EMUS
Expressivity in Music and Speech
http://recherche.ircam.fr/equipes/analyse-synthese/EMUS/

Fourth Conference
September 2008
Thursday 25 and Friday 26

Expression of emotions in Speech and Music
Microgenesis and semiotics of perceptual process

Place : RISC, 28, rue serpente, Paris 6ème, métro Odéon ou Saint-Michel (plan http://www.risc.cnrs.fr./plan.php), salle S35

Scientific Committee : Antoine Auchlin, Greg Beller, Didier Bottineau, Anne Lacheret, Aliyah Morgenstern, Nicolas Obin

Organasing Committee :Didier Bottineau, Greg Beller, Anne Lacheret, Nicolas Larousse, Aliyah Morgenstern, Nicolas Obin

Problématique
Expressions et émotions dans le langage et la musique
Microgénèse et sémiotique des processus perectifs

Un processus perceptif constitue un fait à la fois holistique, associé à l'expérience immédiate, et microgénétique de différenciation et de développement. En d'autres termes, toute expérience perceptive, même à l'échelle du temps présent, suit son propre parcours de développement. Il en va ainsi de la perception des phénomènes expressifs, qu'il s'agisse de traiter des événements verbaux ou non verbaux. Et, plus complexe : des événements où interagissent le verbal et le non verbal. Les journées proposées prennent comme point d'ancrage la perspective microgénétique des formes pour venir clore une série d'événements consacrés à la thématique de l'expressivité dans le langage parlé et le langage musical. En pratique, ces journées à l'interface de la musique et de la parole s'inscrivent dans un dialogue multidisciplinaire entre linguistique, modélisation informatique, neurosciences, psychologie et philosophie.
Il s'agira de présenter les méthodes et les concepts qui peuvent être mobilisés afin de faire le point sur l'apport mutuel des uns et des autres pour l'enrichissement des connaissances relatives

Presentation
Expressivity and emotions in Speech and Music
Microgenesis and semiotics of perceptual process

A perceptual process must be considered both as a holistic fact that is attached to immediate experience and as a microgenetic one in terms of differentiation and development. In other words, any perceptual experience, even within the limits of the present moment, follows its own course of development. This applies to the perception of expressive phenomena of verbal and non-verbal nature, but also to that of more complex events in which the verbal and the non-verbal elements tend to interact. The workshop, based on the microgenetic perspective, conclude a series of scientific events concerning expressivity in speech and music. In this context, the workshop tends to convey a multidisciplanary dialogue between linguistics, computer models, neurosciences, psychology and philosophy.
The goal of the event is to present the methods and the concepts of each discipline so as to fathom the contributions from those various domains and coordinate the respective expertise in order to improve each knowledge in the domain of the perception of expressive facts in spoken and musical language.
Which methods can be implemented in each of

à la perception des faits expressifs dans le langage parlé et musical. Quelles méthodes peut-on mettre en œuvre, quel que soit le niveau d'analyse impliqué (neurosciences, traitement du signal et phonétique, sémiotique, composition musicale, musicologie), et les domaines explorés (acquisition et apprentissage des formes, modélisation des structures et des systèmes, cognition située) pour comprendre les stratégies cognitives impliquées dans la différenciation et la construction des formes malgré le caractère a priori immanent des faits perçus ? Que dire sur le contenu sémiotique de ces formes et leur organisation temporelle ? Comment le sens émerge-t-il en parole et en musique ? Les formes sont-elles au départ des coquilles vides ou sont-elles d'emblée pourvues d'un contenu sémiotique ? Qu'est-ce qui relève dans ce traitement sémiotique de la dénotation d'un côté, de la métaphore et de l'objet fictionnel construit en fonction de ses propres repères culturels ? Comment aborder cette problématique dans une perspective contrastive : langage parlé vs. langage musical ? Par exemple, quel est le rôle de la mémoire dans les processus mis en œuvre ; dans quelle mesure les mécanismes mémoriels associés aux stimuli verbaux et non verbaux pourraient-ils expliquer des traitements sémiotiques et émotionnels distincts également. Autant de questions et certainement beaucoup d'autres pour lesquelles il semble judicieux de solliciter l'éclairage des sciences expérimentales et qu'il paraît légitime de soumettre à la réflexion linguistique, philosophique et musicologique, à l'intelligence artificielle et à la modélisation informatique.	those disciplinary fields (neuroscience, signal processing and phonetics, semiotics, musical composition, musicology), and in the corresponding domains (machine learning of semiotic patterns, models of structures and systems, situated cognition) to understand the cognitive strategies involved in the differentiation and construction of forms in spite of the immediacy of the perceivable facts? What can we say about the semiotic content of those forms and their temporal organization? How does meaning emerge in speech and music? Are forms first and foremost empty shells or do they have any semantic content in the first place? In this semiotic treatment what belongs to denotation on the one side, to metaphor and a fictional object on the other, constructed against the background of its own cultural landmarks? How should this problem be tackled in a contrastive perspective: spoken language vs musical language? For example, what is the role of memory in the processes considered; to what extent could the memorial mechanisms that are associated with verbal and non-verbal stimuli also account for the semiotic and emotional treatments? To answer all these questions, along with many others, the contribution of experimentaal sciences is needed, and it seems legitimate to submit them to the reflections of the linguist, the philosopher, the musicologist, the artificial intelligence and computer sciences.

Speakers : Antoine Auchlin (phonetics & linguistics), Mireille Besson (neurosciences), Didier Bottineau (linguistics), Christophe D'Alessandro (computer sciences & music), Michel Imberty (psychology), Anne Lacheret & Dominique Legallois (phonetic and linguistics), Valérie Pasdeloup & David Piotrowski (phonetics & linguistics), Xavier Rodet (signal processing), Victor Rosenthal (psycholinguistics), Daniel Schon (neurosciences), Jean-Luc Schwartz (computer sciences), Barbara Tillman (neurosciences).

Programme

Thursday September 25

9.30-10 Anne Lacheret: (MODYCO, UMR 7114, Paris X, Nanterre, Institut Universitaire de France)**:** ***Introduction to the workshop***

10-11 Victor Rosenthal (MODYCO, UMR 7114, Paris X, Nanterre, France)**:** ***Microgenesis and the expressive form of life***

The theory of microgenesis describes immediate experience (perception, thought, gesture, imagination) as a dynamical process of form *development* occurring on a *present-time* scale. This development founds the unity of lived experience. The dynamics of the whole process is described in terms of stabilization, categorization and differentiation from general underspecified to more definite and specific. Microgenesis is said to be a *living process* that dynamically creates a structured coupling between a living being and its environment and sustains a knowledge relationship between that being and its world of life (*Lebenswelt*). This knowledge relationship is protensively embodied in a *readiness for action*, and thereby has practical meaning and value. Microgenetic development is thus an essential form of cognitive process: it is a dynamical process that brings about readiness for action. This readiness for action instantiates the anticipatory structure of all lived experience: an anticipation of upcoming meaning. Indeed, the process of dynamic categorization which sets out in the earliest phases of microgenetic development imposes a horizon of *generic meaning* that guides any further differentiation and identification of forms. Emotional, cultural and socio-symbolic factors can thus affect the whole course of perceptual and cognitive processes.
The theory of microgenesis lets us also account for the expressive character of experience. Expressivity ought to be viewed as the most "primitive" semiotic regime of experience: the primary mode of controlling intersubjectivity which unifies its perceptual, affective, motivational, axiological, cognitive and symbolic facets. This expressive form of life is embodied in the *physiognomic* character of perception where any form or configuration primarily appears as an animated tone, a spontaneous manifestation of life. Perception is thus primarily qualitative (in the sense of affective valence) and semiotic; its physiognomic character builds upon the dynamics of constitution of perceptual configurations so that experience turns out to be *expression of its own process of constitution*.

11-12 Antoine Auchlin (Department of Linguistics, University of Geneva)**:** ***Meu su, voyons!* - from *meant* meaning to *meaning* meaning. Notes on enaction, microgenesis and *experiential blending***

Prosody is embodying meaning through vocalization. Prosodic variations engage permanent structural couplings between the variations of state (psycho-corporal) of the speaker and what is said, when it is said. It leads to experiencing discourse, not just handling abstract concepts, as in Cartesian linguistic tradition. "Expressivity" is a dimension of that coupling.
In this communication, we will discuss different prosodic phenomena pleading for an experiential and enactive approach to discourse and communication. Experiential model defines communication as a co-experienciation process. This model attempts to shed light on dynamic integration of sensori-motor activity and linguistic construction of meaning. This integration occurs within affective valence and interest arousal regulation conditions.
Various kinds and levels of blending (Fauconnier & Turner) are at work at the same time in speech; some of them consist in blending perceptive and linguistic inputs, and the resulting output is experiential.
Fónagy's *meu su, voyons* exemplifies such a case of complex experiential blending: harmonics - formants transformation (mais [me]->meu [mø]; si [si]->seu [sø]) is blended onto phono-articulatory posture, which in turn is blended with linguistic (instructional) content "mais si, …". The outcome is experiencing "mais si voyons" articulated with protruded lips; this experiential, embodied, sensori-motor outcome is as evident as its conceptualization is fuzzy. In order to conceptualize what is "expressed" or manifested by protruded articulation (some kind of *hypocoristic sorrow*), the hearer needs introspection, in trying by him/herself to imitate the lips movement. In other words, basic evidence is not conceptual.
As for more complex prosodic levels of integration, we first will examine how far experiential blending can explain, in prominence detection, mismatches between automatic and expert detection, as

reported by recent publications (Obin & al. 2008, Simon, Avanzi, Goldman 2008) and work in progress.
The reminder of the communication discusses cases of lowered register that blend various prosodic ingredients into complex attitudinal information. The attitude, we argue, is not processed conceptually, it is presented in its embodied and pre-conceptual manifestation, and is experienced as such; transitions between successive enacted attitudes show the temporal elaboration of vocal enactive microgenesis processes.

12-13 **Anne Lacheret[1] & Dominique Legallois[2]** (1MODYCO, ParisX Nanterre & IUF, Paris, 2CRISCO, université de Caen) **:** ***Expressivity and emotion in spoken language: what does grammar have to tell us?***

In this presentation the notion of expressivity in its semiotic dimension is regarded as the manifestation of an affective and emotional relation of the perceiving subject to a content through the prosodic and semantic modalities.
First, starting from I. Fonagy (1983)'s schema of the double coding of communication, we will show that even if prosody does indeed convey emotional patterns that can be phonetically characterized, these patterns are underlain by memorial constraints that determine their production and interpretation. We will distinguish three types of memories: discursive, interlocutive and referential. We will see that in every case, the verbal material, or syntactico-semantic domain, that accompanies the prosodic constructions, intrinsically conveys emotional interpretations and strictly constrains the prosodic patterns instantiated in the spoken message. From this point of view, what is denoted brings about connotation; in other words, there cannot be any double or parallel coding: connotation is pre-encoded in grammar.
The second part of this paper will be devoted to the role of grammar. Although expressivity has usually been relegated to a secondary role by the prevailing formal grammatical approach of language, it has indeed been granted some importance by some linguists for a long time. We will present Ch. Bally's fundamental statements about expressivity in his book *Le langage et la vie* (1913/1926). These statements (or proposals) tend to orient linguistics towards an analysis of discursive facts in relation with what the author calls *le mode vécu* "the experience mode" (subjective and affective experience) as opposed to *le mode pur* "the pure mode" (intellectual experience). After presenting and commenting on Bally, we will try to show the effectiveness of expressivity in the domain from which it has systematically been excluded: grammar. If grammar is considered not just as a system of production of linguistic forms, but as a mode of instantiation of preconstructed and conventional forms, it can be shown from illustrative examples that grammar, like the lexicon, is fraught with expressivity and records the emotional patterns that are perceived by or suggested to the speaking subject.
To summarize: grammatical expressivity tends to regain some of the ground it had lost in linguistics on account of the increasing significance that is ascribed to

- the notions of subject and interlocution (the enunciative models)
- the cognitive dimension of language.

13.15-14.30 : Lunch

14.45-15.45 **Mireille Besson[1], Mitsuko Aramaki[1], Daniele Schön[1], Aline Frey[2]** (1Institut de Neurosciences Cognitives de la Méditerranée, CNRS-Marseille Universités, Marseille, 2 Laboratoire Cognitions Humaine et Artificielle; Université Paris 8, France)**:** ***An interdisciplinary approach to the semiotics of sounds***

In this presentation, we will present 3 series of experiments aimed at exploring the similarities and differences when processing the meaning of linguistic and non-linguistic sounds. To this aim, we used both behavioral (percent errors and Reaction Times, RTs) and electrophysiological approaches (Event-Related brain Potentials or ERPs). In the fist series experiments we compared priming effects (unrelated vs related) for words and for environmental sounds (the sounds were specifically built so that the source of the sounds was difficult to identify). In the second series of experiments we presented both typical and ambiguous sounds from material categories (e.g., wood, metal and glass). Sound continua were built between two material categories so that typical sounds were at the extreme of the continua and ambiguous sounds in the middle. We compared priming effects for typical and for

ambiguous sounds of material categories and for words, pseudo-words and non-words. Finally, in the third series of experiments, we used short musical excerpts, called Semiotic Temporal Units (TSUs) that convey specific musical concepts. We compared priming effects for congruent and for incongruous TSUs that is for TSUs that conveyed same or different concepts. In all 3 series of experiments, results revealed higher error rates and slower RTs for unrelated than related stimuli (typical priming effects) and enhanced negativity in the ERPs. However, results also show some differences (in latency and scalp distribution) in the priming effects for linguistic and non-linguistic sounds. The functional significance of these results will be discussed.

15.45-16.45 Christophe d'Alessandro, Sylvain Le Beux, Albert Rilliard (LIMSI, Orsay, France)**:** ***Towards kinematical modelling of expressive speech prosody: experiments in computerized chironomy***

Although various intonation models have been proposed for a variety of languages, the question of expressive intonation representation is still wide open. The approach defended in this conference is based on the hypothesis that intonation shares a lot of common features with other types of expressive human movements or gestures. New insights in intonation research could be gained addressing the question of intonation representation in terms of prosodic movements, inspired by musical representation in terms of hand movements (chironomy).
A system for "computerized chironomy", i.e. real-time intonation and duration modifications driven by hand gestures using a graphic tablet is presented. This system is tested in a reiteration task, where the subjects reproduce intonation contours using a pen and the graphic tablet. The subjects also produce vocal imitation of the same corpus. Correlation and distances between natural and reiterated intonation contours are measured. The results indicate that vocal intonation reiteration and chironomic intonation reiteration give comparable intonation contours in terms of correlation and distance.
This shows that computerized chironomy can be used for expressive speech analysis, as expressive intonation contours can be produced and represented by the hand-made tracings. Intonation modelling in terms of movements is discussed. A kinematical description of prosody using velocity, target position and rhythmic patterns is proposed. The kinematics of speech prosody is compared those of other human skilled movements (like writing or playing musical instruments).

16.45-17.45 Michel Imberty (Département de psychologie, université de Paris X, Nanterre)**:** ***Voice, musicality and temporality***

It is known today that the voice plays a part completely separated and at the same time central in the emergence of the interactive conduits of communication : that it thus appears the large mediator between the biological nature of the musicality and its cultural sources.
It is not a question obviously of dealing here with the whole of the phenomena of the voice, but to only encircle in what the voice, by its natural musicality, organizes the interactions and the individual experiment of time. Thus will be mixed the natural uses of the voice in the language and the expression of the emotions, and the uses culturalized which are the sung voice, in particular in recitative as well the baroque as contemporary. Because, in all these fields, the voice psychologically concretizes what, in a more general way, many authors call today the *"proto-narrative envelope*" of the human experiment.
The basic idea is that the interpersonal temporal experiment continues is cut owing to a capacity or an aptitude of narrative thought. According to D. Stern the narrative thought is a universal means by which everyone, including the newborns, perceives and organizes the expressive human behavior. The proto-narrative envelope, at the same time former to the verbal language and developing out of its own sphere, is organized around two interdependent aspects which are on the one hand the *intrigue*, i.e. what connects "which, where, why, how" of the human activity, and on the other hand, the *line of dramatic tension* which is the contour of the feelings, such as they emergent at the present moment. *The proto-narrative envelope* is thus a form proto-semiotics of the interior experiment of time, a matrix of the "account" of the tensions and relaxations related to the "intrigue" (or "quasi-intrigue") during the search for a satisfaction. It is what gives to the experiment its total unit, whatever the degree of complexity is, which gives the major feeling of the unit of the self in the change ceaseless one of the temporality of the life.

Friday September 26

9.30-10.30 Jean Luc Schwartz (GIPSA, Grenoble, France)***: From auditory patterns to speech "patterned" by perceptuo-motor interactions***

We shall begin with gestalt perception, from visual patterns which drive perception and guide action, or auditory streams and auditory scene analysis, to the multistability phenomenon (the famous Necker's cube) with perceptual switches which enhance our understanding of decision and consciousness.
Then we shall come back towards old questions in psychology and philosophy, about the reality of perception, asking if these "patterns" already exist in the physical world, or are a pure perceptual creation.
Then we shall move towards speech, with its multistability phenomenology associated with the "verbal transformation effect". This "language game" in which a stimulus uttered in loop may transform into another one ("life life life" becoming "fly fly fly") might be at the basis of the creation of "verlan" in French (uttering words in the reverse sense, or rather in loop, to create a new word). We shall describe the "phonological loop", a perceptuo-motor system inside the human brain, enabling to store, analyse and process phonological patterns.

We shall finally evoke what could be the ingredients of a "speech morphogenesis": or how the speech patterns (vowels, consonants, syllables and words) might emerge from the perceptuo-motor interaction between communicating human brains.

10.30-11.30 Barbara Tillmann* & W. Jay Dowling ** (* CNRS-UMR 5020, Lyon, France; ** University of Texas at Dallas, USA.)**:** ***Memory of Music and Poetry: Keeping Details over Time***

It seems to be well-established that short-term memory for detailed information declines over time, especially with additional material presented during the delay. For music, we have reported experiments showing lack of decline, and even improvement, for the memory of musical information. Listeners heard the beginnings of musical pieces, of which one of the initial phrases was tested later. The music continued, and memory was tested after delays up to 30 sec with a repetition of the target, a similar, or a different lure. Discrimination performance (particularly discrimination between targets and similar lures) remained strong and even improved with increasing delay. This effect disappeared when the delay was silence or filled with sound patterns breaking the musical continuity. Based on this data of music, we investigated memory of fine surface details for poetry and prose materials with the same experimental methods. For prose, short-term memory performance for detailed information declines over time, replicating previous findings.
For poetry, we observed a lack of decline for memory for surface details, similar to the data obtained for musical material. The data for music and poetry suggest a particular role played by temporal organization and rhythmic structure in short-term memory.

11.30-12.30 Xavier Rodet (IRCAM, Paris, France) **:** ***Voice transformation: methods, means and applications in music, cinema and multi-media***

In the term "expressivity" of speech, one gathers various aspects like emotion, affect, intention, nuances, etc, which are carried by acoustic variations of the sequence of phones, i.e. variations of pitch, duration, intensity, articulation and timbre. Ircam has developed knowledge and means that allows such sorts of transformation. These means make it possible to change many aspects of a voice: one can modify the perception of the age of the person, of its gender, its height, of other dimensions of the timbre of its voice, like breathiness, but as well of the prosody, and finally of the expressivity of a spoken sentence. Thus a statement perceived as neutral can be transformed to be perceived like sad or astonished or, more lively. We will talk about acoustic variations that one can do, results obtained and research in progress. The examples of applications go from the music (sound installation of J.B. Barrière) to cinema (voice of G. Depardieu in the film Vatel, voice of A. Gillet in last film of E. Rohmer) and to multi-media in general (video games, dubbing, avatars, etc)

12.45-14 : Lunch

14.15-15.15 **Daniele Schon** (Institut de Neurosciences Cognitives de la Méditerranée, CNRS-Marseille, France) :
Music and language: cerebral functions or cultural artifacts?

Music and language, so similar and so different at the same time. Cognitive neurosciences also participate to this debate by studying brain regions involved in music and language processing. Are these regions similar or different? And can the functioning of these regions be modified by external factors? In order to try to give a tentative answer to these complicated questions I will persent data from neuroimaging studies comparing music and language processing as well as musicians and non-musicians.

15.15-16.15 **Valérie Pasdeloup[1] & David Piotroswki[2]** (1 LPL, Aix-en-Provence, 2CREA, Paris, France):
On some aspects of sign perception: the case of the temporal structure of speech

The question of the temporal structure of speech is a very controversial one. The reason is that speech flow duration can be analyzed at least at three levels : (i) phonetic, (ii) phonemic, and (iii) sign. Each level showing a particular time course organization.
At the phonetic level, the speech duration is approached as a pure physical phenomena.
Concerning the phonemic level, and discussing the particular example of rate sensitivity of stressed and unstressed vowels, we will show that the duration structure at this phonemic level is different from the temporal structure at the infra-phonetic level, and that it can be described as a form controlled by parameters of the phonetic level.
At the sign level, the difficulty is linked to the phenomenological ambivalence of the "signifiant" face of sign, which merges the strata of simple phonological perception with that of full sign perception where intentionality of meaning constitutes an essential phenomenological character - ambivalence which makes Saussure hesitating between, on one side, the axiom of "signifiant linearity", which is correlated with a conception of speech time course ordered as a succession of items, and, on the other side, the fact that the speaking subject does not perceive any signs succession : "he is in front of a state". We will show that the phenomenological analysis of sign developed by Husserl takes account of such an ambivalence and leads to a conception of sign perception where the simple phonemic perception is, in terms of intentional modalities, a "floating data" background from which raises intentional objects of meaning in whose linguistic awareness "fully lives".

16.15-17.15 **Discussion**