

FORESIGHT COGNITIVE SYSTEMS INITIATIVE

SPEECHNET

Interdisciplinary Research Towards a Unified Model of Speech Pattern Processing

Although spoken language processing is a broad yet well defined concept for both human-human and human-machine interaction, the fact that it can be described and simulated at so many different levels of abstraction - from acoustic and visual signals through to cognitive and neural activity - means that over many years it has attracted the attention of a large and disparate range of scientific disciplines. This situation has led to a variety of alternative and partial explanations as to how such a faculty might evolve and operate.

However, whether it be a human brain or an electronic machine, for the purposes of scientific investigation speech could arguably be viewed as a *single* 'process' that mediates the expression and communication of ideas, concepts and information between different physical entities through a regularity of behaviour - 'patterning' - that is to some extent shared (and hence 'understood') by each participant. It is this patterning that is the central object of study in all areas of spoken language research but, hitherto, there is no *unified* model of 'speech pattern processing' that is capable of spanning across the different disciplines involved.

The Foresight Cognitive Systems Initiative provides a unique opportunity to compare and contrast some of the main results drawn from a wide variety of speech-related disciplines - specifically between the engineering-based models used in speech technology and the psycholinguistic/neuro-cognitive models used in the speech sciences - with a view to (a) significantly advancing the level of knowledge in each area and (b) deriving a unified model of speech pattern processing that can sustain a more cohesive approach to future scientific progress in the general area.

In order to initiate the cross-fertilisation and integration of these diverse research results, what is needed is the creation of an active research network - 'SPEECHNET' - that links the main UK research centres by the provision of funds for the necessary research and scientific exchanges (including the establishment of working collaborative relationships with key research facilities outside the UK). The aim of the Network would be to create a world-leading research community that is tasked with cross-comparing the key paradigms across the breadth of speech-related disciplines at both the theoretical and practical levels with a view to advancing our knowledge in this key area of human and automatic behaviour.

The key areas of interdisciplinary research are (a) whole-system approaches from the perspective of 'communicative interface agents' as well as (b) core components covering 'speech recognition/perception' (from audio-visual sensors to linguistic and paralinguistic interpretation), 'speech synthesis/production' (from linguistic and paralinguistic expression to audio-visual realisation) and 'spoken language interaction' (including the modelling, planning, execution and orchestration of speech-based cooperative behaviour). In each case the aim would be to identify the 'information' that each system brings to bear as a constraint on the overall process (in terms of priors and data exposure), the 'representation/encoding' mechanisms for the constraints (in terms of modelling paradigms) and the computation that is performed to achieve the necessary constraint satisfaction (algorithms).

Two further research areas that are of particular importance to such a network relate to (a) the acquisition/learning of spoken language skills in both humans and machines and (b) the derivation of *predictive* computational models based on parameterised characterisations and measurements of suitably instrumented systems.

Prof. Roger K. Moore