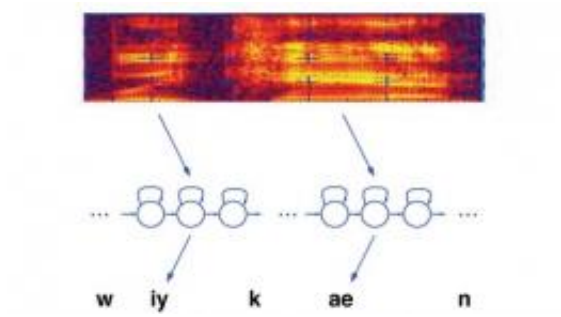
Smart listeners and smooth talkers

November 17 2011



Sounds recognised from an audio recording. Credit: Xunying Liu

Human-like performance in speech technology could be just around the corner, thanks to a new research project that links three UK universities.

Human conversation is rich and it's messy. When we communicate, we constantly adjust to those around us and to the environment we're in; we leave words out because the context provides meaning; we rush or hesitate, or change direction; we overlap with other speakers; and, crucially, we're expressive.

No wonder then that it's proved so challenging to build machines that interact with people naturally, with human-like performance and behaviour.

Nevertheless there have been remarkable advances in speech-to-text technologies and speech synthesizers over recent decades. Current

devices speed up the transcription of dictation, add automatic captions to video clips, enable automated ticket booking and improve the quality of life for those requiring assistive technology.

However, today's speech technology is limited by its lack of ability to acquire knowledge about people or situations, to adapt, to learn from mistakes, to generalise and to sound naturally expressive. "To make the technology more usable and natural, and open up a wide range of new applications, requires field-changing research," explained Professor Phil Woodland of Cambridge's Department of Engineering.

Along with scientists at the Universities of Edinburgh and Sheffield, Professor Woodland and colleagues Drs Mark Gales and Bill Byrne have begun a five-year, £6.2 million project funded by the Engineering and Physical Sciences Research Council to provide the foundations of a new generation of speech technology.

Complex pattern matching

Speech technology systems are based on powerful techniques that are capable of learning statistical models known as Hidden Markov Models (HMMs). Trained on large quantities of real speech data, HMMs model the relationship between the basic speech sounds of a language and how these are realised in audio waveforms.

It's a complex undertaking. For speech recognition, the system must work with a continuous stream of acoustic data, with few or no pauses between individual words. To determine where each word stops and starts, HMMs attempt to match the pattern of successive sounds (or phonemes) to the system's built-in dictionary, assigning a probability score as to which sounds are most likely to follow the first sound to complete a word. The system then takes into account the structure of the language and which word sequences are more likely than others.

Adapt, train and talk

A key focus for the new project is to build systems that are adaptive, enabling them to acclimatise automatically to particular speakers and learn from their mistakes. Ultimately, the new systems will be able to make sense of challenging audio clips, efficiently detecting who spoke what, when and how.

Unsupervised training is also crucial, as Professor Woodland explained: “Systems are currently pre-trained with the sort of data they are trying to recognise – so a dictation system is trained with dictation data – but this is a significant commercial barrier as each new application requires specific types of data. Our approach is to build systems that are trained on a very wide range of data types and enable detailed system adaptation to the particular situation of interest. To access and structure the data, without needing manual transcripts, we are developing approaches that allow the system to train itself from a large quantity of unlabelled speech data.”

“One very interesting aspect of the work is that the fundamental HMMs are also generators of speech, and so the adaptive technology underlying speech recognition is also being applied to the development of personalised speech synthesis systems,” added Professor Woodland. New systems will take into account expressiveness and intention in speech, enabling devices to be built that respond to an individual’s voice, vocabulary, accent and expressions.

The three university teams have already made considerable contributions to the field and many techniques used in current speech recognition systems were developed by the engineers involved in the new project. The new programme grant enables them to take a wider vision and to work with companies that are interested in how speech technology could transform our lives at home and at work. Applications already planned

include a personalised voice-controlled device to help the elderly to interact with control systems in the home, and a portable device to enable users to create a searchable text version of any audio they encounter in their everyday lives.

Provided by University of Cambridge

Citation: Smart listeners and smooth talkers (2011, November 17) retrieved 9 April 2024 from https://phys.org/news/2011-11-smart-smooth-talkers.html

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.