

Face of the future rears its head: 'Zoe' uses a basic set of six simulated emotions (w/ video)

March 19 2013



(Phys.org) —Meet Zoe: a digital talking head which can express human emotions on demand with "unprecedented realism" and could herald a new era of human-computer interaction.

A virtual "talking head" which can express a full range of human emotions and could be used as a digital <u>personal assistant</u>, or to replace <u>texting</u> with "face messaging", has been developed by researchers.



The lifelike face can display emotions such as happiness, anger, and fear, and changes its voice to suit any feeling the user wants it to simulate. Users can type in any message, specifying the requisite emotion as well, and the face recites the text. According to its designers, it is the most expressive controllable avatar ever created, replicating human emotions with unprecedented realism.

The system, called "Zoe", is the result of a collaboration between researchers at Toshiba's Cambridge Research Lab and the University of Cambridge's Department of Engineering. Students have already spotted a striking resemblance between the disembodied head and Holly, the ship's computer in the British sci-fi comedy, <u>Red Dwarf</u>.

Appropriately enough, the face is actually that of Zoe Lister, an actress perhaps best-known as Zoe Carpenter in the Channel 4 series, Hollyoaks. To recreate her face and voice, researchers spent several days recording Zoe's speech and <u>facial expressions</u>. The result is a system that is light enough to work in <u>mobile technology</u>, and could be used as a personal assistant in smartphones, or to "face message" friends.

The framework behind "Zoe" is also a template that, before long, could enable people to upload their own faces and voices - but in a matter of seconds, rather than days. That means that in the future, users will be able to customise and personalise their own, emotionally realistic, digital assistants.

If this can be developed, then a user could, for example, text the message "I'm going to be late" and ask it to set the emotion to "frustrated". Their friend would then receive a "face message" that looked like the sender, repeating the message in a frustrated way.

The team who created Zoe are currently looking for applications, and are also working with a school for autistic and deaf children, where the



technology could be used to help pupils to "read" emotions and lip-read. Ultimately, the system could have multiple uses – including in gaming, in audio-visual books, as a means of delivering online lectures, and in other user interfaces.

"This technology could be the start of a whole new generation of interfaces which make interacting with a computer much more like talking to another human being," Professor Roberto Cipolla, from the Department of Engineering, University of Cambridge, said.

"It took us days to create Zoe, because we had to start from scratch and teach the system to understand language and expression. Now that it already understands those things, it shouldn't be too hard to transfer the same blueprint to a different voice and face."

As well as being more expressive than any previous system, Zoe is also remarkably data-light. The program used to run her is just tens of megabytes in size, which means that it can be easily incorporated into even the smallest computer devices, including tablets and smartphones.

It works by using a set of fundamental, "primary colour" emotions. Zoe's voice, for example, has six basic settings - Happy, Sad, Tender, Angry, Afraid and Neutral. The user can adjust these settings to different levels, as well as altering the pitch, speed and depth of the voice itself.

By combining these levels, it becomes possible to pre-set or create almost infinite emotional combinations. For instance, combining happiness with tenderness and slightly increasing the speed and depth of the voice makes it sound friendly and welcoming. A combination of speed, anger and fear makes Zoe sound as if she is panicking. This allows for a level of emotional subtlety which, the designers say, has not been possible in other avatars like Zoe until now.



To make the system as realistic as possible, the research team collected a dataset of thousands of sentences, which they used to train the speech model with the help of real-life actress, Zoe Lister. They also tracked Lister's face while she was speaking using computer vision software. This was converted into voice and face-modelling, mathematical algorithms which gave them the voice and image data they needed to recreate expressions on a digital face, directly from the text alone.

The effectiveness of the system was tested with volunteers via a crowdsourcing website. The participants were each given either a video, or audio clip of a single sentence from the test set and asked to identify which of the six basic emotions it was replicating. Ten sentences were evaluated, each by 20 different people.

Volunteers who only had video and no sound only successfully recognised the emotion in 52% of cases. When they only had audio, the success rate was 68%. The two together, however, produced a successful recognition rate of 77% - slightly higher than the recognition rate for the real-life Zoe, which was 73%! This higher rate of success compared with real life is probably because the synthetic talking head is deliberately more stylised in its manner.

As well as finding applications for their new creation, the research team will now work on creating a version of the system which can be personalised by users themselves.

"Present day <u>human-computer interaction</u> still revolves around typing at a keyboard or moving and pointing with a mouse." Cipolla added. "For a lot of people, that makes computers difficult and frustrating to use. In the future, we will be able to open up computing to far more people if they can speak and gesture to machines in a more natural way. That is why we created Zoe - a more expressive, emotionally responsive face that human beings can actually have a conversation with."



Provided by University of Cambridge

Citation: Face of the future rears its head: 'Zoe' uses a basic set of six simulated emotions (w/ video) (2013, March 19) retrieved 10 May 2024 from <u>https://phys.org/news/2013-03-future-rears-zoe-basic-simulated.html</u>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.