

<b>Institution:</b> University of Edinburgh		
<b>Unit of Assessment:</b> 26 (Modern Languages and Linguistics)		
<b>Title of case study:</b> Personalised speech synthesis improves quality of life, changes health legislation, and leads to new commercial products		
<b>Period when the underpinning research was undertaken:</b> 2007 – 2016		
<b>Details of staff conducting the underpinning research from the submitting unit:</b>		
<b>Name(s):</b>	<b>Role(s) (e.g. job title):</b>	<b>Period(s) employed by submitting HEI:</b>
Simon King	Professor of Speech Processing	Oct 1997 – present
Junichi Yamagishi	Senior Research Fellow	Apr 2007 – Aug 2020
<b>Period when the claimed impact occurred:</b> August 2013 – December 2020		
<b>Is this case study continued from a case study submitted in 2014?</b> N		
<b>1. Summary of the impact</b> <p>Research on personalised speech synthesis has led to improved quality of life and commercial impact. A new spin-out company (SpeakUnique Ltd, attracting over GBP510,000 investment), established in collaboration with clinicians, provides personalised synthetic voices for users even when their speech is already degraded (e.g. as a result of motor neuron disease). As well improving as the <b>quality of life</b> of people who have lost their voice, the research has led to <b>new legislation</b> guaranteeing access to communication support for all patients suffering voice loss. King's research has also underpinned <b>commercial products and services</b> offered by leading technology companies ObEN and Papercup.</p>		
<b>2. Underpinning research</b> <p>The multidisciplinary (Linguistics and Informatics) Centre for Speech Technology Research (CSTR) is a world leader in the automatic conversion of written language into speech, known as Text-To-Speech (TTS). Founded in 1984, CSTR currently houses 13 academic staff and 15 PhD students. <b>King</b> has been director of the Centre since 2011. Working collaboratively with a number of CSTR academics, notably <b>Yamagishi</b>, King has made important contributions to the development of personalized synthetic speech, where the resulting voice sounds like a particular individual (rather than an “off the shelf” voice). Key research insights include:</p> <p><b>Adaptive statistical parametric speech synthesis</b>  King and colleagues have developed novel mathematical methods and implemented them in their free-to-use TTS toolkits Festival [3.1] and Merlin [3.3] which adapt an “average voice” synthesis model (trained using speech from multiple speakers) to the voice of a new target speaker using much less speech from the target speaker compared with previous approaches [3.2].</p> <p><b>Personalised speech synthesis</b>  Using the adaptive framework (above), King and colleagues have developed algorithms and software that can automatically create a personalised synthetic voice for a target speaker using just a few minutes of data (“voice cloning”). King and colleagues demonstrated this approach by creating thousands of personalised synthetic voices [3.4] including for children [3.5]. Furthermore, the techniques developed in [3.2] work even with lower quality recordings (e.g. web videos) than was previously feasible for speech synthesis.</p> <p><b>Voice reconstruction</b>  King and colleagues built on the above research to develop personalised speech synthesis methods which enable voice reconstruction even when the target speakers already have disordered speech due to a neurological condition such as motor neuron disease [3.6]. The</p>		

resulting synthetic speech repairs the disordered aspects, resulting in normal-sounding, intelligible, personalised speech.

### 3. References to the research

- [3.1] Clark, R. A. J., Richmond, K., & **King, S.** (2007). Multisyn: Open-domain unit selection for the Festival speech synthesis system. *Speech Communication*, 49(4), 317–330. <https://doi.org/10.1016/j.specom.2007.01.014>
- [3.2] **Yamagishi, J.**, Nose, T., Zen, H., Ling, Z., Toda, T., Tokuda, K., **King, S.**, & Renals, S. (2009). Robust speaker-adaptive HMM-based text-to-speech synthesis. *IEEE Transactions on Audio, Speech and Language Processing*, 17(6), 1208–1230. <https://doi.org/10.1109/TASL.2009.2016394>
- [3.3] Wu, Z., Watts, O., & **King, S.** (2016). Merlin: An open source neural network speech synthesis system, in *9th ISCA Workshop on Speech Synthesis (SSW9) Proceedings*. September 2016, Sunnyvale, CA, USA. <https://doi.org/10.21437/SSW.2016-33>
- [3.4] **Yamagishi J.**, Usabaev, B., **King, S.**, Watts, O., Dines, J., Tian, J., Hu, R., Guan, Y., Oura, K., Tokuda, K., Karhila, R., & Kurimo, M. (2010). Thousands of voices for HMM-based speech synthesis – analysis and application of TTS systems built on various ASR corpora. *IEEE Transactions on Audio, Speech and Language Processing*, 18(5), 984–1004. <https://doi.org/10.1109/TASL.2010.2045237>
- [3.5] Watts, O., **Yamagishi, J.**, **King, S.**, & Berkling, K. (2010). Synthesis of child speech with HMM adaptation and voice conversion. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(5), 1005–1016. <https://doi.org/10.1109/TASL.2009.2035029>
- [3.6] **Yamagishi, J.**, Veaux, C., **King, S.**, and Renals, S. (2012). Speech synthesis technologies for individuals with vocal disabilities: Voice banking and reconstruction. *Acoustical Science and Technology*, 33(1), 1–5. <https://doi.org/10.1250/ast.33.1>

### 4. Details of the impact

#### Impact on quality of life for people who are losing their voice

Research in CSTR, led by King, has made it possible to blend a range of donor voices to best approximate an individual's own voice. This allows people to create a personalised synthetic voice, even if their own speech has already degraded. This development is especially useful for people with conditions like motor neuron disease (MND), who often experience degenerative speech loss.

King's research developed into a collaboration with the Anne Rowling Regenerative Neurology Clinic at University of Edinburgh. The **Speak:Unique** project – which included a Scottish Government funded (GBP200,000) trial run in partnership with 4 NHS Scotland health boards – provided personalised voices to 168 MND patients using Assistive and Augmentative Communication (AAC) devices. This figure represents approximately 42% of those living with MND in Scotland in any given year [5.1]. Feedback from patients confirms that being able to use a personalised synthetic voice enabled them to **retain a sense of self and dignity** in the face of a devastating and incurable disease: they felt more “like themselves”, less controlled by their condition, more independent, socially capable, and closer to their loved ones. Patients reported:

“Where someone has lost their voice through a degenerative condition, um, it [the personalised voice] has got to create a more powerful link if it sounds something like the person. Because the emotional bond you have with someone you know, is their voice. It feels different to somebody else's voice” [5.2, p. 5].

“My grandchildren said ‘It sounds just like Gramps!’” [5.2, p. 5].

“I mean, you are your voice, aren't you? I mean you can sit in a wheelchair, but if you can still communicate then, it's still you that's doing the talking. So yeah, it's really just that it's ... it's so your personality as well... it's a huge thing to be able to still communicate and

people know that it's you that's doing the talking and not a machine really. With it being your own voice, I think makes it even more you..." [5.2, p. 5].

Based on this initial application of King's research in Scotland, **SpeakUnique Ltd** launched as a standalone spin-out company in June 2020. SpeakUnique Ltd aims to provide access to personalised synthetic speech for all AAC device users around the world. Users of their service include not only patients with MND, but also sufferers of multiple sclerosis (MS), cerebral palsy, Huntington's disease, and vocal cancer [5.1]. Using King's research, SpeakUnique Ltd is able to simulate all accents of English, including those of people for whom English is a second language, allowing it to be available to any English speaker globally. As of its launch, SpeakUnique Ltd employs 8 people and has attracted over GBP510,000 of investment (including approximately GBP200,000 in private investment), as well as awards from Innovate UK (GBP245,000); a Royal Society of Edinburgh Enterprise Fellowship (GBP65,000); an Emerging Innovation Award from Edinburgh Innovations (GBP1,500); and semi-finalist award in the 2019 Converge Challenge (winning prize value GBP50,000) [5.1].

Two major MND charities/organisations (MND Association and MND Scotland) as well as the UK-wide charity for people with Progressive Supranuclear Palsy (PSPA) have agreed to cover the cost of SpeakUnique's personalised speech synthesis for people in the UK [5.1]. The cost of a synthetic voice can also be reclaimed through NHS Scotland for individuals in Scotland who have or will lose the ability to speak, regardless of their medical condition (*see change to legislation below*) [5.1]. In addition, SpeakUnique offers a service whereby currently healthy people can record and "bank" their voices, which can subsequently be used to create an accurate synthetic equivalent should they experience vocal loss in the future. From June 2020 to December 2020, SpeakUnique has created voices for people in 8 countries, trained over 150 healthcare professionals; it cannot disclose how many voices it created, but its website received over 35,000 visits and social media adverts have been viewed over 200,000 times [5.1]. Individual patients confirm significant benefits to their quality of life:

"I feel like I've saved an important part of me. Every time I use it [my SpeakUnique synthesised voice] it makes me smile." [5.1]

"It's so hard to lose speech, so anything that reduces that sense of loss helps." [5.1]

"Several years ago, I participated in a voice banking research project [Speak:Unique], and spent an interesting afternoon recording phrases in a soundproof studio. I was motivated to do this partly because it seemed such a great innovation, and partly because I enjoyed public speaking and using my voice. I had no idea then that I would have a personal need of those recordings. But, last year I was diagnosed with Motor Neuron Disease, and my speech has been badly affected. It has been hugely important to me to have been able to get help from the SpeakUnique development team, and to acquire a synthetic voice based on my own voice recordings. It makes such a difference to be able to retain some of my personal identity." [5.1]

### Impact on Scottish health legislation and public awareness

One of Speak:Unique's earliest participants was MND patient and celebrated campaigner Gordon Aikman. Aikman, who died from MND in 2017, described how important it is "that patients don't just get a voice, but get *their own* voice back" [5.2, p. 2]. Aikman was so impressed with the concept and the technology behind it that he lobbied to bring communication support and voice banking to wider public and policy attention. The result was an **amendment to the Health (Tobacco, Nicotine etc. and Care) (Scotland) Act 2016** to include routine provision of Communication Aids to patients of all conditions that cause difficulty speaking [5.3]. The Scottish Government, in its 2019 progress report on the provision communication equipment and support resulting from the Act, estimated the potential reach of the policy, as of February 2020, as "... around 27,000 benefiting from some type of AAC, with 2,700 benefiting from powered communication aids". The significance of this policy change has been directly recognised by the Scottish Government, as captured by this tweet:



*Figure 1 Tweet from Scottish Government (March 2019) "Communications equipment and support for those who have difficulty speaking has changed the lives of people like Craig. From 19 March, a new law means anybody who needs it must receive it"*

Addressing the Scottish Parliament's Health and Sport Committee on 26 January 2016, the Scottish Government Minister for Public Health said: "I also highlight the on-going work on voice banking, which is an important development in augmentative and alternative communication [...]. We thank Gordon Aikman for bringing the [Speak:Unique] research work to our attention" (Maureen Watt, Scottish Minister for Public Health, November 2014 to May 2019) [5.4, p. 18].

The resulting **media coverage** followed not only Aikman, but also other SpeakUnique users. BBC2's documentary *My Year with MND*, showed SpeakUnique recreate the voice of Rob Burrow, a former Rugby League player diagnosed with MND in 2020; BBC2's 2017 documentary *MND and 22-Year-Old Me* followed self-confessed "chatterbox" Lucy Lintott; BBC1's *Breakfast* programme (2 June 2017) featured the creation of a synthetic Yorkshire voice for a man with MND; BBC1's *The One Show* (20 June 2016) contained a feature in which Dr Michael Mosley described how Speak:Unique was helping MND sufferers by voice banking [5.5]. At least 20 major media articles have covered SpeakUnique's work, including *The Times*; *The Guardian*; *The Huffington Post*; *The Metro*; *ITV News* and *BBC News* websites. This coverage brought personalised speech synthesis research, its benefits to AAC users, and the difficulties faced by people who have lost their voice through illness, to wider public attention [5.5].

Since 2016, a digital interactive display on Speak:Unique has featured as part of the **National Museum of Scotland's** permanent collection (over 2 million visitors per year), highlighting the role of personalised speech in the development of communication. The Museum's Principal Curator of Communications describes how Speak:Unique allowed the museum to meet several aims, including bringing the story of speech synthesis up to date [5.6].

### **Impact on commercial product development**

In 2016, King was approached by USA-based company **ObEN** (100 employees; established 2014; attracted more than USD23.7million of investment). ObEN provides personalised digital avatars that "look, sound, and behave like users". Personalised speech synthesis is "a key part" of their products [5.7]. King's research, including the Merlin and Festival toolkits, "has been crucial in helping us achieve our aim of building a personalised voice interface to work with our projects" [5.7]. King's work on expressive speech synthesis helped them "realise and improve



our voice personalisation using a relatively small sample of audio recordings” [5.7]. ObEN’s work has been covered in *Forbes*, *Venture Beat*, *Gizmodo*, and *MIT Technology Review*. In 2019, ObEN used technology drawing on King’s research to produce digital avatars for the hosts of China Central Television’s *Spring Network Gala* (approximately 1.8 billion viewers). Personal AI (PAI) News (launched May 2019) is an iOS and Android news application with content delivered by the world’s first virtual anchor; PAI Care (launched March 2019) is a virtual assistant designed to monitor patients with congestive heart failure. “The fact that we were able to create these experiences is partly due to King’s research and his expertise in helping us improve our systems” [5.7].

King’s research also informed London-based company **Papercup Ltd** (67 employees; established 2017; attracted more than GBP10million of investment). Papercup allows online content creators to translate videos by generating personalised Spanish-language voices that sound similar to the original speakers. Since their launch, King has provided technical advice, based on his research, on systems and implementation. Co-founder and Chief Technology Officer confirms that they “would not have been able to build our state-of-the-art speech synthesis system without King’s research and expertise [...]” King’s work on “prosody and expressiveness in text-to-speech systems gave us inspiration for our own proprietary systems and models” [5.8]. As of October 2020, Papercup has translated many thousands of hours of video for its clients, generating over 60 million unique views via digital content channels such as YouTube. Clients include Sky News, with whom Papercup partnered during 2020 to bring a Spanish language version of their news channel to YouTube. In its first 12 months, the new channel has attracted over 96,000 subscribers and over 26 million views, with metrics indicating a long average channel watch time, evidencing significant engagement. Sky News said, “The overall average watch time and completion on our new Spanish Sky News channel is so far above and beyond what we had expected. That’s testament to the quality of the Papercup solution [...] We now can get more bang for our buck using our existing content. And translating to Spanish is only just the first step. And it doesn’t stop with news; it can expand to sports, entertainment and educational content” [5.8].

## 5. Sources to corroborate the impact

- [5.1] Statement from SpeakUnique Ltd CEO, 2021
- [5.2] Speak:Unique executive summary, project report for the Scottish Government, 2019
- [5.3] Health (Tobacco, Nicotine etc. and Care) (Scotland) Act 2016
- [5.4] The Scottish Parliament, Health and Sport Committee, official report, January 2016
- [5.5] Press coverage of Gordon Aikman’s campaign and Speak:Unique
- [5.6] Statement from the National Museum of Scotland communications curator, 2019
- [5.7] Statement from ObEN, 2019
- [5.8] Statement from Papercup and media coverage of the company