A Place for Prosody in a Unified Model of Cognition and Emotion

Branka Zei Pollermann
Liaison Psychiatry
University Hospitals of Geneva, Switzerland
branka.zei@hcuge.ch

Abstract

The relationship between cognition and emotion has been one of major stumbling blocks in defining the concept of emotion. We argue that a unified model of cognition and emotion based on interactionist epistemology would allow to unambiguously define emotion and thus facilitate research in affective aspects of speech communication. The main tenet of the model is that emotion needs clarification. We believe that this can be accomplished within a unified theory which integrates emotion, cognition and behaviour into one model of the organism's adaptive functioning.

1. Introduction

Over the past hundred years, the psychology of emotions has mainly concentrated on various components of the organism's emotional response such as: cognitive, physiological, neurological, motor, expressive, behavioral, linguistic, and the subjective feeling.

We are currently facing a remarkable diversity of emotion theories, each focusing on a different component of the organism's response system. Most theories largely overlap in terms of hypothesized or empirically found components of emotions. Yet, there is no consensus on the criteria for distinguishing between emotional and non-emotional phenomena inherent to human behaviour.

This situation is well illustrated in The Nature of Emotion [11] where twenty-three eminent scholars answered twelve fundamental questions regarding the nature of emotions. In the concluding chapter, the editors deny the existence of a gold standard for knowing that an emotion has occurred. Nonetheless the contributors seem to agree that emotion can be characterised as a phenomenon which involves organism totally, including the shaping of personality. Now, the question remains: how do we distinguish between emotional and non-emotional responses? Scherer [32] believes that the difference rests upon one distinctive feature: temporary synchronisation of the functioning of all or most of the organismic subsystems which occurs when the response is emotional. However, we believe that most adaptive reactions to relevant stimuli require a certain degree of synchronisation of the work of organismic subsystems without necessarily being recognised as emotional.

One of the stumbling blocks in defining emotion has been the relationship between emotion and cognition. We believe that a change of paradigm is required in that the question must be addressed from an epistemological point of view. Unlike some scholars [15] we believe that the relationship between cognition and emotion is a fundamental epistemological issue and not just a semantic controversy.

The model we are proposing is based on two complementary epistemological approaches: Piaget's genetic epistemology [22] and Prieto's semiological theory of knowledge [25]. We shall thus first present an outline of a unified model of cognition and emotion, which will allow us to distinguish between emotional and non-emotional reactions, next we shall address the question of the semiological status of prosodic indicators of affective-cognitive states in terms of: expression vs. communication of affect, and we shall end with two examples from our own research.

2. Cognition is in emotion and emotion is in cognition

It is now generally accepted that for an emotion to occur, the subject must cognitively evaluate the significance of the occurring external or internal stimuli. Most often, this is done unconsciously. It is also accepted that cognition plays a part not only in the appraisal of the stimulus [13] [27] [30] but also in the awareness of the overall subjective feeling [29]. On the other hand, there has been evidence [7] that emotional processes are required for certain types of decision-making to occur. In addition, studies in cognitive neuroscience of emotion [12] point out that “there is no evidence that neural processes subserving cognition differ in fundamental ways from those subserving emotion” [12 p.5].

It thus appears that the relationship of cognition to emotion needs clarification. We believe that this can be accomplished within a unified theory which integrates emotion, cognition and behaviour into one model of the organism's adaptive functioning.

2.1. General Theoretical Background of the Model

The proposed model applies to organisms described as open systems with dynamic teleology [2] which have the following characteristics:

1. A system is defined as a composition of elements in interaction which is not random.
2. Behaviour is determined by the foresight of the goal.
3. The system is characterised by equifinality i.e. the same state can be reached from different initial conditions and in different ways.
4. The system has self regulatory features which aim at maintaining a steady state.
5. Temporary changes in the environment provoke reversible fluctuations of the organism's steady state.
6. The system is an intrinsically and autonomously active system with its parts and processes being in mutual interaction.

The main tenet of a unified model is that cognition is a process whose mechanism necessarily includes the three basic dimensions traditionally attributed to emotions: valence, arousal and power. In other words, emotional dimensions are inherent to cognition. The model is best understood when...
placed in the context of interaction between the subject and the environment.

Given that all interaction requires the use of knowledge of the outer world and that of self, the model posits that both kinds of knowledge form a semiotic structure which subserves the subject's adaptive behaviour: the knowledge of the outer world is constructed with regard to its significance for the subject's interests and behaviours [25]. Ultimately, all adaptive behaviour aims at: maintaining or achieving a positive state (positively valenced state) and/or avoiding a negative state (negatively valenced state). It is assumed that the target states can concern any or all of the three dimensions of Self: intra-subject (personal interests, including own body), inter-subject (relational and social interests), and trans-subject (interests related to the nature as a whole, or religious pursuits).

The organism's cognitive and sensori-motor functioning, can constitute positive target states on its own. These are often related to creativity in scientific or aesthetic pursuits and to self-realisation in sports performances. According to the unified model, the inner mechanism of the cognitive process involved in adaptive interaction - in which the subject identifies the stimulus, and constructs relevant knowledge about it - can be described as follows:

In the initial phase of stimulus perception and identification of its characteristics, the subject constructs or evokes the already existing knowledge about the stimulus and its identity. This knowledge is attributed a value in terms of its significance for the subject [24], including the intra, the inter and the trans subject dimensions. This value can be conveniently described as a value-tag*. We hypothesise that the process of stimulus identification and its appraisal includes attribution of three kinds of value-tags as three dimensions of cognitive-affective space. Each value tag can be conceived of as a scalar ranging from positive to negative values.

- A first value-tag, termed valence tag, is attributed with regard to the appraised beneficial or detrimental character of the stimulus. This kind of value is considered as the exchange value because it is related to what it can be exchanged for, as far as the subject's wellbeing is concerned. Caccioppo and colleagues [4] showed that the function of valence tags is to regulate approach-avoidance behaviour. Valence tags thus represent the subjective meaning that the stimulus have for each of the constituents of Self. Valence tagging is considered to be inherent to the subject's interactive behaviour.

  - b) A second value-tag, termed power tag, is related to the subject's estimate of his/her coping and/or decisional power. Prieto attaches a great deal of importance to this point, which he links to the constitution of Self. A power tag thus denotes the degree of the subject's power to decide and to act upon the stimulus and/or its consequences. It refers to the estimated relative power i.e. the relation between available power and the power needed to cope with the situation.

  - c) A third type of value-tag, termed activation tag, denotes the amount of energy expenditure involved in motoric and autonomic-physiologic changes automatically triggered or estimated as required to handle the stimulus and/or its consequences. It is related to the regulation of bodily, and computational energy expenditure. Piaget calls this kind of value "value de rendement" - the yield value which is also influenced by a kind of cost-benefit calculus [23 p.595]. Piaget considers bodily energy management to be an affective dimension of behaviour.

* The term “tagging” has been borrowed from Ohman [19] who uses it to denote the mainly unconscious process of assigning emotional meaning.

In summary, each moment's implicit and explicit knowledge of the environment, of self, and of one's own action, can be placed into a three dimensional affective space of valence, power, and arousal. There is reason to believe that valence value is primary [8] in that it recruits and drives responses including the appraisal of power, and arousal. In summary, our unified model suggests that cognitive processes involved in interaction with environment necessarily include dimensions which are typically known as emotional, that is: valence, activation, and coping/decisional power. The inseparability of cognitive and affective aspects of adaptive action is clearly expressed by Piaget: "There is no reason to separate intelligence from affectivity, nor is it necessary to ask which of the two precedes and conditions the other. It is a matter of two inseparable aspects of mental development." [24 p. 41].

Inspired by insightful works of Elisabeth Duffy in the physiology of behaviour [10], and more recent work in neurological aspects of emotions [5] we suggest the following criteria for distinguishing between emotional and non-emotional organismic responses:

If the organism's monitoring system reports critical values (that is too low or too high) for valence, arousal or power, the state becomes cognitively dominant. Such a moment's dominant body-mind landscape is likely to penetrate consciousness and be conceptualised as emotional. This is in agreement with Ledoux's proposal regarding the neurological triggers of consciousness which "turn subjective experiences into emotional experiences" [14 p. 146]. Once triggered, an emotional response can then involve a resetting of procedural priorities and the regulation of the speed of task execution (which can be set to 0 as in case of a freeze-reaction) thus producing a cascade of organismic changes. It goes without saying that the sensitivity thresholds, as well as ceiling values will vary from one person to the other depending on genetic, physiological, contextual and social factors.

The boundary between an emotional and non-emotional response is thus set by each individual's critical-range values in a three-dimensional affective space. As long as the person’s monitoring system reports values of valence, arousal and power within his/her range of normal variation, the state will not be subjectively experienced as emotional. Behaviourally speaking adaptive action/reaction patterns have a range of foreseeable fluctuations which represent the shock absorption capacity of a particular system without it being endangered in its structural integrity. In other words the state of the organism - related to each moment's psycho-physiological configuration - is always affectively coloured.

We suggest that the concept of emotion can thus be defined as a cognitively dominant body-mind configuration presenting one or more out-of-normal-range values on at least one of the three dimensions: valence, arousal and power. There is reason to assume that the shifts in the overall configuration are non-linear.

We believe that a unified model of cognition and emotion can constitute a suitable conceptual framework for studying acoustic correlates of affective states. It can explain why
emotions, attitudes, moods, personal stances and even personality traits can be considered as different types of affective states or processes. It is because the three elementary affective dimensions are inherent to all of them.

The advantage of using such a three-dimensional approach over a categorical labelling approach lies in the fact that one can establish a continuity from non emotional to emotional states and processes. This is in particular appropriate for studying natural corpora.

2.2. A place for affective prosody in a unified model of cognition and emotion

Speech communication, just as any other interactive behaviour, necessarily reflects the three dimensions of the affective space. These dimensions are then encoded in many aspects of speech and in prosody in particular. Prosody partakes at the same time of the physical, the physiological, the mental, the individual and the social. It reflects the configuration of the speaker's affective space in two ways: the underlying configuration extending over a longer period of time (moods, stable attitudes, and personality traits) and each moment's on-line configuration (emotions, interpersonal stances and communication strategies). In natural settings the on-line configuration is influenced by ever changing discursive context including the speaker's and the receiver's own non-verbal reactions, the discourse content and the conversational interaction pattern itself.

We are now faced with a number of fundamental questions:

What exactly does affective prosody indicate? What meaning does the receiver attribute to prosodic variations? What is the semiological status of vocal affect signalling? Is affective prosody a symptom, a symbol or a conventional sign? This poses a further -seemingly terminological- problem of expression vs. communication of affect whereby expression = external manifestation of an inner state (Darwin's meaning), while communication implies intentional signalling [25].

To answer these questions let us consider the following semiological distinctions regarding basic types of semiotic entities involved in the mechanism of indication.

1. **Spontaneous indicators** [25] are double-faced entities where the link between the signifier and the signified is naturally given. It is said to be motivated [21 pp. 134-135] because the link reflects relations such as: spatial or temporal contiguity, causality, implication or pars pro toto relationship. Example: footprints in snow, temperature as a symptom of illness. In other words, spontaneous indicators have an informative value without having been produced to this purpose.

2. **Falsely spontaneous** indicators [3] are those that are purposely produced in order to appear as natural or spontaneous. Example: a foreign accent produced by a native speaker wanting to appear as a foreigner.

3. **Intentional indicators** are double-faced entities which are produced in order to provide information other than their own existence. These include:
   - **symbols**, where an originally natural link between the signifier and the signified has been conventionalised for purposes of communication. Example: a picture of a snake symbolising a pharmacy. In the course of the history, a symbol may lose its previously motivated nature.
   - **icons**, where the link between the signifier and the signified is motivated by topological similarity between the two [35]. Examples of icons: maps, images, diagrams, metaphors
   - **signs**, where the relationship between the signifier and the signified has been arbitrarily established and conventionalised for purposes of communication [28]. Example: words of a language.

We believe that the semiological status of affective prosody can be that of: a spontaneous indicator (a symptom), a falsely spontaneous indicator, or a symbol but never a sign as defined above. The non arbitrary nature of prosodic indicators is supported by cross-cultural studies showing that the subjects accurately infer emotional states from acoustic cues produced in other cultures and languages [6]. Parallelisms in animal affect signalling is of interest here as well.

So, what exactly does vocal expression indicate? Scherer's model of vocal-affect-signalling provides precise predictions for phonetic and macro-prosodic changes related to the outcome of each of the components of his model of appraisal [31]. It therefore follows that, emotionally induced phonetic and prosodic variations primarily have a status of spontaneous, natural indicators or symptoms of physiological reactions automatically triggered by cognitive appraisal outcomes. They reflect the "push" force in affect signalling [34]. An example would be low pitch variability, and low overall intensity usually found in sadness. We believe that in the case of intense primary emotions (anger, fear, joy, sadness and disgust) the push effects are dominant and the speaker may have little freedom in influencing his vocal expression. A special case may be made for highly distressful situations where vocal characteristics are intentionally modified for purposes of better transmission over a distance and as cues for the sender's localisation.

By contrast the "pull" force in vocal affect signalling is related to norms or expectations imposed by the physical or social environment which require production of specific acoustic features allowing the sender to achieve a particular effect. The pull force may restrict or enhance some of the surface features of the push effects. It therefore follows that, when push and pull factors are blended, vocal signals have a two-fold status of symptoms and symbols. When push effects dominate, vocal signals may have a status of falsely spontaneous indicators or symbols. An example would be the use more highly-pitched voices in signalling submission and appeasement in friendly or submissive encounters.

The semiological status of affect signalling can thus vary from symptom to a conventionalised or ritualised symbol. This also raises the issue of speaking styles. Is any individual speaking pattern a style? For Prieto a style is related to the possibility of choosing a means (in this case a vocal pattern) for achieving an aim [26 p.99] . It therefore follows that if there is only one possible means of achieving an aim, the subject cannot have a style. Affectively marked vocal expression will be a style as long as it is a result of the speaker's choice. In this context Prieto distinguishes between reactions and behaviours. The latter (speech behaviour included) involve engaging in action by one's own decision in a situation of choice. For Prieto, it is the power to choose how to act, (i.e. the possibility to have a style) that allows the subject (the speaker) to build his own Self. Prosodic choice thus appears to be a psychologically important factor.
Another refinement may be necessary: the distinction between the unconscious purely physiologically driven push factors and the more psychologically driven pull factors. In the latter case the choice of expressive style may be regarded as expression of the speaker’s identity. An example would be persistent usage of speech patterns related to one’s own native regional dialect.

The relationship between push and pull factors varies according to context. It is highly probable that blends of push and pull forces are more frequent than pure dominance of one or the other. In the Table 1, we present the hypothesised relationships between the two forces for five categories cognitive-affective configurations with corresponding values of affective space tags and the semiological status of vocal signals. The five cognitive-affective configurations correspond to the five different types of affective states and traits presented by Scherer as examples of design feature delimitation of different affective states. [33].

<table>
<thead>
<tr>
<th>Affective-cognitive configuration</th>
<th>Affective space tags' values</th>
<th>Push-Pull relationship in speech behaviour</th>
<th>Semiological status of vocal signalling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotions</td>
<td>in critical range</td>
<td>push dominant</td>
<td>mainly symptom (possibly symbol)</td>
</tr>
<tr>
<td>Moods</td>
<td>within normal to border line range</td>
<td>push dominant or blended</td>
<td>mainly symptom</td>
</tr>
<tr>
<td>Attitudes</td>
<td>within normal range</td>
<td>push dominant or blended</td>
<td>mainly symptom (possibly symbol)</td>
</tr>
<tr>
<td>Interpersonal stances</td>
<td>within normal range</td>
<td>pull dominant</td>
<td>mainly symbol</td>
</tr>
<tr>
<td>Personality traits</td>
<td>within normal range</td>
<td>push dominant</td>
<td>symptom</td>
</tr>
</tbody>
</table>

In comparison with other interactive behaviours, speech communication is specific in that it implies the existence of a receiver able to decode the signal and attribute a meaning to the speaker's utterance by taking into account the context in its physical, social and discursive dimensions. Because of the intrinsically natural relationship between prosody, and the speaker’s cognitive-affective and physical state, we believe that affective prosody can be regarded as a parallel natural code which provides contextual information thus helping the receiver to disambiguate the meaning of the utterance. The interpretation of the speech signal is mainly probabilistic, and driven by physical and social context including stereotypical representations of normal vs. affectively marked speech patterns. The greater the mismatch between the expected and the actual, the more likely it is that it triggers the search for interpretation or connotative meaning attribution.

2.3. Prosody as a symptom of arousal dimension: example of flattening of vocal arousal due to autonomic neural lesions in diabetic patients

2.3.1. Hypotheses and method

Given that basic autonomic responsivity influences emotional reactions, we hypothesised that the subjects with autonomic lesions would display diminished emotional reactions resulting in a diminished vocal arousal and consequently a flattening of vocal differentiation between high and low arousal emotional states. The subjects were 40 diabetic patients. The autonomic responsivity was assessed by means of two standard autonomic function tests based on the quantification of heart rate variability (HRV) [36]. These provided a HRV index indicating each patient's level of autonomic responsivity. It was also hypothesised that low autonomic responsivity would be related to a diminished subjective experience of emotions. The latter was assessed through verbal self report.

Emotional states were induced through verbal recall of personal emotional experiences of anger, joy, and sadness. At the end of each recall the patients were asked to pronounce, on a mood congruent, the sentence: “ALORS TU ACCEPTES CETTE AFFAIRE” (“So you accept the deal”). The sentence was presented in capital writing without punctuation so as not to suggest any tone of voice. The patients were then asked to what degree they had subjectively felt the emotion described during their recall. The results were coded on a 4-point scale, ranging from not at all to very much. The patients’ voices were acoustically analysed for standard vocal parameters related to F0, energy and delivery rate. Upon inspection of partial correlations with HRV index (controlling for age, anxiety state and extraversion), three vocal parameters appeared as significantly correlated with HRV index. These were: F0 max/min ratio, voiced energy range and the rate of delivery. To obtain an index of the overall degree of vocal arousal related to each condition (anger, joy, sadness), we calculated a summary score composed of z values of the three above mentioned vocal parameters. The summary score was named Vocal arousal index. As we expected the subjects with high HRV index to exhibit higher vocal arousal in anger than in sadness, we then calculated the delta between the Vocal arousal index in anger and that in sadness. Each subject was thus characterised by his or her Vocal differential index reflecting the degree of his/her vocal differentiation between anger and sadness.

2.3.2. Results

As we expected the Vocal differential index to be correlated with autonomic responsiveness, we performed liner multiple regressions (stepwise method) with Vocal differential index as dependent variable and HRV index, age, sex, state anxiety, and extraversion scores as independent variables. The results of the regressions showed a highly significant effect for HRV index (T = 7.19; p < 0.001) and a much lesser effect for state anxiety (T = -2.052; p = 0.05). The HRV index alone explained 58% of data variance with the multiple R = .79. The subjects with higher degree of autonomic responsivity displayed higher prosodic differentiation of emotions. None of the other variables contributed significantly.

From the above results we conclude that, in our subjects, poor prosodic differentiation between anger and sadness could be interpreted as a symptom of poor autonomic
responsivity and marginally as a symptom of the state anxiety.

To test our second hypothesis regarding the relation between autonomic responsiveness and the degree of self reported subjective feeling, we performed Mann-Whitney U tests on 2 groups obtained by median split on HRV index. The results showed significant differences in the degree of felt sadness (Z = -3.3; P = .0009), and anger (Z = -2.4; P = .02). The groups with higher HRV reported higher degree of subjective feeling for both sadness and anger than those with lower HRV. For details and discussion, see [37].

These results lend support to the hypothesis that subjects with poorer autonomic responsivity experience flattening of emotional reactions on both the level of vocal expression, and the subjective feeling.

2.4. Prosy as a symptom of power dimension: a pilot study of adaptive vs. maladaptive coping style in breast cancer patients

2.4.1. Hypotheses and method

Active coping style has been defined as “tonic readiness to act upon an event”, while passive coping was described as “waiting for and enduring and attending to stimuli” [18]. It was hypothesised that coping power, as an attitudinal dimension related to the appraised power, could be reflected in the prosody of spontaneous speech. Ten breast cancer patients were interviewed by trained clinicians who rated their adjustment to the disease and treatment as adaptive vs. maladaptive. The definition of adaptive coping style corresponded to Obrist’s concept of active coping. The interview was semi-structured and adapted from an interview evaluating mental adjustment in patients with early breast cancer [17]. Samples of the patients’ voice recordings were taken at various moments of the interview that is: a moment of maximum vocal arousal, one of minimum vocal arousal as well as in passages related to specific topics in the course of the interview. They were acoustically analysed for the following voice parameters: mean F0, F0 range expressed as max/min ratio, F0 coefficient of variation, mean intensity, intensity range expressed as max/min dB ratio, and the rate of delivery. For each of these parameters we measured the difference between the values for high and low arousal conditions.

2.4.2. Results

Of particular interest is a remarkable higher difference between F0 range in high and low arousal conditions (ΔF0 range) in cases with adaptive (median = 55.2, 90% CI = 46.0-60.4) versus those with maladaptive adjustment (median = 35.2, 90% CI = 8.0-42.4), with Wilcoxon-Mann-Whitney U = 25.0; P = 0.009. In spite of our small sample, we believe that the results indicate a link between prosodic features and the power dimension inherent to attitudes related to coping styles and mental adjustment [1].

3. Discussion

Although the concept of three dimensional affective space is not novel [9] [16] [20], we believe that so far it has not been integrated into a unified model of cognition and affect whereby the very process of cognition is functionally related to a three-dimensional affective space, which in turn determines the subject’s subjective feeling as well as his actions involved in adaptive behaviour. We are also aware of the fact that, space not permitting, the model has been presented in a schematic way and further sophistication is much needed. We feel that research in affective prosody can benefit from a unified model as it offers a continuity from the states and processes traditionally regarded as cognitive (doubt, certainty) to emotionally coloured interpersonal stances (friendly), and to full fledged emotional experiences. For an example of application of valence, power and arousal as dimensions of emotional meaning see the work of Cécile Pereira [20].

4. References


