UNDERSTANDING HUMAN COMMUNICATION - A DATA-CENTRIC APPROACH

Steve Renals
University of Edinburgh

We wish to understand how humans are able to communicate robustly. It may be the case that the development of such an understanding will translate to advances in the engineering of richer, less fragile spoken language interfaces.

We can start to frame the problem through some general principals of human communication:
- multimodality: people receive rich sensory data through their ears and eyes
- attention and selectivity: human perception is able to filter inputs in different ways, and can focus on specific aspects of a communication scene
- interaction: communication is not a solitary thing, it is embedded in real world interactions (which can also serve to combine sensory inputs between people)

It requires an interdisciplinary effort to make a serious attack on the challenge of understanding human communication. A major problem of mounting a large multidisciplinary effort is that the researchers come from different intellectual traditions, have different theoretical frameworks, etc. What commonalities can we build on which a cognitive systems project can be built? Two possibilities are:
- the data
- the search for statistical regularities in the data.

The development of a set of facilities for multilevel data collection aimed at exposing the processes of human communication could provide a natural coming together point for such a project. The sort of data that could be collected might include fMRI/MEG data, articulatory dynamics (eg collected by EMG), multichannel audio data from microphone arrays and binaural mannikins, video data from stereoscopic imaging and multicamera data collection, and data relating to how people interact via technology, such as PC intraction data, instrumented pens, smart whiteboards, etc. The collection of such data requires investment in various resources (eg MEG machines, anechoic chambers), and poses several major challenges in software organization, distribution, data analysis, annotation and the collection of multiple modalities from one communication task. Distributed processing, storage and annotation schemes are probably necessary to make full use of such a data collection.

It is unrealistic to assume that it is possible to collect the full spectrum of communicative data modalities from a single task (eg, it is probably difficult to fully participate in a meeting when stationed in an fMRI machine). However it should be possible to develop experimental protocols and scenarios which allow the collection of significantly multimodal data for a variety of communication tasks.

The key to dealing with these large quantities of multimodal data lies in machine learning. In particular some generic machine learning problems are appropriate for processing multimodal and multisensory human communication data:

- Data fission: apportioning responsibility across a single data stream to multiple sources or generating processes
- Multistream models: modelling the hidden structure of human communication and its manifestation across several streams
- Feature selection: selecting the most informative features for a task within possibly huge potential feature spaces

Finally there is a real social benefit of working on the same data - even if it is hard to collaborate on modelling or analysis due to very different backgrounds, collaborations can begin by performing different analyses, but on the same data. Good examples of this are the Edinburgh map task (would have been even better if speech recognition people had been involved) and the current work on meetings (eg the EC AMI project) which involves organizational psychologists, discourse modellers, speech recognition researchers and vision scientists.