De Nederlandse Vereniging voor Fonetische Wetenschappen nodigt u uit voor een bijeenkomst in Tilburg met als thema:

MULTIMODALE COMMUNICATIE

Tijd: **vrijdag 20 juni, 13.00 - 16.30u**

Plaats: zaal AZ8, gebouw A
        Universiteit van Tilburg,
        Warandelaan 2
        5037 AB Tilburg

Programma:

13.00 - 13.30: Ledenvergadering NVFW
13.30 - 14.00: Recalibration of auditory speech by lip reading
            Jean Vroomen, Tilburg University
14.00 - 14.30: Multimodale en perceptieve "user interfaces"
            Jacques Terken, Technische Universiteit Eindhoven
14.30 - 15.00: Human Factors issues in multi-modal interaction in
            complex design tasks
            Stephan Rossignol, University of Nijmegen (in English)
15.00 - 15.30: Pauze
15.30 - 16.00: More about brows
            Emiel Krahmer, Tilburg University
16.00 - 16.30: Audiovisual cues to uncertainty
            Marc Swerts, Tilburg University

(alle presentaties, behalve die van Stephan Rossignol, zijn in het Nederlands)
Zie ook: http://www.fon.hum.uva.nl/FonetischeVereniging/

Routebeschrijving:
Zie http://www.uvt.nl/universiteit/wegwijzer.html
Trein

Reizigers die met de stoptrein vanaf de richting Eindhoven, Breda of 's-Hertogenbosch naar de UvT komen, kunnen uitstappen op station Tilburg-West. De universiteit, Theologische Faculteit Tilburg en het Sportcentrum UvT liggen op loopafstand (5 minuten) van het station. Reizigers die met de sneltrein of intercity reizen, kunnen uitstappen op Tilburg Centraal Station. Vandaar kan men per stoptrein, bus (lijnen 46, 47, of 48) of taxi (ongeveer 12 euro) verder reizen naar de universiteit, Theologische Faculteit Tilburg of Sportcentrum UvT.

Auto

Vanuit Breda, Eindhoven, en 's Hertogenbosch (A58 / E312)

Neem de afslag Goirle Tilburg Turnhout (afrit 11). Bij de verkeerslichten op de T-kruising richting Tilburg aanhouden. Rijdend richting Tilburg linker rijbaan aanhouden. Bij de verkeerslichten links voorsorteren. Volg vervolgens de ANWB-borden 'universiteit'.

Vanuit Waalwijk, Kaatsheuvel en Loon op Zand

U nadert Tilburg vanuit noordelijke richting. Rechtdoor blijven rijden, richting centrum. Ook bij de rotonde rechtdoor. Vervolgens de ANWB-borden 'universiteit' blijven volgen.

RECALIBRATION OF AUDITORY SPEECH BY LIP READING

Jean Vroomen, Psychonomy Faculty of Social and Behavioural Sciences, Tilburg University

The kind of after-effects, indicative of crossmodal recalibration that are observed after exposure to spatially incongruent inputs from different sensory modalities have not been demonstrated so far for identity incongruence. We show that exposure to incongruent audiovisual speech (producing the well-known McGurk effect) can recalibrate auditory speech identification. Exposure to an ambiguous sound, intermediate between /aba/ and /ada/, dubbed onto vision of a face articulating either /aba/ (or /ada/), increased the proportion of /aba/ (or /ada/) responses during subsequent sound identification trials. In contrast, fewer /aba/ (or /ada/) responses occurred when a congruent non-ambiguous sound was dubbed onto the face articulating /aba/ (or /ada/), revealing selective speech adaptation. When submitted to separate forced-choice identification trials, the bimodal stimulus pairs producing these contrasting effects were identically categorized, which makes a role of post-perceptual factors in the generation of these effects unlikely.

MULTIMODALE EN PERCEPTIEVE "USER INTERFACES"

Jacques Terken, Industrial Design, Technische Universiteit Eindhoven

Door ontwikkelingen op het gebied van computer- en taal- en spraaktechnologie is de toepasbaarheid van spraak in de interface aanzienlijk toegenomen. Tegelijkertijd leiden ontwikkelingen in de informatie- en communicatietechnologie (ICT) ertoe dat de computer (of afgeleiden daarvan) in alle facetten van het dagelijks leven doordringt. Op het gebied van de aansturing van de systemen door de gebruiker is er behoefte ontstaan aan meer natuurlijke vormen van interactie ("weg van de desktop"). De combinatie van spraak met andere interactietechnologieen speelt daarbij een prominente rol. In de presentatie zal ik ingaan op factoren die aan de orde zijn bij de toepassing van spraak in de interface. Een en ander zal concreet worden gemaakt aan de hand van voorbeelden uit de literatuur en uit eigen onderzoek in het kader van twee samenwerkingsprojecten met andere universiteiten: MATIS (Multimodal Access to Transaction and Information Services) en CRIMI (Creating Robustness in Multimodal Interaction).

HUMAN FACTORS ISSUES IN MULTI-MODAL INTERACTION IN COMPLEX DESIGN TASKS

Stephan Rossignol, Nijmegen Institute for Cognition and Information (NICI), University of Nijmegen,

For human-computer interaction, the naturalness of the dialog is a paramount factor for the acceptance of dialog systems by a large public. It is generally believed that the use of multi-modal communication channels facilitates for natural interactions between a human subject and artificial collocutor. In the European-IST COMIC project (2002-2005, http://www.hcrc.ed.ac.uk/comic/), various scientific and pragmatic aspects of multi-modality are being studied through human factors research and the integration of research findings in a real-life interactive system. The two input modalities considered here are pen gestures (including pointing, handwriting, labeling and sketching) and (continuous) speech. The task of the system is to guide the unexperienced user through a wide range of design options in the highly complex professional domain of bathroom design. The University of Nijmegen is one of the COMIC-participants and focuses on the improvement of the pen and speech recognizers, the use of these recognizers in a real-time application, and the study of human-factors related to the multi-modal human-system interaction. In COMIC, human-factors experiments are carried out to obtain more insight in the complex behavior of humans in interactive design specification tasks. The task of the subjects was to specify shape, size and location of details in bathrooms by using pen input on a Wacom Cintiq 15X LCD tablet and speech through a close-talk microphone. The results of the experiments are presented in this talk. It is shown that there are behavioral differences between subjects, and that the subjects' behavior is critically dependent on the exact task description. The experimental results yield more insight in the classes of speech dialog acts and pen dialog acts, and in the impact of mode combinations in interaction strategies. These experiments advance the knowledge of cognitive principles underlying the natural use of parallel modalities, and result in concrete guidelines for the design and the development of multi-modal dialog systems.

MORE ABOUT BROWS

Emiel Krahmer, Communication & Cognition, Faculty of Arts, Tilburg University (Joint work with Marc Swerts)

In this talk we present a series of experiments which try to assess the usefulness of eyebrow movements for the perception of focus in two languages, namely Dutch (a Germanic language) and Italian (a Romance language). The first group of experiments is based on an analysis-by-synthesis method, where claims from the literature are directly implemented in an existing talking head (i.e., a combination of speech with

computer graphics). Three aspects are investigated in these experiments: (1) what are the preferences of Dutch/Italian subjects concerning the placement of eyebrow movements, (2) do eyebrow movements influence the perceived prominence of words and (3) to what extend are eyebrow movements functional for the way Dutch/Italian subjects process incoming utterances. The advantage of the analysis-by-synthesis method is that results can immediately be implemented in synthetic characters. Nevertheless, the approach is arguably incomplete. If we also want to make claims about human communication, the analysis-by-synthesis technique should be supplemented with data-oriented approaches. To make this point more concrete we present results of an ongoing study based on analysis-by-observation, in which subjects are asked to utter nonsense words (/mamama/ and /gagaga/) with the focus on one syllable. It turns out that some subjects indeed use eyebrow movements to signal prominence, although various other audio-visual cues are relevant as well.

AUDIOVISUAL CUES TO UNCERTAINTY

Marc Swerts, Communication & Cognition, Faculty of Arts, Tilburg University (Joint work with Emiel Krahmer)

Uncertainty is an inherent element of human-machine communication. Uncertainty is often not made explicit in the system's output, nor have there been many efforts to detect uncertainty in users' reactions. Our work concerned with the expression of (degree of) uncertainty in spoken interactions. It focuses on the production and perception of auditory and visual cues, in isolation and in combination. The ultimate goal is to implement possible audiovisual cues to uncertainty in a synthetic talking head of an embodied conversational agent (ECA). We conjecture that a user's acceptance of incorrect system output is higher if the system made it clear in its self- presentation that it is not sure about the answer. Our approach builds on previous studies on the so-called Feeling-of-Knowing (FOK) (e.g. Smith and Clark, 1993), be it that we also include possible visual cues to FOK. Following earlier procedures, our study A consists of three parts: first, in an individually performed test, subjects are instructed to answer 40 factual questions (e.g. what is the capital of Switzerland?, who wrote Hamlet?); questions are posed by the experimentor whom the subjects cannot see, and the responses by the subject are videotaped (front view of head). After this test, the same sequence of questions is again presented to them, but now they have to express on a 7-point scale how sure they are that they would recognize the correct answer if they would have to find it in a multiple-choice test. The final test consists of this actual multiple-choice test. All utterances from the first test of Study A (800 in total) were transcribed orthographically and manually labelled regarding a number of auditive and visual features by four independent transcribers on the basis of an explicit labelling protocol, which included various double-checks. On average, subjects knew the answer to 30 of the 40 questions. When they did not know or could not remember the answer, they sometimes made a guess or gave a non-answer. It appears that their FOK scores not only correlate with their performance in the third multiple-choice test, but also with particular features of the utterances of the first test: lower scores, in line with previous results, correlate with long delay, the occurrence of filled pauses and question intonation. In addition, it appears that speakers tend to use more words, when they have a lower FOK. Regarding the visual cues, low FOK is reflected in averted gaze, more head movements, eyebrow frowning, and overall more body movement. Also, a puzzling look appears to correlate with low FOK, whereas a self-congratulatory expression is more typical for high FOK answers. The goal of our Study B is to explore whether observers of the speakers' answers of Study A are able to guess these speakers' FOK scores. In particular, we are interested in whether a bimodal presentation of stimuli leads to better FOK predictions than the unimodal components in isolation. To test this, we are currently preparing a perception test in which a subset of the utterances of Study A will be presented to subjects. These are instructed to guess what the speaker's FOK was when s/he gave an answer (cfr. Brennan and Williams, 1995). Stimuli will be presented in three conditions: image only, sound only, both image and sound. From the original 800 responses, we select 60 utterances, with an equal amount of answers and non- answers, and an even distribution of high and low FOK scores. While we expect that we get the best performance for bimodal stimuli, it remains an interesting empirical question whether the auditory or the visual features from the unimodal stimuli will turn out to be more informative for FOK predictions. The experiment is currently taking place. Results of this additional test will also be discussed in my talk.