Institution: University College London	
Unit of Assessment: 11 - Computer Science and Informatics	
Title of case study: Synthesia: cheaper and more accessible presenter-led videos	
Period when the underpinning research was undertaken: 2013 – 2020	

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:
Lourdes de Agapito Vicente	Professor of 3D Vision	2013 - present

Period when the claimed impact occurred: 2017 – 2020

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

1. Summary of the impact

Advances from the 3D Vision team at UCL, led by Prof. Agapito, have enabled new ways to synthesise video of photorealistic human faces in speech. This technology has been commercialised by Synthesia, a spinout co-founded by Agapito in 2017, via the launch of services with personalised and localised AI presenters that fit within professional content creation pipelines, as well as products including automatic text-to-video synthesis with a framework for ethical control. Synthesia has rapidly grown to be one of the top UK AI companies in terms of investment, revenue, and customer base, serving large companies such as Facebook, Google, Fedex and Tesco across a wide range of sectors. [TEXT REMOVED FOR PUBLICATION]. As a result of using Synthesia's product, the high-profile campaign Malaria No More (featuring David Beckham) has raised USD14,000,000,000 for the cause.

2. Underpinning research

Synthesising photorealistic, expressive human faces in speech has been a long-standing challenge in computer vision and graphics. For decades, this technology has been the exclusive domain of the film and TV industries, with multi-million budgets needed to build specialised and complex multi-camera 3D capture studios to create digital 3D doubles of humans, and for manual post-production by visual effects artists. While the recent emergence of deep fake technology, based on 2D generative adversarial networks (GANs), has enabled easy creation of videos of talking people, bypassing the 3D capture process comes at the cost of a lack of any form of explicit control or 3D interpretability over the synthesis.

For over 15 years, Agapito's team has been at the forefront of research in non-rigid 3D modelling from monocular video, a technology that has enabled to fully automate the process of capturing vivid 3D models of humans in motion, directly from videos captured casually with a single commodity camera, without the requirement for expensive studios, or hours of manual editing. Perhaps more importantly, these algorithms do not require expensive 3D supervision, or vast amounts of data, operating in a self-supervised fashion.

Agapito's team pioneered the first algorithms to demonstrate full dense tracking and 3D reconstruction of deformable surfaces, such as human faces, from monocular sequences – video clips captured with a single camera. Existing monocular algorithms were simplistic and severely limited to only handle a small set of sparse points. On the other hand, fully dense modelling of non-rigid surfaces had only been shown before for specialised multi-camera setups or depth cameras.

The next level to enable truly light-weight, low-cost, scalable, fast and accurate 3D capture of faces in speech was to enable frame-to-frame sequential operation and to couple tracking and 3D reconstruction into a single inference (**R1**). This research (**R1**) resulted in the first sequential method to simultaneously track and reconstruct deformable surfaces in motion directly from an input video at close to real-time operation. The innovation in (**R1**) was to estimate 3D deformations directly from photometric consistency losses and resulted in the most accurate, fully automated method to reconstruct 3D models of non-rigid surfaces directly from a single video. Agapito's team pushed this method further in (**R2**) to model textures and changes in appearance due to deformations and illumination changes over time.

Beyond faces, Agapito's team has also pioneered weakly supervised methods for 3D human pose estimation from single images that only require 2D image annotations (**R3**) which are cheaper and easier to harvest than the 3D annotations required by other methods.

These breakthrough algorithms for monocular non-rigid 3D reconstruction by Agapito's team at UCL (**R1-R3**) form the underpinning technology that made 3D-driven, photorealistic and low-cost AI video synthesis finally possible and form an integral part of Synthesia's technology. It is this ability to create photorealistic digital doubles of humans at scale, automatically and directly from casually captured videos (even from a mobile phone), that has enabled Synthesia to incorporate 3D reasoning into the synthesis process to provide the explicit control and interpretability that other 2D generative models (such as GANs) completely lack. In turn, this 3D reasoning and control are responsible for the high quality and photorealism of the synthesised videos that sets Synthesia apart from their competitors.

3. References to the research

R1. R Yu, C Russell, NDF Campbell, **L Agapito** (2015) Direct, dense, and deformable: Template-based non-rigid 3d reconstruction from rgb video, Proceedings of the IEEE International Conference on Computer Vision, ICCV 2015, 918-926. DOI: 10.1109/ICCV.2015.111
R2. Q Liu-Yin, R Yu, **L Agapito**, A Fitzgibbon, C Russell (2016) Better Together: Joint Reasoning for Non-rigid 3D Reconstruction with Specularities and Shading, British Machine Vision Conference (BMVC). DOI: 10.5244/C.30.42
R3. D. Tome, C. Russell, **L. Agapito** (2017) Lifting from the Deep: Convolutional 3D Pose Estimation from a Single Image, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. DOI: 10.1109/CVPR.2017.603

4. Details of the impact

The ability to model 3D faces in speech from video alone (described in **R1-R3**) eliminated the need for complex capture studios and 3D scanners, and became the cornerstone for generating 'synthetic media' such as AI-generated realistic, human-like avatars. Agapito co-founded Synthesia with other researchers and entrepreneurs to provide commercial solutions for a range of applications for this new technology, from lip-sync dubbing for content localisation to personalised video messages and corporate training. Synthesia technology allows users to create professional looking videos by simply typing a message, using an automated, 3D-driven AI process to synthesize photorealistic results that are indistinguishable from real video but without the need for cameras, actors or expensive film studios.

Agapito's research has 1) **enabled the commercial viability and rapid growth of Synthesia**; 2) **benefited Synthesia's customers by providing new and more cost-effective services**; and 3) **increased public understanding of synthetic media** through its high-profile work.

Impact on enabling the commercial viability and rapid growth of Synthesia

Regarding the impact of Agapito's research on founding Synthesia, CEO and co-founder, says: "The algorithms proposed in her research for accurate 3D non-rigid shape estimation from video based on photo-consistency, were pivotal to the creation of Synthesia and now photometric tracking lies at the core of our technology. The ability to capture the 3D geometry and appearance of a human face in speech from a single video with algorithms building on Agapito's research has been transformational in allowing us to build a low-cost solution to create high fidelity 3D avatars of humans for animation and synthesis" (**S1**). It is this breakthrough in capturing 3D geometry from short video clips or even a single image which allows the low-cost, controllable synthesis that is so valuable in applications.

3D understanding means that the actions of the synthesized faces can be fully decoupled from the input clip, and **controlled** by smart software, to open up the new applications which are at the heart of Synthesia's **rapid growth**. As the CEO continues: "By making it easier to reconstruct 3D photorealistic faces, the algorithms permitted Synthesia to provide new services to clients" (**S1**). Synthesia has transitioned from offering high profile video-to-video services towards offering a "Software as a Service" platform where users can create videos simply by writing the speech of the digital 'actor'. This technology has a wide range of applications, from corporate training to in-house communication to sales. As such, Synthesia's services have been used by diverse clients, including Reuters, WPP, Dixa, Just Eat, Tesco, FedEx, Facebook and Google.

With this technology in place, Synthesia was extremely well placed to grow rapidly during the recent boom in the use of online services. Synthetic media has become a cheaper and quicker way to produce video, an advantage amplified by the Covid-19 pandemic. With studios and other facilities out of action, companies with needs in corporate communication, training or advertising used synthetic video generation for the first time. In an interview with TechRepublic, the CEO said, "Using AI, we've digitized the video production process and enabled our customers to create a video in 5 to 10 minutes, without the need for any cameras, actors or studios" (**S1**).

During 2020, [TEXT REMOVED FOR PUBLICATION]. As a result of its world-leading technology, Forbes magazine named Synthesia one of its "fearless five" Tech companies (**S3**).

Reshaping famous faces for global commercials and charity campaigns
Synthesia has provided the technology behind many of the most high-profile uses of photorealistic video face synthesis, as the ease of use and low cost of this technology were combined with the need for extremely high professional standards of quality and ethical use.

For example, the company completed a highly successful project for the food delivery company Just Eat. After filming a campaign with the rapper Snoop Dogg in 2020, the company wanted to extend it to its Australian subsidiary MenuLog. However, rerecording the advert with the new name would have been prohibitively expensive. By using Synthesia's rendering, they were able to simply edit the original, reaching an entirely new audience without making a new advert, which created substantial savings for the firm. In addition to saving significant costs for the firm, this advert was crucial in helping MenuLog reach new audiences, and the campaign went on to receive over 10,000,000 views (**S4**).

The technology was also crucial in a 2019 Malaria No More campaign that raised USD14,000,000,000 to help end the world's three biggest preventable killer diseases: AIDS, Tuberculosis and malaria. It was used to make David Beckham speak 9 languages as part of a 2019 campaign video, and because this video localised the campaign to suit specific global audiences, it created 700,000,000 online impressions and resulted in the disease's peak awareness in almost 3 years. The campaign attracted over 1,800 pieces of media coverage, resulting in internet searches for malaria reaching an all-time high (**S5**). It won marketing awards, as well as an award for social good in AI, and was instrumental in winning commitments from world leaders to increase funding for fighting malaria. As one of

the Malaria No More team said, "*This magic wouldn't have been possible without the talented team at Synthesia*" (**S5**).

Cutting costs and increasing engagement for training videos and online assistants

Synthesia's development of a highly automatic and scalable SaaS platform, which is key to its current growth, has enabled easy-to-use Text to Video services which customers can use to generate video of synthetic actors as easily as typing text.

For example, the multinational communications company WPP used Synthesia in 2020 to provide their corporate training videos in multiple languages, without having to reshoot using different actors and scripts. As WPP's chief technology officer told Wired magazine, this saved them a considerable amount of money: "*A company-wide internal education campaign might require 20 different scripts for WPP's global workforce, each costing tens of thousands of dollars to produce. With Synthesia we can have avatars that are diverse and speak your name and your agency and in your language and the whole thing can cost USD100,000*" (**S6**). The former Global Director of WPP said that "Synthesia allowed us to transform the way we think of training materials" (**S6**).

Life Extension Europe, a nutritional supplement supplier, utilised Synthesia to improve its online sales. The success in this approach was reflected in the increased time that visitors spent on their webpages; for instance, with the new videos, the average session duration in the UK increased to 9 minutes 37 seconds from the previous average of 3 minutes and 36 seconds, an increase of +167%. Similarly, the average number of page views per session increased to 6.16 (versus 4.24 previously), an increase of 45.3%. The campaign recorded an additional 101 transactions (**S7**).

Improving public understanding of and engagement with synthetic media

The final impact of Agapito's research, via Synthesia, has been increased awareness of the positive potential of synthetic media. While there has been widespread concern about 'deepfakes' and the potential for misinformation, Synthesia has used Agapito's research to demonstrate the many beneficial applications of synthetic media. Their work, and the code of ethics they have developed for the use of their work, has been widely covered in news media in 2019 - 2020, illuminating ways in which the changed landscape of AI-generated video can be regulated and turned to socially responsible ends. For instance, the *MIT Technology Review* praised Synthesia for only working with vetted clients in its 2019 article, "Making deepfake tools doesn't have to be irresponsible. Here's how." Similarly, *TechCrunch* expressed excitement that Synthesia's products "*could also be used to expand the reach of creators around the world*" (**S8**).

Synthesia's code of ethics includes a commitment only to create synthetic video of people who have given their explicit permission. This debate has also been promoted by other industry leaders such as Samsung, who have highlighted Synthesia's work in their 2020 list of "5 companies leading the creation of AI-enabled images & videos" (**S9**). As the Managing Director of Samsung Next (Samsung's synthetic media initiative) said in relation to the Synthesia's work, "it looks like AI will actually democratize creativity" (**S9**).

In addition to raising awareness of the potentially positive impact of AI, Synthesia created engaging AI videos for public consumption, such as the Synthesia Santa, which allowed users to easily create a video from Santa speaking their text to friends and family. The site made 90,000 cards in the first three weeks of launch, and in September 2020, Synthesia was voted the #2 product by Product Hunt (**S10**).

5. Sources to corroborate the impact

S1. Testimonial from Victor Riparbelli, CEO and co-founder of Synthesia.
S2. Confidential information on the finances of Synthesia, available upon request.
S3. Media Coverage of the Synthesia's commercial prospects:

- Forbes : 2019's Boldest Media & Tech Companies (The "Fearless Five")
  https://www.forbes.com/sites/petercsathy/2019/11/29/2019s-boldest-media--tech-companies-the-fearless-five/#521be2ae3d4c

S4. Discussions of the Just Eat campaign.
- Forbes covers the making of Snoop dogg MenuLog advert, powered by Synthesia
- Video by Synthesia demonstrating how the edit of the advert was done.
  https://www.synthesia.io/post/snoop-dogg

S5. Discussions of Synthesia's work for no More Malaria:
- Synthesia case study:
  https://malarianomore.org.uk/you-spoke-leaders-have-listened-and-committed-14bn-global
- TechCrunch covers David Beckham Malaria-no-more campaign
- ABC News, David Beckham 'speaks' 9 languages for new campaign to end malaria
  https://abcnews.go.com/International/david-beckham-speaks-languages-campaign-end-malaria/story?id=62270227
- Sky News, David Beckham 'speaks nine languages' in malaria campaign's new video
  https://news.sky.com/story/david-beckham-speaks-nine-languages-in-malaria-campaigns-new-video-11688600
- The Drum, How Malaria No More UK inspired world leaders to commit $4bn to defeat deadliest-ever disease
  https://www.thedrum.com/news/2019/11/04/how-malaria-no-more-uk-inspired-world-leaders-commit-4bn-defeat-deadliest-ever

S6. Discussions of WPP's use of Synthesia
- WIRED: Deepfakes Are Becoming the Hot New Corporate Training Tool
  https://www.wired.com/story/covid-drives-real-businesses-deepfake-technology/
- Futurism: This Company is Making Corporate Training videos using Deepfakes
  https://futurism.com/the-byte/corporate-training-videos-using-deepfakes

S7. Zesta NY Resolution video report for Life Extension Europe.

S8. Media discussions of the ethics of 'deepfakes'
- MIT Technology Review Making deepfake tools doesn't have to be irresponsible. Here's how.
  https://www.technologyreview.com/2019/12/12/131605/ethical-deepfake-tools-a-manifesto/
- TechCrunch   An optimistic view of Deepfakes
  https://techcrunch.com/2019/07/04/an-optimistic-view-of-deepfakes/

S9. Samsung Next's 'Landscape of Synthetic Media and its discussions of Synthesia
- Samsung NEXT – Landscape of synthetic media, featuring a synthetic video of the author (Managing Director and General Manager of Samsung NEXT) powered by Synthesia:
  https://www.syntheticmedialandscape.com/
- https://www.samsungnext.com/blog/5-companies-leading-the-creation-of-ai-enabled-images-videos

S10. Public engagement with Synthesia products
- Santa video page usage, 90,000 Santa cards made within 3 weeks of launch:
  https://www.synthesia.io/santa
- Synthesia voted 2nd place of 'product of the month' in Sep 2020 on product hub:
  https://www.producthunt.com/posts/synthesia-2?utm_campaign=producthunt-api&utm_medium=api&utm_source=Application:+IFTTT+(ID:+2742)