doi 10.4436/JASS.91005

JASs *Proceeding Paper*

Journal of Anthropological Sciences
Vol. 91 (2013), pp. 185-200

The evolutionary origin of multimodal synchronization in emotional expression[§]

Klaus R. Scherer

Swiss Center for Affective Sciences, University of Geneva, Switzerland
e-mail: Klaus.Scherer@unige.ch

Summary - *Darwin has pioneered the functional perspective in the study of emotions which currently dominates the field. Pursuing and expanding on this tradition, I hold that emotions should be defined as dynamic episodes characterized by a high degree of coordination between several organismic subsystems in the interest of optimal adaptation to relevant events. The constitutive feature of emotions, synchronization of response channels, is linked to the evolutionary origin of emotional expression by demonstrating that it is an essential aspect of spontaneous affect bursts in animals and humans. The significance of this mechanism to the signalling function in emotional expression communication is explored.*

Keywords - *Evolution, Synchronization, Emotional expression.*

Introduction

Thinking about and research on emotion during the last century has been largely informed by Darwin. Many of the discrete emotion models which have dominated the field in the last 50 years are derived from Darwin's volume on "The expression of emotion in man and the animals" (1872/1998). In this path-breaking work, Darwin had taken a number of major emotion terms in the English language as chapter headings and demonstrated for each of these the functionality, the evolutionary history, and the universality across species, ontogenetic states, and different cultures, focusing on motor expression. Darwin suggested three fundamental principles to explain specific motor activities resulting in observed expressions: 1) expressive movements, or parts thereof, are functional for the organism's response or are rudiments of formerly adaptive movements; 2) general arousal can produce unspecific expressive movements; and 3) some expressions are the opposites - or antitheses - of specific functional expressions. In highlighting the functionality of specific expressive movements, Darwin was influenced by the mid-18th century French scholars Duchenne de Boulogne and Pierre Gratiolet who, rather than assuming coherent, emotion-specific programs of muscle movements, suggested that particular motor movements have clearly circumscribed adaptive functions.

The theorist most responsible for the rediscovery of Darwin's seminal work for psychology, Tomkins (1962), extended Darwin's theorizing to argue that a number of basic or fundamental emotions could be conceived of as phylogenetically stable neuro-motor programs. While Tomkins did not describe the nature of these programs in detail, the assumption was that specific eliciting conditions (which Tomkins sought in different gradients of neural firing) would automatically trigger a pattern of reactions ranging from peripheral physiological responses to muscular innervation, particularly in the face (which Tomkins considered as the primary differentiating effector system). This concept has been popularized by

[§] This paper was presented during the International Conference *Darwin 2009: Emotions from an evolutionary perspective*, organized by the Istituto Italiano di Antropologia (Rome, April 28-30, 2009).

two scholars whose theorizing and research has been strongly influenced by Tomkins, Ekman and Izard, who extended the theory and attempted to obtain pertinent empirical evidence, particularly with respect to early ontogenetic onset of the discrete emotion patterns (Izard, 1994; Izard *et al.*, 1995), the discrete patterning of prototypical facial expressions for a number of basic emotions, and the universality of these patterns (Ekman 1972, 1973, 1992, 1999, 2003; Izard, 1971, 1992). Given the limited number of such basic or discrete emotions, theorists in this tradition have to postulate a mechanism of emotion mixing or blending to explain the large variety of emotional states that are popularly described by laymen and poets alike. In recent years, both Ekman & Izard have elaborated their theoretical ideas to account for the large variety of emotional states (thus Ekman talks about "families of emotion", Ekman, 1994, 2003) or the effects of the environment and culture on emotional development (Izard, 1994). On the whole, the major theorists in this area have focused mostly on the prototypical reaction patterns (in particular facial expression and physiological responses) that are considered to be characteristic for a particular basic emotion. There has been much less concern with the elicitation and differentiation of the respective emotion, generally thought to be determined by eliciting situations that are again considered to be characteristic for the respective emotion (such as death of a close person in the case of sadness or encountering a severe threat for life or well-being in the case of fear). Furthermore, the essentially dynamic unfolding of emotion processes has been largely ignored.

The dominant position of the basic, discrete, or differential emotion view has been increasingly challenged. Unfortunately, there is no agreed upon definition of emotion, even within in the central domain of the psychology of emotion, l*et al*one across the many disciplines that study emotional phenomena; rather the issue remains hotly debated (Frijda, 2007; Scherer, 2005). However, there is some convergence on at least some of the central elements which Frijda & Scherer (2009) have summarized as follows:

1) Emotions are elicited when something relevant happens to the organism, having a direct bearing on its needs, goals, values, and general well being. Relevance is determined by the appraisal of events on a number of criteria, in particular the novelty or unexpectedness of a stimulus or event, its intrinsic pleasantness or unpleasantness, and its motivational consistency, i.e., its conduciveness to satisfy a need, reach a goal, or uphold a value, or its "obstructiveness" to achieving any of those (Scherer, 2001).
2) Emotions prepare the organism to deal with important events in their lives and thus have a strong motivational force, producing states of action readiness (Frijda, 2007).
3) Emotions engage the entire person urging action and/or imposing action suspension and are, in consequence, accompanied by preparatory tuning of the somatovisceral and motor systems. This means that emotions involve several components, subsystems of the organism that tend to cohere to a certain degree in emotion episodes, sometimes to the point of becoming highly synchronized (Scherer, 2005).
4) Emotions bestow control precedence (Frijda, 2007) on those states of action readiness, in the sense of claiming (not always successfully) priority in the control of behavior and experience.

I suggest that these four central features jointly define what is generally meant by emotion, both in lay and scientific terminology. These features also allow distinguishing emotions from other affective states such as preferences, moods, attitudes, interpersonal stances or affective dispositions or traits. Scherer (2005) proposed a design feature system according to which emotions are specific in that they 1) are focused on specific events, 2) involve the appraisal of intrinsic features of objects or events as well as of their motive consistency and conduciveness to specific motives, 3) affect most or all bodily subsystems which may become to some extent synchronized, 4) are subject to rapid change due to the unfolding of events and reappraisals, and 5) have strong impact on behavior due to the generation of action readiness and control precedence. Based these premises, emotion will be considered here as a bounded episode in the life of a system that is characterized as an emergent pattern of component synchronization preparing adaptive responses to relevant events as defined by their behavioral meaning.

Such a componential approach attempts to decompose the phenomenon or process to be explained into its essential elements or constituents and then predict the effect of the determinants on the components and their interaction. In the affective sciences, there is now a convergence between a family of componential emotion theories that define emotion, as defined above, as a process that involves changes in several subsystems - cognitive activity, motor expression, physiological arousal, action tendencies, and subjective feeling state - as components of the theoretical construct "emotion" (see Scherer, 2005, 2009, for an overview). Most of these theories assume that these components are jointly driven by a set of common determinants and interact during the emotion process in a recursive fashion, resulting in a high degree of coherence or synchronization. Componential theorists generally do not endorse the idea of a small number of tightly organized basic emotions (or affect programs), but rather opt for the notion of a large number of highly differentiated emotions, assuming that some of these occur more frequently because of the ubiquity of certain situational outcomes.

Appraisal-driven functional subsystem coherence as a central defining criterion

In this contribution I focus on the central feature of *subsystem coherence and consequent synchronization of emotional response channels*, and, in particular, the special relevance of this synchronization phenomenon on motor expression and emotional communication. In order to explain the fundamental assumption underlying this notion, I will briefly describe my Component Process Model (CPM) of emotion (Scherer, 1984, 2001, 2009). In line with the functional tradition pioneered by Darwin, the component process model is based on the idea that during evolution, emotion has been optimized to serve the following functions: (a) evaluation of objects and events, (b) system regulation, (c) preparation and direction of action, (d) communication of reaction and behavioral intention, and (e) monitoring of internal state and organism–environment interaction (see Scherer, 1984, 2001). The model predicts that the results of sequential appraisal checks will generate appropriate response patterns.

The fundamental assumption is that organisms constantly scan and evaluate their environment and particularly significant events give rise to a process of reactive adaptation that involves functional changes in most, if not all, subsystems of the organism. Figure 1 shows the architecture of the model, including the dynamic, recursive emotion processes following an event that is highly pertinent to the needs, goals, and values of an individual. Emotion is seen as a reaction to significant events that prepares action readiness and different types of alternative, possibly conflicting, action tendencies but not as a sufficient cause for their execution (see also Frijda, 2007b). The assumption is that even highly emotional behaviors such as aggression or flight are prepared by emotions such as anger or fear but that their execution is multiply determined, with emotion being only one, albeit potentially important, factor.

As shown in the flow diagram, the CPM suggests that the event and its consequences are appraised with a set of criteria on multiple levels of processing (the appraisal component). The result of the appraisal will generally have a motivational effect, often changing or modifying the motivational state before the occurrence of the event. Based on the appraisal results and the concomitant motivational changes, efferent effects will occur in the autonomic nervous system (e.g., in the form of cardiovascular and respiratory changes) and in the somatic nervous system (in the form of motor

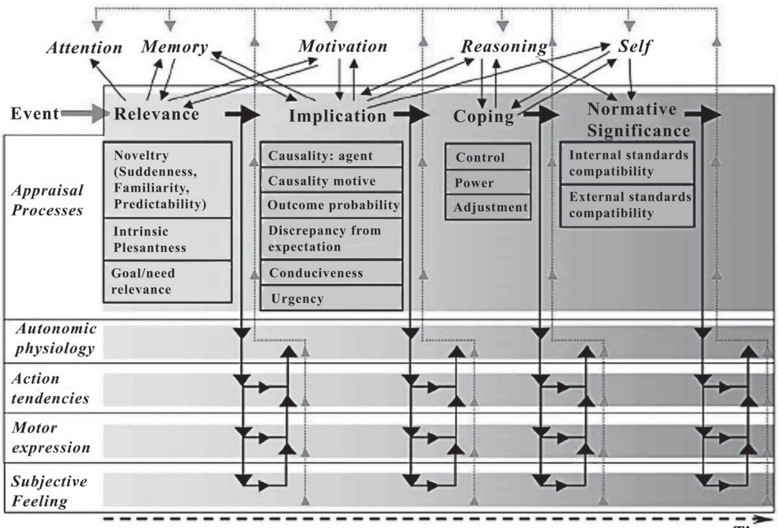

Fig. 1 - Graphical representation of the Component Process Model. The five components are listed vertically on the left of the picture. The appraisal processes are organized in four subsequent groups of stimulus (event) evaluation checks: Relevance, Implication, Coping and Normative Significance. The appraisal component triggers the emotion episodes and has efferent effects on all the other components (descending arrows). These components have reciprocal effects between them and feedback effects on the appraisal processes (ascending arrows) (Reproduced from Scherer, 2009).

expression in face, voice, and body). All of these components, appraisal results, action tendencies, somatovisceral changes, and motor expressions are centrally represented and constantly fused in a multimodal integration area (with continuous updating as events and appraisals change). Parts of this central integrated representation may then become conscious and subject to assignment to fuzzy emotion categories as well as being labeled with emotion words, expressions, or metaphors (recent descriptions of the model can be found in Scherer, 2001, 2004, 2005, 2009).

Multi-modal motor expression provides direct evidence for subsystem coherence in the emotion process

Here I focus on the model predictions that concern the emotion function concerning the communication of reaction and behavioral intention, which covers emotional expression in face, voice, and gestures. Further details on the predictions as well as pertinent empirical findings can be found for facial expression (Aue *et al.*, 2007; Kaiser & Wehrle, 2001; Smith & Scott, 1997; Mortillaro *et al.*, 2013; Wehrle *et al.*, 2000; Scherer, 1992; Scherer & Ellgring, 2007; van Reekum *et al.*, 2004) and vocal expression (Banse & Scherer, 1996; Goudbeek & Scherer, in 2010; Johnstone *et al.*, 2005; Juslin & Laukka, 2003; Scherer, 1985, 1987). The predictions described and tested in these articles have been elaborated on the basis of several classes of determinants: (a) the effects of the physiological change (see Johnstone *et al.*, 2001; Scherer, 1987), (b) the preparation of specific instrumental motor actions, and (c) the production of socio-communicative signals. The first two determinants can be subsumed under what the first author has called “push effects,” that is, internal changes that affect the expressive motor system. These are essentially biologically determined

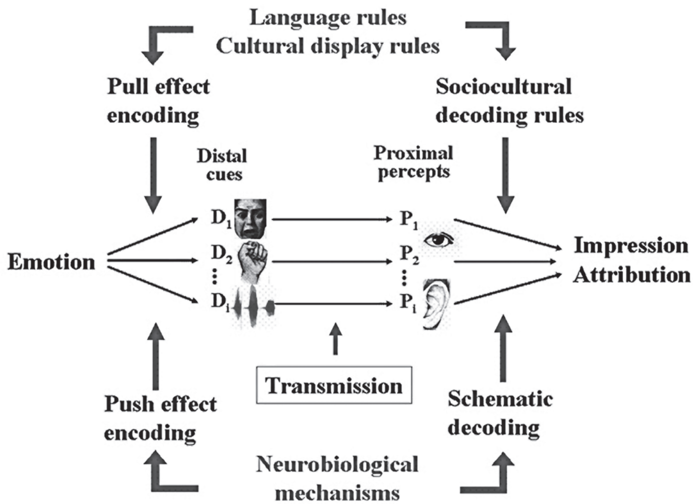

Fig. 2 - Modified version of the Brunswikian Lens Model. Starting from the left of the picture, an emotion is encoded by the sender through a number of multimodal cues which are the result of push effects (due to neurobiological mechanisms) and pull effects (due to language and cultural rules). These distal cues are transmitted through a channel that, along with the positioning of the receptors, determines how similar to the original distal cues the proximal percepts will be. The receiver employs socio-cultural and schematic rules to decode the proximal cues and to form an impression and attribute an emotion intention to the sender (Reproduced from Scherer & Bänziger, 2010).

externalizations of naturally occurring internal processes of the organism, particularly information processing and behavioral preparation. For these effects we expect relatively strong interindividual differences in the expressive patterns (as the underlying biological processes are dependent on both the idiosyncratic nature of the individual and the specific nature of the situation).

In contrast, the social signaling function is served by "pull effects," that is, particular visual or auditory signal configurations that are part of a socially shared communication code. These effects follow linguistic rules for the encoding of syntactic, semantic, and pragmatic aspects of meaning, or socioculturally determined norms or moulds concerning the signal characteristics required by the socially shared codes for the communication of internal states and behavioral intentions. They are characterized by a high degree of symbolization and conventionalization, and thus show comparatively few and small individual differences.

These two classes of determinants closely interact in both production and perception mechanisms (Scherer, 1985; Scherer & Kappas, 1988). Expression is one half of communication, perception is the other, and the two must always go together. This is why we have consistently suggested using a modified version of the Brunswikian lens model (Brunswik, 1956; Scherer, 2003) to better understand the overall process. Figure 2 shows a graphic representation of the dual process using the Brunswikian lens model.

The first class of determinants (push effects) can be subdivided into three major instrumental functions of the facial organs (lips, nose, ears) and the vocal tract (mouth, pharynx, larynx): (a) passing matter (light, air, liquids, solids) to and from internal organs (e.g., in the service of respiration, metabolism, and glandular secretion); (b) positioning sensory organs for optimal reception of stimulation (e.g., raising eyebrows, flaring nostrils); and (c) acting directly on objects and other organisms (biting, licking, kissing). Given

Tab. 1 - CPM Predictions for Facial and Vocal Expressions of Individual SEC Outcomes (modified from Scherer, 2001, 2009).

APPRAISAL DIMENSION	SEC OUTCOME	FACIAL EXPRESSION	VOCAL EXPRESSION
Novelty	Novel	Brows up, lids up; or brows lowered, jaw drop, open mouth, open nostrils, gaze directed	Interruption of phonation, ingressive (fricative) sound with glottal stop (noise-like spectrum)
	Not Novel	No change	No change
Intrinsic Pleasantness	Pleasant	Lids up, jaw drop, open mouth, open nostrils; or lip corners pulled upwards, lips parted, gaze directed	Faucal and pharyngeal expansion, relaxation of tract walls, vocal tract shortened due to lip corners pulled upwards (wide voice)
	Unpleasant	Brows lowered, lid tightened, eye closed, nose wrinkling, upper lip raised, lip corner depression, chin raised, lip press, nostril compression; or lower lip depressed, tongue thrust, lips parted, jaw drop; gaze aversion	Faucal and pharyngeal constriction, tensing of tract walls, vocal tract shortened due to lip corner depression (narrow voice)
Goal Conduciveness	Relevant and consistent	Relaxation of facial muscle tone	Relaxation of vocal apparatus (relaxed voice)
	Relevant and discrepant	Brows lowered, lids tightened, lips tightened, chin raising; gaze directed	Tensing of vocal apparatus (tense voice)
Coping Potential	No Control	Hypotonus of facial musculature, lip corner depression, lips parted, jaw drop, lids droop, eyes closed; if tears, inner brow raised, brows lowered, gaze aversion.	Hypotonus of vocal apparatus (lax voice)
	Control and High power	Brows lowered, lids up; or lids tightened, lips tightened and parted; or lips tightened and pressed together, nostril dilatation; stare	Chest register in phonation (full voice)
	Control and Low Power	Brows up, lids up, jaw drop, mouth stretched and corner retraction, nostril dilatation, switching between gaze direction and aversion	Head register in phonation (thin voice)
Norms	Respect	Elements of pleasantness and high power response	Elements of pleasantness and high power response
	Violation	Elements of unpleasantness and low power response	Elements of unpleasantness and low power response

the multifactorial determination of the muscles in the oropharyngeal and orofacial systems, and the fact that different demands upon the system may be more or less prevalent in particular situations, only approximate predictions can be made. Table 1 shows the component process model predictions for the facial and vocal expressions predicted to occur as a result of individual appraisal checks. The model suggests that the cumulative results of a sequential series of checks

(1, relevance of the event; 2, implications for major needs, goals, and values; 3, ability to deal with these consequences or coping potential; and 4, normative significance of the event) produce a wide variety of complexly patterned emotion episodes. Despite this variability, a number of modal emotions, such as anger, fear, or joy, can be identified (Scherer, 1994a). The final expression resulting from the sequential cumulative process described by the CPM can be predicted on the basis of 1) the assumptions shown in Table 1 and 2) the theoretically predicted appraisal profiles for major modal emotions (see Scherer, 2001). The facial and vocal configurations predicted as final outcomes for a number of modal emotions, as based on the mechanism described, are shown in Table 2 (see Mortillaro *et al.*, 2013; Scherer, 1987, 2001, 2009, for further details).

Given the importance and pervasiveness of multimodal coherence, one would expect to find copious evidence for coherence or synchronization between the channels described above in the research literature. Unfortunately, so far this is not the case. The reasons are twofold:

1) Most research groups in the emotional expression field have specialized in a particular modality, in almost all cases the face, and thus have not generated empirical data pertinent to the issue. A first effort was recently made by Scherer & Ellgring (2007b) who analyzed the multimodal coherence in a large set of actor portrayals of emotional states (for details see Banse & Scherer, 1996; Scherer & Ellgring, 2007a). Using cluster analyses to find coherence between facial action units (AUs; analyzed with the Facial Action Coding system, FACS) and digitally extracted voice parameters, they found some cross-modal clusters. Thus, at a higher level of agglomeration, a cluster characterized by high vocal arousal (high pitch and energy) is joined by AU27 (mouth stretch) and arms stretching sideways (a pattern dominated by arousal-related indices). At a still higher level, one detects a super-cluster combining the latter, a gestural illustrator cluster, and AU5 (upper lid raiser), yielding a total pattern one might describe as agitation (Multimodal Agitation). Furthermore, a positive facial cluster (smiling; AU6 + AU12) is joined by AU26 (jaws drop), fast speech rate, and head shaking (likely to consist of lateral head movements rather than the NO-emblem shake), a general pattern that suggests joyful surprise (Multimodal Joyful surprise). Finally, there is a pattern in which the low vocal arousal cluster combines with, successively, slow speech rate, upper body collapsed, AU14 (dimpler), AU41 (eyelids drop), back of hands pointing forward, and self-manipulators (the latter two variables may be linked as for many self-manipulators, such as scratching or rubbing frontal parts of the body; the back of the hand would point forward). The authors suggest calling this super-cluster Multimodal Resignation.

These results suggest that there is a strong probability to find the hypothesized coherence or synchronization between different expression modalities and channels, once one starts to analyze different modalities – facial, vocal, gestural – of expression for the same corpus of emotional expressions, using appropriate analyses. The latter still need to be developed. While cluster analysis can be used to obtain a first idea on what behavioural units tend to occur, more fine-grained methods need to be developed to examine the temporal contiguity and synchronization of these patterns.

2) As argued above, multimodal synchronization is a fundamental aspect of emotional expression. Given the nature of the underlying mechanisms, driven by appraisal, involving information gathering and action tendencies, one can assume a high degree of automaticity of the underlying processes, as the determinants mostly work on an unconscious lever in extreme rapidity and thus do not allow to control the synchronization of multimodal expression voluntarily, as the origin of the synchronization consists of the functional requirement to prepare coordinated behavioral readiness *very rapidly*. However, most emotion episodes are in fact subject to *emotion regulation*. This is almost always the case when a speech utterance is being produced which requires a high degree of intentionality as well as of precise motor control. This is exactly where the pull effects described above will have a major impact and introduce

Tab. 2 - CPM Predictions of Expressive Configurations for some Modal Emotions.

BEHAVIOR	ENJ/ HAP	ELA/ JOY	DISP/ DISG	SAD/ DEJ	DESP	ANX/ WOR	FEAR	IRR/ COA	RAGE/ HOA
Inner brow raiser		>		>	>>		>>		
Outer brow raiser		>					>>		
Brow lowerer			>	>	>>	>	>	>	>>
Upper lid raiser	>	>				>	>		>
Cheek raiser	>	>							
Lid tightener			>	>	>>	>	>>	>	>
Nose wrinkler			>						
Upper lip raiser			>						>
Lip corner puller	>	>>							
Lip corner depressor			>	>	>>				
Lower lip depressor			(>)						>
Chin raiser			>		>			>	>>
Tongue thrusting			(>)						
Lip stretcher						>	>>		
Lip tightener					>	>		>>	>
Lip pressor			>		(>)			>	
Lips part	>	>	(>)	>	>		>		>
Jaw drops	>	>>	(>)	>	>		>		>
Mouth stretches					>				
Nostril dilator	>								>
Nostril compressor			>						
Lids droop				>					
Eyes slit								>	
Eyes close				>					

elements that are not driven by underlying automatic emotion efference but by linguistic rules and social norms, depending on situational context. Obviously, such regulation of the spontaneous emotion response will tend to disrupt the automatic synchronization processes. As push and pull effects are always interacting, especially in verbal utterances, multimodal coherence is complexly organized and is, in consequence, difficult to demonstrate empirically.

I argue that there is one class of expressive behaviours that provides the primary example for the type of appraisal triggered multi-system coherence described above, allowing to study the phenomenon in its purest form -- affect bursts (see Scherer, 1994b for a detailed treatment).

Tab. 2 - continued

BEHAVIOR	ENJ/ HAP	ELA/ JOY	DISP/ DISG	SAD/ DEJ	DESP	ANX/ WOR	FEAR	IRR/ COA	RAGE/ HOA
Gaze	Dir	Up	Avert	Down	Var	Down	Dir	Dir	Dir
F0 Perturbation	<=	>		>	>		>		>
F0 Mean	<	>	>	<>	>		>>	><	><
F0 Range	<=	>		<	>	>	>>	<	>>
F0 Variability	<	>		<	>		>>	<	>>
F0 Contour	<	>		<	>	>	>>	<	=
F0 Shift Regularity	=	<			>		<		<
F1 Mean	<	<	>	>	>	>	>	>	>
F2 Mean			<	<	<	<	<	<	<
F1 Bandwidth	>	><	<<	<>	<<	<	<<	<<	<<
Formant precision		>	>	<	>	>	>	>	>
Intensity mean	<	<	>	<<	>		>	>	>>
Intensity range	<=	<=		<			>	>	>
Intensity variability	<	<		<			>		>
Frequency range	>	>	>	>	>>		>>	>	>
High-frequency energy	<	<>	>	><	>>	>	>>	>>	>>
Spectral noise				>					
Speech rate	<	>		<	>		>>		>
Transition time	>	<		>	<		<		<

Note - > indicates increase; (>) indicates potential increase; double symbols indicate the predicted strength of the change; joint use of two symbols pointing in the opposite direction refer to cases where antecedent voice types exert influence in opposite direction. ANX/WOR = Anxiety/Worry; DISP/DISG = Displeasure/Disgust; dir = directed; ELA/JOY = Elation/Joy; ENJ/HAP = Enjoyment/Happiness; IRR/COA = Irritation/Cold anger; RAGE/HOA = Rage/Hot anger; SAD/DEJ = Sadness/Dejection; var = variable.

The evolutionary continuity of affect bursts and the role of synchronization

While facial and vocal expression co-occur in the act of speaking, the different patterns of expression are determined, as discussed above, by a multitude of push and pull effects related to different types of marking or read-out functions as well as different communicative functions. For example, various parts of the upper face can be shown to have very important paralinguistic or conversational functions (Ekman, 1979). Similarly, many of the vocal changes related to intonation and pausing also carry a large variety of syntactic and pragmatic, and in some languages even semantic functions. Because of this huge number of determining factors, it has been almost impossible to study the relative role of facial and vocal expression and their interrelation in ongoing social interaction. Early attempts to look at nonverbal behaviour in terms of a multichannel partition (see, for example, Scherer & Wallbott, 1985, pp. 218-222) have remained rather sterile demonstrations of the technical feasibility of such an approach without having encouraged any substantial research activity.

In consequence, the vocal/facial synchronization I focus on here is of a more nonverbal nature -- namely "affect bursts", i.e. very brief, discrete, nonverbal expressions of affect in both face and voice as triggered by clearly identifiable events. One of my preferred examples of an affect burst (see Scherer, 1988) is a facial/vocal disgust expression upon seeing a hairy black worm emerging from an oyster shell one is about to bring closer to one's mouth. While reactions may differ, for most people there will be a brief burst of facial and vocal activity that is directly triggered by the visual information and the evaluation of the significance of the worm's appearance. With respect to the effect of push and pull effects in this burst, I assume a strong predominance of push effects during the initial appearance of the worm which will probably be quickly superseded by a predominance of pull effects, especially in situations of the communal eating of oysters (particularly in "good" company).

The phenomena I am labeling here as facial/vocal affect bursts are not well described in the literature. This is probably due, at least in part, to the difficulty of clearly delimiting the phenomenon in time, especially in the facial domain. These are neither micro-momentary expressions (as described by Haggard & Isaacs, 1966), nor necessarily full blown prototypical expressions of fundamental emotions (the facial phenomenon that has been most intensively treated in the literature), nor conversational markers as described by Ekman (1979). Because of the on/off characteristics of vocal signals, it is easier to isolate and identify these affect bursts in the vocal domain. In fact, the description of the vocal part of these affect bursts goes back to the pioneers in the study of speech and language. The term often used in this literature is "interjection". I will briefly survey some of the earlier literature in this area.

Essays dealing with the origins and functions of interjections are frequently found in earlier writings on the psychology of language, particularly in those concerned with speculations about the origins of human language. Kleinpaul (1888), for example, claimed the reflexive "nature and feeling sounds" belong to the household of the organism, so to speak, and therefore sound very much the same when uttered by speakers in different cultures. He tried to verify this conjecture by citing a number of examples of interjections in diverse languages whose sounds correspond to each other. Kleinpaul also referred to the fact that the same sounds, for example "o", can be used not only as an interjection, but also as a deliberate vocal signal. However, he insisted on a sharp distinction between spontaneously occurring interjections or exclamations expressing an emotional state and calls and shouts intentionally uttered for communicative reasons.

In his ethnopsychology, Wundt (1900) also discussed the "sounds of nature" (Naturlaute) and their role in language in great detail, tracing these back to inarticulate screams and cries accompanying very intense feelings of aversion, rage, and fear. Wundt distinguished between 1) primary interjections, defined as nature sounds which, being isolated remnants of a prelinguistic period, only serve to interrupt the continuity of speech and 2) secondary interjections which become assimilated into the language and eventually, in the course of cultural development, replace the primary interjections. He pointed out that the number of primary interjections found in any one language depends not only on the degree of civilization in a culture, but also on the restraints on affect expression imposed by custom or social norms. For example, one would expect cultures with strong control of emotional expression to develop secondary interjections like "ouch" to replace primary, "raw" pain cries like "aghghgh".

Kainz (1962) also differentiated between primary and secondary interjections and contrasted them with amorphous screams and sighs which he considered to be sheer reflexive vocalisations serving only to release emotions. Kainz assumed that prelinguistic nature sounds would most likely be uttered in emotionally charged situations. These in turn are presumably intuitively and sympathetically understood by the listener and are therefore of value for communication. He made a distinction between these sounds and true interjections, however, for two reasons: 1) the nature sounds would not be intentionally uttered and 2) they lacked the required phonetic symbol constancy due to their large intra- and interindividual variance. This constancy is expected for interjections considered to be components of language, due to the stylising and conventionalising which has moulded these sounds to correspond to the phonological requirements of a given language. Kainz differentiated between expressive and informative interjections (yuck, oh, ah, aua) and eliciting interjections (he, holla, pst) and assumed a diachronic development from expression to information and finally to elicitation in a process by which reflexive sounds of expression acquire communicative and linguistic significance.

Like Wundt, Kainz claimed that the "civilized individual" expresses emotions less and less frequently by means of pure nature sounds as civilization becomes more advanced. Instead, interjections which have been assimilated into language are used. This progressive transformation and symbolization of reflexive affect symptoms is attributed to increased cortical control of behaviour (a view which reminds one of the central role Elias, 1977, assigned to affect control in the process of civilization).

Thus far, I have been focusing on early contributions in the German tradition of "Psychologie der Sprache". As one might expect, linguists interested in speech have also discussed these "prelinguistic fragments" in the flow of speech. James (1974) reviews the respective writings of some of the classic authors in the field (Bloomfield, Fries, Jespersen, Sapir) who all agreed on the affective significance of interjections, their "primitive" and non-communicative nature, and their lack of grammatical structure. As in the German literature, the emphasis is generally on an evaluation of these vocalizations with respect to the structural criteria of language.

The social anthropologist Goffman (1978) has provided an exquisitely written analysis of interjections from an interactionist view. He defines exclamatory, non-lexicalised, discrete interjections as "response cries", expressions which he sees as "a natural overflowing, a flooding up of previously contained feeling, a bursting of normal restraints, a case of being caught off-guard" (p. 800). He discusses the following standard cries: 1) the transition display (e.g. Brr! upon escaping from adverse weather), 2) the spill cry (e.g. Oops! after having dropped something), 3) the threat startle (e.g. Eek! or Yipe! in response to facing a potentially dangerous situation), 4) revulsion sounds (e.g. Eeuw! upon seeing a disgusting sight), 5) the strain grunt (lifting or pushing something heavy), 6) the pain cry (e.g. Oww! or Ouch!), 7) the sexual moan, 8) floor cues (e.g. a deprecatory sound while reading the newspaper in company), 9) audible glee (e.g. Oooo! or Wheee!). Goffman insists that the significance of such response cries does not lie in their expressiveness which they share with most other talk ("Naked feelings can agitate a paragraph of discourse almost as well as they can a solitary imprecation. Indeed, it is impossible to utter a sentence without colouring the utterance with some kind of perceivable affect..." p. 813). Instead, it lies in their social interaction function which "allows and obliges us momentarily to open up our thoughts and feelings and ourselves, through sound, to whoever is present. Response cries, then, do not mark a flooding of emotion outward, but a flooding of relevance in." (Goffman, 1978, p. 815)

Poggi (1981), in one of the most exhaustive treatments of interjections to date (with mostly Italian examples), approaches the topic from the standpoint of a pragmalinguist, focusing on the actual use of different interjections in different contexts. She provides the most systematic and comprehensive effort toward a classification of the different sounds with respect to the underlying state of the speaker, distinguishing between information, interrogation, and request interjections.

This brief review of the literature reflects some of the attempts to define and classify interjections or affect vocalizations from the vantage point of language, communication, and social interaction.

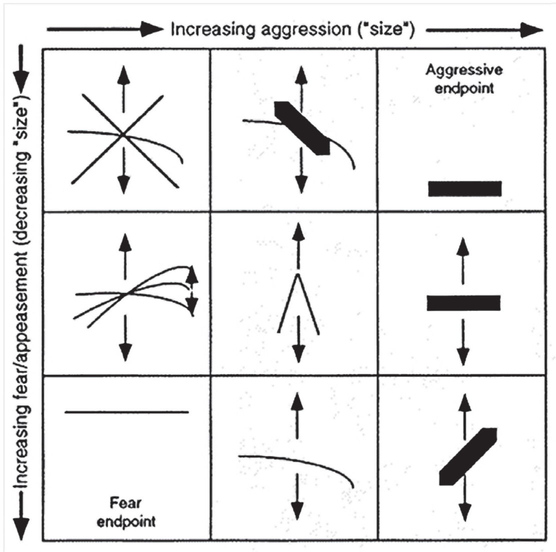

Fig. 3 - An illustration of the motivational structural rules postulated by Morton (1977) that describe the emotional modification of vocalizations in a similar fashion for many species of mammals.

The authors cited above have approached these issues from the vantage point of their particular approach and disciplinary predilection. I propose to study affect bursts from the vantage point of a theory of emotion. This will allow us to engage in a more systematic study of this simple but fundamental case of facial/vocal expression integration. Its special significance lies in the fact that it may be one of the phylogenetically oldest constituents of our communication system and seems to be directly related to the display and call systems in many mammalian species, a link that Darwin has strongly insisted upon in his pioneering work on expression. The evolutionary continuity of affect bursts, especially with respect to the vocalization part, has been shown by the very pertinent analyses of Morton (1977), who identified a number of the motivational structural rules that describe the emotional modification of vocalizations in a similar fashion for many species of mammals (see Fig. 3).

As shown above, in the component process model emotion is defined as a sequence of interrelated synchronized changes in the states of all organismic subsystems (information processing, support, execution, action, monitoring) in response to the evaluation of an external or internal stimulus event as relevant to central concerns of the organism (see Scherer, 1984, 2001). I believe that highly emotionally charged affect bursts are one of the best illustrations of this definition, in particular with respect to the strong synchronization of various organismic systems, particularly the various expressive channels, over a very brief period of time. Concretely, I argue that the results of the stimulus evaluation checks will produce functionally determined changes in the different organismic subsystems, including the ANS and the SNS, changes that will serve to accommodate the needs of information processing and behavioural adaptation.

I suggest using the term "affect bursts" for the extreme push pole of the dimension or continuum suggested above, i.e. those behaviour elements that are almost exclusively determined by the effects of physiological changes and that are therefore highly synchronized (with a graduated onset, following the sequential cumulative model), and barely conventionalized in form. In consequence, their occurrence should be quite universal, but their form should be variable over

Tab. 3 - Synopsis of the Mechanisms Involved in Emotion Expression (A) and Perception (B).

(A) EXPRESSION MECHANISMS	**(B) PERCEPTION MECHANISMS**
Hard-wired neuro-motor programs	Hard-wired patterns of feature detection
Appraisal-driven responses	Inference of underlying appraisals
Response regulation	Detection of regulation strategies. Inference of intention
Symbolic signalling	Schematic decoding of symbolic meaning
	Motor-mimicry
	Appraisal of the contextual information

individuals and situations. I suggest using "affect emblem" (based on Efron, 1941/1973, and Ekman & Friesen, 1969, 1972) as the term for the extreme pull pole, brief facial/vocal expressions that are almost exclusively determined by socio-cultural norms or models and that in consequence show a high degree of conventionality (with a simultaneous onset of all components and very little real synchronization among them).

Between these two extremes, we would expect a large number of intermediate cases, i.e. nonverbal facial/vocal expressions that are triggered by a particular affect arousing event and that show at least some degree of componential synchronization, but are at the same time subject to shaping by pull effects, as evident in control and regulation attempts. It is difficult to give examples for these different types on paper. But we can again use the example of the black hairy worm crawling out of the oyster shell. An affect burst would be the spontaneously triggered production of a nature sound of disgust accompanied by a rapid turning of the head away from it and a facial expression likely to help avoid further visual or olfactory exposure to the stimulus. An affect emblem is most likely to occur in a narrative concerning this event where a person may talk about the black worm by producing a highly culturally standardized vocalization, such as "yuck" in the United States or "igittigitt" in Germany (or "burck" in French) and a conventionalized disgust face. In this case, it is unlikely that the facial, vocal and gestural modalities are not well integrated or synchronized (with respect to onset, duration, intensity, or covariation of the various components) as predicted by a sequential-cumulative model.

While there has been some recent research on affect bursts (Belin *et al.*, 2008; Hawk *et al.*, 2009; Schröder, 2003), these are studies are again exclusively focused on the vocal part of affect bursts. In order to examine the central theses of this contribution, namely that the evolutionary continuity of affect bursts and its adaptational function make this class of emotional expression an ideal speciment to test the central notion of emotion as subsystem synchronization. To begin with, the very nature of the phenomenon requires integrative research approaches involving the assessment of all components of emotion. Even though this contribution emphasized facial and vocal expression, one should also assess antecedent cognitive or subcognitive appraisal, autonomic response patterning, motivational or action tendencies, and subjective feeling state. In line with the definition of emotion as a synchronization of all organismic subsystems (as presented above), all response modalities need to be studied to understand the special nature of affect bursts as an interruption of normal functioning, however brief. The clear temporal delimitation or bondedness of affect bursts make them easier to study in comparison to longer emotional states where the endings are often difficult to determine (see Frijda *et al.*, 1991). The need to study simultaneously all components of an emotional episode, which requires collaboration of researchers specialized in the study of different response modalities, could help to break up the splendid isolation with which much emotion research is confined to a single aspect of the multicomponential phenomenon of emotion.

References

Aue T., Flykt A. & Scherer K. R. 2007. First evidence for differential and sequential efferent effects of stimulus relevance and goal conduciveness appraisal. *Biological Psychology, 74*: 347-357.

Banse R. & Scherer K. R. 1996. Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.,* 70: 614-636.

Belin P., Fillion-Bilodeau S. & Gosselin P. 2008. The Montreal affective voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behav. Res. Methods*, 40: 531-539.

Brunswik E. 1956. *Perception and the representative design of psychological experiments*. University of California Press, Berkeley.

Darwin C. 1998/1872. *The expression of the emotions in man and animals* (3rd ed.). HarperCollins, London.

Ekman P. 1972. Universals and cultural differences in facial expression of emotion. In J. R. Cole (ed): *Nebraska Symposium on Motivation, Vol. 19*, pp. 207-283. University of Nebraska Press, Lincoln.

Ekman P. 1973. *Darwin and facial expression: a century of research in review*. Academic Press, New York.

Ekman P. 1979. About brows: emotional and conversational signals. In M. von Cranach, K. Foppa, W. Lepenies, & D. Ploog (eds): *Human Ethology: Claims and limits of a new discipline*, pp. 169-202. Cambridge University Press, Cambridge.

Ekman P. 1992. An argument for basic emotions. *Cogn. Emot.,* 6: 169-200.

Ekman P. 1999. Basic emotions. In T. Dalgleish & M. J. Power (eds): *Handbook of cognition and emotion,* pp. 45-60. John Wiley & Sons Ltd, Chichester.

Ekman P. 2003. *Emotions revealed* (2nd ed.). Times Books, New York.

Ekman P. & Friesen W.V. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica,* 1:49-98.

Ekman P. & Friesen W.V. 1972. Hand movements. *Journal of Communication,* 22*:* 353-374.

Elias N. 1977. *Der Prozeß der Zivilisation*. Suhrkamp, Frankfurt.

Frijda N. H. 2007a. What emotions might be? Comments on the comments. *Social Science Information*, 46: 433-443.

Frijda N. H. 2007b. *The laws of emotion*. Lawrence Erlbaum Associates, Mahwah NJ.

Frijda N. H. & Scherer K.R. 2009. Emotion definitions (psychological perspectives). In D. Sander & K. R. Scherer (eds): *The Oxford Companion to Emotion and Affective Sciences,* pp. 142-144. Oxford University Press, New York.

Frijda N.H., Mesquita B., Sonnemans J. & Van Goozen S. 1991. The duration of affective phenomena or Emotions, sentiments and passions. In K.T. Strongman (ed): *International Review of Studies on Emotion*, *vol. 1*, pp. 187-225. Wiley, New York.

Goffman E. 1979. Response cries. In M. von Cranach, K. Foppa, W. Lepenies & D. Ploog (eds): *Human ethology,* pp. 203-240. Cambridge University Press, Cambridge.

Goudbeek M. & Scherer K. R. 2010. Beyond arousal: Valence and potency/control in the vocal expression of emotion. *J. Acoust. Soc. Am.*,128: 1322-1336.

Grandjean D. & Scherer K. R. 2008. Unpacking the cognitive architecture of emotion processes. *Emotion,* 8: 341-351.

Haggard E.A. & Isaacs F.S. 1966. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. In L.A. Gottschalk, & A.H. Auerbach (eds): *Methods of research in psychotherapy*. Appleton-Century-Crofts, New York.

Hawk S. T., van Kleef G. A., Fischer A. H. & van der Schalk J. 2009. "Worth a thousand words":

Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion,* 9: 293-305.

Hess U., Kappas A. & Scherer K. R. 1988. Multichannel communication of emotion: Synthetic signal production. In K. R. Scherer (ed): *Facets of emotion: Recent research,* pp. 161-182. Lawrence Erlbaum Associates, Hillsdale, NJ.

Izard C. E. 1971. *The face of emotion*. Appleton-Century-Crofts, New York.

Izard C. E. 1992. Basic emotions, relations among emotions, and emotion-cognition relations. *Psychol. Rev.*, 99: 561-565.

Izard C. E. 1994. Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychol. Bull.*, 115: 288-299.

Izard C. E., Fantauzzo C. A., Castle J. M., Haynes O. M., Rayias M. F. & Putnam P. H. 1995. The ontogeny and significance of infants' facial expressions in the first 9 months of life. *Dev. Psychol.*, 31: 997-1013.

James D. 1974. *The syntax and semantics of some English interjections*. Papers in Linguistics, University of Michigan, Ann Arbor.

Johnstone T., van Reekum C. M., Hird K., Kirsner K. & Scherer K. R. 2005. Affective speech elicited with a computer game. *Emotion,* 5: 513-518.

Johnstone T., van Reekum C. M. & Scherer K. R. 2001.Vocal correlates of appraisal processes. In K. R. Scherer, A. Schorr & T. Johnstone (eds): *Appraisal processes in emotion: Theory, Methods, Research,* pp. 271-284. Oxford University Press, New York and Oxford.

Juslin P. N. & Laukka P. 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychol. Bull.,* 129: 770-814.

Kainz F. 1962. *Psychologie der Sprache. I. Band. Grundlagen der allgemeinen Sprachpsychologie*. (3rd ed.). Stuttgart.

Kaiser L. 1962. Communication of affects by single vowels. *Synthese,* 14: 300-319.

Kaiser S. & Wehrle T. 2001. Facial expressions as indicators of appraisal processes. In K. R. Scherer, A. Schorr & T. Johnstone (eds): *Appraisal processes in emotion: Theory, methods, research,* pp. 285-300. Oxford University Press, Oxford.

Kleinpaul R. 1888. *Sprache ohne Worte. Idee einer allgemeinen Wissenschaft der Sprache*. Leipzig. (New printing The Hague: Mouton, 1972).

Mortillaro M., Mehu M. & Scherer K. R. 2013. The Evolutionary Origin of Multimodal Synchronisation and Emotional Expression. In E. Altenmüller, S. Schmidt & E. Zimmermann (eds): *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man,* pp. 3-25. Oxford University Press, Oxford (in press).

Morton E.S. 1977. On the occurrence and significance of motivational-structural rules in some bird and mammal sounds. *Am. Nat.,* 111: 855-869.

Poggi I. 1981. *Le interiezoni: Studio del linguaggio e analisis della mente*. Boringhieri, Torino.

Scherer K. R. 1984. On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (eds): *Approaches to emotion,* pp. 293-317. Lawrence Erlbaum Associates, Hillsdale NJ.

Scherer K. R. 1985. Vocal affect signalling: A comparative approach. In J. Rosenblatt, C. Beer, M. C. Busnel & P. J. B. Slater (eds): *Advances in the study of behavior, Vol. 15*, pp. 189–244. Academic Press, New York.

Scherer K. R. 1987. Toward a dynamic theory of emotion: The componential process model of affective states. *Geneva Studies in Emotion and Communication,* 1: 1-98.

Scherer K. R. 1988. On the symbolic functions of vocal affect expression. *J. Lang. Soc. Psychol.,* 7: 79-100.

Scherer K. R. 1992. What does facial expression express? In K. Strongman (ed): *International Review of Studies on Emotion*, *Vol. 2*, pp. 139-165. Wiley, Chichester.

Scherer K. R. 1994a. Toward a concept of "modal emotions". In P. Ekman & R. J. Davidson (eds): *The nature of emotion: Fundamental questions*, pp. 25-31. Oxford University Press, New York.

Scherer K. R. 1994b. Affect bursts. In S. H. M. van Goozen & N. E. Van de Poll (eds): *Emotions: Essays on emotion theory*, pp. 161-193. Lawrence Erlbaum Associates, Hillsdale NJ.

Scherer K. R. 2001. Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr & T. Johnstone (eds): *Appraisal* *processes in emotion: theory, methods, research,* pp. 92-120. Oxford University Press New York.

Scherer K. R. 2003. Vocal communication of emotion: a review of research paradigms. *Speech Commun.,* 40: 227-256.

Scherer K. R. 2005b. What are emotions? And how can they be measured? *Social Science Information*, 44: 693-727.

Scherer K. R. 2009. The dynamic architecture of emotion: Evidence for the component process model. *Cogn. Emot.*, 23: 1307-1351.

Scherer, K. R. & Bänziger, T. 2010. On the use of actor portrayals in research on emotional expression. In Scherer K. R., Bänziger, T. & Roesch E. B. (Eds): Blueprint for affective computing: A sourcebook, pp. 166-178. Oxford University Press, Oxford.

Scherer K. R. & Ellgring H. 2007a. Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion,* 7: 113-130.

Scherer K. R. & Ellgring H. 2007b. Multimodal expression of emotion: affect programs or componential appraisal patterns? *Emotion,* 7: 158-171.

Scherer K. R. & Kappas A. 1988. Primate vocal expression of affective states. In D. Todt, P. Goedeking & E. Newman (eds): *Primate vocal communication,* pp. 171-194. Springer, Heidelberg.

Scherer K. R. & Wallbott H.G. 1985. Analysis of nonverbal behavior. In T.A. van Dijk (ed): *Handbook of discourse analysis, Vol. 2,* pp. 199-230. Academic Press, London.

Schröder M. 2003. Experimental study of affect bursts. *Speech Communication,* 40: 99-116.

Smith C. A. & Scott H. S. 1997. A Componential Approach to the meaning of facial expressions. In J. A. Russell & J. M. Fernandez-Dols (eds): *The psychology of facial expression*, pp. 229-254. Cambridge University Press, New York.

Tomkins S. S. 1962. *Affect, Imagery, Consciousness. Vol. 1, The positive affects*. Springer, New York.

van Reekum C. M., Johnstone T., Banse R., Etter A., Wehrle T. & Scherer K. R. 2004. Psychophysiological responses to appraisal dimensions in a computer game. *Cogn. Emot.,* 18: 663-688.

Wehrle T., Kaiser S., Schmid, S. & Scherer K. R. 2000. Studying the dynamics of emotional expression using synthesized facial muscle movements. *J. Pers. Soc. Psychol.,* 78: 105-119.

Wundt W. 1900. *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte. Band 1. Die Sprache.* Kröner, Leipzig.

Associate Editor, Rita Vargiu