Institution: The University of Edinburgh

Unit of Assessment: B11 — Computer Science and Informatics

Title of case study: Speech Graphics Ltd: Audio-driven animation

1. Summary of the impact

Speech Graphics Ltd is a spinout company from the University of Edinburgh, building on research into the animation of talking heads during 2006–2011. Speech Graphics' technology is the first high fidelity lip-sync solution driven by audio. Speech Graphics market a multi-lingual, scalable solution to audio-driven animation that uses acoustic analysis and muscle dynamics to drive the faces of computer game characters accurately matching the words and emotion in the audio. The industry-leading technology developed by Speech Graphics has been used to animate characters in computer games developed by Supermassive games in 2012 and in music videos for artists such as Kanye West in 2013.

This impact case study provides evidence of *economic impacts* of our research because:
  i)   a spin-out company, Speech Graphics Ltd, has been created, established its viability, and gained international recognition;
  ii) the computer games industry and the music video industry have adopted a new technology founded on University of Edinburgh research into a novel technique to synthesize lip motion trajectories using Trajectory Hidden Markov Models; and
  iii) this led to the improvement of the process of cost-effective creation of computer games which can be sold worldwide because their dialogue can be more easily specialised into different human languages with rapid creation of high-quality facial animation replacing a combination of motion capture and manual animation.

2. Underpinning research

Speech Graphics Ltd was founded by Gregor Hofer and Michael Berger, PhD students of Dr Hiroshi Shimodaira (lecturer, 2004–present). The company is based on research carried out by Hofer, Berger, and Shimodaira, in the School of Informatics from 2005–2012 together with colleagues Junichi Yamagishi and Korin Richmond in the Centre for Speech Technology Research an interdisciplinary research centre at The University of Edinburgh.

The underpinning research concerns audio-driven facial animation. Speech animation, or lip synchronization, is a significant research challenge as it is highly interdisciplinary, involving expertise in Speech Technology, Phonetics, and Computer Graphics. The founders of Speech Graphics have conducted basic research in all three areas.

The research of Hofer, Berger, and Shimodaira has three main facets:

1. A novel technique to synthesize lip motion trajectories based on an audio speech signal was developed, based on Trajectory Hidden Markov Models (HMMs). The Trajectory HMMs may be estimated from training data using maximum likelihood estimation, and the trajectory HMM parameter generation algorithm can be used to produce the optimal smooth motion trajectories that drive control points on the lips directly. A perceptual evaluation of this work was carried out with human subjects. (References: [1, 2, 3, 4].)
2. The combination of research on muscle dynamic modeling of speech production by Michael Berger (this was work which was not published in order to protect the value of potentially commercialisable IP) and the use of HMM-based research in modelling of speech and lip motion conducted at the University of Edinburgh.
3. The development of Carnival, an object-oriented environment for integrating speech processing with real-time graphics. Carnival is comprised of modules that can be dynamically loaded and assembled into a mutable animation production system. Carnival takes the output from the speech processing and applies it in real time to a 3D facial model. (References: [5, 6].)

The goal is to automatically animate a 3D facial model by using acoustic and phonetic information. The main processing steps involved are: *acoustic analysis*, where audio is converted into acoustic parameters to find corresponding speech categories; *motion synthesis*, that uses the timing information of the speech categories to produce muscle dynamic parameters; and *adaptation*, which maps the muscle dynamic parameters to be rendered on a particular 3D facial model. Figure 1 below shows the different processing steps required to produce facial animation from audio.

Figure 1: Processing in the Speech Graphics pipeline (reproduced from [6]).
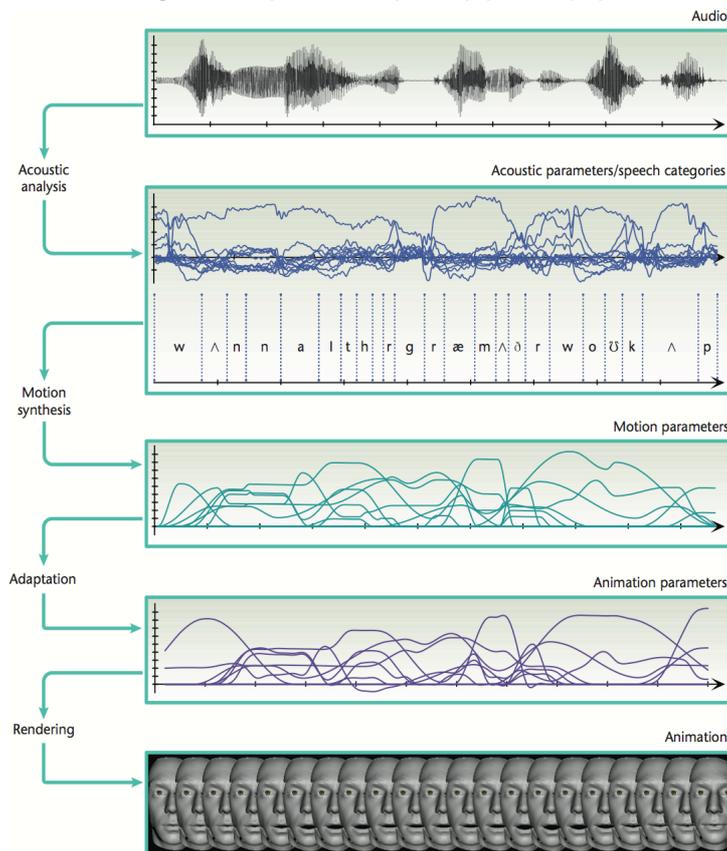
3. References to the research

1. Gregor Hofer and Hiroshi Shimodaira. *Automatic head motion prediction from speech data*. In Proc. Interspeech 2007, Antwerp, Belgium, 2007. *http://www.era.lib.ed.ac.uk/handle/1842/2006*
2. Gregor Hofer, Junichi Yamagishi, and Hiroshi Shimodaira. *Speech-driven lip motion generation with a trajectory HMM*. In Proc. Interspeech 2008, pages 2314-2317, Brisbane, Australia, September 2008. *http://www.era.lib.ed.ac.uk/handle/1842/3883*
3. Michal Dziemianko, Gregor Hofer, and Hiroshi Shimodaira. *HMM-based automatic eye-blink synthesis from speech*. In Proc. Interspeech, pages 1799-1802, Brighton, UK, September 2009. *http://www.cstr.ed.ac.uk/downloads/publications/2009/dziemianko_interspeech2009.pdf*
4. Gregor Hofer and Korin Richmond. *Comparison of HMM and TMDN methods for lip synchronisation*. In Proc. Interspeech, pages 454-457, Makuhari, Japan, September 2010. *http://www.era.lib.ed.ac.uk/handle/1842/4558*
5. Michael Berger, Gregor Hofer, and Hiroshi Shimodaira. 2010. *Carnival: a modular framework for automated facial animation*. In ACM SIGGRAPH 2010 Posters (SIGGRAPH '10). ACM, New York, NY, USA, Article 5, 1 page. *http://doi.acm.org/10.1145/1836845.1836851*
6. Michael A. Berger, Gregor Hofer, and Hiroshi Shimodaira. *Carnival — combining speech technology and computer animation*. IEEE Computer Graphics and Applications, 31:80-89, 2011. *http://dx.doi.org/10.1109/MCG.2011.71*.

Papers 1, 2, 3 and 4 were presented at Interspeech in the year 2007 to 2010. Interspeech is one of the two major annual speech-processing conferences. Paper 5 is a presentation at ACM SIGGRAPH 2010, which was awarded a medal in the ACM Student Research Competition *http://www.siggraph.org/s2010/for_attendees/acm_student_research_competition* Paper 6 is based on this presentation. References [2], [3] and [6] are most indicative of the quality of the underpinning research.

4. Details of the impact

4.1. Formation of the company
Hofer and Berger have commercially exploited the research described above through the formation of a start-up company, Speech Graphics Ltd., formed in 2010 [A]. Speech Graphics provide a service that automatically analyses a speech audio signal, and then automatically moves an animated character's face in synchrony with the audio. The techniques used derive directly from the doctoral research of Hofer and Berger, under the supervision of Shimodaira. The key scientific novelty is the Trajectory HMM approach pioneered for audio animation by Hofer *et al* (paper 2) in combination with research on muscle dynamic modeling of speech production. Finally, the company exploits a novel software framework, developed at the University of Edinburgh, called Carnival (paper 6). Speech Graphics Ltd subsequently extended the Carnival software framework to manage large numbers of files using database software.

Speech Graphics provide these technologies as a service aimed at computer games development companies. Clients provide their facial models and audio assets, and Speech Graphics produce synchronized animation curves. Output is provided in industry standard formats, Maya or 3dsMax. Carnival provides the backbone for efficient production work on several thousands of files.

The company was launched at the Game Developers Conference Expo in San Francisco in March 2012, demonstrating the technology producing high-fidelity lip synchronization in a wide variety of languages.

4.2. Awards and recognition
Speech Graphics Ltd won a John Logie Baird innovation award in 2010, as the "Knowledge Transfer Champion" [B]. Speech Graphics is supported by the High Growth Startup Unit at Scottish Enterprise. The unit grants the company access to a number of support systems and finance-based resources. Criteria for acceptance into the pipeline focus on growth potential and global business outlook, specifically that Scottish Enterprise judge that the company's Intellectual Assets (know-how or IP) will generate £5 million in revenue in five years or will be worth £5 million or more in three years, and that they have the potential for global trade.

Speech Graphics began to attract more media attention and publicity [C]. Speech Graphics won a prize in the national Santander Entrepreneurship Awards in July 2011 [D,E]. They were nominated in the Tools and Technology award category for the TIGA Games Industry Awards presented in November 2012. Their channel on YouTube contains a selection of videos representing their product [F]. Together these videos have more than 113,000 views.

Scottish Development International described the company as offering ``unprecedented quality at a price point that is scalable to today's cinematic, dialogue-rich games", adding that instead of artists spending ``hundreds of hours of lip sync, doing motion capture cleanup or key framing, they can spend time on art and polish'' [G].

4.3. Customers and more details of the impact
Speech Graphics Ltd has been working with Supermassive Games based in England, developing a game for one of the largest multi-national games developers. Computer games now can have thousands of lines of dialogue delivered by game characters. For a game marketed internationally, these lines of dialogue need to be re-animated against alternative deliveries of the dialogue in different languages.

The game that uses the Speech Graphics technology features substantial portions of dialogue written by writers who have previously written for Hollywood movies and US TV. The game was announced at a Gamescom media briefing in Cologne, Germany in 2012, and described as a highly realistic video game with a story-driven adventure with multiple player perspectives. The game features eight characters in an integrated story. Decisions made by the player affect the participation of the characters in later chapters of the game. Speech Graphics worked with Supermassive Games in 2012 to animate the dialogue in the game [H]. Specifics of their work are presently the subject of a non-disclosure agreement.

In addition Speech Graphics has also won its first major US contract in 2013 from a major media company providing facial animation for one of the biggest entertainment franchises, which includes non-human characters like Orcs. The company has also entered into a marketing partnership with Havok, a wholly owned subsidiary of Intel Inc [I].

In July 2013 Speech Graphics were contacted by Def Jam Records to animate Kanye West's face in a music video just days before the deadline for release of the video and the accompanying website. Their industry-leading technology won Speech Graphics the contract. The company was recommended to Def Jam records for the Kanye West video because David Bennett, the facial animation lead on the Avatar movie, told Def Jam records that "the only way we can get this done in this time frame is with Speech Graphics." Speech Graphics completed the work on the video in 36 hours [J].

4.4. Company involvement and community engagement
Speech Graphics Ltd sponsored the Third International Symposium on Facial Analysis and Animation (*http://faa2012.ftw.at*) held in Vienna in September 2012. This meeting brought together researchers and practitioners from both academia and industry interested in visual effects and games, with a particular focus on aspects of facial animation and related analysis.

5. Sources to corroborate the impact

A. Speech Graphics Ltd, company website: *http://www.speech-graphics.com*
B. John Logie Baird award winners, 2010. *http://bit.ly/1cJeXNa*
C. *Develop*, the online information source for global games development sector (monthly readership of over 300,000) profiled Speech Graphics Ltd in 2012: *http://www.develop-online.net/features/1595/Evolving-facial-animation*
D. *Students awarded top award for smooth talking*, Scottish Television News website, July 2011, *http://local.stv.tv/edinburgh/21239-students-awarded-top-award-for-smooth-talking/*
E. *Edinburgh lip-synch spin-out Speech Graphics wins national entrepreneurship award*, July 2011, *http://startupcafe.co.uk/2011/07/21/edinburgh-lip-synch-spin-out-speech-graphics-wins-national-entrepreneurship-award/*
F. The Speech Graphics channel on You Tube *http://www.youtube.com/user/SpeechGraphics*
G. Scottish Development International, *Scottish Games Industry Profiles 2012*, *http://bit.ly/1gpvGH7*
H. *http://www.supermassivegames.com/index.php/about/partner-list*
    Supermassive Games lists Speech Graphics as one of their partners.
I. *http://www.havoksimulation.com/?q=corporate-relationships* Havok, a wholly owned subsidiary of Intel Inc., lists Speech Graphics as one of its partners.
J. The Scottish startup that made animated Kanye rap in his Black Skinhead video, Wired, July 2013. *http://www.wired.co.uk/news/archive/2013-07-25/kanye-west-speech-graphics*

Copies of these web page sources are available from *http://ref2014.inf.ed.ac.uk/impact*