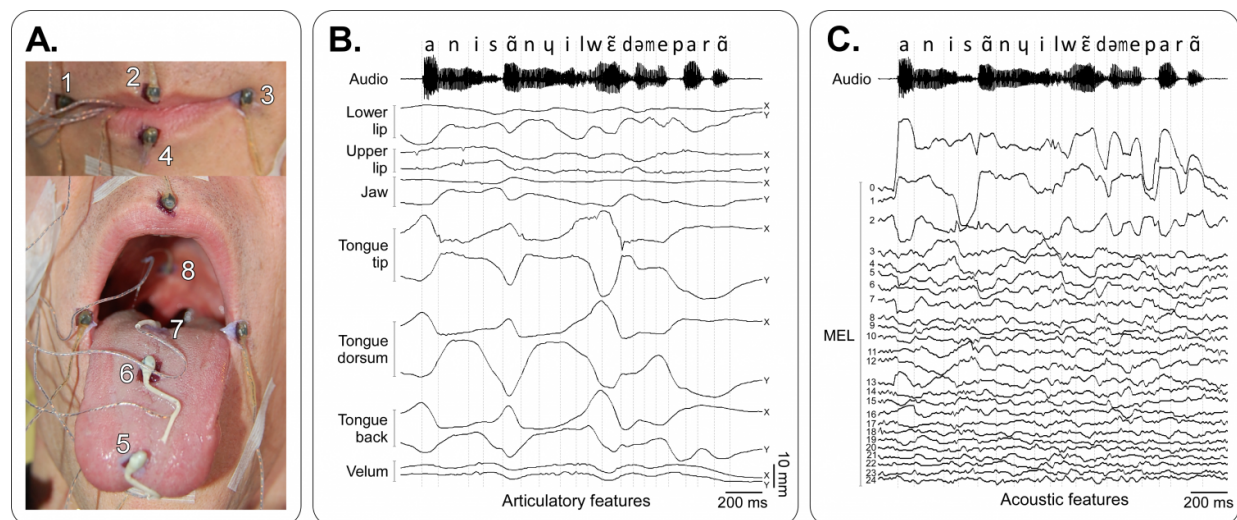# Speech synthesizer designed to work out mouth movements into words

November 29 2016, by Nancy Owano

Articulatory and acoustic data. A–Positioning of the sensors on the lip corners (1 & 3), upper lip (2), lower lip (4), tongue tip (5), tongue dorsum (6), tongue back (7) and velum (8). The jaw sensor was glued at the base of the incisive (not visible in this image). B–Articulatory signals and corresponding audio signal for the sentence "Annie s'ennuie loin de mes parents" ("Annie gets bored away from my parents"). For each sensor, the horizontal caudo-rostral X and below the vertical ventro-dorsal Y coordinates projected in the midsagittal plane are plotted. Dashed lines show the phone segmentation obtained by forced-alignment. C–Acoustic features (25 mel-cepstrum coefficients—MEL) and corresponding segmented audio signal for the same sentence as in B. Credit: *PLOS Computational Biology* (2016). DOI: 10.1371/journal.pcbi.1005119

(Tech Xplore)—French scientists have worked on a speech synthesizer designed for people who have vocal cord paralysis. They have put nine sensors to work capturing lip movements, tongue, jaw, soft palate.

A neural network was a key factor; it learned to convert data into speech emitted from a vocoder. Sam Wong in *New Scientist* commented, "Vocoders just got a serious upgrade."

(What is the role of a vocoder? One explanation is that "A vocoder aims to replace the carrier of your voice with another carrier from another source. Thus, it changes the sound of the voice but not the message when you speak.")

A video shows the researchers' work in action. Thing is, what you hear is not a person's speech but "robot" speech in monotone.

Wong in *New Scientist* remarked that "Although the synthesiser might not be immediately useful, it's a first step towards building a brain-computer interface that could allow paralysed people to talk by monitoring their thought patterns."

How easy will it be to make progress?

Wong said in *New Scientist* that "Recent research has shown that the speech area of the motor cortex contains representations of the various parts of the mouth that contribute to speech, suggesting it might be possible to translate activity in that region into signals like the sensor data used in the synthesiser."

Abigail Beall in *Daily Mail* said Monday that takes away the need for any voicebox, as the translation from mouth to speech is direct

For anyone interested in learning more about their methods and results,

the researchers' work has been published in *PLOS Computational Biology*. This is a journal of the International Society for Computational Biology (ISCB).

The article is "Real-Time Control of an Articulatory-Based Speech Synthesizer for Brain Computer Interfaces." The authors, from France, are Florent Bocquelet, Thomas Hueber, Laurent Girin, Christophe Savariaux and Blaise Yvert.

The authors stated that "Restoring natural speech in paralyzed and aphasic people could be achieved using a Brain-Computer Interface (BCI) controlling a speech synthesizer in real-time."

The authors said their synthesizer is based on a machine-learning approach. The data recorded by electro-magnetic articulography (EMA) is converted into acoustic speech signals using deep neural networks.

They showed that "intelligible speech could be obtained in a closed-loop paradigm by different subjects controlling this synthesizer in real time from EMA recordings while articulating silently, i.e. without vocalizing. Such a silent speech condition is as close as possible to a speech BCI paradigm where the synthetic voice replaces the actual subject voice."

They said nine 3-D coils were glued on the tongue tip, dorsum, and back, as well as on the upper lip, the lower lip, the left and right lip corners, the jaw and the soft palate.

The authors said that all speakers were silently articulating and were given the synthesized acoustic feedback through headphones.

  More information: Florent Bocquelet et al. Real-Time Control of an Articulatory-Based Speech Synthesizer for Brain Computer Interfaces, *PLOS Computational Biology* (2016). DOI:

Abstract

Restoring natural speech in paralyzed and aphasic people could be achieved using a Brain-Computer Interface (BCI) controlling a speech synthesizer in real-time. To reach this goal, a prerequisite is to develop a speech synthesizer producing intelligible speech in real-time with a reasonable number of control parameters. We present here an articulatory-based speech synthesizer that can be controlled in real-time for future BCI applications. This synthesizer converts movements of the main speech articulators (tongue, jaw, velum, and lips) into intelligible speech. The articulatory-to-acoustic mapping is performed using a deep neural network (DNN) trained on electromagnetic articulography (EMA) data recorded on a reference speaker synchronously with the produced speech signal. This DNN is then used in both offline and online modes to map the position of sensors glued on different speech articulators into acoustic parameters that are further converted into an audio signal using a vocoder. In offline mode, highly intelligible speech could be obtained as assessed by perceptual evaluation performed by 12 listeners. Then, to anticipate future BCI applications, we further assessed the real-time control of the synthesizer by both the reference speaker and new speakers, in a closed-loop paradigm using EMA data recorded in real time. A short calibration period was used to compensate for differences in sensor positions and articulatory differences between new speakers and the reference speaker. We found that real-time synthesis of vowels and consonants was possible with good intelligibility. In conclusion, these results open to future speech BCI applications using such articulatory-based speech synthesizer.

November 29) retrieved 4 May 2024 from https://techxplore.com/news/2016-11-speech-mouth-movements-words.html