Sensible machines

We've probably all uttered the phrase 'stupid machine' when technology has let us down but what if we could communicate our frustrations to the machine so it could learn and we could truly interact with it. If machines could process multi-sensory information then this might be possible. **Matthew Casey** of the Department of Computing at the University of Surrey believes that this day is not far off

At a noisy party our senses are remarkable – even above the music and the clink of glasses, we can understand what people around us are saying by combining what we hear with lip reading. This unique ability of humans and animals to combine senses is something that machines are just beginning to emulate. Imagine an ATM that could ask the customer what service they required, listen to and action their response, and all above the noise of a busy road. But what about fraud I hear you say? Well, the ATM would of course accurately identify the customer from their face, voice and other biometrics first.

Far fetched? Not really. This is a simple example of how we can improve our interactions with machines. Some of what I describe here is established technology, such as will be used with the forthcoming UK identity card scheme. However, these technologies are not advanced enough to replace the use of a keyboard, mouse and screen to 'interact' with a computer. To move beyond this and to have a machine 'perceive' using a number of senses simultaneously will require new approaches. I believe by taking inspiration from biology on how we sense however, that within 15 years we could have machines that can truly sense and therefore communicate with us much better.

We use our senses to perceive our environment, from identifying potential threats to helping us find food. We also use sight, sound, touch and even smell to communicate. A large portion of our brains is dedicated to processing sensory information and our understanding of this has traditionally been that each sense is processed individually before being combined. Mimicking this, we have developed sufficiently robust computerised voice and handwriting recognition techniques (to name but two). Voice recognition, for example, has become popular for phone-based customer services, avoiding the need to listen to a long list of menu items. Yet, whilst machines may be able to recognise a series of spoken words, those words must be spoken clearly and without background noise – a long way from a noisy party.

So, how do we help machines to sense? First, I believe we have to understand how we process sensory information better, not just in a single modality, but in combination. Studies of animals and humans have established that we combine sensory information early during processing. Take for example the part of our brains called the superior colliculus, thought to help orient our head and eyes to something we've either seen or heard, such as to a person's lips as they are talking. This structure clearly shows that integration is important; indeed visual information can even influence what we think we hear, such as at a noisy party. So if human senses are processed in combination, why don't we do this in machines?

Limited work has been done on this – for example Kismet at the MIT Computer Science and Artificial Intelligence Laboratory, in the US is a 'sociable machine' that has been developed to study face-to-face interaction between robots and humans. Interestingly, in other areas, combining information is an established paradigm used to improve, say, pattern classification for anti-fraud measures. An EU project we undertook at Surrey combined the automatic analysis of news with numerical data to help quantify 'market sentiment'. Despite these programmes, we still haven't managed to develop human-level sensory systems, let alone combine them. Whilst we have adaptive neural techniques that have been used to computationally model aspects of vision and other senses, so far these are very limited. However, biological modelling is starting to show some promise – for example, guiding visual attention is a technique increasingly being applied to CCTV images.

I believe that for machines to sense, we must take inspiration from biology, start with a simple model of integrated sensory processing and then embed this in a system that can bring together video, audio and other information. Academic work on this is already underway, such as discussed at the recent workshop on biologically inspired information fusion held at Surrey. Small beginnings perhaps, but our hope is that this will lead to machines that can truly sense their environment. Once they can sense, communication should be easier. You'll know when we've got somewhere when you next go to an ATM and it asks you "how can I help?"