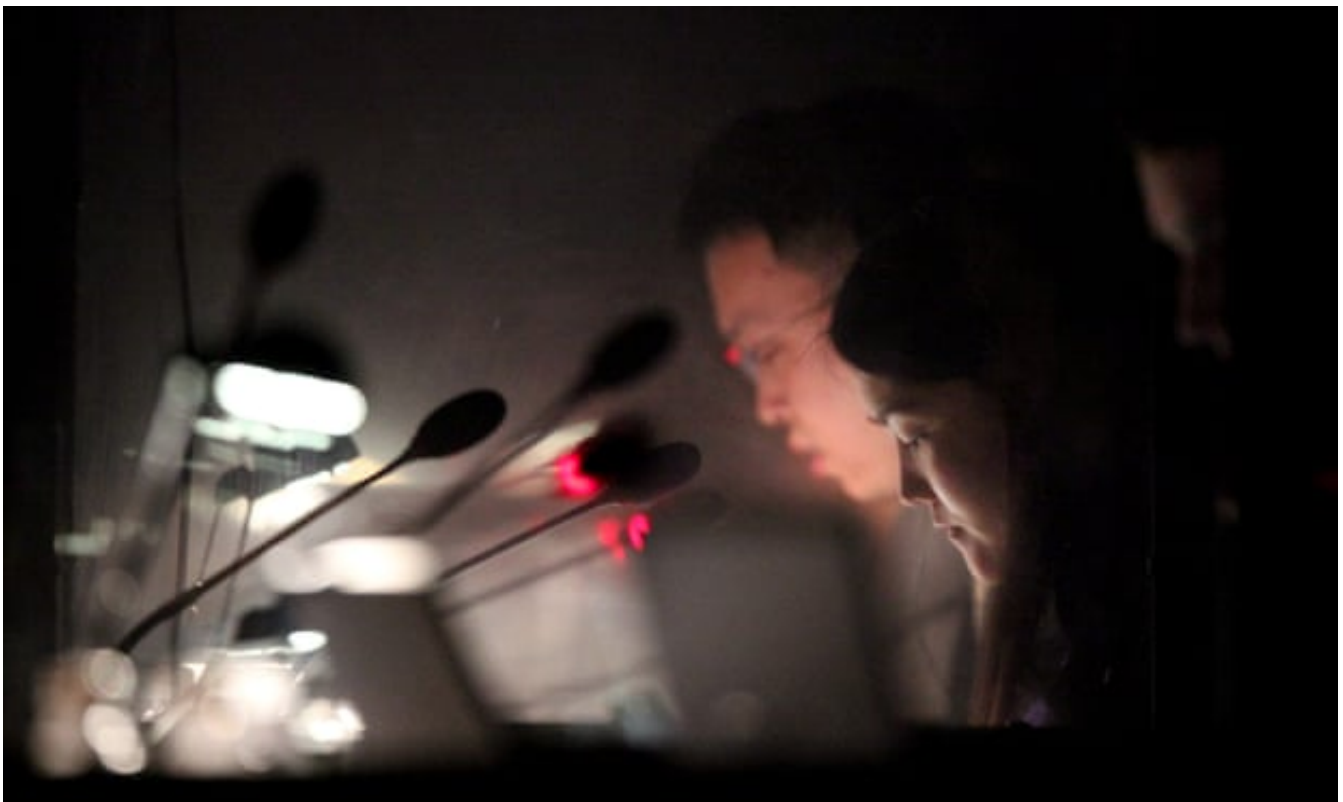# The Guardian

Is the era of artificial speech translation upon us?

Once the stuff of science fiction, technology that enables people to talk using different languages is now here. But how effective is it?

Interpreters work in a sound booth while China's President Hu Jintao speaks at the 2009 Asia-Pacific Economic Cooperation summit in Singapore. Photograph: Julian Abram Wainwright/EPA

Marek Kohn

Sun 17 Feb 2019 09.00 GMT

Noise, Alex Waibel tells me, is one of the major challenges that artificial speech translation has to meet. A device may be able to recognise speech in a laboratory, or a meeting room, but will struggle to cope with the kind of background noise I can hear surrounding Professor Waibel as he speaks to me from Kyoto station. I'm struggling to follow him in English, on a scratchy line that reminds me we are nearly 10,000km apart – and that distance is still an obstacle to communication even if you're speaking the same language. We haven't reached the future yet.

If we had, Waibel would have been able to speak in his native German and I would have been able to hear his words in English. He would also be able to converse hands-free and

seamlessly with the Japanese people around him, with all parties speaking their native language.

At Karlsruhe Institute of Technology, where he is a professor of computer science, Waibel and his colleagues already give lectures in German that their students can follow in English via an electronic translator. The system generates text that students can read on their laptops or phones, so the process is somewhat akin to subtitling. It helps that lecturers speak clearly, don't have to compete with background chatter, and say much the same thing each year.

The idea of artificial speech translation has been around for a long time. Waibel, who is also a professor of computer science at Carnegie Mellon University in Pittsburgh, "sort of invented it. I proposed it at MIT [Massachusetts Institute of Technology] in 1978." Douglas Adams sort of invented it around the same time too. *The Hitchhiker's Guide to the Galaxy* featured a life form called the Babel fish which, when placed in the ear, enabled a listener to understand any language in the universe. It came to represent one of those devices that technology enthusiasts dream of long before they become practically realisable, like portable voice communicators and TVs flat enough to hang on walls: a thing that ought to exist, and so one day surely will.

Waibel's first speech translation system, assembled in 1991, had a 500-word vocabulary, ran on large work stations and took several minutes to process what it heard. "It wasn't ready for prime time," he acknowledges. Now devices that look like prototype Babel fish have started to appear, riding a wave of advances in artificial translation and voice recognition. Google has incorporated a translation feature into its Pixel earbuds, using Google Translate, which can also deliver voice translation via its smartphone app. Skype has a Translator feature that handles speech in 10 languages. A number of smaller outfits, such as Waverly Labs, a Brooklyn-based startup, have developed earpiece translators. Reviews in the tech media could reasonably be summarised as "not bad, actually".

The systems currently available offer proof of the concept, but at this stage they seem to be regarded as eye-catching novelties rather than steps towards what Waibel calls "making a language-transparent society".

One of the main developments driving artificial speech translation is the vogue for encouraging people to talk to their technology.

"We're generally very early in the paradigm of voice-enabled devices," says Barak Turovsky, Google Translate's director of product, "but it's growing very rapidly, and translation will be one of the key parts of this journey."

Last month, Google introduced interpreter mode for its home devices. Saying: "Hey, Google, be my French interpreter" will activate spoken and, on smart displays, text translation. Google suggests hotel check-in as a possible application – perhaps the obvious example of a practical alternative to speaking travellers' English, either as a native or as an additional language.

You can do this already if you have the Translate app on your phone, albeit using an awkwardly small screen and speaker. That kind of simple public interaction accounts for much usage of the app's conversations feature. But another popular application is what

Turovsky calls "romance". Data logs reveal the popularity of statements such as "I love you" and "You have beautiful eyes". Much of this may not represent anything very new. After all, chat-up lines have been standard phrasebook content for decades.

Waverly Labs used the chat-up function as a hook for its Indiegogo funding drive, with a video in which the company's founder and CEO, Andrew Ochoa, relates how he got the idea for a translator when he met a French woman on holiday but couldn't communicate with her very well. Trying to use a translation app was "horrible". Phones get in the way – but earpieces are not in your face. The video shows what might have been: he presents a French woman with an earpiece, and off they go for coffee and sightseeing. The pitch was spectacularly successful, raising $4.4m (£3.4m) – 30 times the target.

Waverly Labs' Pilot earpiece (red and white) and Google's Pixel earbuds (black). Photograph: Observer Design desk

One customer said the company's Pilot earpiece had enabled him to speak to his girlfriend's mother for the first time. Some even report that it has enabled them to speak to their spouses. "Every once in a while, we'll receive an email from someone who says they're using this to speak with their Spanish-speaking wife," says Ochoa. "It baffles me how they even got together in the first place!" We might surmise that it was through the internet and an agency. Ochoa acknowledges that "the technology has to improve a bit before you'll really be able to find love through the earbud, but it's not too far away".

Many of the early adopters put the Pilot earpiece to entirely unromantic uses, acquiring it for use in organisations. Waverly is now preparing a new model for professional applications, which entails performance improvements in speech recognition, translation accuracy and the time it takes to deliver the translated speech. "Professionals are less inclined to be patient in a conversation," Ochoa observes.

The new version will also feature hygienic design improvements, to overcome the Pilot's least appealing feature. For a conversation, both speakers need to have Pilots in their ears. "We find that there's a barrier with sharing one of the earphones with a stranger," says Ochoa. That can't have been totally unexpected. The problem would be solved if earpiece translators became sufficiently prevalent that strangers would be likely to already have their own in their ears. Whether that happens, and how quickly, will probably depend not so much on the earpieces themselves, but on the prevalence of voice-controlled devices and artificial translation in general.

Here, the main driver appears to be access to emerging Asian markets. Google reckons that 50% of the internet's content is in English, but only 20% of the world's population speak the language.

"If you look at areas where there is a lot of growth in internet usage, like Asian countries, most of them don't know English at all," says Turovsky. "So in that regard, breaking language barriers is an important goal for everyone – and obviously for Google. That's why Google is investing so many resources into translation systems."

Waibel also highlights the significance of Asia, noting that voice translation has really taken off in Japan and China. There's still a long way to go, though. Translation needs to be simultaneous, like the translator's voice speaking over the foreign politician on the TV, rather than in packets that oblige speakers to pause after every few remarks and wait for the translation to be delivered. It needs to work offline, for situations where internet access isn't possible – and to address concerns about the amount of private speech data accumulating in the cloud, having been sent to servers for processing.

Systems not only need to cope with physical challenges such as noise, Waibel suggests, they will also need to be socially aware – to know their manners, and to address people appropriately. When I first emailed him, aware that he is a German professor and that continental traditions demand solemn respect for academic status, I erred on the side of formality and addressed him as "Dear Prof Waibel". As I expected, he replied in international English mode: "Hi Marek." Etiquette-sensitive artificial translators could relieve people of the need to be aware of differing cultural norms. They would facilitate interaction while reducing understanding. At the same time, they might help to preserve local customs, slowing the spread of habits associated with international English, such as its readiness to get on first-name terms.

Professors and other professionals will not outsource language awareness to software, though. If the technology matures into seamless, ubiquitous artificial speech translation – Babel fish, in short – it will actually add value to language skills. Automated translation will deliver a commodity product: basic, practical, low-prestige information that helps people buy things or find their way around. Whether it will help people conduct their family lives or romantic relationships is open to question – though one noteworthy possibility is that it could overcome the language barriers that often arise between generations after migration, leaving children and their grandparents without a shared language.

Whatever uses it is put to, though, it will never be as good as the real thing. Even if voice-morphing technology simulates the speaker's voice, their lip movements won't match, and they will look like they are in a dubbed movie.

The contrast will underline the value of shared languages, and the value of learning them. Making the effort to learn someone's language is a sign of commitment, and therefore of trustworthiness. Sharing a language can also promote a sense of belonging and community, as with the international scientists who use English as a lingua franca, where their predecessors used Latin. Immigrant shopkeepers who learn their customers' language are not just making sales easier; they are showing that they wish to draw closer to their customers' community, and politely asserting a place in it.

When machine translation becomes a ubiquitous commodity product, human language

skills will command a premium. The person who has a language in their head will always have the advantage over somebody who relies on a device, in the same way that somebody with a head for figures has the advantage over somebody who has to reach for a calculator. Though the practical need for a lingua franca will diminish, the social value of sharing one will persist. And software will never be a substitute for the subtle but vital understanding that comes with knowledge of a language. That knowledge will always be needed to pick the nuances from the noise.

Marek Kohn's *Four Words for Friend: Why Using More Than One Language Matters Now More Than Ever* is published by Yale University Press (£20). To order a copy go to guardianbookshop.com or call 0330 333 6846. Free UK p&p over £15, online orders only. Phone orders min p&p of £1.99

Since you're here...

... we have a small favour to ask. We have a core mission which is driving the work that we do. It's about ensuring our journalism is kept open and accessible to readers all around the world.

More people are reading and supporting our independent, investigative reporting than ever before. And unlike many news organisations, we have chosen an approach that allows us to keep our journalism accessible to all, regardless of where they live or what they can afford.

Readers' support powers our work, giving our reporting impact and safeguarding our essential editorial independence. This means the responsibility of protecting independent journalism is shared, enabling us all to feel empowered to bring about real change in the world. Your support gives Guardian journalists the time, space and freedom to report with tenacity and rigor, to shed light where others won't. It emboldens us to challenge authority and question the status quo. And by keeping all of our journalism free and open to all, we can foster inclusivity, diversity, make space for debate, inspire conversation - so more people, across the world, have access to accurate information with integrity at its heart.

The Guardian is editorially independent, meaning we set our own agenda. Our journalism is free from commercial bias and not influenced by billionaire owners, politicians or shareholders. No one edits our editor. No one steers our opinion. This is important as it enables us to give a voice to those less heard, challenge the powerful and hold them to account. It's what makes us different to so many others in the media, at a time when factual, honest reporting is critical.

Every contribution we receive from readers like you, big or small, goes directly into funding our journalism. This support enables us to keep working as we do - but we must maintain and build on it for every year to come. **Support The Guardian from as little as £1 - and it only takes a minute. Thank you.**

Support The Guardian

Topics

- Technology

- The Observer
- Artificial intelligence (AI)
- Language resources
- features