

**NATO ADVANCED STUDY INSTITUTE (ASI) on**  
*The Fundamentals of Verbal and Non-verbal Communication and  
the Biometrical Issue*  
*Vietri sul Mare, Italy, 2-12 September 2006*

## Abstracts



### Age as a Disguise in a Voice Identification Task

Ruth Huntley Bahr

Dept. of Communication Sciences and Disorders, University of South, Florida, Tampa, USA

**ABSTRACT:**

Listeners have been shown to be quite reliable in identifying speaker age; however, little is known about how speakers may use age as a voice disguise. Previous research has shown that fundamental frequency and speech rate may be powerful predictors of speaker age. Yet, it is not known if talkers manipulate these parameters when attempting to produce an older voice or if they will rely on voice stereotypes to guide their productions. The following investigation is designed to study how manipulation of speaker age might affect speaker identification.

Four young actors (2 males and 2 females) provided voice samples that represented young, middle-aged, and older voices. Four healthy elders also provided voice samples. In the first experiment, naïve listeners were asked to estimate talker age of both the young adult simulations of vocal age and the "real age" samples. In the second task, listeners were asked to determine if the presented voice pairs were from the same or different speakers.

Results indicated that talkers were able to produce voices that were perceived as different from their chronological age. However, these vocal disguises were not perceived as being representative of actual older voices. In terms of disguise effectiveness, the listeners did not confuse the age-disguised voice for another individual, but these same auditors had much difficulty matching an age-disguised voice to its owner. These results will be discussed in terms of aging stereotypes and cues to speaker identity. The influence of these findings on the speaker identification process will be described.



### Virtual talking heads and ambient face-to-face

#### Communication

Gérard BAILLY, Frédéric ELISEI and Stephan RAIDT

Institut de la Communication Parlée, 46 av. Félix Viallet, 38031 Grenoble - France

**ABSTRACT:**

We describe here our first effort for developing a virtual talking head able to engage a situated face-to-face interaction with a human partner. This paper concentrates on the low-level components of this interaction loop and the cognitive impact of the implementation of mutual attention and multimodal deixis on the communication task.



### Detection of Faces and Recognition of Facial Expressions

Nikolaos Bourbakis

ITRI, Wright State University, Dayton, Ohio, USA

**Abstract**

Face detection is the foremost task in building vision-based human-computer interaction systems and in particular in applications such as face recognition, face identification, face tracking, expression recognition and content based image retrieval. A robust face detection system must be able to detect faces irrespective of illuminations, shadows, cluttered backgrounds, facial pose, orientation and facial expressions. Many approaches for face detection have been proposed. However, as revealed by FRVT 2002 tests, face detection in outdoor images with

uncontrolled illumination and in images with varied pose (non-frontal profile views) is still a serious problem. In this talk, we describe a Local-Global Graph (LGG) based method for detecting faces and for recognizing facial expressions accurately in real world image capturing conditions both indoor and outdoor, and with a variety of illuminations (shadows, high-lights, non-white lights) and in cluttered backgrounds. The LG Graph embeds both the local information (the shape of facial feature is stored within the local graph at each node) and the global information (the topology of the face). The LGG approach for detecting faces with maximum confidence from skin segmented images is described. The LGG approach presented here emulates the human visual perception for face detection. In general, humans first extract the most important facial features such as eyes, nose, mouth, etc. and then inter-relate them for face and facial expression representations. Facial expression recognition from the detected face images is obtained by comparing the LG Expression Graphs with the existing the Expression models present in the LGG database. The methodology is accurate for the expression models present in the database.



## Image Chromatic Adaptation for Face Skin Color Detection

Nikolaos Bourbakis and Praveen Kakumanu  
ITRI, Wright State University, Dayton, Ohio, USA

### Abstract

The goal of image chromatic adaptation is to remove the effect of illumination and to obtain color data that reflects precisely the physical contents of the scene. We present in this talk an approach to image chromatic adaptation using neural networks (NN) with application for detecting - adapting human skin color. The network is trained on randomly chosen color images containing human subject under various illuminating conditions, thereby enabling the model to dynamically adapt to the changing illumination conditions. The proposed network predicts directly the illuminant estimate in the image so as to adapt to the human skin color. The comparison of our method with Gray World, White Patch and Neural Network on White Patch algorithms is presented. We also present our results on detecting skin regions in NN color corrected test images. The results are promising and suggest a new approach for adapting human skin color using NN's. The skin detect technique presented here is the first part of an integrated methodology-tool used for detecting human face and facial expressions of emotion.



## Nonverbal Communication as a Factor in Linguistic and Cultural Miscommunication

Maja Bratanić, Professor, University of Zagreb, Croatia

### ABSTRACT:

The presentation discusses two major assumptions:

- that a great deal of human communication is culturally molded and conditioned
- that people convey meanings not only through language but through various aspects of nonverbal communication as well.

Nonverbal behavior is to a great extent universal but in many ways also marked by culture-specific patterns. Being less obvious than misunderstandings in verbal communication, nonverbally induced miscommunication is far more difficult to detect. Furthermore, the line between verbal and nonverbal components of communication is often hard to delineate precisely.

Main categories of nonverbal behavior and its role in communication will be briefly discussed with the focus on proxemics - the study of the human use of space within the context of culture. The concept of proxemics and its implications will be elaborated on examples from American cultural patterns.

Further examples of culturally-conditioned miscommunication will draw on an aviation-related context.

The presentation will be accompanied by a video *A World of Differences: Understanding Cross-Cultural Communication* by D. Archer.



## Face recognition from 2D still images

Paola Campadelli  
Università di Milano, Italy, [campadelli@dsi.unimi.it](mailto:campadelli@dsi.unimi.it)

### Abstract:

In the past two decades a lot of research work has been devoted to the development of automatic methods aimed at recognizing people from images;

such systems are attractive since this type of identification does not require any interaction with the subject. However, the problem is very difficult especially when very few assumptions are done on the images to be treated.

In this talk the most interesting methods developed for face recognition from still images will be presented and compared. Open problems will be dealt with, and the contribution that 3D information might provide will be discussed.



## Technology for Non-Verbal Speech Processing

Nick Campbell,

ATR Science Labs, Kyoto, Japan, [nick@atr.jp](mailto:nick@atr.jp)

### Abstract:

This pair of lectures will focus on the technological needs for the processing of non-verbal speech in a dialogue context. It is based on an analysis of a very large corpus of spoken interactions captured under extremely natural situations. The talks will present a model of speech interaction as not only facilitating the exchange of linguistic or propositional information, but also facilitating the display of affect and interpersonal or social relationships.

### Part I: Speech Synthesis and Discourse Information

This talk presents some recent work towards a conversational speech synthesis system for use in interactive dialogues, such as might take place between a person and an information system, a robot, or a speech translation device. The talk describes several types of response utterances that are currently very difficult to implement using traditional speech synthesis methods, and shows how these non-verbal speech sounds function to provide feedback and status-updates in an interactive discourse. The lecture will be illustrated with examples of such practical utterances, including laughter and grunts as well as common phrases and idioms, showing how their variety can reveal several types of information about the speaker- (i.e., listener) states. The proposed model of information exchange through non-verbal speech shows how this feedback from the listener can help the speaker to deliver content more efficiently, and at the same time to be reassured of success in information transmission.

### Part II: Towards Recognising Speech Gestures in Discourse

This talk describes how the lowest level of information can be processed in a speech signal for annotation of discourse progress and speaker participation status. In a semi-formal round-table meeting situation there is typically only one main speaker at any given moment, but several participants may be speaking simultaneously, expressing (dis-)agreement, chatting, translating, etc., in addition to the main speaker. We are currently performing research into technology to process this audio landscape in order to detect the main speaker and to categorise the competing forms of speech. Several speech gestures such as laughter, agreement, and feedback-responses can be recognised, isolated, and used to determine the progress of the meeting and the degrees and types of participation status among the members present. This talk will describe the current state of the technology and will present examples of the frequent gestures with descriptions of their typical usage.



## The Amount of Information on Emotional States Conveyed by the Verbal and Nonverbal Channels: Some Perceptual Data

Anna Esposito

Dipartimento di Psicologia, Seconda Università di Napoli, and IIASS Italy

### ABSTRACT:

In a face-to-face interaction, the addressee exploits both the verbal and nonverbal communication modes to infer the speaker's emotional state. Is such an informational content redundant? Is the amount of information conveyed by each communication mode the same or is it different? How much information about the speaker's emotional state is conveyed by each mode and is there a preferential communication mode for a given emotional state? This work attempts to give an answer to the above questions evaluating the subjective perception of emotional states in the single (either visual or auditory channel) and the combined channels (visual and auditory). Results show that vocal expressions bring the same amount of information as the combined channels and that the video alone brings poorer emotional information than the audio and the audio and video together. Interpretations of these results (that seem to not support the data reported in the literature proving the dominance of the visual channel in the

emotion's perception) are given in terms of cognitive load, language expertise and dynamicity. Also, a mathematical model inspired to the information processing theory is hypothesized to support the suggested interpretations.



## Analyzing and modelling verbal and non-verbal communication for talking animated interface agents

David House and Björn Granström  
Royal Institute of Technology, Sweden

### ABSTRACT:

The use of animated talking agents is a novel feature of many multimodal spoken dialogue systems. The addition and integration of a virtual talking head has direct implications for the way in which users approach and interact with such systems. However, understanding the interactions between visual expressions, dialogue functions and the acoustics of the corresponding speech presents a substantial challenge. Some of the visual articulation is for obvious reasons closely related to the speech acoustics (e.g. movements of the lips and jaw), while there are other articulatory movements affecting speech acoustics that are not visible on the outside of the face. On the other hand, many facial gestures used for communicative purposes do not affect the acoustics directly, but might nevertheless be connected on a higher communicative level in which the timing of the gestures could play an important role. The context of much of our research regarding these questions is to be able to create an animated talking agent capable of displaying realistic communicative behaviour and suitable for use in conversational spoken language systems.

The focus of these lectures is to look into the communicative function of the agent, both the capability to increase intelligibility of the spoken interaction and the possibility to make the flow of the dialogue smoother, through different kinds of communicative gestures, such as visual prosodic gestures (e.g. focal accent and emphatic stress) and gestures for different expressive states, turntaking and negative or positive system feedback. We will give some examples of recent work, primarily at KTH, involving the collection and analysis of databases for audiovisual prosody. We will report on methods for the acquisition and modelling of visual and acoustic data, and provide some examples of analysis of e.g. head nods and eyebrow settings related to communicative functions. We will also demonstrate how this analysis can be implemented to generate useful and realistic expressive audiovisual synthesis using a combination of data-driven and rule-based methods.



## Individual Speech Rhythm Variation Within the Plosive Structure of Speech

Eric Keller, IMM, University of Lausanne, Switzerland

### Abstract:

The "perceived humanness" of speech revolves centrally around the issues of regularity and variation. Within an utterance's temporal structure, we subjectively experience humans to speak with a certain regularity -- which creates perceived rhythm within speech -- at the same time as we expect them to display variation, mostly for emphasis and to satisfy personal preferences. Synthesized speech that does not exhibit these perceptual qualities is often classified as "robotic" and "unnatural".

The search for the objective bases of the perceived regularity in speech is old and has produced less than satisfactory results. In fact in 1977, Ilse Lehiste, in an extensive review of the issue of isochrony (acoustic evidence for rhythmicity in speech) came to the conclusion that there were no direct acoustic correlates of rhythmicity, a view that has formed the consensus for spontaneously produced speech since then, despite a number of further studies performed on the issue.

However, we have data to show that regularity may actually be directly dependent on what might be called the "plosive structure" of the speech chain. If one considers vowel onsets in terms of the suddenness and the relative strength of voice onset, it turns out that human speakers exhibit considerable inter-speaker agreement with respect to the placement of sudden ("strong") vowel onsets, but that this inter-speaker agreement is gradually reduced as vowel onsets "weaken". The "strength" or "weakness" of the vowel onset can be determined automatically from the acoustic signal, and is thus likely to correlate with both motor and perceptual saliency within the utterance. "Strong" vowel onsets (i.e., those that resemble plosive sounds) appear to set a "frame" for speaking, and between those onsets, we are much freer to choose the timing values appropriate to the specific semantic and personal context.

Current statistical or neural network models for temporal structuring of speech may thus well be flawed. Currently, we model all aspects of speech timing with rigid, totally predictable statistical structures. Instead, temporal prediction systems should probably provide a set of main "temporal anchor points" within the utterance, and introduce "motivated variation" for the remaining aspects of temporal structure. We will pass in review the different types of psycholinguistic and pragmatic events that can motivate such variation, and we will consider a prediction system that can handle these new requirements.



## TWO LECTURES ON GESTURE

ADAM KENDON

### Abstract:

We begin with the question: what is 'gesture'? Can we identify, in a theoretically coherent manner, a domain of human action to be called 'gesture' which is to be distinguished from other forms of human bodily expression? After answering this question, we shall proceed to look at the problem of how meaning is attributed to gestural actions. Then we shall look at gesture in relation to speech and consider the different ways it may be employed whenever it is used in conjunction with speech within an utterance. Examples will be presented that show that speaker's appear to employ speech and gesture as partners in a common enterprise of utterance construction. Gesture, it will be shown, must be seen to be as much the 'final product' of a speaker's utterance as are the speaker's words. Discussion will then turn to a consideration of how gesture and speech interact semantically as these two modalities create unified expressions. It will be shown that the ways in which gestures contribute to the overall meaning of an utterance are quite diverse. No simple generalizations are possible. Here we shall first consider the contributions gestures may make to the propositional content of utterances. This will be followed by a discussion of the 'meta-discursive' or pragmatic functions that gestures also frequently can be shown to have. Throughout these lectures numerous examples will be presented, drawn from video-recordings of conversations in many different circumstances in southern Italy (especially in the provinces of Naples and Salerno) and in central England.

Recommended reading: Adam Kendon *Gesture: Visible Action as Utterance*. Cambridge University Press, 2004. Especially Chapters 7-13.



## A Methodological and Theoretical Framework for the Study of Psychological and Biometrical Characteristics in Verbal and Nonverbal Communication

Dominic W. Massaro, Ph.D

Perceptual Science Laboratory, Department of Psychology, University of California  
Santa Cruz, CA 95060 U.S.A, 1-831-459-2330, FAX 1-831-459-3519, [massaro@fuzzy.ucsc.edu](mailto:massaro@fuzzy.ucsc.edu),  
<http://mambo.ucsc.edu/psl/dwm/>

### Abstract:

The goal of the lectures is to provide an empirical and theoretical overview of a paradigm for inquiry on the Psychological and Biometrical Characteristics in Verbal and Nonverbal Communication. A persistent theme of our approach is that humans are influenced by many different sources of information, including so-called bottom-up and top-down sources. Understanding spoken language, for example, is constrained by a variety of auditory, visual, and gestural cues, as well as lexical, semantic, syntactic, and pragmatic constraints. We will first present a theoretical framework for language processing, and the methodological implications of this framework. We will then review experimental evidence in support of this framework while inconsistent with other frameworks. We will indicate some limitations in other current studies, and suggest changes for improvement. Finally, we will extend the analyses to cross-linguistic studies of speech perception, as well to studies of emotion.

Research questions for psycholinguists and speech and reading scientists include the nature of the sources of information; how each source is evaluated and represented; how the multiple sources are treated; whether or not the sources are integrated; the nature of the integration process; how decisions are made; and the time course of processing. Research in a variety of domains and tasks supports the conclusions (for summary see Massaro, 1998) that a) perceivers have continuous rather than categorical information from each of these sources; b) each source is evaluated with respect to the degree of support for each meaningful alternative; c) each source is treated independently of other sources; d) the sources are integrated to give an overall degree of support for each

alternative; e) decisions are made with respect to the relative goodness of match among the viable alternatives; f) evaluation; integration; and decision are necessarily successive but overlapping stages of processing; and g) cross-talk among the sources of information is minimal. The fuzzy logical model of perception (Massaro, 1998; FLMP) will be described and contrasted with other models of speech and emotion processing. These models will be tested against speech perception experiments involving cross-linguistic comparisons and experiments on emotion and gesture perception.



### Abstract of McNeill lectures

David McNeill University, Chicago, USA, [dmcneill@uchicago.edu](mailto:dmcneill@uchicago.edu)

1. Introduction to gesture study - how and why it is done.
2. Some basic facts of speech-synchronized gestures.
3. The growth point, context, catchment, imagery-language dialectic, dynamic and static dimensions and how they relate.
4. Gesture in social interaction.
5. Gesture and culture, linked in non-obvious ways.
6. Gesture and brain, relevance to the origin of language.
- 7.



### Multimodal expressive ECAs

Catherine Pelachaud,

Universite de Paris 8, France, [c.pelachaud@iut.univ-paris8.fr](mailto:c.pelachaud@iut.univ-paris8.fr):

#### Abstract:

Embodied Conversational Agents (ECAs) are human-like entities capable of communicating with other ECAs and/or users. They exhibit synchronized verbal and nonverbal behaviors (facial expression, gesture, body movement and gaze). During these lectures we will present an ECA system architecture. We will introduce the taxonomy of communicative functions developed by Isabella Poggi. Following this taxonomy each communicative function is represented as a pair where the first element corresponds to the meaning of the communicative function while the second element is a description of the signal that is used to transmit this meaning. A representation language, Affective Presentation Markup Language, APML has been elaborated. It is used to drive the animation of the agent and ensures synchrony between the multimodal signals. Behaviors are defined not only by the signals that composed them but also by how they are displayed. We will present an expressivity model where six parameters have been designed to change the gesture and face expressivity.



### Intonation, Accent and Personal Traits

Michelina Savino

Dept. of Psychology, University of Bari, ITALY

[m.savino@psico.uniba.it](mailto:m.savino@psico.uniba.it)

#### Abstract:

Speaking with an accent reveals the sociolinguistic background of locutors, and it is a widely shared belief that intonation plays a crucial role in characterising regional varieties of spoken languages. This is attested by the large amount of descriptive studies in the literature, and also from the perceptual point of view a number of experiments have been carried out to test the hypothesis that language varieties can be identified by pitch information alone (even though it is currently not clear to what extent segmental information can also play a role in such identification task).

An exemplar case is represented by Italian speakers, for whom the spoken language represents a reliable way for identifying their sociolinguistic traits in verbal interactions, as they always speak with an accent, in both formal and informal situations. This is a consequence of the particular status of Italian with respect to other languages: for historical reasons, in fact, the process of standardisation has been successfully achieved for the written form but not for the spoken language, which is presently characterised by quite strong regional accents. In fact, standard Italian has never been taught in any level of the Italian education system, and its use has been restricted to a small number of professional speakers and actors, and therefore in very specific contexts.

In this lecture, a discussion on the role of intonation in conveying information on the speakers' sociolinguistic background as personal traits will be presented, basing mainly on examples of Italian varieties.



## Blind Signal Pre-Processing for MPEG-7 based Multimedia Metadata applications: an Assisted Living use-case

Giovanni Tummarello, Stefano Squartini, Francesco Piazza  
Università Politecnica delle Marche, Italy

### ABSTRACT:

Metadata extracted from Multimedia or live sensing is set to play a major role in any intelligent and multimodal interactions between humans and computers. Furthermore, it is generally required that such metadata are structured and encoded according to well agreed standards. This is fundamental to enable interoperability and create complex applications as a mesh of heterogeneous services and components. On purpose, the MPEG-7 standard for dealing with multimedia metadata and the tools developed within the Semantic Web initiative are providing today the basic framework. Their application to real world problems, however, is made problematic by the fact that the data are often captured from difficult live conditions. It is therefore of primary importance to enhance the quality of the observable signals before the metadata extraction algorithms are employed. In particular, for the case of audio signals, it is important to perform separation and deconvolution of audio signals captured in real environments and in blind conditions. In this work a full featured real world multimedia metadata assisted living scenario is constructed using a combination of Blind Signal Processing and MPEG-7 based metadata techniques. In such example, an array of microphones captures speech and audio signals and thanks to MPEG-7 technologies the user can select multimedia content to be played.



## Visual Phrasemes and Pragmaphrasemes in English, Polish and Croatian

Neda Pintaric  
University of Zagreb, Croatia

### ABSTRACT:

The author writes about etimological and semantic meaning of an eye as the organ with the biggest capacity among all human organs. Eyes, ears and tactil receptors are main receptors in human communication. The author claims that nonverbal code is a pracode based on these receptors, therefore it has been developed earlier than the verbal code.

Nonverbal code is multicode consisting of kinetic, tacezic, deiktic, proxemic and prosodic signs which we use consciously and unconsciously. Our unconscious informations couldn't be hidden and the other person in communication can read it in our eyes.

In various cultures people operate with visual culturemes, such as a custom of eye-contact which can mean sincerity (e.g. among Croats) or sexual affection (e.g. among Poles).

The main part of the paper consists of lexical and phrasematic analyse of pragmatic items called pragmemes and phrasopragmemes. The author compares different linguistic systems using examples in English, Polish and Croatian eye-signs.



## Videocaptured Verbal and Nonverbal Foreign Language Teacher Feedback

Leticia Vicente-Rasoamalala  
Aichi Prefectural University, JAPAN

### ABSTRACT:

Videotaping for research purposes in the field of Second Language Acquisition (SLA) classroom context is still quite experimental and a minority practice. Apart from the difficulties for getting the consent of the participants (i.e. the school authorities, the teachers and the parents of the students) for videorecording classroom interactions, there are not very defined research lines and instruments to work with the obtained data. Specifically, the present poster will highlight different issues concerning the collection, the identification and the analysis of the videocaptured verbal and the nonverbal foreign language teacher feedback in classroom context. The focus on this area is at identifying from the collected database the teachers' strategies that appear to be more successful in dealing with L2 learner oral output containing deviant forms. From the last decade, a number of SLA studies influenced by *Long's Interaction Hypothesis* (1996) are studying in detail the negative implicit forms of teacher feedback under the assumption that they might be more beneficial for acquisition. Such works have adopted diverse qualitative and quantitative paradigms. For instance, elements of ethnographic research,

Conversational Analysis and the Neo-Vygotskian perspective (Vygotsky, 1968). The ultimate goal of many studies in this area is finding out the instructional sequences that might optimize foreign language teaching and learning. Nevertheless, there are some shortcomings relating to the lack of consensus for the existing schemes analyzing corpora and the types of verbal and non verbal annotations.

Originally, most approaches in classroom discourse research have built analytic frameworks almost exclusively dealing with audiotaped linguistic data designed to build didactic models. Significantly, most instruments have neglected the audio-visual data which might be captured in videorecordings. Thanks to the new technologies the notations of teacher gestures and the manipulation of tools are being incorporated in some classroom studies. Some works have suggested that non verbal elements might enhance verbal feedback and often regulate classroom exchanges among teachers and learners. Additionally, those outcomes suggest that using extensively video recordings is necessary for future comprehensive studies of classroom discourse.

#### References

1. Block, D (2003) *The Social Turn in Second Language Acquisition*. Edinburgh University Press.
2. Gardner, R. & Wagner, J. (Eds.) (2004) *Second Language Conversations*. London: Continuum.
3. Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In Bhatia and Richie, (Eds.), *Handbook of second language acquisition*, San Diego: Academic Press, Inc.
4. Lyster, R. (2001) Negotiation of form, recasts and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, 51, 265-301.
5. Vicente-Rasoamalala, L. (2006) Elementos No Verbales en la Retroalimentación del Docente de L2". *The Journal of the Faculty of Foreign Studies. Aichi Prefectural University* 38, 159-188.



## Effectiveness of Short-Term Prosodic Features for Speaker Verification

Iker Luengo, Eva Navas, Inmaculada Hernáez

University of the Basque Country, Spain, [ikerl@bips.bi.ehu.es](mailto:ikerl@bips.bi.ehu.es)

#### Abstract:

In this work a traditional MFCC based system is combined with a prosody based one to determine whether simple short-term prosodic information is useful for improving current state-of-the-art ASV. Results do not show significant improvement



## A Partially Observable Markov Decision Process approach to Affective Dialogue Modeling

Bui Huu Trung,

Human Media Interaction, Department of Computer Science, Faculty of Electrical Engineering Mathematics and Computer Science, University of Twente, P.O. Box 217 7500AE Enschede the Netherlands

#### Abstract:

We propose a novel approach to developing a dialogue model which is able to take into account some aspects of the user's emotional state and acts appropriately. The dialogue model uses a Partially Observable Markov Decision Process approach with observations composed of the observed user's emotional state and action. A simple example of route navigation is explained to clarify our approach and preliminary results & future plans are briefly discussed.



## A Systemic Approach to Enhance Writing, Analysis, and Presentation Skills

Aly N. El-Bahrawy, Professor

Faculty of Engineering, Ain Shams University, Cairo, Egypt

#### Abstract:

The paper discusses the systemic of technical communication - for researchers and professionals - which includes writing, analysis and presentation. Each of the three components has sub-components related to conveying the technical message clearly to the receiver. The first component 'Technical writing' is related to language rules in general, and technical writing guidelines in particular. The second component 'Data analysis' is related to

statistical and database principles in addition to graphics basics. The third component 'Professional Presentation' is related to organization of material, audio-visual equipment, the body movement of the speaker, which includes voice, hand gestures, facial expressions, etc. Another encompassing factor is the use of computers to help the three components send the respective message clearly. As an example, the Microsoft programs WORD, EXCEL, and PowerPoint are powerful tools to write, analyze and present the technical message. In WORD, features like formatting, language tools, and automatic generation of reference tables are very appealing. In EXCEL, data analysis and graphics are very elaborate. In PowerPoint, organization and animation tools can be used to enhance the presentation significantly. The paper presents examples of the use of such approach to execute successful training courses for engineers and researchers. Finally, the paper stresses the importance of the three components to form the technical and academic character of professionals.



## Can Word prime gestures?

Paolo Bernardis

Università di Bologna, Scuola superiore di Studi Umanistici. Università di Parma, Dipartimento di Neuroscienze

### Abstract:

Can words prime gestures?

Evidence for language and action relationships was recently highlighted in both behavioral (Glenberg, 2002) and neurophysiological research (Gentilucci, 2001), thus reinforcing the already well established link between language and gestures for communicative purposes (Bellugi, 1979; 1987; McNeill, 1992; 2000). However, at present there is no clear evidence of the direct interaction of the two systems.

Aim of this study was precisely to check for evidence of this interaction with the priming effect, i.e. to investigate whether words could prime the recognition of a gesture with the same meaning. The participants were presented with a word and then asked to recognize a gesture either having the same or a different meaning. The words chosen as primes were of different kinds: 1) a noun referring to a simple object with (1a) no specific action required (e.g., 'clock') or (1b) a specific action required (e.g., 'gun' -to shoot); 2) a verb referring to (2a) a direct simple action (e.g., 'knock') or (2b) an action to be performed with a tool (e.g., 'write'). The gestures presented were video-clips showing the upper half-body of an actor, which was blurred, performing the gestures with his arms and hands. Response times and errors were recorded.

Two main results were obtained. The first was a clear priming effect of meaning (i.e., a semantic priming effect of word on gesture). The second result showed a greater priming effect when both words and gestures referred to objects and actions requiring tools.



## The study of similarities in learning foreign languages

Daboveanu Diana-Cristina, Nicolae Carmen - Eufrosina

Romania, Bucharest, Sector 3, 37, Mircea-Voda Boulevard, Block M29, Scara D, 3rd Floor, Flat 113,

### Abstract:

Today student mobility has already become reality. It is supported by numerous education programmes at national and international levels and in particular within the framework of EU funded programmes. There is, however, a clear need for support programmes to assist exchange students to prepare for their studies. The EUROMOBIL project (72139-CP-2-2000-1-FI-L2) the development of a multimedia language learning and information programme on CD-ROM for DE, EN, HU and FI) was started in 1999 with the aim of developing a self-study course, which would enable exchange students to prepare, both in terms of the host language and knowledge of the culture, for their visit to universities in DE, UK, HU and FI. The project was expanded to CZ, FR, PL, PT and RO.

In a needs analysis at the beginning of the project differences in the requirements for studies abroad were noted. Starting from this need analysis and using rank distance as a measure for similarity our goal was to research how related are the problems that appear when studying a foreign language and to expand this result to see the differences and similarities between the cultures and languages that are part of the project.



# Tongue motor cortex excitability is modulated by the observation of the type of grasp action

Dalla Volta R<sup>1</sup>, Bernardis P<sup>1</sup>, Buonocore A<sup>1</sup>, Sato M<sup>1</sup>, Palumbo D<sup>1</sup>, Gentilucci M<sup>1</sup>  
<sup>1</sup>Dept. Neuroscience, University of Parma, Italy, gentiluc@unipr.it

## Abstract:

Voice spectrum and lip kinematics during pronunciation of syllables are affected by the simultaneous execution or observation of transitive actions, like the grasp of different objects, according to the type of involved hand grip. The study aimed to verify whether in humans the observation of hand grasping actions onto objects requiring different types of grip affects tongue motor cortex excitability. We recorded motor evoked potentials (MEPs) from the tongue of 16 right handed healthy subjects after delivering single pulses by using transcranial magnetic stimulation (TMS) over the tongue left motor cortex. While stimulating the participants looked at the PC monitor where video-clips showing either hand grasping of fruits of different size that required different types of grip or the same fruits alone were presented. The syllable DA appeared on the fruits in both cases. Power grip of large fruits was linked to MEPs significantly greater than those linked to precision grip of small fruits. In a control experiment we presented either different tools approaching geometric solids or the same solids alone. Neither when presenting the same fruits alone nor in the control experiment any MEP modulation was observed. We conclude that observation of different biological hand actions onto edible objects specifically modulates tongue motor cortex excitability. These data support the hypothesis that a motor resonance system is activated by hand action observation. This resonant circuit sends double motor commands to both hand and mouth.



# Machine Translation Evaluation: a Case study of Croatian-English and Russian-English MT Systems

Ivana Simeon

Department of Linguistics Faculty of Philosophy, University of Zagreb, Ivana Lučića 3, HR-10000 Zagreb, Croatia, E-mail: [isimeon@ffzg.hr](mailto:isimeon@ffzg.hr)

## Abstract:

From the earliest days of machine translation (MT), evaluation has been an inherent and significant part of efforts invested into machine translation research. In this paper, an overview of the history of MT evaluation is presented, with emphasis on one of the most comprehensive MT evaluation projects, undertaken in the 1960s, namely the Automatic Language Processing Advisory Committee Report.

Furthermore, strategies and problems pertaining to MT evaluation are discussed, with emphasis on the distinction between subjective criteria, such as comprehensibility, and the objective, quantifiable criteria, such as error quantification and analysis.

Within the practical part of the paper, the results of testing four MT systems - one for the language pair Croatian-English, and three for the language pair Russian-English - are shown. The systems were tested on three textual samples belonging to general, fictional and scientific genres. The analysis of the results included a comprehensibility poll which included five informants (native or proficient target language speakers) for each target language, as well as quantification of errors and error type assessment across genres and across individual MT systems. Finally, cumulative results are given for each MT system and for each language pair.

As a conclusion, recent developments in the field of MT evaluation are presented, including automatic evaluation methods, such as IBM's measures BLEU and NIST.



# Using the Wavelet Transform in Real-time Digital Signal Processing

Jan Vlach, Přinosil Jiří

Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, Purkynova 118, 612 00 Brno, Czech Republic,

## Abstract:

The new method of segmented wavelet transform (SegWT) makes it possible to exactly compute the discrete-time wavelet transform of a signal segment-by-segment. This means that the method could be utilized for wavelet-type processing of a signal in "real time", or in case we need to process a long signal (not necessarily in

real time), but there is insufficient memory capacity for it (for example in the signal processors). Then it is possible to process the signal part-by-part with low memory costs by the new method. The method is suitable for universal utilization in any place where the signal has to be processed via modification of its wavelet coefficients (e.g. signal denoising, compression, speech segmentation, music processing, alternative modulation techniques for xDSL systems). It is also possible to use SegWT in wavelet-processing (e.g. compression, selective area processing) of large images. In the paper, the principle of the forward segmented wavelet transform is described.



## The Integrative and Structuring Function of Speech in Face-to-Face Communication from the Perspective of Human-Centered Linguistics

Krzysztof Korzyk

Jagiellonian University Kraków, Poland

### Abstract:

This paper illustrates the need for study of the interdependencies between verbal and nonverbal behavior treated as a unified form of activity, manifesting itself in face-to-face communication. Invoking the principles of human-centered linguistics (see Yngve 1996, 2000), the author treats verbal communication not as something passed on via language, but rather as something to which language merely contributes. One of the consequences of such an approach to this issue is a reassignment of focus. Rather than attention being drawn to linguistic phenomena, the spotlight is on the communicative properties of the interlocutors, creatively utilizing various elements of the interactional "symbolic space."

With reference to the above, this text presents a realistic account of the interpretational activity taking place between communicating subjects. The action is perceived as a function of choices correlating verbal, prosodic, and kinesthetic signs and signals. Concurrently, taking the pragmatic and interactional aspects of these multimodal choices under consideration, the author discusses typical situations in which the structuring role of speech is particularly evident. Light will also be shed on the crucial interconnections between the above-mentioned systems of signs and signals, as well as on the advantages stemming from an integrated modeling of communicative phenomena.

1. Yngve V.H. (1996) *From Grammar to Science. New Foundations for General Linguistics*, Amsterdam: John Benjamins.
2. V.H. Yngve and Z. Wąsik (2000) *Exploring the Domain of Human-Centered Linguistics from a Hard-Science Perspective (Workshop)*, Poznań: Motivex.



## Research on Speech Synthesis and Speech Recognition of Croatian language on the Faculty of Humanities and Social Sciences

Lazic Nikolaj,

Faculty of Humanities and Social Sciences University of Zagreb- Ivana Lucica 3 10000 Zagreb- Croatia,  
[nlazic@ffzg.hr](mailto:nlazic@ffzg.hr);

### Abstract:

Speech synthesis and speech recognition are processes in need of multidisciplinary approach. Faculty of Humanities and Social Sciences in Zagreb has departments that can aid in the processes of synthesis and recognition, namely Information sciences, Linguistics, Phonetics, Croatian language.

One of the problems in Croatian language is accurate word accent, distinguishing orthographically identical, but differently sounding words. Proper word accentuation is therefore essential for accurate speech synthesis of Croatian language. Different word accentuations may be completely wrong or "dialectally coloured". Accurate word accent recognition in speech recognition systems is needed for semantics in case of later machine translation. Different approaches to speech synthesis may ease or complicate production of correctly sounding synthesized speech. Speech synthesis based on concatenation needs all variants of word accents present in the language repertoire for synthesis, but makes synthesized speech more natural. Formant synthesis, on the other side, produces whatever accentuation needed, but it is harder to describe all acoustically relevant sound segments for speech synthesis.



# Intercultural Differences in Vocal Communication of Emotions: An Experimental Comparison Between Chinese and Italian Young Adults

Fabrizia Mantovani, Luigi Anolli, Lei Wang, Alessandro De Toni

CESCOM\_Centre for Studies in Communication Sciences University of Milan-Bicocca P.za Ateneo Nuovo,1 20123 Milan Italy, [mantovani@unimib.it](mailto:mantovani@unimib.it);

## ABSTRACT:

The poster presents an experimental study comparing the vocal communication of emotions between Chinese and Italian young adults. Main goal of the study is to investigate whether:

- (a) the vocal expression of eight emotions (joy, sadness, anger, fear, contempt, pride, guilt, shame) is characterized by distinguished patterns of paralinguistic features;
- (b) the vocal patterns of emotional expressions - produced in reaction to comparable eliciting situations - differ between members of two different cultures (Chinese and Italian);
- (c) specific cultural configurations exist in vocal expression of emotion.

Forty-eight undergraduates (29 Chinese and 19 Italian) were asked to read aloud short stories inducing different emotions via scenario approach. The short stories had been prepared and validated in a preliminary phase: in each text a standard sentence was included in order to carry out subsequently acoustic comparisons. Acoustic analyses were carried out through the Computerized Speech Lab (CSL) 4300B. Different acoustic parameters were considered referring to time (total duration, partial duration, duration of pauses, speech rate and articulation rate), fundamental frequency (mean, standard deviation, range, minimum and maximum of F0) and intensity (mean, standard deviation, range, minimum and maximum).

Results from statistical analyses confirmed the importance of vocal production in generating distinctive emotional patterns, as well as the presence of both similarities and differences between the vocal emotional patterns of Chinese versus Italian participants. The theoretical implications of these findings will be discussed.



## Non-verbal Interaction and Ambient Entertainment

Anton Nijholt, Dennis Reidsma, and others

University of Twente Department of Computer Science PO Box 217 7500 AE Enschede, The Netherlands, [anijholt@cs.utwente.nl](mailto:anijholt@cs.utwente.nl);

## Abstract:

In future Ambient Intelligence (AmI) environments we assume intelligence embedded in the environment, its objects (furniture, mobile robots) and its virtual, sometimes visualized agents (virtual humans). These environments support the human inhabitants or visitors of these environments in their activities and interactions by perceiving them through their sensors (proximity sensors, cameras, microphones, etc.). Support can be reactive, but also and more importantly, pro-active, anticipating the needs of the inhabitants and visitors.

Health, recreation, sports and playing games are among these needs. Sensors in these environments can detect and interpret bodily activity and can give multimedia feedback to invite, stimulate, guide and advise on bodily activity. Rather than aiming at improving user task efficiency, in the environments we investigate the aim is to improve physical and mental health (well-being) through exercise and through play. Exercises can be done in order to improve fitness, to prevent certain injuries (e.g., RSI), or to recover from an accident (e.g., physiotherapy exercises). Other exercises may aim at improving certain capabilities related to a profession (ballet, etc.), some kind of recreation (juggling, etc.), or sports (fencing, etc.). Fun, just fun, achieved from interaction (e.g. dancing or physical gaming) can be another aim of such environments.

In this presentation we look at our research on bodily and gestural interaction with environments equipped with some simple sensors (cameras, microphones, dance pads), some application-dependent intelligence (allowing reactive and pro-active activity), and an embodied virtual agent employed in the display of reactive and pro-active activity. Dance, music, and associated movements in human and virtual agents are the main modalities that are used in our environmental installations.



## Towards an all-inclusive cross-media relations framework

Katerina Pastra

**Abstract:**

While there is a growing demand for developing Intelligent Multimedia Interfaces and Systems, one still strives to find a descriptive framework of how different media and modalities interact with one another. The significance of the latter becomes evident, when one attempts to build multimedia systems or intelligent agents, where multimedia content integration decisions are to be made. In this paper, we identify two important parameters in developing such a framework: the use of multiple and clearly stated criteria for defining interaction relations across media and the integration of findings from the analyses of the interaction of as many different media-pairs as possible. In correlating our own corpus-based work on image-language interaction with existing work on image-language and gesture-language interaction, we identify three such criteria and corresponding interaction relations. We further suggest a way of validating the applicability and expressiveness of these interaction relations, which involves a set of simple metrics for computing them in a multimedia corpus. Therefore, we lay the bases for a descriptive framework that will be closer to an ``all-modalities'' and an ``all-perspectives'' inclusive one.



## "Unseen gestures" and the Mind of the Speaker: An analysis of co-verbal gestures in map-task activities

Nicla Rossini

Dipartimento di Linguistica Teorica e Applicata Università degli Studi di Pavia, [tattvamasi@libero.it](mailto:tattvamasi@libero.it);

**Abstract:**

The analysis of co-verbal gestures in map-task activities is particularly interesting for several reasons: on the one hand, the speaker is engaged in a collaborative task with an interlocutor; on the other hand, the task itself is designed in order to place a cognitive demand on both the speaker and the receiver, who are not visible to one another. The cognitive effort in question implies the activation of different capabilities, such as self-orientation in space, planning (which can also be considered a self-orientation task concerning the capability of organising successful communicative strategies for the solution of a given problem), and communication in "unnatural" conditions.

The co-verbal gestures performed during such a task are quantitatively and qualitatively different from those performed in normal conditions, and can provide information about the Mind of the Speaker (Poggi & Magno Caldognetto, 1997). In particular, the recursive pattern of some metaphors (McNeill, 1992 and following) can be interpreted as a reliable index of the communicative strategy adopted by the speaker: recurrent metaphors indicating the adoption of a plan, its abandonment, or its confirmation will be shown and analysed. Moreover, cases of gestures indicating the opposition between Given and New (Halliday, 1985), and other basic psycholinguistic phenomena centred on collaborative speech acts, such as awaiting feedback, frustration, wrong-footing, etc., will be discussed and compared with the co-verbal gesticulation of subjects intent on a face-to-face interaction.



## On the analysis of fundamental frequency control characteristics of nonverbal utterances and its application to communicative prosody generation

Ke Li, Yoko Greenberg and Yoshinori Sagisaka

Waseda univ. GITI 29-7 building 1-3-10 Nishi-Waseda Shinjuku-ku Tokyo 169-0051 Japan,  
[yoshinori.sagiska@atr.jp](mailto:yoshinori.sagiska@atr.jp);

**Abstract:**

Aiming at communicative speech generation, control characteristics of nonverbal utterances were analyzed. From the analyses using FO generation model, utterance specific control characteristics were observed. Their prosodic characteristics are linked to the multi-dimensional vectors expressing listener's subjective impression. A quantitative prosody control scheme is newly proposed to test the validity of FO generation and their effectiveness is conformed by perceptual evaluation tests.



## Face recognition using sparse meshes: a promising approach

Samokhval Vladimir

**Abstract:**

The sparse meshes are considered as the suitable tool for construction of the classifier in recognition problems of human faces. Their use for face modeling is based on a number of remarkable properties of such data presentation and additional opportunities to increase a level of authentic recognition. In particular, representation of area of interest in the form of a 2,5-dimensional mesh potentially allows to receive missing foreshortenings of the image of human face, that essentially increases recognition rate as it is shown with use of PCA and discriminant analysis methods. For successful work of PCA method is necessary to receive the frontal image of the face, and it is possible to rotate a mesh in depth on a certain angle. At the synthesis of discriminant filters their functioning is directly connected with the volume of training sample. In this case mesh rotations enable to receive additional views of facial image and to expand training set. Besides, the degree of mesh sparseness in itself is the parameter, that influences both on classification results, and on the volume of calculations, and finally on speed and system functioning. In this research we consider some aspects of the performance of PCA and discriminant analysis methods for which input data are 2,5-dimensional models of the face in the form of sparse meshes, and also we establish a degree of sparseness of these meshes for optimum performance of recognition system.



## Verbal and nonverbal resources in constructing the topical flow in early interaction in picture book environment

Sari Karjalainen

Department of Speech Sciences, Siltavuorenpenger 20 A/F, P. O. Box 9, 00014 University of Helsinki, Finland , [sari.karjalainen@helsinki.fi](mailto:sari.karjalainen@helsinki.fi)

**Abstract:**

The gestures, particularly pointings, will be analyzed in the process of topical co-operation between adult and child at preverbal stage and, especially, how these resources are used in making the shifts within the topic when looking at picture books. The method for the study is qualitative and data driven CA (conversation analysis). The data base is composed of videotaped naturalistic picture book conversations between child (at the age from 1 to 2 years) and adult. Video data from 6 Finnish families is transcribed and analyzed and the events are presented with different text-based transcriptions, and also a computer-aided visualization of these annotations to supplement the analysis. The micro-analysis is focused on the sequential organization of the participants' verbal and nonverbal action (pointing and other gestures, gaze, vocalizations and adult's speech) and, especially, on the sequences where the topic is extended from the referent in the picture book to the noticeable or non-noticeable referents outside the book. The child's acts, for example, the pointing at the window, get different meanings in different sequential contexts. The topical sequences of different kind will be presented focusing on how both the verbal and nonverbal resources used in referring to picture referents and related referents outside the picture book reveal the use of already existing shared knowledge or constructing the shared knowledge between the participants.



## Visual Search, Baggage Screening, and the Assessment of Mental Workload through the Analysis of Eye-Movements

Michela Terenzi & Francesco Di Nocera

Cognitive Ergonomics Laboratory, Department of Psychology, University of Rome "La Sapienza"

**Abstract:**

In light of the events of September 11, 2001, many efforts have been made to support security officers in identifying potential threats. Most of them are technological aids used by airports' personnel when performing security operations such as baggage screening. However, automation support is known to alleviate some tasks and, at the same time, to create new forms of workload. For this reason, also Human Factors / Ergonomics researchers addressed the vast range of technical challenges that, because of these acts of terrorism, now face society (Hanckock & Hart, 2002). The most important issues in this field, is the analysis of the operators' mental workload, which is a key factor in determining human error. Therefore, it is crucial to find viable strategies for minimizing the cognitive load, for optimizing work schedule, and for managing automation support by workload-matched procedures.

Indeed, one approach for improving performance could be to support the limited human information processing capabilities through the use of adaptive aids, triggered by variations in human physiology and behavior. Previous

studies (Di Nocera et al, 2006a; 2006b) showed a relation between the distribution of eye fixations and workload, providing a real-time measure of the operator's load. In the present study, a typical visual search task was used. Subjects were requested to find a target among a set of distractors. Eye movements were recorder during the task. Results showed sensitivity of the proposed index to variations in mental workload, thus confirming the utility of fixations patterns as triggers for adaptive automation.



## The socio-cultural differences and the personal traits in Ukraine's scientific life

Oksana Udovik and Oleg Udovik

National University "Kyiv Mohyla Academy" and National Institute for Strategic Studies  
Kyiv, Ukraine, E-mail: [xenna\\_2003@ukr.net](mailto:xenna_2003@ukr.net) and [oleg\\_udovik@hotmail.com](mailto:oleg_udovik@hotmail.com)

### Abstract:

Ukraine suffers from an identity crisis that is inhibiting its scientific, as well as its economic and political, development. The 47 million inhabitants of the former Soviet republic are deeply divided between pro-European and pro-Russian factions. The celebrated 'orange revolution' of November 2004 did less to bridge this divide than is commonly thought.

The nation's research system broadly reflects this wider societal divide. On the one hand, there are many young, well-educated and highly motivated researchers and a network of increasingly independent universities. On the other, there's the National Academy of Sciences of Ukraine, a leviathan of militant senility that retains just enough power to control critical aspects of Ukraine's scientific life.

The academy employs 47,000 permanent staff in a network of largely unproductive research establishments. Given the advanced age of its senior management, time alone will eventually resolve the issue. But that won't happen soon enough for those young Ukrainians currently in search of a productive scientific career.

Integrating the Ukraine into the Framework research programme of the European Union (EU) would allow this generation far greater interaction with its peers abroad. The European Commission supports the idea, which could also help open the way to future EU membership for Ukraine. But the leadership of the academy, deeply rooted in Soviet traditions, seems to be thwarting such integration through a mixture of contrariness and lack of interest.

A high-level EU-Ukrainian steering committee on scientific cooperation, for example, was established on paper four years ago but has yet to actually meet. When it does, the academy's leaders are expected to obstruct collaborative steps that might bring an infusion of foreign influences into the country — including respect for the value of independent peer review.

Ukrainian science has potential in several spheres, including materials sciences, radioastronomy, theoretical physics and agricultural research. The nation badly needs to focus its scarce resources in those areas where its scientists can compete, and dispose of some of its anachronistic scientific heritage. That will require a rigorous external evaluation of the performance of hundreds of the academy's institutes.

The government needs to identify these reforms as a priority and then act with determination to overcome the academy's likely resistance to them. The oligarchy that has controlled Ukrainian science since Soviet times may then lose out. But the nation's economic potential and its prospects for integration into the EU, as well as science itself, can only benefit. Reform of Ukraine's archaic research system is needed sooner rather than later.



## Recognizing the effects of voluntary facial activations using heart rate patterns

Toni Vanhala and Veikko Surakka

Research Group for Emotions, Sociality, and Computing Tampere Unit for Human-Computer Interaction,  
Dept. of Computer Sciences, FI-33014 University of Tampere, Finland, Email: [Toni.Vanhala@cs.uta.fi](mailto:Toni.Vanhala@cs.uta.fi)

### Abstract:

Continuously measured physiological signals have the potential to act as non-invasive, real time indicators of human psycho-physiological phenomena. Recently, several non-intrusive, wireless, and discrete measurement devices have

been developed. For these reasons, there has been growing interest for using physiological signals for estimating emotions and other psychological processes during human-computer interaction, as well as for person identification [e.g. 1]. Due to the interaction of the human physiological and psychological systems there are several unique challenges for analyzing these signals. In the current work, we present the first steps towards constructing an online system that automatically identifies heart rate responses and estimates subjective experiences during voluntary facial activations. The preliminary results of our study showed that voluntarily produced facial expressions had an effect on subjective emotional experiences and physiological processes. Further, our results suggest that heart rate responses to facial activations can be detected in order to develop face detection systems for more accurate, online person-identification and emotion recognition.

1. Poulos, M. Rangoussi, M., Chrissikopoulos, V., Evangelou, A. (1999) Parametric person identification from the EEG using computational geometry. In Proceedings of ICECS '99, 1005-1008.



## On the analysis of disfluencies in large spontaneous speech corpora: the case of autonomous fillers

Ioana Vasilescu

Limsi-Cnrs Spoken Language Processing Group Bat.508 Bp 133 F-91403 Orsay Cedex France,  
[ioana@limsi.fr](mailto:ioana@limsi.fr)

### Abstract:

The hesitation or "edition" phenomena, such as filled pauses, silent pauses, word lengthening etc. are widely encountered in world's languages. They are to be distinguished from the lexical level. Consequently, they have been for decades considered as "speech disfluencies", i.e. articulatory events without a role in building the verbal message. Recently, the research of cognitivists such as Clark and Fox Tree, brought into light a new decoding of the presence of those phenomena in speech [1]. The authors focused more particularly on the autonomous fillers in English ("uh", "um"), which are defined as long and stable vocalic segments, potentially inserted at any moment within spontaneous speech. According to the authors, those items play a role in communication, i.e. "to announce the initiation of what is expected to be a [...] delay in speaking".

My work focuses on the analysis and modeling of the speech disfluencies in the framework of automatic speech processing in a multilingual context. In this purpose I analyzed fillers from a multilingual corpus of broadcast news in Arabic, Mandarin Chinese, French, German, Italian, European Portuguese, American English and Latin American Spanish. I addressed so far the question of the specificity of acoustic fillers models: generic across languages or language-dependent. I will present some acoustic and perceptual findings supporting the second hypothesis. I will also mention the effects of number of external factors which influence the acoustic and prosodic patterns of fillers in different types of speech corpora (prepared, conversational etc.). Among those factors language, gender, speaking style and language proficiency engender significant variation needing to be taken into account in order to accurately model the phenomenon.

1. H.H., Fox Tree J.E., Clark, Using uh and um in spontaneous speaking, *Cognition* 84, 73-111, 2002.



## Prosodic Cues for Automatic Phrase Boundary Detection in ASR

Klara VICSI - Gyorgy SZASZAK

Budapest University for Technology and Economics, Dept. for Telecommunications and  
Mediainformatics,  
Budapest, Hungary.

### ABSTRACT:

This article presents a cross-lingual study for Hungarian and Finnish about the segmentation of continuous speech on word and phrasal level based on prosodic features. A word level segmentation has been developed which can indicate the word boundaries with acceptable accuracy for both languages. The ultimate aim is to increase the robustness of Automatic Speech Recognizers (ASR) by detection of word and phrase boundaries, and thus significantly decrease the searching space during the decoding process, very time-consuming in case of agglutinative languages, like Hungarian and Finnish. They are however fixed stressed languages, so by stress detection, word beginnings can be marked with reliable accuracy. Algorithms based on data-driven (HMM) approach were developed and evaluated. The best results were obtained by time series of fundamental frequency and energy together. Syllable length was found to be much less effective, hence was discarded. By use of supra-segmental features, word boundaries can be marked with high correctness ratio, if we not going to find all of

them. The method we evaluated is easily adaptable to other fixed-stress languages. To investigate this we adapted our data-driven method to the Finnish language and obtained similar results.



## Experiments in Assessing the Validity and Reliability of Item N-gram Distributions in Texts as Proxy for Textual Fingerprints

Carl Vogel

Computational Linguistics Group, Trinity College, University of Dublin, Dublin 2, Ireland.

### Abstract:

My research in this area is associated with text rather than speech or other visual cues, and is driven by hypotheses within forensic linguistics that there are valid and reliable methods which can be used for authorship attribution tasks. Behind these hypotheses is the claim that individuals essentially unconsciously fingerprint themselves in the texts that they produce. There are many proposals in the literature and includes analysis of consciously manipulated author style, and other aspects of text that are extremely difficult for an author to manipulate. Orthography is one such aspect of texts.

While a great deal of deliberation may go into selection of lexical items from open-class categories, somewhat less reflection tends to be involved in closed class categories, and therefore it is interesting to explore distributions of words in closed class categories used by an author across texts and genre, and in comparison with other authors as a complement to any analysis of lexical richness or quiriness. However, orthographic analysis crosses both categories, and is basically driven by the fact that while one can choose one's words, one does not generally choose how words are spelled.

Along these lines I and my research group have been experimenting with character n-gram analyses of texts and using similarity of character n-gram distributions to guide the assessment of similarity among texts. The research involves comparative analysis, rating the character approach with various values of n with other methods of tokenizing texts -- e.g. word n-grams, n-grams of part of speech tags, accounting for stop-words, etc. Some of the experiments have been on closed systems of texts in which authorship is actually known, others involve partly open systems of texts in which the claim of single authorship is disputed (e.g. whether one author is responsible for the entire Shakespeare corpus, including the apocrypha), and still more open ended explorations in sentiment analysis -- using essentially bag-of-character analysis to estimate similarity among non-governmental political party positions on the basis of election manifestos.

The research relates to the scope of the meeting in that the communication of information about authorship is assumed to be unintended by the author (although it is acknowledged that some authors do directly manipulate orthography for effect --- lipograms provide a relevant example), yet which can potentially be used to identify the author. I intended to present and obtain feedback on the experiments reported, as well as directions of ongoing research which include tracking language change in individuals over time (an issue perhaps of interest in research on aging and early detection of neuro-degenerative disorders which impinge on language production).

A couple of relevant papers:

1. Van Gijssel, Sofie and Carl Vogel (2003) Inducing a Cline from Corpora of Political Manifestos, International Symposium on Information and Communication Technologies, edited by Markus Aleksy, et al., pp 304 - 310.
2. O'Brien, Cormac and Carl Vogel (2003) Spam Filters: Bayes vs. Chi-Squared; Letters vs. Words, International Symposium on Information and Communication Technologies, edited by Markus Aleksy, et al., pp 298 - 303.



## Empty pauses detection in a noisy speech conditions

Vojtich Stejskal<sup>a</sup>, Zdenek Smékal<sup>a</sup>, Anna Esposito<sup>b</sup>

<sup>a</sup>Dep. of Telecommunications, Brno University of Technology, Purkynova 118, 612 00, Brno CZ,  
[stejskal@kn.vutbr.cz](mailto:stejskal@kn.vutbr.cz);

<sup>b</sup>Dep. of Psychology, Second University of Naples and IIASS, Italy

### Abstract:

Nowadays, the most important role in a process of speech recognition plays successful pause detection. There is need of robust detection algorithm, if we consider that most of speech recordings are taken under very adverse conditions. This paper presents a comparison of several algorithms for empty pauses detection on a spontaneous

speech records gained in noisy environments. Input signal is transformed into log spectral energy and divided into specific frequency bands. Each band is smoothed and tracked by dynamically adjusted threshold based on pause (noise) energy estimation. Then the post processing edges correction follows. All proposed algorithms are capable to process a real time input.



## Images-Signes and Cognitive Scenes to Anchor Comprehension in Language Learning: An Experiment on the Relationship between Verbal and Non Verbal Communication through Italian Film.

Rosa Volpe

148 rue Saint Honoré, 75001 Paris, France, [rvolpe@univ-orleans.fr](mailto:rvolpe@univ-orleans.fr);

### Abstract:

This study has as a central focus a foreign language classroom environment that makes daily use of film discourse in the target language in order to provide « in-context » and « situational training » as well as « anchored » target language input. More specifically this study explores (1) whether first-semester, first-year college students of Italian can and will understand film narrative (2) if and how, the comprehension of the film narrative will affect, if any, their written production. The first experimental probe consists in a comprehension task. The results suggest that the Anchored Learning Group performed better than the Basic-Skills Learning Group in the comprehension of the two film segments. The second experiment probe consists in a production task and shows that compared to the Basic-Skills Learning Group, the written production of the Anchored Learning Group reflects far better the structure of the narrative discourse they have been exposed to. In the occasion of this production task, students were required to write their essay using at least 10 verbs they were familiar with. The performance of the Anchored Group was different from the performance of the Basic-Skills Group both in quality and quantity. The anchored Group wrote more correct verbs, resulting in the production of more correct sentences, and the structure of their discourse was closer to the structure of the discourse of film narrative.



## Language and Communication. An Eco-Anthropological Point of View

GALATCHI Liviu-Daniel

Ovidius University of Constanta, Str. Dezrobirii, nr. 114, bl. IS7, sc. A, et. 4, apt. 16,  
RO 900241, Constanta - 4, Romania, [galatchi@univ-ovidius.ro](mailto:galatchi@univ-ovidius.ro) or [liviugalatchi@yahoo.com](mailto:liviugalatchi@yahoo.com)

### Abstract:

Although wild primates have only call systems, chimps and gorillas can understand and manipulate non-verbal symbols based on language. Primates emit calls only in the presence of particular environmental stimuli. Calls cannot be combined when different stimuli are present simultaneously. At some point in human evolution, our ancestors became capable of displaced speech. Other contrasts between language and call systems include productivity and cultural transmission. Over time, our ancestral call systems developed into true language. Call systems grew too complicated for genetic transmission and began to rely on learning. Language is the main system humans use to communicate, although we also use nonverbal communications, gestures, and body stances and movements.

No language includes all the sounds that the human vocal apparatus can make.

There are culturally distinctive as well as universal relationships between language and mental processes. The lexicons and grammars of particular languages can lead speakers to perceive and think in particular ways. Speakers of different languages categorize their experience differently. However, language does not tightly restrict thought, because cultural changes can produce changes in thought and in language.

People vary their speech on different occasions, shifting styles, dialects, and languages. As linguistic systems, all languages and dialects, are equally complex, rule-governed, and effective for communication. However, speech is used, is evaluated, and changes in the context of political, economic and social forces. The linguistic traits of a low-status group are negatively evaluated (often by the members of the group) not because of their linguistic features but because they are associated with and symbolize low social status. One dialect, supported by the dominant institutions of the state, exercises symbolic domination over the others.

Cultural similarities and differences often correlate with linguistic ones. Linguistic clues can suggest past contacts between cultures. Related languages descend from an original protolanguage. Relationships between languages don't necessarily mean that there are biological ties between their speakers, because people can learn new languages.



## Face Recognition Experiments on AR Database

Marco Grassi<sup>1</sup>, Marcos Faundez<sup>2</sup>

<sup>1</sup>Ingegneria Elettronica dell'Università degli Studi di Ancona, [margra75@hotmail.com](mailto:margra75@hotmail.com)

<sup>2</sup>Escuela Universitaria Politecnica de Mataro, Spain

### Abstract:

Biometric recognition and authentication based on face recognitions can actually be used in many real-time applications such as: surveillance, security systems, access control, and much more. For these purposes the system has to grant a fast computation speed but also robustness to illumination and facial expressions variations. The main objective of this paper is to implement a face recognition system using the DCT (Discrete Cosine Transform) method for characteristics extraction. Applying the DCT to the image results possible to concentrate the information reducing the dimensionality of the problem. For the classification has been used nearest neighbour classifiers using MAD (Mean Absolute Difference) and MSE (Mean Square Error), that grant a very fast computation, and a RBF (Radial Basis Function) Neural Network, that presents a faster training than a classic Neural Network. Simulation results, over the AR face Database, show that the proposed system have very good performances with very fast computation speed, high recognition rate and good robustness.



## Overcomplete Blind Separation of Speech Sources in the Post Non Linear case through Extended Gaussianization

S.Squartini, S.Cecchi, E.Moretti, F.Piazza  
A3Lab-DEIT Università Politecnica delle Marche  
Via Breccie Bianche 31, 60131, Ancona, Italy

**Abstract** - This work deals with the blind separation problem in presence of more sources than sensors and Post-Nonlinear (PNL) mixing. The addressed method is made of three separate steps: compensation of nonlinearity, mixing matrix recovery and final unknown source estimation. It has been recently proposed and successfully evaluated in the case of synthetic mixtures of real world data (like speech signals). Here, the Extended Gaussianization approach is employed to perform the first step instead of the common Gaussianization one in order to reduce the approximation error on the linearized mixture pdfs. Computer simulations allowed to achieve a significant improvement of separation performances over the previous approach.



## The relationships between gestures and prosody: A preliminary investigation on Italian

A. Esposito<sup>1</sup>, D. Esposito<sup>1</sup>, M. Refice<sup>2</sup>, M. Savino<sup>3</sup>, S. Shattuck-Hufnagel<sup>4</sup>

<sup>1</sup>Department of Psychology, Second University of Naples (SUN), Italy

<sup>2</sup>Department of Elettrotecnica, Politecnico di Bari, Italy

<sup>3</sup>Department of Psychology, Università di Bari, Italy

<sup>4</sup>MIT, Research Laboratory of Electronics, Cambridge, MA, USA

### Abstract.

This work investigates on the relationships between gestures and prosody, exploiting a class of gestural movements named HITS and defined by Yasinnik, Renwick, Shattuck-Hufnagel (2004) as: "An abrupt stop or pause in movement, which breaks the flow of the gesture during which it occurs" Our analysis show that, as in American English, also in Italian, these gestural entities are correlated with high level prosodic units.

[1] Y. Yasinnik, M. Renwick, S. Shattuck-Hufnagel: The Timing of Speech-Accompanying Gestures with Respect to Prosody. Proceedings of the International Conference From Sound to Sense, C97-C102, MIT, Cambridge, June 10-13, 2004.