

Norwegian researchers succeed in creating an artificial child's voice

February 21 2012

It is very difficult to get a PC to recognise the voice of a child. Equally problematic is using a computer to synthesise speech in a child's voice. Norwegian researchers have found simple, effective solutions to both challenges.

“Synthesised speech has grown more and more similar to human speech. Yet children communicating via a speech device are still forced to use a synthetic adult voice,” explains Magne Lunde, Managing Director of Media LT, a company developing tools to assist disabled persons.

This drawback was the driver behind a collaborative research project involving MediaLT and Lingit, a software company. Together they are developing Norway's first synthesised childlike voice.

Using funding granted under the Research Council programme ICT for the disabled (IT Funk), they are putting an entirely new method to the test.

Converting the master voice into a comprehensible child's voice

“We start with what is known as a master voice, which is the product of three or four adult speakers recording several thousands of phrases. Then we record a single child reading a smaller number of phrases aloud. We use this recording to modify the master voice, making it sound like a

child's voice," relates Torbjørn Nordgård from Lingit. Dr Nordgård is also a professor of linguistics at the University of Nordland.

The phrases recorded by the child have been selected to include a number of the most essential sounds found in Norwegian.

"The master voice still carries the intonation, i.e. a phrase's melody. The result sounds rather like a child with unusual elocution skills, but it's still much better than the voice of an adult," says Mr Nordgård.

Far ahead of the rest of the world

Very little research has been carried out on this subject internationally. MediaLT and Linget's innovative method of synthesising a child's voice is bringing them to the forefront of their field.

Everything is now in place to start testing trial versions of the child's voice.

"We hope to have a beta version in place this summer," says Magne Lunde.

PC's need to understand children's speech

Mr Lunde and his colleagues are also researching voice control such as use of verbal commands to operate a PC.

In order to operate a computer by means of speech, the machine must successfully decipher what is being said. Interpreting the speech of individuals on both the young and the older end of the scale is especially challenging since the distance from their vocal cords to their lips is shorter than that of the average adult.

“Teaching a speech recognition program to understand the pronunciation of the various sounds of a language requires a relatively large amount of recorded speech. Unfortunately, insufficient data exist today in terms of actual children’s speech,” states Professor Torbjørn Svendsen from the Norwegian University of Science and Technology.

Professor Svendsen and his research partners have come up with a very elegant, yet simple method of overcoming the challenges associated with speech recognition and children – they have synthesised children’s voices and used the results to compile a collection of data.

A vast improvement in quality

The length of the vocal tract affects the frequency distribution of the speech energy. The researchers are using technology to render the energy distribution of adult speech so that it more closely resembles that of a child.

“The converted adult speech resembles the way children speak in terms of sound as well. Thus, we could apply our conversion technique to a large database of adult speech and generate a functional database of artificial childlike voices. We then used this to train a separate speech recognition program for children,” explains Professor Svendsen.

“This greatly improved the recognition fidelity of children’s speech. The error rate was reduced by 50 to 70 per cent,” he states.

Activities are being carried out in cooperation with the researchers in the [Voice](#) control in multimodal dialogue (SMUDI) project, which received funding from the Research Council’s Large-scale programme Core Competence and Value Creation in ICT (VERDIKT) and the Ministry of Education and Research.

Norwegian a tough language

The Norwegian language poses a number of especially steep challenges to speech recognition experts.

“In general, the degree of variation in any language is large enough to make it difficult to model. But Norwegian is especially tricky; there are two distinct written forms of the language, countless dialects and a wide range of accepted alternatives for words, declensions and compounds. On top of all this, there is no single pronunciation standard,” stresses Torbjørn Svendsen.

Dr Svendsen also points out that people can experience considerable difficulty when faced with voice-controlled devices.

“It is easy to get caught up in our fascination with [speech recognition](#) and the many possibilities it holds, so it’s important not to replace existing technology when it remains the best option for getting something done – like using buttons to operate a lift,” he concludes.

Provided by The Research Council of Norway

Citation: Norwegian researchers succeed in creating an artificial child's voice (2012, February 21) retrieved 27 July 2024 from <https://phys.org/news/2012-02-norwegian-artificial-child-voice.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.