekman and Friesen; Ekman, Friesen, and Tomkins; Ermiane and Gergerian- Frois-Wittman; Fulcher; Izard. In the studies that used a selective technique it is not possible to know whether there might have been other facial actions not included in the scoring technique that might have predicted the emotion poses or judgments just as well or better The two comprehensive techniques ─ Ekman and Friesen and Ermiane and Gergerian ─ provided that information. They were able to show that it was the movements they specified as emotion-relevant, not other movements, that were signs of particular emotions.

Ekman and Friesen's FACS also predicted not only *which* emotion was posed or judged, but the *intensity* of emotion as well.

Poses, however, by definition are artificial. Although they may resemble spontaneous facial expressions in some respects (Ekman & Friesen, 1982), one difference is that they are likely to be easier to score. The onset may be more coordinated and abrupt, the apex frozen, and the scope very intense or exaggerated. The velocity of smile onsets in relation to intensity also appears to differ markedly between posed and spontaneous smiles (Cohn & Schmidt, 2004). Evidence that a technique is a valid measure of emotion cannot rest just upon measurement of poses; it is necessary to determine that the measurement will be valid when it measures spontaneous emotional expression.

Spontaneous expressions. A number of studies have shown the validity of Ekman and Friesen's FACS in measuring the occurrence of spontaneous emotional expressions. Ancoli (1979) studied autonomic nervous system (ANS) responses when subjects watched a pleasant or stress-inducing film. A different pattern of ANS response during the two films was found only during the times in each film-viewing period when the face registered maximal emotional response. In another study of that data, Ekman et al. (1980) found that FACS accurately predicted the subjects' retrospective reports of their emotional experience while watching the films: the intensity of happy feelings, the intensity of negative feelings, and, specifically, the intensity of the emotion disgust. Ekman, Friesen, and Simons (1985) differentiated the specific facial actions that signify a startle reaction from the emotional reactions subsequent to being startled. Both the type of actions and the onset time were crucial to this distinction. They also were able to differentiate a genuine from a simulated startle accurately. Ekman, Hager, and Friesen (1981; Hager & Ekman, 1985) examined the differences between deliberate facial movements and spontaneous emotional expressions. Scoring the intensity of each specific facial action on each side of the face, they found that requested facial movements were asymmetrical more often than spontaneous emotional expressions: The actions usually were more intense on the left side of the face for the deliberate, but not for the spontaneous, emotional expressions. Krause (1978) utilized FACS to measure facial actions during conversations among stutterers and non-stutterers. As he predicted, the facial actions specified in FACS as relevant to anger occurred more often among the stutterers. There is little or no comparable evidence that the other facial measurement techniques listed in Table 3 can be used to measure spontaneous emotional expressions.

The only exception is Izard's use of his MAX technique to study infants. He found that observers scoring brief segments of videotape showing infant expressions *selected* to correspond to adult posed expressions could reliably identify the actions making up those expressions. This shows that his technique can be used to identify at least those particular expressions when they occur in spontaneous behavior. At this point, however, there is no evidence to support Izard's claim that an infant producing a particular expression is experiencing a particular emotion or blend of emotions (Oster et al., 1992). The evidence suggests that emotion-specified expressions in infants may commonly occur in the absence of the hypothesized emotion (Camras, 1992; Camras, Lambrecht, & Michel, 1996), and hypothesized emotions may occur in the absence of expression-specified expressions (Klaus R. Scherer, Zentner, & Stern, 2004). Infant expression also appears to be less differentiated than claimed by Izard (Matias & Cohn, 1993). Because Izard has not described infants' facial behavior comprehensively, he cannot even specify how representative the selected expressions are in the behavior of infants of a given age and in a variety of situations.

Oster (1978) (Oster & Ekman, 1978) has provided more complete information about the range of facial muscle activity observed in infants and the infant's capacity for coordinated facial movement. Unlike Izard, she began not by looking for adult posed expressions but by analyzing the configurations and sequences of facial actions actually produced by infants in a variety of situations. Oster found that almost all of the single facial actions included in FACS are apparent early in life. Though certain combinations of facial actions common in adult facial expression can be observed in the newborn period, others have not been observed in infants. Oster (1978) (1978) has argued that the only way to determine the affective meaning and signal function of infants' facial expressions is by a detailed description of the expressions themselves ─ including their timing and sequencing ─ combined with a thorough functional analysis of their behavioral correlates and stimulus context. Though far from complete. Oster's work has provided evidence that complex, spontaneous facial actions observed in infants (e.g., smiling, brow knitting, pouting) are not random but represent organized patterns and sequences of facial muscle activity that are reliably related to other aspects of the infants' behavior (e.g., looking at or away from the care giver, motor quieting or restlessness, crying). Such relationships can provide insights into the infant's affective state and cognitive processes.

Stable individual differences

Several studies have found moderate stability in FACS action units and predictive validity for a wide range of personality and clinical outcomes. Cohn, Schmidt, Gross, and Ekman (2002) found moderate to strong stability in FACS action units over a 4-month interval; stability was sufficiently robust as to suggest that facial behavior could function as a biometric. Person recognition from FACS action units was comparable to that of a leading face recognition algorithm. Harker and Keltner (2001) found that FACS action units predicted adjustment to bereavement, teacher ratings of problem behaviors, and marital adjustment over periods as long as 30 years. Malatesta, Culver, Tesman, and Shephard (1989) found low to moderate stability in infant facial behavior over several months using MAX. There is no comparable evidence of stability or predictive validity for personality related measures for the other measurement techniques.

Costs

This last criterion for evaluating measurement techniques was not included in Table 3 because Ekman and Friesen was the only study to provide information about time costs for learning to measure and for scoring a specified sample of behavior. It takes approximately 100 hours to learn FACS. More than half of the time is spent scoring practice materials (still photographs and video) included in FACS at the end of each chapter in the instructional manual. Ekman and Friesen do not know whether people will still achieve high reliability if they skip such practice; they do know that high reliability was achieved when all the instructional steps were followed.

The costs for using a measurement technique once it is learned are much more difficult to estimate. For FACS and probably any other technique, the costs depend upon how densely the facial behaviors are packed in the time sample to be scored. Consider first comprehensive scoring in which FACS is used to measure *all* visible facial activity in a 15-second period. This could take as little as 1 minute if only one or two easily distinguished actions occurred and the investigator wanted only to locate start-stop points for each action. It could take as long as 10 hours, however, if the behavior was as densely packed as it is in the facial activity of deaf persons signing, and if onset-apex-offset was scored for every action. Ekman and Friesen have not observed any other instances in which facial behavior is so densely packed over so many seconds.

If selective rather than comprehensive scoring is done, the costs are lower. Presume that the investigator wants to score only actions that are said to be indicative of disgust, and they select the actions listed in the *Investigator's Guide to FACS* (Ekman & Friesen, 1978) (Ekman et al., 2002) that are predicted to be prototypic for that emotion. A 2:1 ratio, 30 seconds of scoring time for every 15 seconds of live action, is probably a reasonable estimate. Ekman and Friesen developed a more economical system for measuring the occurrence of single emotions, based on FACS. Occurrences of actions considered to be the most common signs of anger, fear, distress and/or sadness, disgust and/or contempt, surprise, and happiness are noted. In what they call EMFACS (Ekman & Friesen, 1982) (EM standing for emotion), time is saved in 3 ways:

1. Scoring does not extend to the particular action, but only to whether a member of a group of specified actions occurred. For example, there are seven signs grouped together that Ekman and Friesen consider relevant to disgust. EMFACS does not differentiate among nose wrinkling, nose plus upper lip raising plus lower lip depression, nose wrinkling plus lower lip elevation, and so on. If any of these is seen, a check is made for that grouping. All actions not in one of the groupings are ignored.
2. Intensity of action is not scored, although intensity is included in the requirements for particular actions within a grouping. For example, a slight depression of the lip corners with slight pushing up of the lower lip is included in the sad grouping, but when those two actions are moderate or strong they are not included.

3. The timing of actions is not measured; only a frequency count is taken. EMFACS takes one-fifth the time of FACS, but of course it suffers from all of the problems already discussed in detail for selective as compared to comprehensive measurement techniques.

For a similar method of identifying action unit composites in infants, see (Camras, Oster, Campos, Miyake, & Bradshaw, 1992).

Izard's MAX technique is similar to Ekman and Friesen's EMFACS. It, too, combines actions presumed to be relevant to the same emotion, and makes no provision for scoring the timing or the intensity of action. Unlike FACS, it requires the scorer to examine different regions of the face separately, and admittedly, it includes in some regions changes in appearance that are due to actions in another region. By contrast, FACS and EMFACS alert the scorer to all the appearance changes resulting from particular muscles. Rather than inspecting an arbitrary division of the face in three regions, the scorer learns where to look in the face for those changes. Izard's MAX technique was developed by collapsing some of the distinctions he had made in his earlier FMCS technique, but FMCS was itself selective, not comprehensive. A benefit of EMFACS and the approach of Camras et al. in defining composites of action in comparison to Izard's MAX and other selective techniques is that what has been excluded is exactly specified.

Facial Electromyography

Facial electromyography (EMG) measures the electrical activity of motor units in the striated muscles of the face. The force and velocity of movement are controlled by the number of motor units and their rate of firing. The size and shape of the waveform represents the movement, which may be visible to the eye or occult depending on the degree of activity and characteristics of the overlying tissue. The signal is recorded using surface electrodes attached to the skin, which is first prepared by a slight scraping and application of paste or solution to enhance electrical contact. Alternatively, fine wire needles are inserted into the muscle, which increases specificity. Thin cables or leads are run from the electrodes to a bio-amplifier.

The electrophysiology of EMG and its acquisition and processing are described in several sources (Cacioppo, Tassinary, & Fridlund, 1990; Fridlund & Cacioppo, 1986; Soderberg, 1992). We discuss here the comprehensiveness, reliability, validity, and utility of facial EMG for measurement of facial motion. Unless otherwise noted, the material presented here refers to surface facial EMG.

Comprehensiveness or selectivity

Facial EMG has relatively low specificity but high spatial and temporal resolution. Because there is more than one muscle in most facial areas, and their fibers interweave or lie on top of each other (Figure 1), placing leads on the surface of the face often has the consequence of picking up activity in more than just the muscle targeted by the investigator. Although investigators using surface EMG have usually been careful to talk about a *region* rather than a muscle, their reasoning and much of their interpretation assumes success in isolating the activity of specific muscles. Ekman and Friesen, in a joint study with Schwartz (1978) found that in the *corrugator* region the activity of many muscles other than *corrugator* itself was recorded by the electrode placed in this region: *orbicularis oculi*; *levator labii superioris alaeque nasi; frontalis, pars medialis*. The activity of these other muscles could be distinguished from that of *corrugator* and from each other, but these distinctions require more electrodes, some of which must be placed in adjacent facial regions. Another way to obtain measurement of specific muscles, as noted above, is to insert fine wires into a muscle, a procedure that, though not as painful as it sounds, requires medical training and certification, and is not practicable for many studies.

Insert Figure 1 About Here

An advantage of facial EMG is its high temporal resolution, which makes it well suited for measuring emotion, which have rapid onset and short duration. An example of the temporal resolution of facial EMG is shown in Figure 2 from (Dimberg, Thunberg, & Grunedal, 2002). Subjects were asked to contract their *zygomatic major* or *corrugator supercilli* muscles (AU 12 in FACS) in response to a picture of a happy or an angry face. Consistent with hypotheses that emphasize automaticity, contraction of *zygomatic major* was facilitated by the happy face while contraction of the *corrugator supercilii* muscle was facilitated by the angry face. The temporal resolution of the recordings was sufficient to discriminate differences in response time within about a half second.

Insert Figure 2 About Here

Types of persons to which the measurement has been applied

With few exceptions, use of facial EMG is limited to older children and adults. Infants and young children are difficult to test with facial EMG because they are less likely to tolerate electrodes attached to their faces. When the method has been used with this population, it has typically been restricted to the *orbicularis occuli* region for measurement of potentiated startle (Balaban, Anthony, & Graham, 1989; L. A. Schmidt & Fox, 1998). In older children, use EMG presents no special problems. We routinely record EMG in the *zygotmatic major*, *corrugator supercilii, levator labii*, and *orbicularis occuli* regions in children age 13 years and older without event (Forbes, Fox, Cohn, Galles, & Kovacs, Submitted).

Reliability

In the past, a problem with facial EMG was lack of a standard system for specifying exactly where to place an EMG electrode in order to detect activity in a particular facial region. The efforts of Fridlund and Cacioppo (1986) to introduce guidelines for EMG placement have let to increasing standardization, which has largely overcome this problem. Method variance due to unknown variation in electrode placement has been reduced with increased adoption of these standards.

Nevertheless, some variation in placement is inherent in the use of electrodes on the face. Consider the use of surface EMG to measure whether there is more or less activity in the *zygomatic major* region on the two sides of the face. Any differences obtained might not be due to the greater involvement of the right or left hemisphere but might to an unknown extent reflect differences in placement of the EMG electrode in relation to the muscle mass on the two sides of the face or to asymmetry in facial structure or tissue (Liu, Schmidt, Cohn, & Mitra, 2003). Between-subjects designs, in which, for example, a measure of *zygomatic major* was correlated with a personality test score, would also be vulnerable to error owing to electrode placement. These problems can be circumvented by utilizing research designs in which EMG activity is compared in two or more conditions for each subject.

When EMG is used to measure change over time, and the leads must be placed on the face more than once, variations in placement of the leads on each occasion can introduce errors. Miller (1981-1982) addressed this problem by devising a template that can be attached to a subject repeatedly to ensure that electrode placement is identical on different occasions.

Reliability for EMG intensity has been shown by comparing EMG and FACS intensity scoring. Persons highly skilled in the ability to activate specific muscles (Ekman and Oster) contracted them on command at different intended intensity levels, while a video record was made and surface EMG was recorded. FACS scoring was later found to be highly correlated with the EMG readings (Pearson *r* = 0.85) (Ekman et al., 1978). Figure 3 shows an example from this data, a plot of the relationship between EMG measures of electrical activity and FACS scoring of the intensity of action for a specific muscle.

Insert Figure 3 About Here

Validity

A number of studies have used surface EMG to measure muscle activity in relation to emotion and found evidence of good concurrent and predictive correlation with self- and observer reported emotion (Cacioppo, Martzke, Petty, & Tassinary, 1988; Cacioppo et al., 1992; Cohn et al., 2002; Dimberg et al., 2002; Fridlund et al., 1990; Tassinary & Cacioppo, 1992). Most of this literature has used facial EMG to discriminate between positive and negative emotion (Cacioppo, Petty, Losch, & Kim, 1986).

An issue is whether EMG can provide measurement of more than just one or two emotional states. Most emotions cannot be identified by the activity of a single muscle. Happiness may be the only exception, but even here evidence (Ekman, Davidson, & Friesen, 1990; Frank et al., 1993) suggests that the differentiation of felt from simulated happiness, of controlled from uncontrolled happiness, and of slight from extreme happiness requires measurement of more than one muscle. Disgust might be measured by the activity of two muscles, and surprise by the activity of three. To measure anger, fear, or sadness, many muscles need to be measured. There are limits, however, to the number of leads that can be placed on a person's face without unduly interfering with the behavior under study. Nevertheless, there have been some successful efforts in discriminating among three or more emotions using facial EMG (Fridlund, Schwartz, & Fowler, 1984; Vrana, 1993).

Utility

Facial EMG has had an important role in certain methodological studies of facial behavior. Mention was made earlier of Ekman and Friesen's use of fine-wire EMG to stimulate and record facial movement in order to discover how the muscles work to change appearance. Facial EMG could be used to help teach people how the muscles work as part of the process of teaching them a visual measurement procedure such as FACS or as part of physical rehabilitation in the case of facial neuromuscular disorders. Facial EMG can be used to calibrate and investigate measurement of visible facial behavior.

Another important use for facial EMG is to measure phenomena that are difficult or impossible to measure with techniques based on visible movements (Tassinary & Cacioppo, 1992). Ekman, Schwartz, and Friesen (Ekman et al., 1978) found that there are reliable electrical changes associated with muscle tonus changes that are not visible. For two muscles studied systematically (*corrugator* and *frontalis*, *pars medialis*), there were significant changes in EMG without any visible sign of activity when the performer was instructed just to think about each muscle. This study also showed that there are visible clues to muscle tension, measurable by EMG, when there is no movement. The persons measuring the faces with FACS guessed which muscle had been tensed when they could not see any movement. Sometimes the person guessing felt that there was no basis for the guess. At other times there seemed to be evidence of very slight tightening or bulging of skin. Analyses showed that when these guesses were correct—

when the scorer predicted which muscle the performer was tensing, even though no movement was visible ─ there was a greater increase in EMG than when the guesses were incorrect.

For measuring visible changes in the face, work reported in the next section suggests that facial EMG has high concurrent validity with visible intensity changes in *zygomatic major*, with average correlation above 0.90.

Stable individual differences

Facial EMG shows moderate test-retest stability over relatively long intervals, comparable to that for self-reported emotion. As one example, in a longitudinal study of emotion regulation, 66 adults viewed short film clips on two occasions separated in time by 12 months or more. On both occasions, EMG was measured in four facial regions. After viewing each film, subjects rated their degree of enjoyment on Likert-type scales. EMG in the *zygomatic major* region was analyzed for the film intended to elicit enjoyment. Stability coefficients for facial EMG and self-reported emotion were nearly identical, 0.58 and 0.56, respectively (Cohn et al., 2002).

Costs

EMG requires specialized equipment and staff trained in psychophysiology, which entails significant laboratory and personnel costs. Data processing is efficient, however, and significantly less time-intensive than manual coding. The need to attach electrodes to the face, on the other hand, is mildly invasive and is a limiting factor in use of EMG. Cabling from the electrodes to an acquisition device effectively confines the wearer's activity to a relatively small area making use in naturalistic settings difficult. Telemetric recording, which dispenses with cabling, could be helpful in this regard (Gerleman & Cook, 1992). Another limitation is that facial EMG may inhibit facial activity. Large or sudden head or facial motion can loosen the electrodes. To prevent these problems, subjects usually have been studied in isolation. Even when subjects have been studied in a social context (Fridlund, 1991), social interaction among subjects tends to be avoided. Subjects typically have been measured when trying to pose, imagine, remember, or create for themselves an emotional experience. Even in these situations, if a subject makes a large expression they will feel the tape that holds the electrode in place pull or tear, which could inhibit large expressions, even if the experimenter does not explicitly discourage large expressions by instruction, choice of task for the subject to perform, or restrictions on context, such as limiting social interaction. The seriousness of these concerns is difficult to evaluate since comparisons between manual coding and facial EMG have been few (Cohn & Schmidt, 2004).

In summary, EMG may be the only method for measuring non-visible changes in muscular tension, and for measuring changes that, while barely visible, involve not movement but bulging of the skin and would be hard to measure with any of the techniques described in Table 1. It also may be useful as a method for automatically measuring quantitative change in facial muscles related to emotion-eliciting stimuli. The need to attach electrodes to the face limits applications to those for which invasive methods are feasible. To automatically measure quantitative change in facial muscles non-invasively, other methods are needed.

Automatic Facial Image Analysis

Within the past 5-10 years, there has been considerable effort toward automatic measurement and recognition of facial expression by computer vision, which is the science of extracting and representing feature information from digitized images and recognizing perceptually meaningful patterns. Early work used markers to enhance facial features (Kaiser & Wehrle, 1992;

Terzopoulos & Waters, 1990), and markers are used still in some applications (Wachtman, Cohn, VanSwearingen, & Manders, 2001). What are referred to as motion capture techniques use reflective markers attached to the skin to facilitate feature extraction. Commercially available systems include those from Vicom™ and Peak Performance™. As with facial EMG, motion capture approaches are expensive and require specialized training and expertise to use; and reflectors attached to the skin may as with electrodes inhibit facial expression. Most current research in automatic facial image analysis requires no markers or other enhancement of facial features. We review progress here in the development of markerless systems for measurement of facial actions.

Most of the work in automatic facial expression recognition has focused on emotion-specified expressions, such as joy and anger (Black & Yacoob, 1994; Essa & Pentland, 1997; Lyons, Akamasku, Kamachi, & Gyoba, 1998; Padgett & Cottrell, 1998; Yacoob & Davis, 1997). Within the last five years, the more challenging task of recognizing facial sign vehicles has received increasing attention. At least four research groups (See Table 4) have reported results for automatic recognition of facial sign vehicles in digitized video without aid of facial markers. All used FACS to define facial sign vehicles due in large part to its descriptive power in modeling facial action.

Insert Table 4 About Here

Each of these four research groups has automatically recognized FACS action units without relying on artificial enhancement of facial features. Comprehensive reviews of the literature in automatic facial expression analysis and recognition can be found in (Fasel & Luettin, 2003; Pantic & Rothkrantz, 2000a, 2003; Tian, Cohn, & Kanade, In press)

Automatic recognition of facial actions must solve four tasks: extraction of facial features, image alignment, action unit recognition, and system integration. We review each of these in turn and then evaluate the current state-of-the art in automatic action unit recognition. Before doing so, we first consider the type of video records required for analysis.

Types of records and persons to which the measurement has been applied

Still or motion records

While image data may consist of either static images (e.g., photographs) or image sequences (video), analysis of the latter is much further advanced and many of the methods (e.g., optical flow for feature extraction and head tracking for recovery of head orientation) require video input[8]. Video may be recorded using either analog or digital recordings. If recorded using analog tape, digitizing prior to analysis will be needed. Digitizing until recently required specialized equipment and training and was costly. As digital video becomes more common, the expense and expertise required in acquiring digital video or converting from analog video is greatly reduced.

Modifications for varying age levels

Most approaches to automatic facial image analysis have been applied only to adults. Analysis of infant facial actions is challenging because infant faces have relatively little texture and head movements are often sudden and large. Facial texture is important to feature extraction methods such as optical flow (described below), and sudden and large head motion is more difficult to track. Large variation in pose across an image sequence is challenging as well. We have some experience with automatic infant facial image analysis and efforts are continuing

(Cohn, Zlochower, Lien, Hua, & Kanade, 2000; Messinger, Acosta, Cassel, Ambadar, & Cohn, 2004b). Other individual differences, such as skin color, racial background, and gender have been examined. Action unit recognition appears to be unaffected by these factors (Cohn, Xiao, Moriyama, Ambadar, & Kanade, 2003; Cohn, Zlochower, Lien, & Kanade, 1999; Moriyama, Kanade, & Cohn, 2004; Tian, Kanade, & Cohn, 2001).

Tasks in automatic facial image analysis

Feature extraction

A number of approaches have been used to extract feature information from face images. These include difference imaging, principal components analysis (PCA), optical flow, and edge detection. A given system may use one or more of these in combination.

Difference imaging. In a digitized grayscale image, each pixel has an intensity value that varies between 0 and 255. Digitized color images have a larger range of intensity variation. Change from one image to the next may be computed by subtracting one image from another. Figure 4a shows an example of an infant with a relaxed facial expression and partially opened lips (AU 25 in FACS). In the next image (Figure 4b) the same infant begins to smile (AU 6+12). The corresponding difference images appear in the next row. Pixels that change from one image to the next appear as white in the difference image. While this method is relatively efficient in identifying areas of motion, it fails to capture pixel-wise correspondence between face images. Different facial actions might produce identical patterns of intensity differences. Difference images also are easily confounded by head motion, which can be seen in the example.

Insert Figure 4 About Here

Principal components analysis (PCA). Another approach, initially developed for face recognition, is principal components analysis of digitized face images. High dimensional face images (e.g., 640 by 480 gray-scale pixel arrays) can be reduced to a lower dimensional set of eigenvectors, or "eigenfaces" (Turk & Pentland, 1991). Under controlled conditions, eigenvectors can capture differences between action units. A generalization of principal components analysis, referred to as independent components analysis (ICA), appears useful when covariation among pixels includes nonlinear relations. Like other approaches, PCA and ICA perform best when face images are viewed from the front and any head motion is small and remains parallel to image plane of the camera. When these conditions are not met, image alignment, as discussed below, becomes a critical issue.

Optical flow. In FACS, each action unit is anatomically related to contraction of a specific facial muscle. AU 12 (oblique raising of the lip corners), for instance, results from contraction of the *zygomatic major* muscle, AU 20 (lip stretch) from contraction of the *risorius* muscle, and AU 15 (oblique lowering of the lip corners) from contraction of the *depressor anguli* muscle. Muscle contractions produce motion in the overlying tissue. Algorithms for optical flow quantify the magnitude and direction of this motion. When optical flow is computed for the entire face image, it is referred to as **dense flow.** Figure 5 shows an example of dense flow extraction. In the initial image, each point represents a selected pixel whose motion will be represented by motion vectors across the image sequence. As the jaw drops, the eyes widen, and the brows are raised, dense flow systematically captures these facial actions.

Insert Figures 5 and 6 About Here

Obtaining dense flow for the whole face image is computationally intensive. In our experience, it is more efficient to compute feature motion for a small set of localized facial features. Tracking specific “feature points” in these regions yields motion that is highly consistent with that obtained from dense flow (Figure 6). For action unit recognition, Lien (Lien, Kanade, Cohn, & Li, 2000) found that the two approaches to optical flow computation achieved similarly high accuracy for action unit recognition.

Edge detection. Facial motion produces transient wrinkles and furrows perpendicular to the motion direction of the activated muscle. These transient features provide information relevant to the recognition of action units. Contraction of the *corrugator* muscle, for instance, produces vertical furrows between the brows, which is coded in FACS as AU 4, while contraction of the medial portion of the *frontalis* muscle (AU 1) causes horizontal wrinkling in the center of the forehead. Some of these lines and furrows may become permanent with age. Permanent crows-feet wrinkles around the outside corners of the eyes, which is characteristic of AU 6 when transient, are common in adults but not in infants. When lines and furrows become permanent facial features, contraction of the corresponding muscles produces changes in their appearance, such as deepening or lengthening. The presence or absence of these lines and furrows in a face image can be found by edge feature analysis or by the use of spatial and frequency filters (Bartlett, Ekman, Hager, & Sejnowski, 1999) (Tian, Kanade, & Cohn, 2000) (Tian, Kanade, & Cohn, 2002). Wrinkles and furrows present at rest may be ‘removed’ by thresholding the edge image. In our work, we detect wrinkles and furrows in the forehead (e.g., AU 1 and AU 2), lateral to the eye corners (AU 6), the nasal root (AU 4), and the nasolabial region (e.g., AU 10 and 12) by a combination of edge detection and spatio-frequency filters.

Image alignment

Facial actions often co-occur with head movement, such as when people raise their head in surprise or turn toward a friend while beginning to smile (Camras et al., 1996; Kraut & Johnson, 1979). Expression may also vary as a result of individual differences in facial proportions (Farkas & Munro, 1994; K. L. Schmidt, Tian, & Cohn, 2003). Head motion, individual differences in facial proportions, and camera orientation are all potential confounds in extracting feature information from digitized face images (Kanade, Cohn, & Tian, 2000). Camera orientation may be frontal (that is, parallel to the image plane of the face) or to the side, which changes the appearance of face images. While variation due to pose and motion may be eliminated by securing the head in a clamp, as is typically done in neuro-imaging studies, or by wearing a head-mounted camera (Pantic & Rothkrantz, 2004), these solutions are not without limitations. We seek accurate and efficient image alignment, which is critical for valid feature extraction, without imposing any constraints on subjects' activity.

When out-of-plane rotation of the head is small, either an affine or a perspective transformation of images can align images so that face position, size, and orientation are kept relatively constant across subjects, and these factors do not interfere significantly with feature extraction. The affine transformation is computationally faster, but the perspective transformation gives more accurate warping for a higher degree of out-of-plane rotation (Lien et al., 2000). For larger out-of-plane motion, it is necessary to model the head as a 3D object. (Xiao, Kanade, & Cohn, 2003) developed a 3-D head tracker using a cylindrical head model. The tracker precisely estimates the 6 degrees of freedom of head motion: movement in the horizontal and vertical planes (i.e., translation), movement toward and away from the camera (i.e., scale), rotation, pitch, and yaw. Once these parameters are estimated, the face image is stabilized, by warping each frame to a common orientation and size. In this way, motion due to expression is not confounded by rigid head motion. Figure 7 shows an example of automatic head tracking, image alignment, and feature localization.

Insert Figure 7 About Here

An alternative to a cylindrical head model is to use either a generic face or person-specific face model. The UCSD group (Bartlett, Braathen, Littlewort, Sejnowski, & Movellan, 2001) used a generic face model to estimate 3D head position and warp face images to a common view. To date, this model requires manual initialization of each frame and so is not yet functional for automatic processing. The CMU/Pitt group has developed a person-specific face model that automatically initializes and recovers full 6 degrees of freedom of head motion as well as tracking facial expression and direction of gaze (Xiao, Baker, Matthews, & Kanade, 2004). Before the person-specific head model may be used, some training is required. Typically, 15 to 20 images are hand labeled prior to use. Like the cylindrical head model, the person-specific head model is robust to occlusion and runs at frame rate (30 frames per second) or faster (Gross, Matthews, & Baker, 2004; Xiao et al., 2004)

Action unit recognition

Once quantitative information is extracted from an image sequence, the measurements can be used to recognize facial actions. The data first are divided into a “training” set and a “testing” set. One is used for training a classifier; the other is used to test its validity and utility in an independent sample. A number of classifiers have been used. The most common are artificial neural networks (NN) and hidden Markov models (HMM). HMM uses temporal information, whereas NN algorithms with few exceptions do not. Lien et al. (Lien et al., 2000) found that

HMM and discriminant analysis produced highly similar results for their data. Bartlett, Movellan, Sejnowski and their colleagues have been especially active in comparing the strengths and weakness of various classifiers (Bartlett et al., 2001; Bartlett, Littlewort, & Movellan, 2004); their findings suggest that system performance may be optimized by careful selection. Whatever classifier is used, to ensure generalizability it is important that training and testing images be independent, preferably with no subjects included in both training and testing image sequences, and that the number of image sequences and samples of target action units in each set be sufficiently large. While some investigators have used upwards of 500 or more sequences from 100 subjects with a minimum of 25 action units of each type (Cohn et al., 1999), others have used much smaller samples of action units and subjects, for which results may generalize poorly to new situations. Fasel and Luettin (Fasel & Luettin, 2000), for instance, used image data from a single subject for training and testing their method of automatic action unit recognition.

System integration

For research purposes, the various components of an automated system need not be integrated. To be useful for theoretical and applied research in behavioral science, ease of use is an important feature. The CMU/Pitt Automated Face Analysis system (AFA) affords an example of how components may be integrated. Shown in Figure 8 is an overview of version 3 of their system (Cohn & Kanade, In press; Cohn, Reed, Ambadar, Xiao, & Moriyama, 2004).

Insert Figure 8 About Here

Given an image sequence, the face and approximate location of individual face features are detected automatically in the initial frame. The contours of the face features and components then are adjusted manually as needed in the initial frame. The image sequence then is processed.

A cylinder-based 3D head model is used to estimate the 6 degrees of freedom of head orientation and to stabilize the face image across the image sequence. Stabilization entails warping each face image to a common frontal view. Both permanent (e.g., brows, eyes, lips) and transient (lines and furrows) face feature changes are tracked in the image sequence using a combination of optical flow, color, and edge detection (Tian et al., 2000, 2002). Facial feature parameters are fed to a neural network-based classifier for action unit recognition. Output from all processing steps is automatically stored in linked database files for export to statistical packages. The feature trajectories may be used to model the timing of facial actions as well as for action unit recognition (Cohn, Reed, Moriyama et al., 2004; K. L. Schmidt, Cohn et al., 2003).

The system uses multiple types of image features (e.g., optical flow and edge information). For some action units, only one or another type of feature may provide useful information. For instance, with AU 14, which causes dimpling lateral to the lip corners, texture information obtained rather than motion is needed. For most action units, the use of multiple features provides convergent information (as when smiling, or AU 12, is indicated by oblique motion of the lip corners and deepening and change in orientation of the nasolabial furrows), which increases precision of measurement and increases accuracy of action unit recognition (see also (Bartlett et al., 1999).

Reliability

Approaches to automatic facial image analysis often entail some manual preprocessing, such as manually marking permanent facial features (e.g., eyes) in the initial image. To evaluate reliability of manual feature marking, Cohn and colleagues (1999) compared the results of pairs of coders for manual feature marking of 33 feature points. Mean inter-observer error was 2.29 and 2.01 pixels in the horizontal and vertical dimensions, respectively. Mean inter-observer reliability, quantified with Pearson correlation coefficients, was 0.97 and 0.93 in the horizontal and vertical dimensions, respectively. Most important, agreement on FACS coding between automated facial image analysis and manual FACS coding in several studies was comparable to that of manual FACS coding (Cohn et al., 2003; Cohn et al., 1999; Tian et al., 2001, 2002) . This finding suggests that any error in feature labeling is unrelated to the accuracy of system performance. As techniques change however, it will be important to continue to assess the reliability of any human preprocessing.

Validity

Concurrent validity for action unit recognition has been evaluated by comparing automatic and manual FACS coding of both directed facial action tasks and spontaneous facial behavior. Concurrent validity for intensity has been evaluated by comparing automatic facial image analysis and both facial EMG and q-sorts by human judges of spontaneous facial behavior. Spontaneous facial behavior included non-frontal orientation to the camera, small to moderate out-of-plane head motion, and occlusion by glasses.

Concurrent validity with manual FACS coding

Directed facial action tasks. Fasel and Luettin (2000) analyzed facial action in a subject who was an expert in FACS. For 9 action units and 7 action unit combinations, they achieved 74% accuracy. Pantic (2000b; Valstar, Pantic, & Patras, 2004) achieved moderate to high accuracy for 29 action units. This result was achieved using dual views (frontal and profile), and facial actions were recorded using a head-mounted camera, which effectively eliminated head motion and pose variation. Others have used a single, tripod-mounted camera.

The most extensive studies of directed facial action tasks have been conducted by the CMU/Pittsburgh, UCSD, and Delft groups. The CMU/Pittsburgh group achieved 81% to 96%

accuracy for 19 action units: six in the upper face (AU 1, AU 2, AU 4, AU 5, AU 6, AU 7) and 13 in the lower face (AU 9, AU 10, AU 12, AU 15, AU 17, AU 20, AU 25, AU 26, AU 27, AU 23. AU 41, AU 42, AU 45) (Tian et al., 2001, 2002). The action units recognized were ones most common in emotion expression and social behavior and represent 19 of 31 action units that have a known anatomical basis (Kanade et al., 2000). Moreover, action units were recognized whether or not they occurred in combinations, many of which involved co-articulation effects, which suggests that the system is capable of making the kinds of complex perceptual discriminations made by human observers. This capability is important because the number of possible action unit combinations numbers in the thousands. If the system had to learn each combination separately, the task would become intractable. These findings suggest that these systems are on course toward achieving the comprehensiveness of manual FACS coding.

Spontaneous facial behavior. Spontaneous facial behavior presents greater challenge to automatic facial image analysis than does directed facial action tasks. Orientation to the camera typically is non-frontal, moderate to large head motion is common, and facial occlusion by glasses, facial jewelry, and hand gesture occurs. In initial tests, we (Cohn et al., 2003) analyzed image data from Frank and Ekman (1997) in which subjects were interviewed about a mock theft as part of a study of deception. Image data from ten subjects were analyzed. The subjects were ethnically heterogeneous, two wore glasses, and small to moderate out of plane head motion was common. All instances of AU 45 (blinking) during 1 minute of each interview were analyzed. Automatic facial image analysis (AFA) and manual FACS coding agreed in 98% of cases. In related work using the same image database of spontaneous facial behavior, AFA achieved 76% agreement between manual FACS coding of action units in the brow region and automatic recognition (Cohn, Reed, Ambadar et al., 2004). These initial findings suggest concurrent validity of AFA with manual FACS coding of AU 1+2, AU 4, and AU 45 in spontaneous facial behavior with variable pose, moderate out-of-plane head rotation, and occlusion.

Concurrent validity with facial EMG for action unit intensity in spontaneous facial behavior

To evaluate concurrent validity for degree of eye closure (AU 45) in the Frank and Ekman image data described above, luminance intensity of the upper eye region as determined automatically was normalized over the range of 0 to 1. Luminance was darkest when the eye was open (normalized luminance = 1) and brightest (luminance = 0) when the eye was closed. The digitized images then were randomly sorted. Two researchers, blind to the results of automatic processing, then manually sorted (i.e., q-sort) each sequence from eye open to closed to open. They next estimated the degree of eye closure on a scale from 0 (eye closed) to 1 (eye open). A representative example is shown in Figure 6. In each of 10 sequences examined, automatic analysis and human judgment were highly consistent. An example is shown in Figure 9.

Insert Figure 9 About Here

To evaluate concurrent validity for contraction of *zygomatic major* (AU 12), Cohn, Schmidt, Gross, & Ekman (2002) collected image and EMG data from subjects while they watched a film clip intended to elicit enjoyment. Contraction of *zygomatic major* was determined by EMG. When visible smiling was observed, it was confirmed by manual FACS coding. Feature vectors from the lip corner were highly consistent with facial EMG recorded from the *zygomatic major* region. In 72% of cases with a distinct EMG and visible smile onset, feature point tracking by optical flow and facial EMG were highly correlated, with an average time lag of 0.23 seconds. An example is shown in Figure 10.

Insert Figure 10 About Here

Utility

AFA has been used to investigate theoretical and applied issues involving facial action. Some of the applications include assessment of facial neuromuscular disorders (Wachtman et al., 2001), facial asymmetry in biometrics (Liu et al., 2003), the timing of spontaneous and deliberate smiles (Cohn & Schmidt, 2004; K. L. Schmidt, Cohn et al., 2003), the relation between head motion, smiling, and direction of gaze (Cohn, Reed, Moriyama et al., 2004), brow raising and lowering (Cohn, Reed, Ambadar et al., 2004) , and facial expression in infants (Cohn et al., 2000; Messinger, Acosta, Cassel, Ambadar, & Cohn, 2004a). The scope of applications in theoretical and applied research can be expected to increase further as development efforts continue and the system becomes available to other investigators.

Remaining challenges

Before AFA and related systems are ready for release, several challenges must be addressed. These include how to parse the stream of behavior, prevent error accumulation, and increase automation. AFA and other systems have assumed that facial actions begin and end from a neutral face. In actuality, facial expression is more complex. Transitions among action units may involve no intervening neutral state. For automatic facial image analysis, parsing the stream of facial action units under these circumstances is a challenge. Human FACS coders meet this task in part by having a mental representation of a neutral face. For automated facial image analysis, parsing will likely involve higher order pattern recognition than has been considered to date.

Many of the methods used in automated facial image analysis so far involve dynamic templates for which estimates are continually updated. With dynamic templates, error tends to propagate and accumulate across an image sequence. So far, most AFA applications have involved relatively short image sequences up to 10 seconds or so, for which error accumulation is not a significant problem. For longer sequences, an appropriate measure is required. The head tracking module in AFA overcomes this problem through a combined use of robust regression and reference images. Robust regression identifies and discounts the effects of outliers, and reference images provide a way to reinitialize estimates so as to reduce error accumulation. For head tracking, this approach has been highly successful. The cylinder model head tracker has performed well for image sequences as long as 20 minutes. Similar capability will be needed for action unit recognition.

Almost all current methods entail some manual initialization, such as labeling permanent facial features (e.g., eyes or mouth) with a compute mouse in one or more face images. This is especially the case when person-specific face models are used. These models may require hand labeling of 20 to 30 images; once this is completed, these models are relatively robust to error accumulation and operate automatically on long sequences. While a fully automated system is not always necessary for all applications, increased automation will reduce the personnel costs of using the system and increase the kinds of applications for which it may be used.

Conclusions

This chapter has reviewed measurement techniques for only one type of signal: rapid, not slow or static. Among these, only one kind of rapid signal ─ visible movement ─ has been considered. Most of the studies that have used one or another technique to measure visible movement were concerned with only one of the many messages rapid signs may convey: information about emotion. Presumably, future research will expand to consider other messages and to develop methods for measuring rapid signals other than movement, as well as the variety of slow and static signals.

A few manual coding techniques have become widely used, especially that of Ekman and Friesen and to a lesser extent Izard. The former was designed to be applicable to the study of any message, not just emotion. Wedding studies of facial sign vehicles to studies using the more traditional message judgment approach should allow discovery of the particular actions that form the basis for correct and incorrect inferences when people judge facial expression (Juslin & Scherer, In press; Oster et al., 1992). These techniques may also allow discovery of particular facial actions that are not customarily known or even knowable by the usual observer, movements that are too subtle and/or complex to notice or interpret when seen once at real time.

As further research is generated by the facial measurement techniques reviewed here, the techniques themselves may undergo further development or be replaced by other measurement approaches. This development may be seen in the system of Ekman and Friesen, which exists now in 3 versions: the initial version, FACS 1978, FACS 1992 (update document based), and FACS 2002, which includes significant improvements in scoring criteria and in didactic materials, including extensive use of hyperlinked cross-referenced text and embedded video links in the CD version. With the release of new versions, such as that of FACS 2002, it becomes critical that those who publish findings using one or the other version identify which version they have used. Even better would be for investigators to use the most current version of a system, as is done routinely in fields such as intelligence testing and clinical diagnosis in which new versions of assessment instruments are common.

In part because of its descriptive power, the technique of Ekman and Friesen has encouraged a wide range of research on facial movement (Ekman & Rosenberg, In press) as well as becoming influential in the fields of computer animation (e.g., (Parke & Waters, 1996)) and automated facial expression recognition, in which fine-grained description of motion parameters is needed.

The development of automated methods of facial expression analysis, in particular, is an exciting development. Automated analysis using computer vision produces both action unit recognition and quantitative measures of feature trajectories (e.g., K. L. Schmidt, Cohn et al., 2003). Initial work suggests that Automated Face Analysis has high concurrent validity for both action unit recognition and intensity variation, as assessed by trained observers and facial EMG. Automatic analysis has several potential advantages. By computing quantitative measures of facial action over time, powerful statistical techniques may be used to asses individual facial behavior and dyadic behavior, such as synchrony and dominance (Boker, Xu, Rotondo, & King, 2002; Cohn & Tronick, 1988). From an information processing perspective, comparisons between automated and human-observer based facial expression analysis would afford a new means of studying social perception. In addition, a system that operates in real time could provide continuous monitoring and feedback for research and clinical applications. While work in this area is still in the early stages, initial applications to theoretical and clinical issues are encouraging.

References

Ancoli, S. (1979). *Psychophysiological response patterns of emotion.* University of California San Francisco, San Francisco, California.

Balaban, M. T., Anthony, B. J., & Graham, F. K. (1989). Prestimulation effects on blink and cardiac responses of 15-month-old infants: Reactions to facial expressions of emotion. *Developmental Psychobiology, 22*, 115-127.

Bartlett, M. S., Braathen, B., Littlewort, G., Sejnowski, T. J., & Movellan, J. R. (2001). *A comparative study of alternative FACS coding algorithms. MPLAB-TR-2001-06*. San Diego, California: Institute for Neural Computation, University of California, San Diego.

Bartlett, M. S., Ekman, P., Hager, J., C., & Sejnowski, T. J. (1999). Measuring Facial Expressions by Computer Image Analysis. *Psychophysiology, 36*, 253-263.

Bartlett, M. S., Littlewort, G., & Movellan, J. R. (2004). Machine learning methods for fully automatic recognition of facial expressions and facial actions. *IEEE Conference on Systems, Man, and Cybernetics*, The Hague, The Netherlands

Birdwhistell, R. (1952). *Introduction to kinesics*. Louisville, KY: University of Louisville Press.

Black, M. J., & Yacoob, Y. (1994). Recognizing facial expressions under rigid and non-rigid facial motions. *International Workshop on Automatic Face- and Gesture Recognition*, Zurick. 12-17

Blurton Jones, N. G. (1971). Criteria for use in describing facial expression in children. *Human Biology, 41*, 365-413.

Boker, S. M., Xu, M., Rotondo, J. L., & King, K. (2002). Windowed cross–correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychological Methods, 7*, 338-355.

Brannigan, C. R., & Humphries, D. A. (1972). Human nonverbal behavior, a means of communication. In N. G. Blurton Jones (Ed.), *Ethological studies of child behavior*. Cambridge: Cambridge University Press.

Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception*. New York, New York: Oxford University.

Cacioppo, J. T., Martzke, J. S., Petty, R. E., & Tassinary, L. G. (1988). Specific forms of facial EMG response index emotions during an interview: From Darwin to the continuous flow hypothesis of affect-laden information processing. *Journal of Personality and Social Psychology, 54*, 592-604.

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H.-S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology, 50*, 260-268.

Cacioppo, J. T., Tassinary, L. G., & Fridlund, A. J. (1990). The skeletomotor system. In J. T. Cacioppo & L. G. Tassinary (Eds.), *Principles of psychophysiology: Physical, social, and inferential elements*. New York, New: Cambridge University.

Cacioppo, J. T., Uchino, B. N., Crites, S. L., Snydersmith, M. A., Smith, G., Berntson, G. G., et al. (1992). Relationship between facial expressiveness and sympathetic activation in emotion: A critical review, with emphasis on modeling underlying mechanisms and individual differences. *Journal of Personality and Social Psychology, 62*, 110-128.

Camras, L. A. (1992). Expressive development and basic emotions. *Cognition and Emotion, 6*, 269-283.

Camras, L. A., Lambrecht, L., & Michel, G. F. (1996). Infant "Surprise" Expressions as Coordinative Motor Structures. *Journal of Nonverbal Behavior, 20*, 183-195.

Camras, L. A., Oster, H., Campos, J. J., Miyake, K., & Bradshaw, D. (1992). Japanese and American infants' responses to arm restraint. *Developmental Psychology, 28*, 578-583.

Clemente, C. D. (1997). *Anatomy: A regional atlas of the human body (4th edition)*: Williams & Wilkins.

Cohn, J. F., Campbell, S. B., Matias, R., & Hopkins, J. (1990). Mother-infant face-to-face interactions of postpartum depressed and non-depressed mothers. *Developmental Psychology, 26*, 15-23.

Cohn, J. F., & Kanade, T. (In press). Use of automated facial image analysis for measurement of emotion expression. In J. A. Coan & J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment*. New York, NY: Oxford.

Cohn, J. F., Reed, L. I., Ambadar, Z., Xiao, J., & Moriyama, T. (2004). Automatic analysis and recognition of facial actions and out-of-plane head motion in spontaneous facial behavior. *IEEE Conference on Systems, Man, and Cybernetics*, the Hague, the Netherlands

Cohn, J. F., Reed, L. I., Moriyama, T., Xiao, J., Schmidt, K. L., & Ambadar, Z. (2004). Multimodal coordination of facial action, head rotation, and eye motion. *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul, Korea. 645-650

Cohn, J. F., & Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing, 2*, 1-12.

Cohn, J. F., Schmidt, K. L., Gross, R., & Ekman, P. (2002). Individual differences in facial expression: Stability over time, relation to self-reported emotion, and ability to inform person identification. *International Conference on Multimodal User Interfaces*, Pittsburgh, PA. 491-496

Cohn, J. F., & Tronick, E. Z. (1983). Three month old infants' reaction to simulated maternal depression. *Child Development, 54*, 185-193.

Cohn, J. F., & Tronick, E. Z. (1988). Mother-Infant face-to-face interaction: Influence is bidirectional and unrelated to periodic cycles in either partner's behavior. *Developmental Psychology, 34*, 386-392.

Cohn, J. F., Xiao, J., Moriyama, T., Ambadar, Z., & Kanade, T. (2003). Automatic recognition of eye blinking in spontaneous facial behavior. *Behavior Research Methods, Instruments, and Computers, 35*, 420-428.

Cohn, J. F., Zlochower, A. J., Lien, J. J. J., Hua, W., & Kanade, T. (2000). Automated face analysis. In C. Rovee-Collier & L. Lipsitt (Eds.), *Progress in infancy research (Vol. 1)* (pp. 155-182). Hillsdale, NJ: Erlbaum.

Cohn, J. F., Zlochower, A. J., Lien, J. J. J., & Kanade, T. (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology, 36*, 35-43.

Darwin, C. (1872/1998). *The expression of the emotions in man and animals (3rd Edition)*. New York, New York: Oxford University.

Dimberg, U., Thunberg, M., & Grunedal, S. (2002). Facial reactions to emotional stimuli: Automatically controlled emotional responses. *Cognition and Emotion, 16*, 449–471.

Donato, G. L., Bartlett, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. (1999). Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21*, 974-989.

Duchenne, B. (1862). *Mechanisme de la physionomie humaine; ou, Analyse electrophysiologique de l'expression des passions*. Paris: Bailliere.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and analysis*. New York, NY: Wiley.

Ekman, P. (1964). Body position, facial expression and verbal behavior during interviews. *Journal of Abnormal and Social Psychology, 68*, 295-301.

Ekman, P. (1965). Differential communication of affect by head and body cues. *Journal of Personality and Social Psychology, 2*, 725-735.

Ekman, P. (1980). Facial asymmetry. *Science, 209*, 833-834.

Ekman, P. (1997). Lying and deception. In N. L. Stein, P. A. Ornstein, B. Tversky & C. Brainerd (Eds.), *Memory for everyday and emotional events*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain Psychology II. *Journal of Personality and Social Psychology, 58*, 342-353.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior. *Semiotica, 1*, 49-98.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1982). *Rationale and reliability for EMFACS.*

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial cues*. Cambridge, Massachusetts: Malor Books.

Ekman, P., Friesen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology, 39*, 1125-1134.

Ekman, P., Friesen, W. V., & Hager, J. C. (Eds.). (2002). *Facial action coding system*: Research Nexus, Network Research Information, Salt Lake City, UT.

Ekman, P., Friesen, W. V., & Simons, R. (1985). Is the startle reaction an emotion? *Journal of Personality and Social Psychology, 49*, 1416-1426.

Ekman, P., Friesen, W. V., & Tomkins, S. S. (1971). Facial affect scoring technique: A first validation study. *Semiotica, 3*, 37-58.

Ekman, P., Hagar, J. C., & Friesen, W. V. (1981). The symmetry of emotional and deliberate facial actions. *Psychophysiology, 18*, 101-106.

Ekman, P., & Heider, K. G. (1988). The universality of a contempt expression: A replication. *Motivation and Emotion, 12*, 303-308.

Ekman, P., & Rosenberg, E. (Eds.). (In press). *What the face reveals (2nd Edition)*. New York, NY: Oxford.

Ekman, P., Schwartz, G., & Friesen, W. V. (1978). *Electrical and visible signs of facial action*. San Francisco, California: Human Interaction Laboratory, University of California San Francisco.

Ermiane, R., & Gergerian, E. (1978). *Atlas of facial expressions; Album des expressions du visage*. Paris: La Pensee Universelle.

Essa, I., & Pentland, A. (1997). Coding, analysis, interpretation and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 7*, 757-763.

Farkas, L. G., & Munro, I. R. (1994). *Anthropometry of the face and head*. New York, New York: Raven Press.

Fasel, B., & Luettin, J. (2000). Recognition of asymmetric facial action units and intensities. *International Conference on Pattern Recognition*, Barcelona, Spain

Fasel, B., & Luettin, J. (2003). Automatic facial expression analysis: A survey. *Pattern Recognition, 36*, 259-275.

Forbes, E. E., Fox, N. A., Cohn, J. F., Galles, S. J., & Kovacs, M. (Submitted). Children's affect regulation during a disappointment: Psychophysiological characteristics, risk for depression, and behavior problems.

Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stakes lies. *Journal of Personality and Social Psychology, 72*, 1429-1439.

Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology, 64*, 83-93.

Fridlund, A. J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology, 60*, 229-240.

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology, 23*, 567-589.

Fridlund, A. J., Sabini, J. P., Hedlund, L. E., Schaut, J. A., Shenker, J. J., & Knauer, M. J. (1990). Audience effects on solitary faces during imagery: Displaying to the people in your head. *Journal of Nonverbal Behavior, 14*, 113-137.

Fridlund, A. J., Schwartz, G. E., & Fowler, S. C. (1984). Pattern recognition of self-reported emotional state from multiple-site facial EMG activity during affective imagery. *Psychophysiology, 21*, 622-637.

Frois-Wittmann, J. (1930). The judgment of facial expression. *Journal of Experimental Psychology, 13*, 113-151.

Fulcher, J. S. (1942). "Voluntary" facial expression in blind and seeing children. *Archives of Psychology, 38*, 1-49.

Gerleman, D. G., & Cook, T. M. (1992). Instrumentation. In G. L. Soderberg (Ed.), *Selected topics in surface electromyography for use in the occupational setting: Expert perspectives* (pp. 44-69). Washington, DC: U. S. Department of Health and Human Services.

Grant, N. B. (1969). Human facial expression. *Man, 4*, 525-536.

Gross, R., Matthews, I., & Baker, S. (2004). Generic vs. person specific AAMs. *British Machine Vision Conference*.

Hager, J. C., & Ekman, P. (1985). The asymmetry of facial actions is inconsistent with models of hemispheric specialization. *Psychophysiology, 22*, 307-318.

Harker, L., & Keltner, D. (2001). Expressions of positive emotion in women's college yearbook pictures and their relationship to personality and life outcomes across adulthood. *Journal of Personality & Social Psychology, 80*, 112-124.

Hess, U., Kappas, A., McHugo, G. J., Kleck, R. E., & Lanzetta, J. T. (1989). An analysis of the encoding and decoding of spontaneous and posed smiles: The use of facial electromyography. *Journal of Nonverbal Behavior, 13*, 121-135.

Hjorstjo, C. H. (1970). *Man's face and mimic language*. Lund: Studentlitterature.

Izard, C. E. (1979a). *Facial expression scoring manual (FESM)*. Newark, Delaware: University of Delaware.

Izard, C. E. (1979b). Personal communication.

Izard, C. E. (1983). *Maximally discriminative facial movement coding system (MAX).*Unpublished manuscript, University of Delaware, Newark, Delaware.

Izard, C. E., Hembree, E. A., Dougherty, L. M., & Spizzirri, C. C. (1983). Changes in facial expressions of 2 to 19-month-old infants following acute pain. *Developmental Psychology, 19*, 418-426.

Juslin, P. N., & Scherer, K. (In press). Vocal expression of affect. In J. A. Harrigan, R. Rosenthal & K. R. Scherer (Eds.), *Handbook of nonverbal behavior research methods in the affective sciences*. New York, NY: Oxford.

Kaiser, S., & Wehrle, T. (1992). Automated coding of facial behavior in human-computer interactions with FACs. *Journal of Nonverbal Behavior, 16*, 67-84.

Kanade, T. (1973). *Picture processing system by computer complex and recognition of human faces.* Kyoto, Kyoto.

Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. *IEEE Conference on Automatic Face and Gesture Recognition*. 46-53

Krause, R. (1978). Nonverbales interaktives Verhalten von Stotterern. *Schweizrische Zeitschrift fur Psychologie und ihre Anwendugen, 3*, 16-31.

Kraut, R. E., & Johnson, R. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology, 37*, 1539-1553.

Landis, C. (1924). Studies of emotional reactions: II General behavior and facial expression. *Journal of Comparative Psychology*, 447-509.

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology, 27*, 363-384.

Lien, J. J. J., Kanade, T., Cohn, J. F., & Li, C. C. (2000). Detection, tracking, and classification of subtle changes in facial expression. *Journal of Robotics & Autonomous Systems, 31*, 131-146.

Lightoller, G. H. S. (1925). Facial muscles: The modiolus and muscles surrounding the rima oris with some remarks about the panniculus adiposus. *Journal of Anatomy, 60*.

Littlewort, G., Bartlett, M. S., & Movellan, J. R. (2001). Are your eyes smiling? Detecting genuine smiles with support vector machines and Gabor wavelets. *Joint Symposium on Neural Computation.*

Liu, Y., Schmidt, K. L., Cohn, J. F., & Mitra, S. (2003). Facial asymmetry quantification for expression invariant human identification. *Computer Vision and Image Understanding, 91*, 138-159.

Lyons, M., Akamasku, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. *International Conference on Face and Gesture Recognition*, Nara, Japan

Malatesta, C. Z., Culver, C., Tesman, J. R., & Shephard, B. (1989). The development of emotion expression during the first two years of life. *Monographs of the Society for Research in Child Development, 54*.

Matias, R., & Cohn, J. F. (1993). Are max-specified infant facial expressions during face-toface interaction consistent with differential emotions theory? *Developmental Psychology, 29*, 524-531.

McGrew, W. C. (1972). *An ethological study of children's behavior*. New York: Academic Press.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 26*, 746-748.

Mead, M. (1973). Review of Darwin and facial expression. *Journal of Communication, 25*, 209-213.

Messinger, D., Acosta, S., Cassel, T., Ambadar, Z., & Cohn, J. F. (2004a). Automated measurement of infant expressions: Are the dynamics emotional or facial.

Messinger, D., Acosta, S., Cassel, T., Ambadar, Z., & Cohn, J. F. (2004b). Automated measurement of infant expressions: Are the dynamics emotional or facial. *International Society for Research in Emotion*, New York, NY

Miller, A. J. (1981-1982). Electomyography in analysis of neuromuscular function. In E. Harvold (Ed.), *Hemifacial microsomia and other related craniofacial anomalies. Cleft Palate Journal Monograph*.

Moriyama, T., Kanade, T., & Cohn, J. F. (2004). Meticulously detailed eye model and its application to analysis of facial image. *IEEE Conference on Systems, Man, and Cybernetics (SMC'04)*, the Hague, the Netherlands. xxx-xxx

Nystrom, M. (1974). Neonatal facial-postural patterning during sleep: I Description and reliability of observation. *Psychological Research Bulletin, 14*, 1-16.

Oster, H. (1978). Facial expression and affect development. In M. Lewis & L. A. Rosenblum (Eds.), *Affect development*. New York: Plenum.

Oster, H., & Ekman, P. (1978). Facial behavior in child development. In A. Collins (Ed.), *Minnesota Symposium on Child Development (Vol. 11)*. Hillsdale, New Jersey: Erlbaum.

Oster, H., Hegley, D., & Nagel, L. (1992). Adult judgments and fine-grained analysis of infant facial expressions: Testing the validity of a priori coding formulas. *Developmental Psychology, 28*, 1115-1131.

Oster, H., & Rosenstein, D. (Undated). *Baby FACS: Analyzing facial movement in infants*.Unpublished manuscript, New York, New York University.

Padgett, C., & Cottrell, G. W. (1998). A Simple Neural Network Models Categorical Perception. *Cognitive Science Conference*, Madison, Wisconsin

Pantic, M., & Rothkrantz, M. (2000a). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*, 1424-1445.

Pantic, M., & Rothkrantz, M. (2000b). Expert system for automatic analysis of facial expression. *Image and Vision Computing, 18*, 881-905.

Pantic, M., & Rothkrantz, M. (2003). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE, 91*, 1371-1390.

Pantic, M., & Rothkrantz, M. (2004). Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*. 1449-1461

Parke, F. I., & Waters, K. (1996). *Computer facial animation.* Wellesley, MA: A.K. Peters.

Rosenthal, R. (In press). Conducting judgment studies. In J. A. Harrigan, R. Rosenthal & K. R. Scherer (Eds.), *Handbook of nonverbal behavior research methods in the affective sciences*. New York, NY: Oxford.

Sayette, M. A., Cohn, J. F., Wertz, J. M., Perrott, M. A., & Parrott, D. J. (2001). A psychometric evaluation of the Facial Action Coding System for assessing spontaneous expression. *Journal of Nonverbal Behavior, 25*, 167-186.

Sayette, M. A., Smith, D. W., Breiner, M. J., & Wilson, G. T. (1992). The effect of alcohol on emotional response to a social stressor. *Journal of Studies on Alcohol, 53*, 541-545.

Scherer, K. R. (1992). What does facial expression express? In K. T. Strongman (Ed.), *International Review of Studies on Emotion* (Vol. 2, pp. 138-165): John Wiley & Sons Ltd.

Scherer, K. R., Zentner, M. R., & Stern, D. (2004). Beyond surprise: The puzzle of infants' expressive reactions to expectancy violation. *Emotion*.

Schmidt, K. L., Cohn, J. F., & Tian, Y. (2003). Signal characteristics of spontaneous facial expressions: Automatic movement in solitary and social smiles. *Biological Psychology, 65*, 49-66.

Schmidt, K. L., Tian, Y., & Cohn, J. F. (2003). The role of structural facial asymmetry in symmetry of peak facial expressions. *European Conference on Facial Expression, Measurement and Meaning*, Rimini, Italy

Schmidt, L. A., & Fox, N. A. (1998). Fear-potentiated startle responses in temperamentally different human infants. *Developmental Psychobiology, 32*, 113-120.

Soderberg, G. L. (Ed.). (1992). *Selected topics in surface electromyography for use in the occupational setting: Expert perspectives*. Washington, DC: National Institute for Occupational Health and Safety.

Tassinary, L. G., & Cacioppo, J. T. (1992). Unobservable facial actions and emotion. *Psychological Science, 3*, 28-33.

Terzopoulos, D., & Waters, K. (1990). Analysis of facial images using physical and anatomical models. *IEEE International Conference on Computer Vision*. 727 732

Tian, Y., Cohn, J. F., & Kanade, T. (In press). Facial expression analysis. In S. Z. Li & A. K. Jain (Eds.), *Handbook of face recognition*. New York, New York: Springer.

Tian, Y., Kanade, T., & Cohn, J. F. (2000). Eye-state detection by local regional information. *Advances in Multimodal Interfaces - ICMI 2000. Third International Conference*, Beijing, China. 143-150

Tian, Y., Kanade, T., & Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*, 97-115.

Tian, Y., Kanade, T., & Cohn, J. F. (2002). Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. *Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition*. 229-234

Tomkins, S. S. (1962). *Affect, imagery, consciousness*. New York: Springer.

Turk, M. A., & Pentland, A. P. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3*, 71-86.

Valstar, M. F., Pantic, M., & Patras, I. (2004). Motion history for facial action detection in video. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*, the Hague, the Netherlands

Vrana, S. R. (1993). The psychophysiology of disgust: Differentiating negative emotinal contexts with facial EMG. *Psychophysiology, 30*, 279-286.

Wachtman, G. S., Cohn, J. F., VanSwearingen, J. M., & Manders, E. K. (2001). Automated tracking of facial features in patients with facial neuromuscular dysfunction. *Plastic and Reconstructive Surgery*, 1124-1133.

Wu, Y. T., Kanade, T., Li, C. C., & Cohn, J. F. (2000). Image registration using wavelet-based motion model. *International Journal of Computer Vision, 38*, 129-152.

Xiao, J., Baker, S., Matthews, I., & Kanade, T. (2004). Real-time combined 2D+3D active appearance models. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*

Xiao, J., Kanade, T., & Cohn, J. F. (2003). Robust full motion recovery of head by dynamic templates and re-registration techniques. *International Journal of Imaging Systems and Technology, 13*, 85-94.

Yacoob, Y., & Davis, L. (1997). Recognizing human facial expression from long image sequence using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18*, 636-642.

Young, G., & Decarie, T. G. (1977). An ethology-based catalogue of facial/vocal behaviors in infancy. *Animal Behavior, 25*, 95-107.

Appendix

<u>How the facial action brow raise is described in each of the 14 measurement techniques.</u>

Birdwhistell

Raised Brows

Blurton Jones

A very conspicuous movement of raising the eyebrows which can be rather difficult to judge on photographs because of the individual variations in the resting position of the brows. One or more of the following criteria could apply:

a) The height of the brow above the eye corner appears to be equal or more than the width of the open eye. (Fig. 3a measure B equal or greater than A).

b) Horizontal lines visible across the forehead above the brows.

c) There is an enlarged area between the brow and the eyelids which is often highlighted (very pale) in photographs.

d) There is a less sharp fall from the brow into the eye socket (orbit) because the brow is raised beyond the edge of the orbit which it normally covers. Therefore there is less shadow between brow and eye than usual.

e) The shape of the eyebrows change, becoming more curved when they are raised (but they are not curved when the brows are slanted or oblique as well as raised).

Brow raising is presumably a result of contraction of the frontal belly of the *occipto-frontalis*, which can occur simultaneously with *corrugator* or *orbicularis occuli* contraction. Thus many oblique brows were also scored as raised.

Brannigan & Humphries

One or both eyebrows are raised and are held, at least briefly, in the raised position. They are not drawn in towards the midline and are not tilted.

Grant

The eyebrows are raised and stop in the raised position for an appreciable time (see plate 10A).

Flash. A quick raising and lower of eyebrows.

These two elements are very similar in use. They seem to have an attractive function, drawing the attention of the other person to the face. They are concerned with regulation and timing of speech.

Nystrom

- horizontal wrinkles
- elevated brows

(*Note*: These are listed by Nystrom as separate scoring items in his technique.)

Young & Decarie

Brow raise stare:

Brow: the eyebrows are raised and held giving them a curved appearance and creating horizontal creases on the brow. There is no inward movement of the eyebrows and no vertical furrow.

Eyes: The eyes may be held wide open but not sparkling, wrinkling at the corners and forming of pouching under the eyes. Blinking may be decelerated, and the head is definitely held in its regular forward position. Visual fixation on a specific target is characteristic of this expression.

Mouth: as in normal face.

Other: as in normal face.

(*Note*: Young & Decarie present this as a total face score. No provision is made for scoring if the brow raise action occurs without the eye action or with some other mouth action.)

Ekman, Friesen, & Tomkins

(*Note*: Two photographs depict this scoring item. The authors' Facial Affect Scoring Technique contains only visual, not verbal, descriptions.)

Izard: MAX (Maximally Discriminative Facial Movement Coding System)

Code 20: The brows are raised in their normal shape. The forehead shows some thickening and the tissue under the eyebrows some thinning out as a result of the eyebrows being raised. The thickening or massing of tissue in the forehead gives way to long transverse furrows with increasing age. The nasal root is narrowed. The skin directly below the eyebrows is stretched upward.

Code 21: One brow is lifted higher than the other.

Code 30: The eyes have a widened and roundish appearance. The furrow above the eyelashes of the upper lid may be visible. The widened, roundish appearance of the eyes is brought about mainly by the eyebrow raise of code 20 that lifts and stretches the tissue between the eyebrow and the eyelid. The upper eyelid is not raised. The artist's drawing for 20 also illustrates 30.

(*Note*: Izard furnishes video examples of this action in addition to the artist's drawing.)

Ekman, Friesen, & Hager: FACS (Facial Action Coding System, 2002 version)

Action Unit Combination 1+2

(Note: This section on brow raise from the FACS manual is preceded in the manual by separate sections on the two components of this action, AU 1 (inner brow raise) and AU 2 (outer brow raise. All sections include still and video examples not included in this Appendix)

A. Appearance Changes due to AU Combination 1+2

The combination of these two Action Units raises the inner (AU 1) and the outer (AU 2) corners of the eyebrows, producing changes in appearance which are the product of their joint action.

1. Pulls the entire eyebrow (medial to lateral parts) upwards.
2. Produces an arched, curved appearance to the shape of the eyebrow.
3. Bunches the skin in the forehead so that horizontal wrinkles appear across the entire forehead. The wrinkles may not appear in infants, children, and a few adults.
4. Stretches the eye cover fold so that it is more apparent.
5. In some people (those with deeply set eyes) the stretching of the eye cover fold reveals their upper eyelid, which usually is concealed by the eye cover fold.

Compare the image 1+2 with image 0. Inspect the video of AUs 1+2.

B. How to do AU Combination 1+2

(Note: FACS teaches learners how to perform each action so that they can utilize their own facial actions to understand the mechanics and appearance of the face.)

This behavior should be easy for you to do. Simply lift your eyebrows up, both ends as high as you can. Note the wrinkling in your forehead. In some people the wrinkling does not occur but the skin is still bunched up. In some people these wrinkles are permanently etched (see 0 and w0) but they deepen noticeably when 1+2 acts. Suppress any tendency you may also have to lift your upper eyelid (AU 5) when performing 1+2. Make sure you are not pulling your brows together (AU 4) when you lift them.

C. Intensity Scoring for AU Combination 1+2

The criteria for AU 1 and those for AU 2 are altered significantly in this combination from the criteria for each alone. Do not use Section C for AUs 1 and 2, you must use the criteria listed below for the total configuration 1+2. The criteria for intensity scoring are described for roughly equal intensities of AUs 1 and 2. Of course, any combination of intensities of AUs 1 and 2 can occur in action unit combination 1+2, and to score these intensities (e.g., 1B+2C), you must consider the relative contribution of the separate AUs in the combination you score against the criteria listed below. When considering whether AU 2 is present when the action of AU 1 is clearly evident, be sure that any lifting of the outer eyebrows is not due merely to the action of AU 1 alone, as can occur with stronger AU 1s.

AUs 1A+2A in AU Combination 1+2

The appearance changes for AUs 1+2 are sufficiently present to indicate AU 1+2, but are insufficient to score 1B+2B (e.g., the entire brow is raised a *trace*).

AUs 1B+2B in AU Combination 1+2

1. Entire brow raised *slightly*.

If you did not see the brows move it must also meet the additional criteria:

2. *Slight* horizontal wrinkles or muscle bunching reaching across forehead. If horizontal wrinkles are evident in the neutral face, change from the neutral appearance must be *slight*. (If you are scoring the face of an infant or child who never shows forehead wrinkles with AUs 1+2 or 1+2+4, then the wrinkling criterion needs to be discounted, and you must rely on the other criteria.)

and

3. *Slightly* more exposure of eye cover-fold than in neutral.

or

4. If there is no wrinkling or bunching in the brow, but the brow raise and exposure of the eye cover fold is *marked*, you can score 1+2.

AU 1C+2C in AU Combination 1+2

Entire brow is raised at least *markedly*, but less than for level 1D+2D. Wrinkling and eye cover fold exposure should both be evident and at least one should be at least *marked*, but the evidence is less than the criteria for 1D+2D.

AU 1D+2D in AU Combination 1+2

Entire brow is raised at least *severely*. Wrinkling and eye cover fold exposure should both be evident and at least one should be at least *severe*, but the evidence is less than the criteria for 1E+2E.

AU 1E+2E in AU Combination 1+2

The entire brow is raised *maximally*.

Frois-Wittmann

Brows raised.

Fulcher

Frontalis which raises the brows wrinkling the forehead transversely.

Ermiane & Gergerian

Frontalis-the eyebrow *levator*. Externalized emotionality.

(Raises the eyebrows).

Letting himself go to an impression.

(*Note*: A few photographic illustrations show this action.)

Landis

Frontalis. This is the vertical sheet muscle of the forehead, the contraction of which produces transverse wrinkles ("the wrinkled brow").

Author Note

Jeffrey F. Cohn, Department of Psychology, 4327 Sennott Square, University of Pittsburgh, Pittsburgh, PA 15260. *http://www.pitt.edu/~jeffcohn, jeffcohn@pitt.edu*

Paul Ekman, University of California, San Francisco, 401 Parnassus Avenue, San Francisco, CA 94143, *http://www.paulekman.com/*

Preparation of this chapter was supported in part by NIMH grants MH 11976 and MH 06092 to Paul Ekman and NIMH grant MH 51435 to Jeffrey Cohn.

Tables

Tables begin on following page.

Table 1 Summary of human-observer based methods for measuring facial behavior for units and comprehensiveness

	Basis for deriving units	*Comprehensiveness*		
		Type of action	*Intensity of action*	*Timing of action*
Linguistically based				
Birdwhistell (1952)	Observation of inter-personal behavior; parallel linguistic units	Not claimed to be comprehensive; 53 actions	No provision	No provision
Ethologically based				
Blurton Jones (1971)	Observation of 500 still photographs of 2-5-year-old children	Measures any child's facial expressions; 52 actions	6 degrees of eye openness; 4 degrees of lip separation; 2 degrees of frowns	No provision

	Basis for deriving units	*Comprehensiveness*		
		Type of action	*Intensity of action*	*Timing of action*
Brannigan & Humphries (1972)	Observation of children and adults	Not claimed to be comprehensive; 70 actions	No provision	No provision
Grant (1969)	Observation of children and adults	Not claimed to be comprehensive; 53 actions	No provision	No provision
McGrew (1972)	Observation of 3-4-year old children	Not claimed to be comprehensive; 31 actions	No provision	No provision
Nystrom (1974)	Observation of 1-month-old infants	Not claimed to be comprehensive; 35 descriptors	No provision	No provision

	Basis for deriving units	*Comprehensiveness*		
		Type of action	*Intensity of action*	*Timing of action*
Young & Decarie (1977)	Observation of 36 infants	Measures 42 facial configurations; selected only to be relevant to emotion in the last quarter of first year in six test situations	No provision	No provision
Theoretically based				
Ekman, Friesen, & Tomkins (1971)	Theory about emotion expression	Measures signs of just 6 emotions; 77 descriptors	No provision	Start-stop

	Basis for deriving units	*Comprehensiveness*		
		Type of action	*Intensity of action*	*Timing of action*
Izard (1983)	Theory about emotion signals; data from posed still photographs	Measures just actions needed to identify emotion in infants; 29 descriptors	No provision	Start-stop
Anatomically based				
Ekman & Friesen (1978) Ekman, Friesen, & Hager, (2002)	Muscular	Measures all visible movements; 44 action units that singly or in combination can score any observed action	Three-point intensity scale for 4 action units in 1978 version increased to five-point intensity scale for all action units in 2002 version	Start-stop and onset-apex-offset

	Basis for deriving units	*Comprehensiveness*		
		Type of action	*Intensity of action*	*Timing of action*
Frois-Wittmann (1930)	Muscular	Not claimed to be comprehensive; 28 descriptors	No provision	No provision
Fulcher (1942)	Muscular	Not claimed to be comprehensive; absence/presence of 16 muscular actions	Amount of movement in each of three facial areas related	No provision
Ermiane & Gergerian (1978)	Muscular	Measures all visible movements; 27 muscle actions	Each action rated only on 3-point intensity scale	No provision
Landis (1924)	Muscular	Not claimed to be comprehensive; 22 descriptors	Each action rated on 4-point intensity scale	No provision

Table 2. Summary of human-observer based methods for measuring facial behavior: unit depiction, inference/description, and application

	Way in which each unit is depicted	*Use of inference or description*	*Types of records and persons to which measurement has been applied*
Linguistically based			
Birdwhistell (1952)	Two or three words	Mixed: e.g., *pout, smile, sneer*	Not known
Ethologically based			
Blurton Jones (1971)	Verbal description of changed appearance of features, a few drawings and illustrative photos	Mostly description but a few inferential terms: e.g., *frown, pout*	Infants and children

	Way in which each unit is depicted	*Use of inference or description*	*Types of records and persons to which measurement has been applied*
Brannigan & Humphries (1972)	Verbal description	Mixed: e.g., *wry smile, angry frown, sad frown, threat*	Children and adults
Grant (1969)	Primarily verbal description, some photos	Mixed: *sad frown, aggressive frown, smile, sneer*, etc.	Children and adults
McGrew (1972)	Verbal description; compared to Grant, Blurton Jones	Mostly description but a few inferential terms: e.g., *pout, frown, grin*	Children
Nystrom (1974)	Verbal description	Description	Neonates
Young & Decarie (1977)	Verbal description	Mixed: *fear face, sad face, shy smile*, etc.	Infants in last quarter of first year

	Way in which each unit is depicted	*Use of inference or description*	*Types of records and persons to which measurement has been applied*
Theoretically based			
Ekman, Friesen, & Tomkins (1971)	Photographs of descriptor	Description	Video and still photos of adults' posed and spontaneous expressions
Izard (1983)	Verbal description, photos, drawings, and video	Description	Video of infants

	Way in which each unit is depicted	*Use of inference or description*	*Types of records and persons to which measurement has been applied*
Anatomically based			
Ekman & Friesen (1978; Ekman, Friesen, & Hager, 2002)	Verbal description, still photos, and video examples of each action and certain combinations of actions	Description	Spontaneous, deliberate, and posed video and photos of neonates, children, adults, deaf stutterers, mental patients
Frois-Wittmann (1930)	Verbal description; very brief	Only one inferential term: *frown*	Still photos of poses by one adult
Fulcher (1942)	Verbal description; very brief	Description	Films of poses by blind and sighted children
Ermiane & Gergerian (1978)	Verbal description, still photos	Description	Adult poses and patients' spontaneous photographs

	Way in which each unit is depicted	*Use of inference or description*	*Types of records and persons to which measurement has been applied*
Landis (1924)	Verbal description	Description	Neonates

Table 3. Summary of human-observer based methods for measuring facial behavior: reliability and validity

	Reliability	*Validity*			
		Descriptive	*Emotional*	*Conversational*	*Other*
Linguistically based					
Birdwhistell (1952)	Not reported	None	None	None	None
Ethologically based					
Blurton Jones (1971)	Data reported on requirements 1, 2, 3*b*, 6*a*	None	None	None	None
Brannigan & Humphries (1972)	Not reported	None	None	None	None

Grant (1969)	Not reported	None	None	None	Predicts severity of mental illness, but no data reported
McGrew (1972)	Data reported on requirements 1, 2, 3*b*, 6*a*	None	Spontaneous	None	Predicts gender differences & relation to agonistic interaction
Nystrom (1974)	Data reported on requirements 1, 2, 3*b*, 6*a*	None	None	None	None
Young & Decarie (1977)	Not determined by authors	None	Spontaneous, but no data reported	None	Said to differentiate infants' response when mother departs and when she frustrates, but no data reported

Theoretically based					
Ekman, Friesen, & Tomkins (1971)	Data reported on requirements 2 and 3*c*	None	Posed and spontaneous: positive vs. negative, stressful vs. neutral film conditions; differentiates patterns of heart rate	None	Predicts attribution of emotion
Izard MAX (1983)	Data reported on requirements 2, 3*a*-*b*, 5 and 6*a*	None	Posed	None	Provides preliminary data on relations to vocalization and body movement in infants

Anatomically based					
Ekman & Friesen (1978), Ekman, Friesen, & Hager (2002)	Data reported on requirements 1, 2, 3*a*-*c*, 4, 5, 6*a* & *c*	Meets performed actions and EMG criteria	Posed and spontaneous; measures intensity and type of emotion; differentiates startle reaction; differentiates certain deliberate from spontaneous expression	Measures syntactic and emphasis signals	None
Frois-Wittmann (1930)	Not reported	None	Posed	None	Predicts developmental changes; compares blind and sighted
Fulcher (1942)	Data reported on requirements 1, 2, 3*b*, 6*a*	None	Posed	None	None

Ermiane & Gergerian (1978)	Data reported only on scoring photos of poses and on requirement 3*c*	None	Posed	None	None
Landis (1924)	Not reported	None	None	None	Predicts individual differences

Table 4. Automatic recognition of facial action units

Research Group	Key publications
Carnegie Mellon University / University of Pittsburgh	(Cohn et al., 1999)
	(Lien et al., 2000)
	(Tian et al., 2001)
	(Tian et al., 2002)
	(Cohn, Reed, Ambadar et al., 2004)
Delft University of Technology	(Pantic & Rothkrantz, 2000b)
	(Pantic & Rothkrantz, 2003)
	(Pantic & Rothkrantz, 2004)
	(Valstar et al., 2004)
Institut Dalle Molle d'Intelligence Artificielle	(Fasel & Luettin, 2000)
University of California San Diego	(Bartlett et al., 1999)
	(Donato, Bartlett, Hager, Ekman, & Sejnowski, 1999)
	(Littlewort, Bartlett, & Movellan, 2001)
	(Bartlett et al., 2004)

Figure Captions

Figure 1. Muscles of the face (Clemente, 1997).

Figure 2. The mean facial EMG response for the *zygomatic major* and *corrugator supercilii* muscles plotted in intervals of 100ms during the first second of exposure when subjects were instructed to react as quickly as possible to a happy face or an angry face. (Dimberg et al., 2002).

Figure 3. Plot of relationship between FACS and EMG measurement of performances of Action Unit 1 (*frontalis, pars medialis*).

Figure 4. Example of difference images. Row *a* shows infant's facial expression changing from neutral to a Duchenne smile (AU 6+12). Row *b* shows the difference between the first and each subsequent image in row a. Areas of white indicate motion caused by change in facial expression and/or head motion.

Figure 5. Example of dense flow extraction using the method of Wu et al.(Wu, Kanade, Li, & Cohn, 2000) from Lien (Lien et al., 2000).

Figure 6. Example of feature-point tracking (Cohn et al., 1999).

Figure 7. 3D head tracking and image alignment (Cohn et al., 2003).

Figure 8. System diagram for CMU/Pitt Automated Facial Image Analysis (AFA) version 3 (Cohn, Reed, Ambadar et al., 2004).

Figure 9. Comparison of manual and automatic ordering of blink sequence (Cohn et al., 2003).

Figure 10. Example of the relation between *zygomatic major* EMG and displacement of the lip corner as determined by AFA (Cohn & Schmidt, 2004).

Figure 1. Muscles of the face (Clemente, 1997).

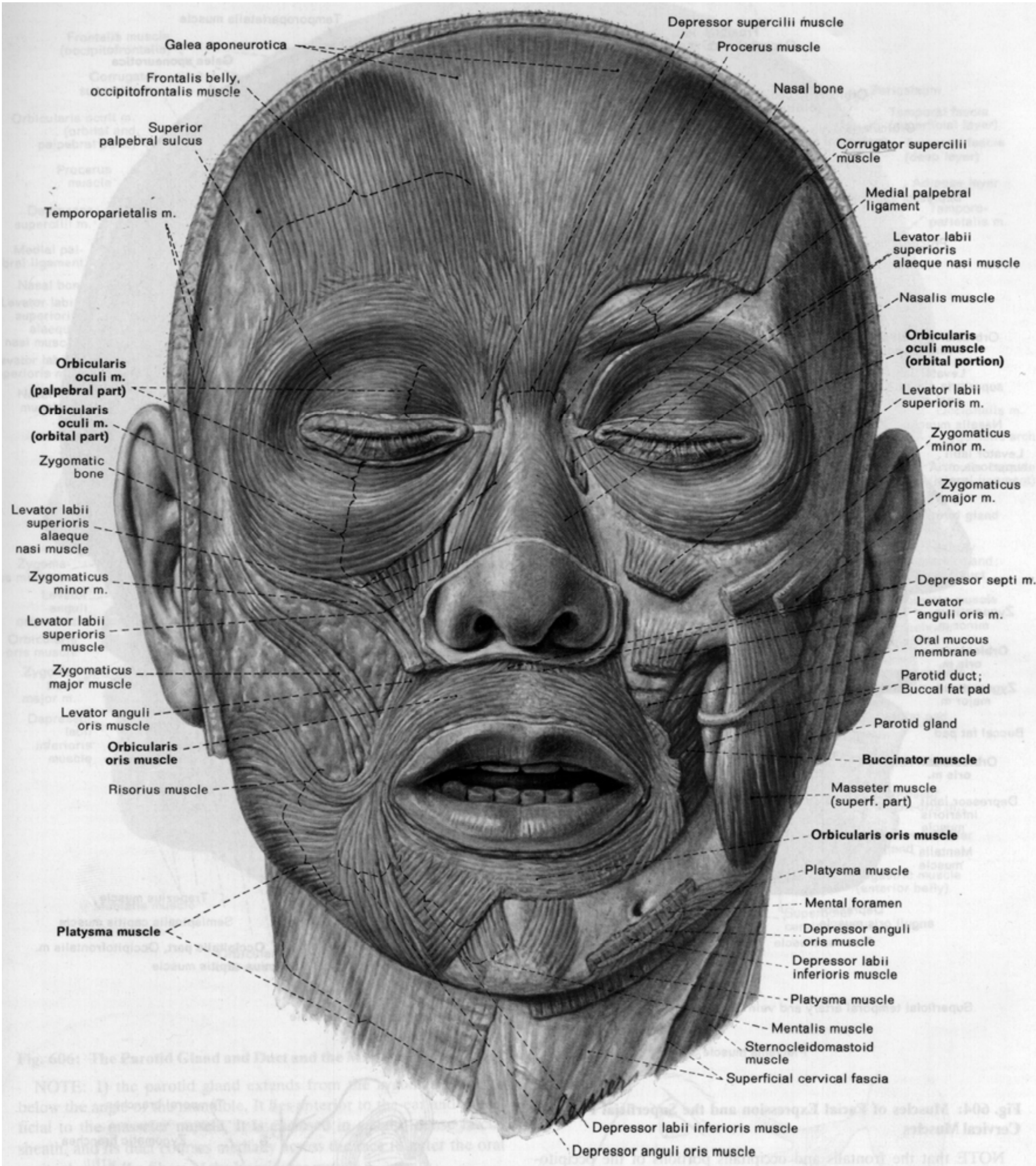


Figure 2. Mean facial EMG response for zygotmatic major and corrugator supercilii muscles plotted in intervals of 100ms (Dimberg et al., 2002).

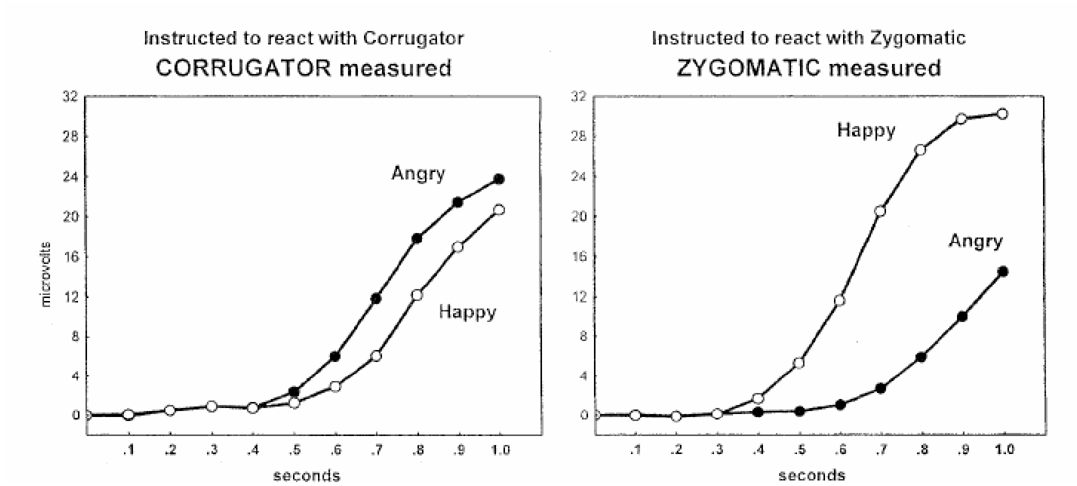


Figure 3. Relation between FACS and EMG for action unit 1.

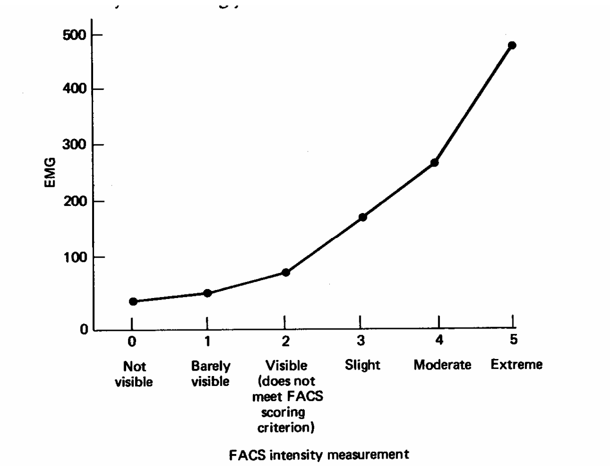


Figure 4. Example of difference images (Lien et al., 2000).

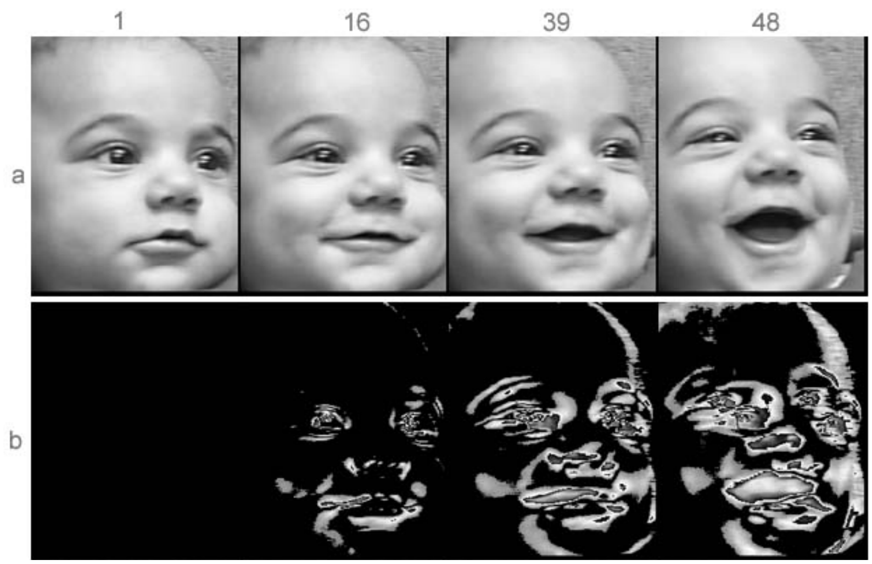


Figure 5. Example of dense flow (Wu et al., 2000).

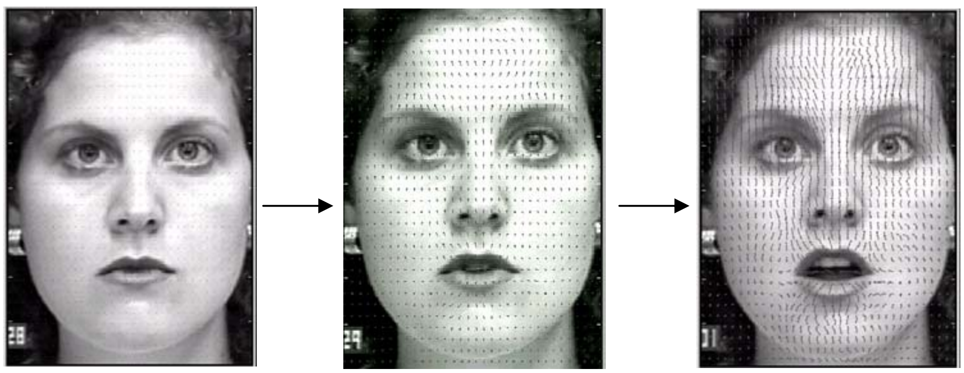


Figure 6. Example of feature-point tracking (Cohn et al., 1999).

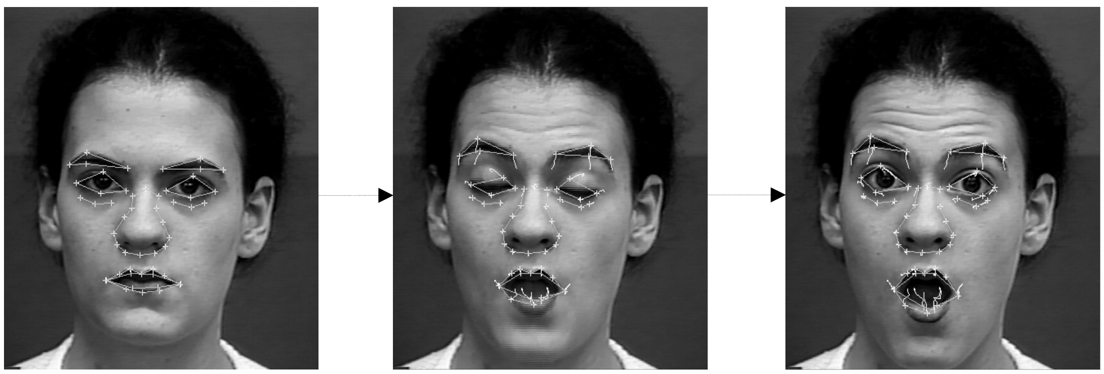


Figure 7. 3D head tracking, alignment, and feature localization (Cohn et al., 2003).

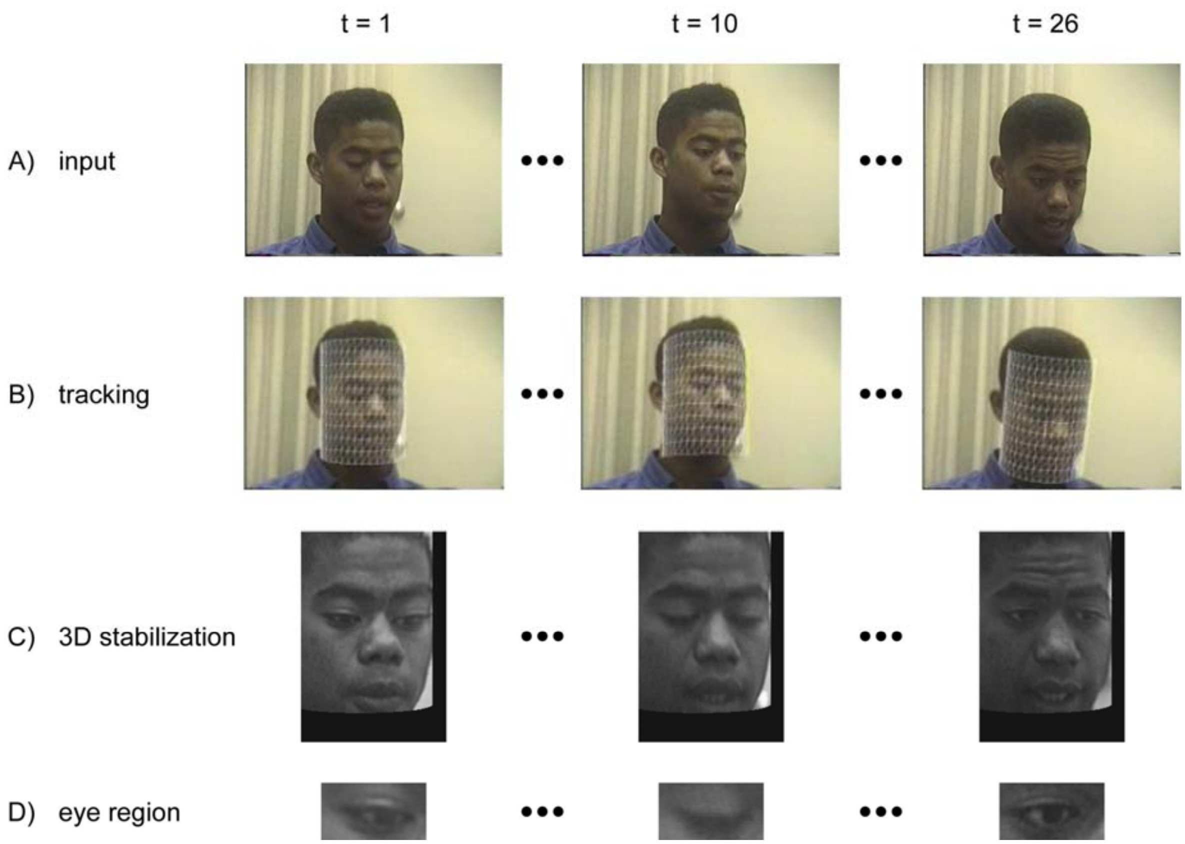


Figure 8. System diagram for CMU/Pitt Automated Facial Image Analysis (AFA) version 3 (Cohn, Reed, Ambadar et al., 2004).

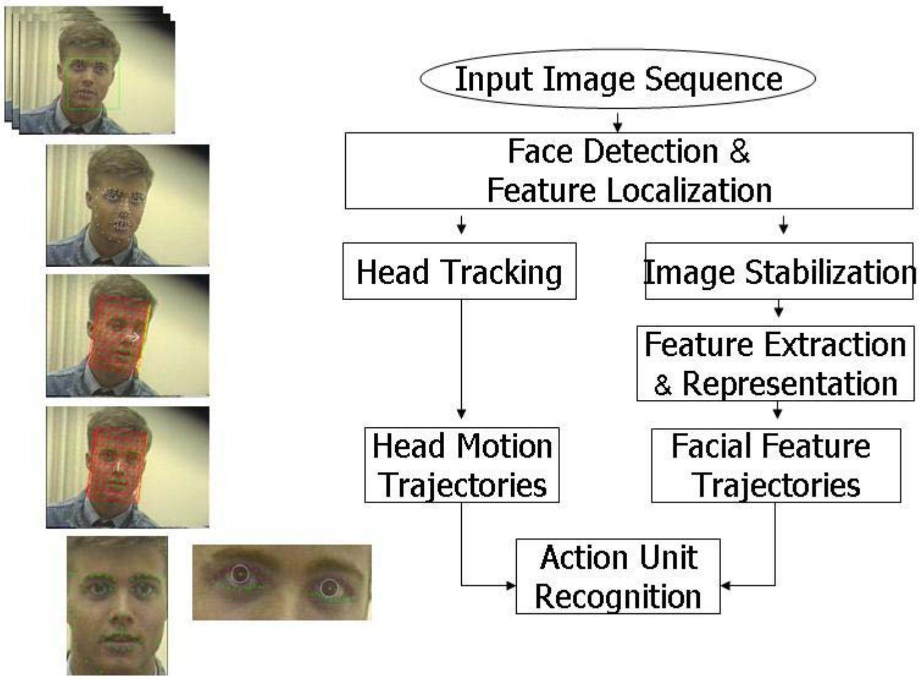


Figure 9. Comparison of manual and automatic ordering of blink sequence.

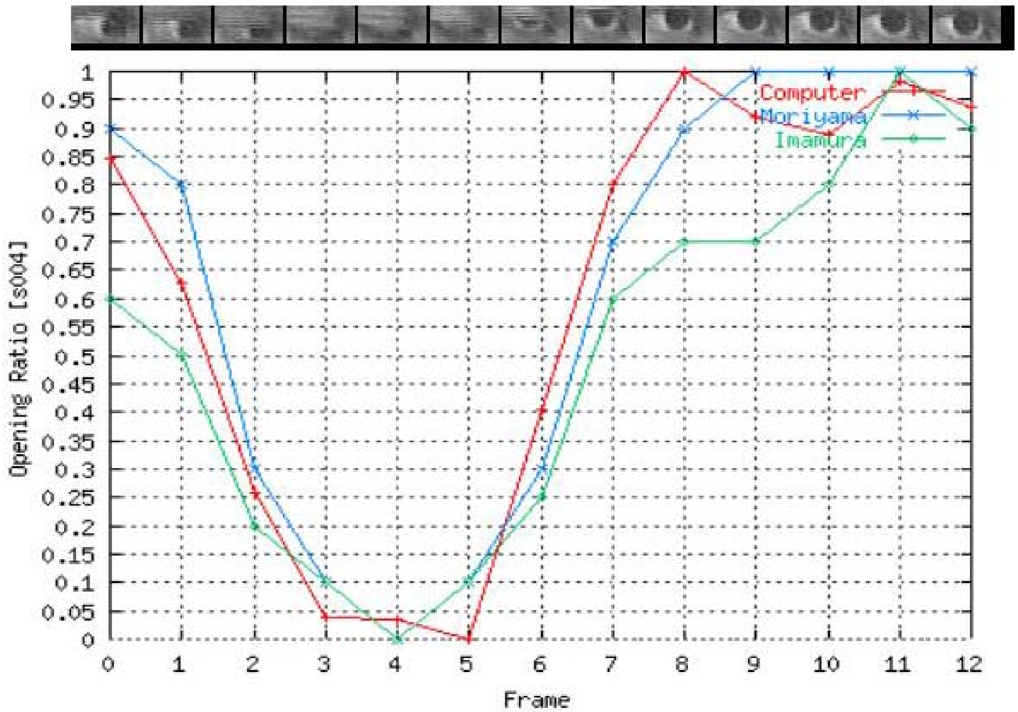
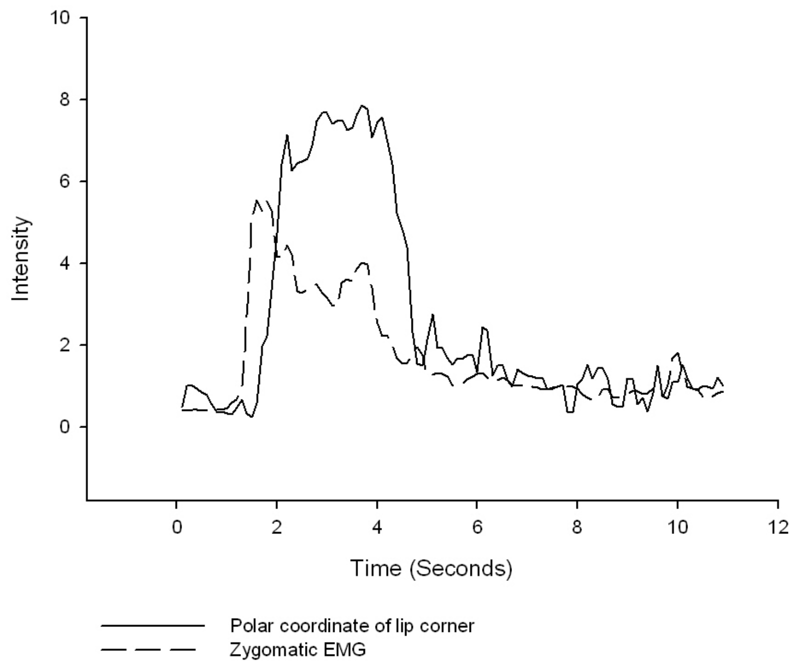


Figure 10. Example of the relation between zygomatic major EMG and displacement of lip-corner (Cohn & Schmidt, 2004).



Notes

[1] Over the years Ekman has proposed a number of different phrases to distinguish these two approaches. In previous discussions the message judgment approach has been labeled the stimulus, communicative, or judgment approach, and the measurement of sign vehicles approach has been labeled the response, indicative, or components approach. It is to be hoped that the present terms, taken from semiotics, allow a more lucid differentiation of these two methods.

[2] The two approaches are complementary. One could use the sign vehicle approach to determine what facial expressions differ among diagnostic groups and the message judgment studies to determine which of those expressions influence message judgments about diagnosis. (Juslin and Scherer, this volume, discuss use of a modified Brunswikian lense model, in this context. See also (Hess, Kappas, McHugo, Kleck, & Lanzetta, 1989)).

[3] Izard (C. E. Izard, 1979b) said that as part of an attempt to establish independent discovery, he deliberately did not examine Ekman & Friesen's Facial Action Coding System, even though it had already been published at the time when he was developing his measurement techniques.

[4] Strangely, Izard excluded specific actions that are said by many theorists to signal emotions and that are shown by Ekman and Friesen’s data to be emotion signals. Izard and Dougherty (1981) say that actions were dropped that were not efficient, but inspection of that article and of earlier versions of Izard’s scoring technique (FMCS) (C.E. Izard, 1979a) suggests instead that Izard never considered a number of facial actions important to differentiating among emotions, especially in infants (Oster et al., 1992).

---

[5] In part because of its very uniformity, Ekman and Friesen consider the startle reaction to be not an emotion but instead a reflex. Some writers about emotion (Tomkins, 1962) disagree and classify startle with the emotion of surprise. For further discussion and data on this issue see Ekman, Friesen, and Simons (Ekman et al., 1985)

[6] They acknowledge that for certain actions — for example, the movements of the tongue — their technique is not complete.

[7] These errors are the product of limited sampling: Izard chose his actions on the basis of what he observed in a set of photographs of posed emotions.

[8] In contrast, until recently almost all research in the related field of automatic face recognition has used static images. This is in part because applications in this area are driven by large databases, numbering millions of images, that already exist and the belief that face and head motion contribute little to person recognition. Evidence for the importance of face motion and video input and a broadening application base contribute to increasing interest in video for automatic face recognition (e.g., First IEEE Workshop on Face Processing in Video, 2004, Washington, DC).