

Computer program recognises any language

August 22 2012, by Norunn K. Torheim/Else Lie



Credit: Unsplash/CC0 Public Domain

If computers are rendered capable of recognising speech it will one day be the norm to give commands by voice rather than via a keyboard. “Speaking” with a mobile phone is already commonplace for many people. The technology can also be used for searching through an audio archive for files or films on the Internet.

But achieving good [speech recognition](#) is a difficult task. [Spoken language](#) differs widely from written language and there is wide variation in spoken language between individuals, such as differences in dialects.

Everyone sounds alike

With funding from the Large-scale Programme on Core Competence and Value Creation in ICT (VERDIKT) under the Research Council of Norway, Professor Torbjørn Svendsen of the Norwegian University of Science and Technology (NTNU) and fellow research colleagues have been testing an innovative approach to creating next-generation speech recognition technology.

The Norwegian researchers have demonstrated that the production of human speech is fundamentally the same across languages. As such, the technology being developed will be applicable to any language without being reliant on speech data for each individual language to train a machine.

The researchers based their approach on phonetics – that is, the study of the sounds of human speech. They have also incorporated additional speech and language knowledge into the system, for example the correspondence between sound frequency and words and how words are put together in forming sentences.

The method developed by Dr Svendsen and colleagues involves training a computer to determine which parts of the speech organs are in activity based on analysis of the pressure of sound waves captured by the microphone.

One language at a time

Up to now, two different approaches to speech recognition systems have been most prevalent. Both are based on the use of speech data and source texts in training a computer to recognise different languages on an individual basis.



Credit: Unsplash/CC0 Public Domain

The one approach involves individuals observing words and sounds and deducing rules which are then entered into the computer. For instance, whether or not a sound is voiced depends on whether the vocal cords vibrate during the production of the sound.

“If we analyse a small speech segment and determine that a specific sound is voiced in speech peaking at resonances of 750 and 1200 hertz (Hz) then the sound is likely an ‘a’. If the peaks range between 350 and 800 Mz, it’s likely to be a ‘u’,” Professor Svendsen points out.

The other approach is to leave the training up to the computer by feeding it large amounts of sample material.

“Initially, a machine perceives all sound events to be equally probable. But as the data-driven learning proceeds, occurrences with higher frequency are interpreted as more likely while less common occurrences decrease on the probability scale,” Dr Svendsen explains.

“This type of approach enables us to process much more speech data than we can using human-based observations. There are just limits to how much data a human can handle.”

Classifying sounds

The research group has chosen an approach somewhere between the two traditional approaches.

“We have great confidence in the statistical approach. However, we also need to give consideration to the patterns of predictability that exist in speech in the real world. The researchers include relevant information about this into the system and combine data-driven learning and a rule-based approach.

Speech patterns differ due to variations in the physiology, dialect, education and health of individuals. This all affects the production of voice and sentence structure. In order for a machine to learn how to understand speech, it must be able to discern among the most common variations in normal speech and language.



Credit: Unsplash/CC0 Public Domain

“We are currently developing a computer program which determines the probability of various distinctive characteristics being present or absent during sound production – for example, if there is vocal cord vibration, this indicates the occurrence of a voiced sound. This is our method of classifying sounds,” Professor Svendsen relates.

Identifying a language in a matter of seconds

The next step for the Norwegian researchers is to develop a language-independent module for use in designing competitive [speech](#) recognition

products.

“The solutions will result in savings both in terms of time and money. It is an important technology, not only for people who are part of a minor language group such as Norwegian. There are a staggering number of languages with only a few million speakers that would benefit greatly from such tools,” concludes Dr Svendsen.

A by-product is that this type of technology can be useful in contexts where several different languages are being used at once. It takes only in 30 to 60 seconds to identify a given spoken language. This can be helpful in instances where, for example, a person giving a presentation in one language cites a quote in another. It can also be significant in investigative work to determine quickly which [language](#) an individual is speaking.

Provided by The Research Council of Norway

Citation: Computer program recognises any language (2012, August 22) retrieved 18 April 2024 from <https://phys.org/news/2012-08-recognises-language.html>

<p>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.</p>
