



Tony Robinson, the founder of Speechmatics: “I still have no formal qualifications in computer science,” he says

CHRIS RADBURN FOR THE TIMES

THE MANIFESTO

# Alexa, tell me who blazed the trail in speech recognition technology

Tony Robinson was in at start of the industry and his Speechmatics firm is a market leader

Katie Prescott, Technology Business Editor

Friday March 10 2023, 5.50pm, The Times

Share



Save 

A photograph of Cambridge University’s engineering department’s Speech and Vision Group from 1988 shows the small cohort of scientists who would revolutionise computers’ understanding of the spoken word — the genesis of devices such as Apple’s Siri and Amazon’s Alexa. It was led by Professor Frank Fallside, who died prematurely a few years after the picture was taken.

As hype around artificial intelligence has gained momentum in recent months, attention has zeroed in on Cambridge, which has built a gold-plated reputation for voice artificial intelligence and automatic speech recognition (ASR), which this group pioneered.

Standing at the back, with a very 1980s mullet, is Tony Robinson, the founder of Speechmatics, who has spent his career working on accurate speech and language processing.

Founded in 2006, the company makes speech recognition technology that creates text from voices, regardless of how the person speaks. It can transcribe audio in real time from 48 different languages and is used in business applications, such as by Ubisoft, the game developer, to live subtitle its content, and by Red Bee Media to caption TV and streaming services, for example on [BT Sport](#). Other customers include contact centres, consumer electronics, security and software, [Deloitte UK](#), the US business Veritone and government agencies.

Robinson credits his academic past, running an engineering division, with teaching him about business. “You’ve got a team, you need to get your own money, hire your own people and get your name out there. I was running this tiny little business where I had PhD students under me and I was competing for grants.”

#### ADVERTISEMENT

It was a winding road to the creation of Speechmatics, which today employs 176 people in its Cambridge base as well as in London, Boston, the Czech Republic and India.

His degree was in physics, as his teacher dissuaded him from computing, his passion, deriding it as not a “real” science. “I went to a secondary modern school, which meant I had pretty

much no chance of getting into university — until it changed to a comprehensive and I had some teachers who were used to teaching in a secondary modern who thought, hey, we've got a chance to do something here. So they helped a few people and one of them was me. I scraped into Cambridge by the skin of my teeth. I still have no formal qualifications in computer science.”

After a brief stint away, his passion brought him back to the engineering department at Cambridge to study for a PhD at an “amazing time” for the subject.

The lightning bolt moment came while he was listening to a seminar by Geoff Hinton, the British-Canadian computer scientist now at Google. “I thought, I know what I want to do and so speech in neural networks was it, and has been it almost to this day.”

SPONSORED



The nineties were a period of rapid development for ASR and progress was driven, Robinson recalls, by tough annual competitions led by the US government's research department, Defense Advanced Research Projects Agency. Cambridge fielded two teams: one led by Steve Young, another pioneer, the other by Robinson.

On a sabbatical in 1995, Robinson decided to turn his research into a business and founded Soft Sound, which made voice data searchable: “There is no point in interesting work staying in universities. You need to get it out to people.”

His real breakthrough came with a pitch to the BBC, where he showed how to convert recordings of programmes into text, which allowed the audio and video to be searched. “In the 1990s you could just type a few words in and the video would start playing the news you were interested in. We also created the first automated subtitling for *EastEnders*.”

A major stake was bought in 2000 by [Autonomy](#), the business run by the Mike Lynch, the discredited Cambridge

entrepreneur, who is facing extradition proceedings by the US over criminal charges related to its sale to Hewlett Packard in 2011. Robinson exited the business in 2006.

#### ADVERTISEMENT

Also controversial was another company he worked for: SpinVox, which turned voicemails into text messages used by carriers such as [Vodafone](#) and Skype. It licensed Robinson's technology and he was seconded to the business, to further automate the process. Subsequently, a BBC investigation in 2009 found that voicemail data was transcribed mainly by humans in centres outside the EU, calling into question the ability of the software and breaching data regulations. Robinson put this down to the "limitations of the technology at the time". SpinVox was sold to Nuance Communications, a US business, for £64 million in 2009.

Keeping up with Robinson's own speech during the interview is not easy. His brain flits between subjects amid a stream of information about the technical capability of his work and the computer models, complex for the layperson to track.

When it comes to his software, the complexity does not come from the content itself but from sounds: from accents and background noise, especially difficult when someone is not speaking in their mother tongue and syntax might get muddled.

Voice to text is being pursued by other companies such as the UK's Trint and the US's Verbit. It has also attracted the interest of [Google](#) and Microsoft and the market is growing exponentially because of the increasing shift to online interactions wrought by the pandemic.

Speechmatics insists it can fend off the competition because of its greater accuracy and it boasts of "self-supervised learning".

Rather than being fed labelled text, it is trained through exposure to hundreds of thousands of voices using millions of hours of voice data without the need for human intervention. This has been made possible, Robinson said, by the huge increase in computer power.

Speechmatics is still loss-making. Its parent is Cantab Research, which turned over £9.5 million in 2021, up from £6.4 million in 2020. According to its latest accounts filed with Companies House, it made a £3.7 million loss in 2021 up from £3.2 million in 2020 as it spends on research and development. Robinson expects it to show a profit by 2025.

The business has more than doubled its headcount in the past two years. In June, the company raised £50 million in a Series B funding round and plans to use the investment to reach the holy grail of accurate transcripts, cutting through speech impediments or noisy football matches. “With something nice like a video call or a broadcast where the audio quality is good, we nail it. When you’re on a train or in a noisy pub, it becomes more of a challenge.”

#### ADVERTISEMENT

Reflecting on the hype around the launch of [the chatbot ChatGPT](#) and the excitement around artificial intelligence, Robinson said the main issue continues to be trust. “There is no ‘intelligence’ behind it, it’s just distilling all the knowledge that we’ve that we’ve written down — it’s very easy for nonsense to come out.”

There is still a long way to go before we are truly conversing with computers.

#### Q&A

**Who is your mentor?**

I've stood on the shoulders of giants but never had a mentor.

**Does money motivate you?** Money is broken. We live in an unsustainable world, and the price we pay does not include fixing the damage we caused.

**What was the most important event in your working life?**

It was something that I don't want to make public to protect the guilty — some people fight dirty, an important business lesson.

**Who do you most admire?** Everyone is very different with their own unique qualities. I admire qualities and achievements, not people.

**What is your favourite TV programme?**

Any clever multi-level sci-fi with plenty of plot twists gets me — like the original Total Recall.

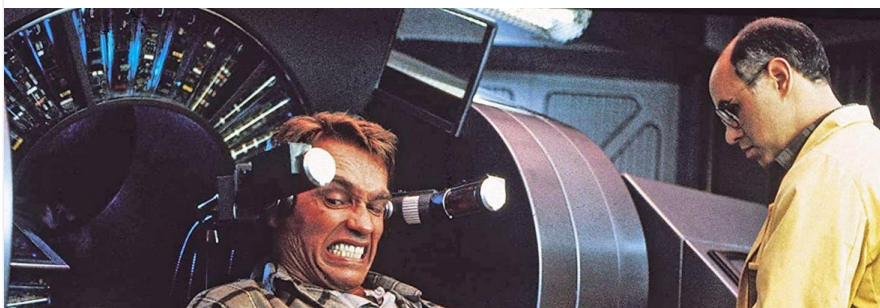
ADVERTISEMENT

**What does leadership mean to you?**

Good management includes good leadership, but leadership does not necessitate management: indeed, good leadership is minimal management.

**How do you relax?**

I love growing trees: there's so much to learn, they'll outlive me and we need more. Longer term I love long-distance walking and wild camping — just me and the mountains.





networks, all at the University of Cambridge  
**Career:** 1995-2006 Founder and managing director, SoftSound Ltd; 1989-2000 Founder of Connectionist Speech Recognition Group, Cambridge University; 2006-present Founder, chief exec, then chief technology officer and now adviser at Speechmatics  
**Family:** Lives with his partner and their son in Cambridge

- Technology
- United States

Related articles

ADVERTISING  
**AI tech tools that make online marketing more accurate**  
March 10 2023, 9.00am  
Richard Tyler

SIR MARTIN SORRELL | COMMENT  
**Why the Mad Men would be crazy to ignore AI**  
March 09 2023, 9.00am  
Sir Martin Sorrell

MATT GLANVILLE | RED BOX  
**Why ChatGPT is an opportunity for schools**  
February 28 2023, 12.01am  
Matt Glanville



Get in touch

- About us
- Contact us
- Help
- The Times Editor
- The Sunday Times Editorial Complaints
- Place an announcement
- Classified advertising
- Display advertising
- The Times corrections
- The Sunday Times corrections
- Careers

More from The Times and The Sunday Times

- The Times e-paper
- Times Currency Services
- Times Print Gallery
- Times Crossword Club
- Times+
- Times Expert Traveller
- Schools Guide
- Best Places to Live
- Times Appointments
- Times Money Mentor
- Sportswomen of the Year Awards

© Times Media Limited 2023.  
Registered in England No. 894646.  
Registered office: 1 London Bridge Street, SE1 9GF.

- Privacy & cookie policy
- Licensing
- Site map
- Topics
- Authors
- Commissioning terms
- Terms and conditions