Institution: University of Edinburgh

Unit of Assessment: 11

Title of case study: World-leading audio animation research enables continued growth of spinout company and novel features in computer gaming

Period when the underpinning research was undertaken: 2005 – 2011

Details of staff conducting the underpinning research from the submitting unit:

Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Steve Renals	Professor	2003 – present
Korin Richmond	Reader	2005 – present
Hiroshi Shimodaira	Senior Lecturer	2004 – present
Junichi Yamagishi	Senior Research Fellow	2007 – 2020

Period when the claimed impact occurred: 2013 – 2020

Is this case study continued from a case study submitted in 2014? Yes

1. Summary of the impact

Research at the University of Edinburgh (UoE) into speech-driven animation of the lips, face, and head has resulted in a novel technology that enables computer-generated "talking heads" to be created using only an audio recording. The technology has been commercialised by spinout Speech Graphics into two software product licenses, leading to the company's global leadership in computer gaming speech animation services. During the REF2021 impact period Speech Graphics has generated over [text removed for publication] worth of contracts for work on games including Fortnite, Star Wars and Tomb Raider, and created never-before-seen features which have been lauded by games critics and professional gamers. Speech Graphics has expanded significantly since 2014 and has seen its revenue double year on year from 2018 onwards.

2. Underpinning research

Since 2005, researchers at the Centre for Speech Technology Research (CSTR) at the University of Edinburgh (UoE) have pursued an interdisciplinary programme to deepen knowledge of, and model the relationship between, speech acoustics and the visual movements of the head, face, and speech articulators (lips and tongue). This work has resulted in significant research findings relating to acoustic-articulatory modelling and the speech-driven computer animation of lifelike characters: so-called "talking heads".

Acoustic-articulatory modelling – sometimes called articulatory inversion, since the speech production process is inverted to give the sequence of speech articulator movements that generated the audio – has focused on the development of machine learning models and algorithms that can infer the sequence or trajectory of articulator movements (movements of the lips, lower jaw, and tongue) from recorded audio. This work focused on the development of statistical and neural generative models, in particular the trajectory mixture density network [3.1] and the trajectory hidden Markov model [3.2]. Both of these techniques resulted in significant improvements over previous state-of-the-art methods developed for articulatory inversion, using standard open datasets for experimental evaluation. The work was supported by the EPSRC grants Data-driven articulatory modelling (01/11/2006 –

31/10/2009) and ESPF (Edinburgh Speech Production Facility; 01/04/2007 – 31/08/2010), as well as the Marie Curie Early Stage Research Training Scheme EdSST (Edinburgh Speech Science and Technology; 01/01/2006 – 31/12/2009).

"Talking heads" require the head and facial animation to be synchronized with the speech spoken by the character. This provides a major challenge to animators, since viewers are very sensitive to any inaccuracies in an animated face. Lip synchronisation is clearly crucial, however the requirements for speech animation are broader than this, since speech affects the whole facial surface below the eyes, as well as head motion.

Using the trajectory modelling techniques developed for articulatory inversion, the team developed machine learning approaches that use acoustic signals to predict head motion [3.3] and for lip synchronisation [3.4, 3.5]. This novel interdisciplinary research combined research findings in human non-verbal behaviour with state-of-the-art machine learning approaches to enable head motion and lip synchronisation to be driven purely by an audio recording of the speech. The work was supported by the EU FP7 Network of Excellence SSPNET (Social Signal Processing Network; 01/02/2009 – 31/01/2014).

The research advances were integrated in Carnival, a software framework developed by the UoE team, which combines articulatory modelling and speech-driven head animation with real-time graphics [3.6]. Carnival defines a pipeline that takes speech processing output and applies it in real time to a 3D facial model to render lifelike facial animations from the spoken audio.

The advances in acoustic-articulatory modelling and the application to speech animation resulted in the Carnival software, which became the core intellectual property of Speech Graphics, a spinout company founded by PhD students Gregor Hofer and Michael Berger, who had carried out the research with Shimodaira, Richmond, Renals, and Yamagishi.

3. References to the research

The underpinning research was presented at INTERSPEECH, the leading international speech processing conference, IEEE Signal Processing Letters (the leading letters journal in signal processing, accept rate of about 20%), and ACM SIGGRAPH.

3.1. Richmond, K. (2006). A Trajectory Mixture Density Network for the Acoustic-Articulatory Inversion Mapping. In *INTERSPEECH 2006 - ICSLP Ninth International Conference on Spoken Language Processing* [1790-Mon3WeS.4] International Speech Communication Association.. https://www.isca-speech.org/archive/interspeech_2006/i06_1790.html **(97 citations)**.

3.2. Zhang, L., & Renals, S. (2008). Acoustic-Articulatory Modeling With the Trajectory HMM. *IEEE Signal Processing Letters*, 15, 245-248. https://doi.org/10.1109/LSP.2008.917004 **(88 citations)**.

3.3. Hofer, G., & Shimodaira, H. (2007). Automatic head motion prediction from speech data. In *INTERSPEECH-2007* (pp. 722-725). ISCA. https://www.isca-speech.org/archive/interspeech_2007/i07_0722.html **(38 citations)**.

3.4. Hofer, G., Yamagishi, J., & Shimodaira, H. (2008). Speech-driven lip motion generation with a trajectory HMM. In *INTERSPEECH 2008, 9th Annual Conference of the International Speech Communication Association, Brisbane, Australia, September 22-26, 2008* (pp. 2314-2317). ISCA. https://www.isca-speech.org/archive/interspeech_2008/i08_2314.html **(40 citations)**.

3.5. Hofer, G., & Richmond, K. (2010). Comparison of HMM and TMDN Methods for Lip Synchronisation. In *INTERSPEECH 2010 11th Annual Conference of the International Speech Communication Association* (pp. 454-457). International Speech Communication Association. http://www.isca-speech.org/archive/interspeech_2010/i10_0454.html **(26 citations)**.

3.6. Berger, M. A., Hofer, G., & Shimodaira, H. (2011). Carnival - Combining Speech Technology and Computer Animation. *IEEE Computer Graphics and Applications*, 31(5), 80-89. https://doi.org/10.1109/MCG.2011.71 (Awarded third place in ACM SIGGRAPH Student Research Competition 2010, http://s2010.siggraph.org/for_attendees/acm_student_research_competition.html **(10 citations)**.

Citations based on Google Scholar, 2020-12-03.

Key research grants

- EPSRC: Data-driven articulatory modelling (EP/E027741/1, GBP358,572); ESPF (EP/E01609X/1, GBP757,184)
- European Commission: EdSST (20568, GBP647,359); SSPNET (231287, GBP587,641)

4. Details of the impact

Speech Graphics (SG) commercialised the University of Edinburgh (UoE) research into world-leading computer animation technology, expediting its own growth as a company and positively impacting the creation of video games. The SG innovations have saved labour for games publishers, and generated gaming features which were previously thought impossible. These features have been met with overwhelming enthusiasm by reviewers and gamers.

The company refined the Carnival software (described in 3.6) into two products now sold as licences: SGX 3.0, designed specifically for the games market; and SG COM, an advanced product for generating real-time animation, used so far in virtual reality and gaming [5.1].

During the impact period the company has grown from 4 to 42 staff members, and has won over 100 contracts to a combined value of [text removed for publication], doubling revenue every year since 2018 [5.2, para. 4]. The company confirms that “Thanks to the rapid improvement of our products, based on the Edinburgh research…we were able to move to a subscription-based business model in 2017 only 4 years after first spinning out” [5.2, para. 6]. Such business models are usually reserved for major tech companies.

Speech Graphics’ has rapidly become a powerhouse in the gaming industry, collaborating with some of the world’s largest publishers. In 2019, the industry was worth an estimated USD152,100,000,000 (06-2019) [5.3, para. 3]. Nine of the top 10 games publishers use Speech Graphics technology, including Epic (the company behind Fortnite), Electronic Arts (whose subsidiary Respawn created the Star Wars games), and household names Sony Playstation and Microsoft Xbox [5.2, para. 4]. As further evidence of the company’s success, Speech Graphics won Scottish Tech Startup of the Year in 2019 [5.1]. In 2020, they were one of 67 tech firms representing the UK at the Consumer Electronics Show in Las Vegas, showcasing Britain’s world class technology sector and cutting-edge innovation to buyers on the global stage, as the Department for International Trade put it [5.4, para. 1].

In addition to the success of Speech Graphics' business, UoE research has improved working practices in the gaming industry. The software has streamlined technical processes and saved labour, contributing to the commercial success of games such as Gears of War, Star Wars and Fortnite. Speech Graphics' clients confirm:

> *Their solution saved us a huge amount of work on facial animation and the results were truly outstanding.*
>
> - Centroid Motion Capture [5.1, 5.2, para. 5]

> *Speech Graphics provided efficient, quality results for our automated in-game speech. The Maya tools and plug-in allowed us to easily iterate on the results of the audio analysis onto our custom face rig until we met our visual target, without the need for animators to spend manual effort on individual face animations.*
>
> - The Coalition [5.1, 5.2, para. 5]

These creative innovations have enhanced the realism and emotion of games characters and have gone on to improve the gaming experience for many of the 2,500,000,000 gamers who play worldwide [5.3, para. 3].

Realism in facial animation is a key feature of the gaming experience, often scrutinised by reviewers, and has a negative impact when done poorly. For instance, in 2017 the game Mass Effect Andromeda became "a laughingstock" in the gaming community for its poor facial animation, widely blamed on its reliance on lower-quality automated animation tools (as opposed to manual animation) [5.5]. The significant advance achieved by Speech Graphics has been the development of realistic automated facial animation, making it feasible to produce games with hundreds of thousands of lines of dialogue [5.2, para. 3], translated into many languages [5.1, under SGX].

Both professional gamers and critics have homed in on the realism in games that utilise SG technology. In 2018 Fortnite, the global phenomenon with 350,000,000 users worldwide [5.6], used SG technology to enable live-streaming gamers' avatars to speak their words with emotional expression in real-time while waiting in the game's virtual lobby for play to begin [5.1 under SG COM, 5.7]. This allowed professional gamers who stream their play live on YouTube to generate entertaining content for followers during time that would usually be used for waiting and set-up.

YouTube gamers Imane Anys, (aka Pokimane, 4,300,000 followers) and Cizzorz (4,370,000 followers) were both emphatic in their live streams about the SG COM update, exclaiming to their viewers: "Oh my god! The Fortnite characters talk when you talk!" and "It literally looks like you're talking!" [5.7]

In 2019, SG animated the speech for different characters in the new Star Wars game, Jedi: Fallen Order. The emotional nuance and realism brought to the game was noted by both the studio and by reviewers:

> *The team at Respawn used Speech Graphics to generate thousands of high-quality facial animations for gameplay lines in Star Wars Jedi: Fallen Order…We were immediately impressed by how the whole face of the in-game character would reflect the tone of the line.*
>
> - Respawn (EA subsidiary) [5.1, 5.2, para. 5]

> *It didn't take long for a cutscene to give us a glimpse at the facial animation in Star Wars Jedi: Fallen Order. Lips were synced incredibly well with the dialogue, characters were able to emote realistically, and overall animation was smooth…Saw Gerrera, Forest Whitaker reprising his role from Rogue One: A Star Wars Story, is a spitting image of the actor.*
>
> - GotGame review [5.8, para. 3]

Speech Graphics have enhanced the experience of close to 400,000,000 gamers. In addition to Fortnite, the company's YouTube reel showcases their work on popular games such as Shadow of the Tomb Raider and Call of Duty WWII [5.9]. Combined with Star Wars, these games have generated sales in excess of 27,500,000 copies [5.10]. Speech Graphics continues to expand its products' capabilities outside of gaming. In 2018, the company attracted GBP2,000,000 equity investment to apply its techniques to the development of virtual assistants [5.2 para. 7, 5.11], and announced a private release of the Rapport digital avatar platform in 2020 [5.12].

5. Sources to corroborate the impact

5.1. Speech Graphics. (2020). Speech Graphics Facial Animation Software and lip sync. Retrieved August 19, 2020, from https://www.speech-graphics.com/
5.2. Letter of corroboration from Speech Graphics
5.3. Wijman, T. (2019, June 18). The Global Games Market Will Generate $152.1 Billion in 2019 as the U.S. Overtakes China as the Biggest Market. Retrieved October 23, 2020 from https://newzoo.com/insights/articles/the-global-games-market-will-generate-152-1-billion-in-2019-as-the-u-s-overtakes-china-as-the-biggest-market/
5.4. Department for International Trade. (2020, January 7). 67 tech firms represent UK at world's biggest trade show. Retrieved September 24, 2020, from https://www.gov.uk/government/news/67-tech-firms-represent-uk-at-worlds-biggest-trade-show
5.5. Zacny, R. (2017, March 27). Animators: Awkward 'Andromeda' Animations Are Automation Amok. Retrieved September 21, 2020, from https://www.vice.com/en_us/article/vvkzwa/animators-awkward-andromeda-animations-are-automation-amok
5.6. Aditya. (2020, May 8). Fortnite now has 350 million registered users - Latest Fortnite player count, April 2020. Retrieved September 15, 2020, from https://www.sportskeeda.com/esports/fortnite-now-has-350-million-registered-users-latest-fortnite-player-count-april-2020
5.7. Your_ Highn3ss. (2018, November 4). Pokimane and Cizzorz REACTS TO *FORTNITE* Characters Talking When You Talk FUNNY & EPIC Moments. Retrieved March 23, 2020, from https://www.youtube.com/watch?v=RFIDMHE369k
5.8. Poole, D. (2019, June 8). E3 2019 Preview: 7 Things We Learned From Star Wars Jedi: Fallen Order. Retrieved March 19, 2020, from https://gotgame.com/2019/06/08/e3-2019-preview-7-things-we-learned-from-star-wars-jedi-fallen-order/
5.9. Speech Graphics. (2018, October 9), Speech Graphics gamereel, YouTube. Retrieved August 28, 2020 from https://www.youtube.com/watch?v=cp21VeuLuhQ
5.10. Collection of games' sale statistics (Star Wars Jedi: Faller Order, Shadow of the Tomb Raider, Call of Duty WWII)
5.11. Archangels. (2018, October 10). Speech Graphics Join the Archangels' Portfolio. Retrieved March 23, 2020, from https://archangelsonline.com/speech-graphics-investment
5.12. Hofer, G. (2020, September 9). Introducing Rapport. Retrieved September 24, 2020, from https://rapport.cloud/blog/introducing-rapport/