Institution: Queen's University Belfast

Unit of Assessment: 11

Title of case study: Applications of Novel Speech and Audio-Visual Processing Research

1. Summary of the impact (indicative maximum 100 words)

Research in robust speech enhancement and audio-visual processing has led to impact on a range of different fronts:

(i)   Collaboration with **CSR**, a leading $1 billion consumer electronics company, has shaped its R&D research agenda in speech enhancement, has inspired ideas for new product improvements, and has helped establish Belfast as an audio research centre of excellence within the company.

(ii)   Our technology has changed the strategic R&D direction of a company delivering **healthcare monitoring systems**, with potential for multi-million pound savings in NHS budgets.

(iii)   Audio-visual speech processing research has led to a proof-of-concept **biometric system**, *Liopa*: a novel, robust and convenient person authentication and verification technology exploiting lip and facial movements (www.liopa.co.uk). A start-up company is in an advanced stage of being established to commercialise this product. The product and commercialisation strategy was awarded First Prize in the 2013 NISP Connect £25K entrepreneurship competition in the Digital Media and Software category. The first commercial partner for Liopa has been engaged.

(iv)   A **system-on-chip** implementation of a version of our speech recognition engine, which was developed through an EPSRC project, was awarded first prize in the High Technology Award in the 2010 NISP £25K Awards competition, and contributed to the founding of a spin-out company, Analytics Engines (www.analyticsengines.com).

2. Underpinning research (indicative maximum 500 words)

The underpinning research of the Speech Processing group spans approximately 2000-2013. The current key researchers are: Professor M Ji, Dr D Stewart (Lecturer), and Professor D Crookes. Ji, Stewart and Crookes were in academic posts at QUB throughout all of this period. The group's research in speech processing has grown to include multi-modal processing, with a particularly novel approach to using lip movements. Though the initial aim was to enhance speech recognition, our current system uses just visual for lip movement analysis for biometric identification. The relevant research includes the following projects. This research was undertaken in, and facilitated by, QUB's research flagship Institute of Electronics Communications and Information Technology (ECIT, www.ecit.qub.ac.uk), based in the Northern Ireland Science Park.

1.   *Novel methods for robust speech and speaker recognition in noisy environments.* This early research developed new statistical methods for modelling fast-varying or unexpected noise assuming minimum information about the noise[1]. The resulting methods, including the Probabilistic Union Model, Missing Feature Theory, and Universal Compensation, were tested using both international standard databases and bespoke test data for mobile applications, and were found to improve upon existing state of the art methods.

2.   *Corpus-based speech separation.* This more recent project considered two challenging problems in signal processing research: (i) restoring clear speech from noisy recordings, and (ii) separating simultaneous crosstalk voices.

      We have tackled the extremely challenging conditions when the recordings are from a single channel, when the noise is fast-varying and unpredictable, and when the crosstalk voices are arbitrary in speaker, language, vocabulary and structure. This project has developed a fundamentally different and effective solution to the above problems. For separating complex mixtures of speech on noise[2], and speech and speech[3], the new method (called CLOSE) has reached a level of accuracy previously unattainable with existing techniques. Our Interspeech 2010 paper was selected as the Best Paper in speech enhancement. The research led to a follow-on EPSRC Knowledge Transfer Secondment (KTS) scheme with CSR, for technology transfer.

3.   *Audio-Visual Biometrics.* A development with particularly exciting commercial potential has

seen the imminent establishment of a start-up company to exploit novel research in lip-based biometric identification. This research originally started out as a multi-modal extension of our speech processing research, by using video of lip and facial movements to improve speech recognition, and for speaker verification [4]. The research discovered that certain features of lip movements are particularly potent for speaker identification. Using facial movements to supplement speech recognition has also resulted in a unique 'liveness' test for secure biometric access. The Liopa project received £50K TSB funding, with our proposal being ranked first out of sixty from within the UK for the "Preventing fraud in mCommerce" funding competition. We have since been invited to apply for further Phase-2 funding. Another novel method for multimodal person recognition has been developed for when there is limited training data [5].

4. *A system-on-chip recognition engine for real-time, large-vocabulary mobile speech*. This EPSRC-funded Shares project implemented our noise-robust speech recognition algorithm on a hardware (SoC) platform for embedded speech recognition applications. The system was one of the first of its kind developed in the world [6]. Prof M Ji was the Computer Science co-investigator.

3. References to the research (indicative maximum of six references)

[1] M Ji, T J Hazen, J R Glass, and D A Reynolds, "Robust speaker recognition in noisy conditions," IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 5, pp. 1711-1723, 2007. [110 Google Scholar citations]
This early research was funded through EPSRC grants: GR/M93734/01: "The probabilistic union model: a model for partially and randomly corrupted speech", 2000-2003; and GR/S63236/01: "Robust speaker recognition in noisy conditions - a feasibility study", 2004-2005.
[2] M Ji, R Srinivasan and D Crookes, "A corpus-based approach to speech enhancement from nonstationary noise," IEEE Transactions on Audio, Speech and Language Processing, vol. 19, pp. 822-836, 2011. [22 Google Scholar citations]
[3] M Ji, R Srinivasan, D Crookes and A Jafari, "Close - a data-driven approach for unconstrained speech separation," IEEE Transactions on Audio, Speech and Language Processing, Vol.21, No.7, July 2013. pp.1355-1368.
This research was funded through EPSRC grant EP/G001960/1: "Corpus-Based Speech Separation", 2008-2011 (EP/G001960/1) to M Ji and D Crookes, and a one-year secondment funded through the EPSRC Knowledge Transfer Secondment (KTS) scheme (2011-2012).
[4] R Seymour, D Stewart and M Ji, "Comparison of Image Transform-Based Features for Visual Speech Recognition in Clean and Corrupted Videos", EURASIP Journal on Image and Video Processing. 2008, p. 1-9. This research was funded through EPSRC grant EP/E028640/1 "ISIS - An Integrated Sensor Information System for Crime Prevention", £1.39m, 2007-2010.
[5] N McLaughlin, M Ji and D Crookes, "Robust Multimodal Person Identification with Limited Training Data", IEEE Transactions on Human-Machine Systems, vol.3, no.2, March 2013. pp. 214-224. DOI: 10.1109/TSMCC.2012.2227959. This research was funded by Intel (2008-2011).
[6] Jianhua Lu, Ming Ji, and Roger Woods, "Adapting noisy speech models — Extended uncertainty decoding", Proc. ICASSP 2010. March 2010. pp. 4322 – 4325. This research was funded through EPSRC grant EP/D048605/1 "SHARES - System-on-chip Heterogeneous Architecture Recognition Engine for Speech", to R Woods, M Ji et al, £503K, 2006-2009.

4. Details of the impact (indicative maximum 750 words)

(i) **Impact on CSR**.
Background: APT was a Queen's University spin-out company, set up in 1989 to exploit innovative research in digital audio technology. The company achieved particular success with its aptX audio compression solutions for professional audio and consumer applications, which is now found in around 85% of all Bluetooth headsets on the world market as well as in mobile phones made by Samsung, Nokia and HTC. In 2010, APT was bought by Cambridge Silicon Radio (CSR), a pioneering designer and developer of silicon and software for the consumer electronics market. It is a $1 billion British company employing nearly 3,000 people around the globe.
    Following the acquisition, ECIT brought the QUB speech enhancement research to the company's attention and after discussions and demonstrations, CSR entered into initial

collaboration with the research group. They gave access to their own test audio data, and the results of the enhancement led to an ESPRC-funded secondment of the research fellow to CSR under the KTS scheme (2011-2012). The Research Fellow on secondment is now employed by CSR. The results were presented to CSR's CEO prior to CSR completing their final report to ESPRC on the secondment.

Although the work is recent, in its final report to EPSRC, CSR rated the significance of this work to their future performance as the maximum **5 out 5**. They said that the research and collaboration has already:

- "*brought about change in the nature of its business by identifying and defining new and enhanced R&D areas within CSR which are expected to result in enhanced products and services;*
- *contributed to company strategy by providing valuable information to shape and prioritise elements of the research agenda related to speech enhancement;*
- *assisted in the R&D of next-generation speech processing techniques for residential and automotive environments;*
- *transferred additional technical knowledge to the company on particular speech-processing techniques.*"

The Director of Advanced Audio Research at CSR has further put it on record that:

"*The collaboration opened our eyes to new possibilities, and has inspired ideas to improve CSR products. The collaboration helped establish Belfast as an audio research centre of excellence within CSR*"

A Patent Application has been filed to support exploitation of the CLOSE method for speech source separation and enhancement. The UK provisional application was filed on 26 August 2011, and the International Application No. PCT/EP2012/066549 "Method and Apparatus for Acoustic Source Separation" was filed on 24th August 2012.

In a separate example of the impact of our speech enhancement research, our methods have been incorporated into an award-winning speech recognition system by NTT, Japan. NTT used our Corpus-based Speech Separation method in its speech recognition entry to the International Competition for Machine Listening in Multisource Environments (CHiME'2011), in which it took 1st place. (See reference in source 3 below).

(ii) **Impact on Health Monitoring**. Vitalograph Ireland (Ennis, Ireland), a world-leading provider of cardio-respiratory diagnostic devices, is using our software for robust speech and audio analysis in a system for automated monitoring of how inhalers are being used in clinical trials. The UK NHS spends £4bn per annum on inhalers, and research has shown that up to 90% of the drugs they dispense is wasted largely because of improper inhaler use. Vitalograph has developed hardware which incorporates an embedded microphone and audio recorder in each inhaler. At the start of the project, the audio data was analysed manually to identify and interpret the various audio events such as inhaling, holding breath and exhaling. Funded by InterTradeIreland (2012-2013), the UoA has developed automated audio analysis software that enables robust automatic analysis of a person's inhaler use. At the start of the project, the audio data was analysed manually to identify the various audio events such as inhaling, holding breath and exhaling. This analysis is used in clinical trials, and to train patients in better inhaler use, but is very time-consuming and a major draw-back. Although this research is very recent, the accuracy of the results of our automated system has led the company to refocus its strategy with a view to delivering an automated analysis and training solution. They have employed the researcher on the project. The system is currently a prototype but when released and deployed, given the huge cost of the drug, the potential for savings could run into millions of pounds. Vitalograph's Director of Operations & R&D has said:

"*The leading-edge research of the Queen's team has caused Vitalograph to change its strategic direction and product development plans for its clinical trials programme for inhalers. The research will find its way in to a Vitalograph product in the not-too-distant future and this element will be the clear differentiator that sets the product apart from its competitors. Our planned product will be more adventurous because of the success of the automatic audio analysis. We have already employed a person to assist with transfer of this research into this product.*

Vitalograph has collaborated with several universities in UK and Ireland over the last ten years, and our interaction with the team at Queen's was one of the smoothest and most productive."

(iii) **Liopa: a Lip-reading Biometric System**. A start-up company is in the process of being established to commercialise our novel lip-reading based biometric system called Liopa (www.liopa.co.uk). So far the Liopa team has attracted £50K of procurement/product development funding in the form of a Small Business Research Initiative grant from the Technology Strategy Board which has resulted in the following component parts of the Liopa system being successfully developed and ready for user trials:

1.   An Authentication/Verification server
2.   An API for system integrators
3.   An Android SDK and proof-of-concept mobile application

In the recent 25K Awards entrepreneurship competition run by the Northern Ireland Science Park (NISP), Liopa was awarded First Prize in the Digital Media and Software category. Several significant corporations have approached us with interest in incorporating this technology within their projects/products/services and we are actively communicating with these potential partners/customers regarding possible engagements. For example, we were recently invited to Canary Wharf to demonstrate Liopa to Infosys and a large grouping of their key partners which was very positively received.

We have also signed an agreement with our first commercial partner (AirPOS Ltd) to carry out trials of Liopa with a large number of users. AirPOS Ltd creates software for small to medium size retailers selling across single and multiple points of sale (POS). AirPOS were recently announced as the first UK POS company to partner with PayPal here on their new payment device. The company was founded in 2010 and now serves 3,120 customers in over 80 countries. Liopa will be used in two ways: firstly, it will be incorporated into the retail ePOS system to enable biometric employee log-in as an anti-fraud and anti-theft measure; and secondly, it will be used for person authentication and verification in order to grant access to ticketless fans and concert goers at sporting and concert venues.

(iv) **A speech recognition chip**. The system-on-chip speech recognition engine developed by our EPSRC Shares project was awarded the High Technology Award in the 2010 NISP 25K Awards. Two of the researchers on this project (Woods and Fischaber) founded the spin-out company Analytics Engines (www.analyticsengines.com), who specialize in high performance data analytics and accelerated computing. A research student from the project was employed by the US-based Nuance Group, one of the largest world leaders in speech technology products.

5. Sources to corroborate the impact (indicative maximum of 10 references)

(1) The Final Report for the EPSRC KTS secondment (completed by CSR).
    Confirmation of the quoted impact on CSR can be obtained from:
    Director of Advanced Audio Research, CSR.
(2) For details of the Patent: International Application No. PCT/EP2012/066549 "Method and Apparatus for Acoustic Source Separation" filed on 24th August 2012.
(3) The paper on NTT's prize-winning speech recognition system, referencing the QUB work, is:
    Marc Delcroix et al. (NTT), "Speech recognition in the presence of highly non-stationary noise based on spatial, spectral and temporal speech/noise modelling combined with dynamic variance adaptation", CHiME 2011 Workshop on Machine Listening in Multi-source environments, Sept, 2011. Corroboration of the CHiME 2011 results is available at: http://spandh.dcs.shef.ac.uk/projects/chime/PCC/results.html
(4) For corroboration of the impact on the work with Vitalograph on health monitoring:
    Director of Operations & R&D
    Vitalograph Ltd.
(5) For confirmation of the two NISP 25K Awards, on the speech recognition chip in 2010, and the Liopa lip reading-based biometric system in 2013:
    Chief Executive Officer, Northern Ireland Science Park.