AISB Quarterly

No. 137   July, 2013

All the latest Society information is available
at the newly renovated AISB website:
http://aisb.org.uk

The cover of this issue is designed by visualising the behaviour of agents aiming to trace the Q of the AISB Quarterly magazine. The agents are powered by two nature inspired swarm intelligence algorithms and a biological mechanism.

The swarm intelligence algorithms are Particle Swarm Optimisation simulating the behaviour of birds flocking, and Stochastic Diffusion Search, which mimics the recruitment behaviour of one species of ants, Leptothorax acervorum.

The biological mechanism is inspired by the behaviour of blood flow and cells in blood vessels. This particular design deployed the concept in Outward Eutrophic Remodelling, where the concept of high and low blood pressure and its impact on the vessel calibre is utilised. The details of the hybridisation can be found in the following book chapter:

Book review: The Voice in the Machine

by *Roger K. Moore* (Univ. Sheffield)

I've known Roberto for more years than probably either of us cares to remember—right back to the early 1980s when we were both young post-doctoral researchers starting out on our respective careers in speech technology—Roberto at the Italian Centro Studi e Laboratori Telecomunicazioni (CSELT) in Turin, and myself at the UK Government Speech Research Unit (SRU) in Malvern. These were exciting times, with several of the most significant advances in our field happening all around us. The first International collaborations (such as the NATO Research Study Group—RSG10) were paving the way for sharing standardised speech corpora, National and International funding agencies were starting to invest heavily in directed speech technology programmes, and the results of years of laboratory work were beginning to establish a tentative foothold in the commercial marketplace. Roberto was (and, I should say, continues to be) very much part of the action, so it was with eager anticipation that I awaited my copy of The Voice in the Machine. Touted as an examination of "six decades of work in science and technology to develop computers that can interact with humans using speech", I was keen to hear about his perspective on all the developments—positive and negative—that many of us 'old hands' have lived through (and participated in).

I was not disappointed. Although the catchy title might mistakenly imply that the main topic is speech synthesis, the subtitle—Building Computers that Understand Speech—makes it clear that the emphasis is in fact very much on automatic speech recognition and understanding with, unsurprisingly, a significant component dedicated to Roberto's own speciality—spoken language dialogue systems. Unusually for a book of this type, Roberto approaches the central topic from the high-level perspective of language and thought, rather than the low-level characteristics of speech signals. This novel structure (at least, novel in a book about speech) provides a compelling framework within which to emphasise the importance of meaning and context in spoken language interaction; a theme that pervades the rest of the book, particularly through a running humorous example of speech communication involving Roberto's dog—Ares—and a neighbour's stolen sausage!

The central theme of the book (and the title of the third Chapter) is the fundamental dichotomy between the GOFAI 'expert' approach and the use of 'brute force' statistical methods to solve the challenges posed by attempting to create automated speech technology systems. By taking a historical perspective, Roberto manages to convey, not just the algorithms that are now deployed in contemporary systems, but also the philosophical and practical arguments that culminated in the use of stochastic state machines and optimal graph-based search

as the core basis for modelling the patterned behaviour in speech, language and dialogue. The book is packed with wonderful sub-headings that will have special meanings for us older folk such as "Finding your Way in Time Warp Land", "The Hidden Models of Markov", "Digit Wars", "Mad Cows" and, a particularly good one, "Invisible Speech"—a nod towards the phonetic writing system known as 'Visible Speech' developed by Alexander Melville Bell (Alexander Graham Bell's father) in the 1860s. Indeed Roberto's crucial insight—"the fate of any good technology is to become invisible"—provides a very satisfactory set of opening and closing parentheses on the entire endeavour.

Of particular importance is the central role that has been played by competitive benchmarking and assessment using shared standardised data. Roberto spells out the long and hard road that had to be taken in order to turn a rather ad-hoc engineering-oriented 'suck-it-and-see' field into the rigorous mathematical discipline that it is today. Of course, the imposition of a strict evaluation framework has its own particular downsides and, as Roberto correctly points out, researchers dependent on funding from Government agencies tended to become risk-averse and true innovations are now rare events in the continuous drive for guaranteed incremental improvements. And the safest and most effective approach to improving performance turns out to be. . . more training data. Indeed, to repeat one of the most famous quotes in the speech technology field—"There's no data like more

data"—a phrase used by Roberto as another of his Chapter headings.

Unsurprisingly, given the career trajectory of the author, the book is very much focused on telephone-based applications, with an inevitable emphasis on developments in the United States, particularly those which took place at Bell Labs/AT&T. However, the style is not overly egocentric or self-promotional, and most of the major speech labs around the world receive a satisfactory acknowledgement. Nevertheless, there are some surprising omissions; some of the significant early speech recognisers are missing, for example there's no mention of GEC-Marconi's SR128 connected word recogniser (the first to fly in an aircraft), the French VecSys systems, NEC's ground-breaking DP100 system, or indeed of Loquendo—the very successful Italian speech company that spun out from Roberto's old lab in Turin. There is also no reference to Dennis Klatt's famously critical review of the ARPA SUR programme, or that HARPY was actually the outcome of a one-man PhD project and, perhaps the biggest omission of all, there is no reference to the Cambridge HTK hidden-Markov model toolkit which could easily be argued to have played the largest role in transforming worldwide speech technology R&D over the past twenty years. Likewise, the section on speech synthesis completely fails to mention recent developments in hidden Markov model based speech generation despite there being a section headed "Brute Force and Statistics"!

Rather less surprising is that Roberto falls into the standard traps when getting down into details about speech it-

self. He presents the usual engineer's description of a 'phoneme' as a "distinct sound" corresponding to a particular configuration of the vocal tract. If only this were true—building speech technology systems would be so much easier! Almost all engineers fail to grasp the significance of the difference between phonetics and phonology, so it's a shame that Roberto missed the opportunity to educate the naïve reader into the importance of the phoneme as an abstract concept deriving from linguistic contrast. Likewise, 'coarticulation' is defined as "the continuous transition between consecutive phonemes", thereby missing the most important property of coarticulation that it can operate over many sound segments (such as nasalisation or lip-rounding occurring several phones in advance of a nasal or lip-rounded sound). Roberto also repeats the popular misconception that "vowel sounds in normal speech reach a point of stability in the middle. . . that doesn't depend on the preceding or following sound" which, I'm sorry to say, reflects the wishful thinking of the optimistic engineer more than the reality of speech.

Overall, *The Voice in the Machine* is a masterclass in communication. It is a highly personal account of the development of a key area of contemporary AI from an individual who has been involved almost from the beginning. It is written from the perspective of someone who has direct experience of both the rigour of the research laboratory and the urgency of the commercial environment. It is an unusual book in as much as the various algorithms and approaches are placed in their historical context rather than simply listed as elements in the speech technology toolbox. For older members of the community (myself included), it will be appreciated as a nostalgic trip down memory lane that reminds us that we are part of a relatively modern field of science and engineering that has by now acquired a reasonable amount of interesting history and given birth to a viable industry. For people who are new to the area, or who are interested in one of the main success stories in data-driven approaches to machine learning, it provides a definitive account of exactly how we came to be where we are today. If you've always wondered how (and why) speech technology people use hidden Markov models, but you've been put off by the more mathematical texts on offer, then this is the book for you. It should be essential reading for all students studying speech and language technology, but it also contains important lessons (both positive and negative) for other areas of AI where rigorous benchmarking has yet to be established.

Finally, this book is not the end of the speech technology story. It was clearly written just before the 2011 surprise release of Apple's speech-based personal assistant for the iPhone—Siri. Luckily, Roberto managed to slip in an epilogue in which he rightly points out that there's nothing dramatically different in Siri—it's just an example of something happening "in the right place at the right time". He leaves open the question as to whether Siri is the beginning or the end of developments in speech technology, but we can probably guess his thoughts from his remark

that "there is a huge distance between what we expect talking machines to do and what they actually can do". In fact no real attempt is made in this book to speculate on where we might go from here. So I look forward to a follow-up to *The Voice in the Machine* in which Roberto draws on his extensive knowledge and experience to take us on another gentle ramble, this time into possible futures for speech-based human machine interaction.

Roger K. Moore
Department of Computer Science
University of Sheffield

Back matter

Editors − aisbq@aisb.org.uk
Dr David Peebles (*Univ. Huddersfield*)
Dr Etienne B Roesch (*Univ. Reading*)

Advertising and Administration
Dr Katerina Koutsantoni (AISB Executive Office)
Institute of Psychiatry, King's College London
Addictions Sciences Building (B3.06)
4 Windsor Walk, Denmark Hill
SE5 8AF, London, United Kingdom
Tel: +44 (0)20 7848 0191, Fax: +44 (0)20 7848 0126

AISB Patron
Prof John Barnden (*Univ. Birmingham*)

AISB Fellows
Prof Harry Barrow (*Schlumberger*), Prof Margaret Boden (*Univ. Sussex*), Prof Mike Brady (*Univ. Oxford*), Prof Alan Bundy (*Univ. Edinburgh*), Prof Tony Cohn (*Univ. Leeds*), Prof Luciano Floridi (*Univ. Hertfordshire*), Prof John Fox (*Cancer Research UK*), Prof Jim Howe (*Univ. Edinburgh*), Prof Nick Jennings (*Univ. Southampton*), Prof Aaron Sloman (*Univ. Birmingham*), Prof Mark Steedman (*Univ. Edinburgh*), Prof Austin Tate (*Univ. Edinburgh*), Prof. Mike Wooldridge (*Univ. Liverpool*), Dr Richard Young (*Univ. College London*)

AISB Steering Committee
Chair: Prof Mark Bishop (*Goldsmiths Univ. London*), Vice-Chair: Prof John Barnden (*Univ. Birmingham*), Secretary: Dr Rodger Kibble (*Goldsmiths Univ. London*), Treasurer: Dr Bertie Müller (*Univ. Glamorgan*), Webmasters: Dr Mohammad Majid al Rifaie (*Goldsmiths Univ. London*) & Dr Kent McClymont (*Univ. Exeter*), Membership: Dr Dimitar Kazakov (*Univ. York*), Publications: Dr Ed Keedwell (*Univ. Exeter*), Public Relations: Dr Nir Oren (*Univ. Aberdeen*), Dr Colin Johnson (*Univ. Kent*), Publicity: Dr Floriana Grasso (*Univ. Liverpool*), School Liaison: Dr Yasemin J Erden (*St Mary's Univ. College*) Prof Slawomir Nasuto (*Univ. Reading*), Science Officer; Dr Manfred Kerber (*Univ. Birmingham*), AISB 2012 co-chair, and the AISBQ editors.

AISB Quarterly – No. 137, July, 2013

Book review: The Voice in the Machine – by *Roger K. Moore* (Univ. Sheffield) ---