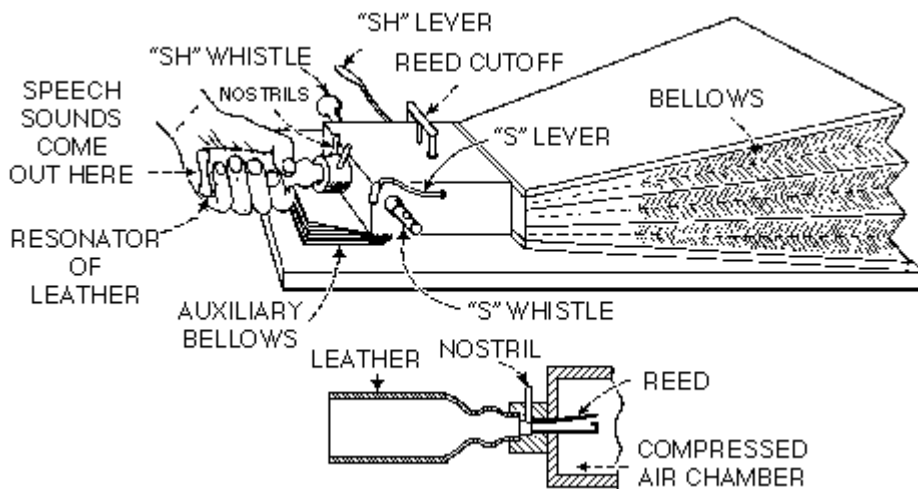


# *the dream of an speaking machine ...*

## **von Kempelen (1791)**



The dream of machines that are similar to us humans, for example, by speaking, is old. Although there are several prehistoric sources that tell of monks that constructed "talking" heads, so is dating of the first well-documented speech machine to the 1791. It was constructed by Wolfgang von Kempelen (1734-1804) who was a talented Hungarian inventor and mechanical engineer. The figure does not show the speech machine in the original, but the reconstruction which was later built by Whetstone (1835).

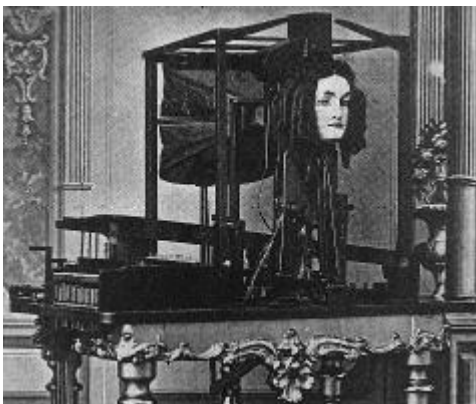
The mechanical speech machine tries to mimic the human body's organs of speech. A bellows replaced the lungs, a reed (from a woodwind) replaced the vocal cords. A resonator of leather replaced the lips. By hand maneuver leather tube could be different vowel sounds are produced. The consonants are obtained by releasing the air by four different hand-operated passages.

Now, if a machine could sound like an human - could we not also build mechanical artificial intelligence? von Kempelen answered this question with "yes!" and built the chess-playing machine "the Turk" exhibited publicly in many cities. Napoleon Bonaparte is said to have played against the machine - and this chess game he lost!

As you probably can guess the machine were a scam, but this is another story ...

[ [the Turk](#) ]

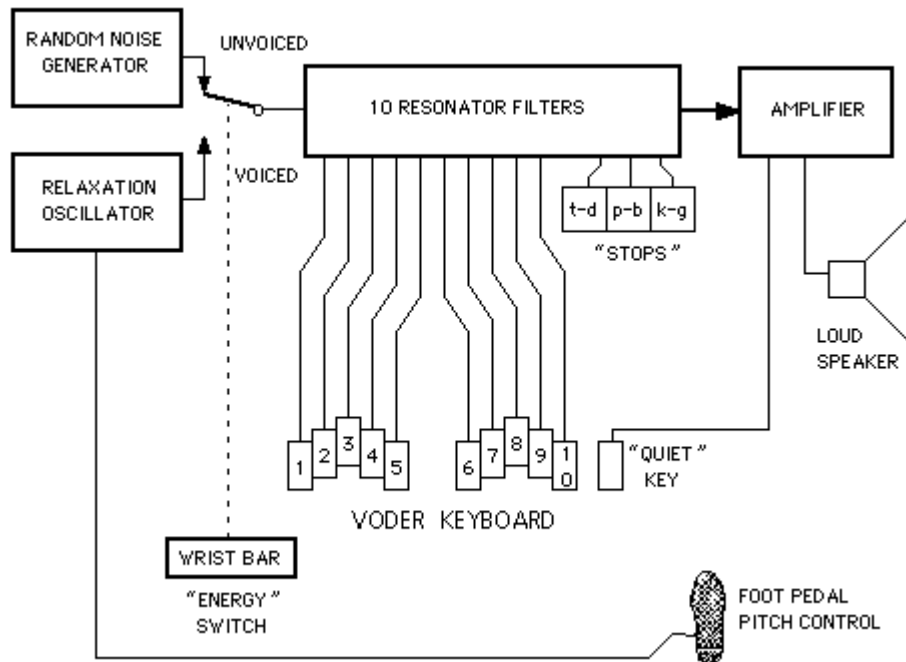
## **Faber's "Euphonia" (1830-1840)**



During the 1830s a German immigrant designed the "talking machine Euphonia". The machine "was operated" with a keyboard (like an organ).

After ten years of work on refining this mechanical instrument so was many of the spoken words and sentences good (With a trained operator at the keys). Unfortunately, the machine brought little interest among the public, although it could also sing "God Save the Queen" (1846).

## H. Dudley the "Voder" (Bell labs 1939)



During the 1930s, the mechanical technology was replaced by electronics in speech machines. Voder is the first electric speech synthesis machine. It consisted of a noise generator and a tone generator and 10 different switchable resonant circuits (Comparable to an amplifier with an "equalizer"). It was operated with a keyboard. Voder was demonstrated by a trained operator at the World Fair in New York in 1939. Although it took relatively long workout (one year or more) to manage the instrument so the operators could "play" on the machine and produce intelligible speech.

An audio sample of "Voder" from a radio program at the World Expo 1939 can be heard on the website [Klatt's "History of speech synthesis"](#)

## G. Fant "Ove I, II, III" (1953 ...)



Speech synthesis is an international research area. Although much is common to all spoken languages, then it's still a lot more that divides them. If you want there to be Swedish synthetic speech, then it will be required with Swedish research on speech synthesis.

Gunnar Fant at [KTH](#) is the one who started and built up the research on speech synthesis in Sweden. The picture shows the electronic synthesizer machine "Ove" which existed in various "Generations" from 1953 onwards. The speech machine was operated with a hand over a "compound table".

An audio sample of "Ove" from 1953 can be heard on the website [Klatt's "History of speech synthesis"](#)

## Speech synthesis with computer

Starting in the 1970s, the computers were so developed that all speech synthesis research has used the computer as a tool to speak and understand speech. The computers did speech synthesis practically useful, and many programs that translated from text to speech were developed. The technology got commercialized use.

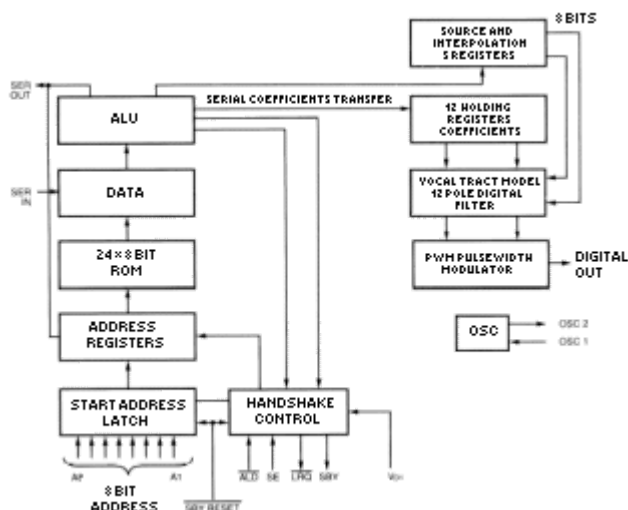
### The chip SP0256 1980 from (Data General) General Instruments/Microchip



Around 1980 speech synthesizer chips were introduced, SP0256 from Data Instruments was used in the hobby computers and video games. The circuit could pronounce all the English phonemes (vowels and consonants) with some variations (allophones) a total of 59. This enabled it to say everything that can be said in the English language.

The chip SP0256 production has stopped since long ago, and now it receives "cult status" and therefore commands a high price in online auctions.

To our lab [the speaking processor](#) we use a cheap PIC-processor (PIC16F690) and an EEPROM with 64Kbyte of recorded samples from an actual SP0256-chip.



## [Inside SP0256 \(reverse engineering\)](#)

Another speech chip from another manufacturer, Texas Instruments, were found in the teaching material/toy Speak & Spell who practiced child in spelling like to do before a spell checker questioning.



An audio sample of "Speak and Spell" from 1980 can be heard on the website [Klatt's "History of speech synthesis"](#)

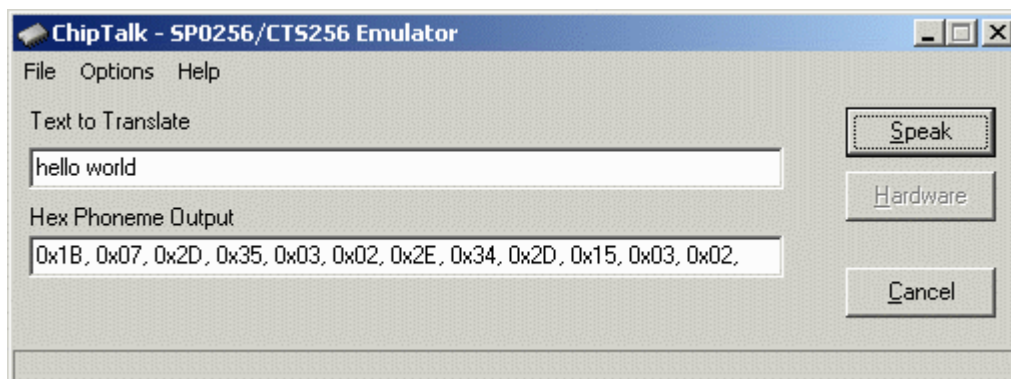
## Text to fonem program

To determine the result of which allophone sequences that will give the best pronunciation of an English word is time-consuming. One has to try it out among allophones given in tabular form.

- [SPO256 Allophone table](#)

At the "text to speech synthesis" it requires a computer program that translates to allophones automatically. It would require a set of rules, as "the Naval Research Labs TTS algorithm" (from the American Navy research in the 1970's).

- For example. Tom Jennings has written a C program (based on a program of John Wasser) which adapts this algorithm to chip SP0256. [T2A: Phoneme text-to-speech software](#) [t2a.zip](#)
- The company [SpeechChips](#) has made an SPO256-AL2/CTS256-AL2 emulator. With it you can directly enter the English texts to hear how it would sound, and see what allophones are to be used. [ChipTalkSetup.exe](#).



## Why does speech-chips sound so bad?

Written language letters correspond to spoken language phonemes. We are all familiar with this principle because we tried to learn to read by "sounding" words. When "sounding" you hear "roughly" how the word sounds and can guess what that is. The relationship between letters and phonemes therefore exists, but this is far from the whole truth of the story!

Allophone theory says that a phoneme has alternative pronunciation depending on where in the word it is (eg. In the beginning, middle or at the end). That it would be enough with the 59 allophones to speak English, however, is wishful thinking, in any case, if you want it to sound natural! This truth we chose to turn a blind eye in the 1980s, because at those days circuit technologies still could not suffice for so much more.

If you want to make a synthesis voice today, One uses diphones, which are phonemes in pairs. Theoretical there are  $43 \times 43 = 1849$  ordinary difones in Swedish language ( all combination of before and after each other ) + 86 phonem-pause combinations. Some combinations can be ruled out but still remains about 1658.



A recording is done by sitting at the computer with a headset and saying the "nonsense words". Recording takes about 20 minutes and those who are interested can learn about the whole process from the researcher *Adina Svenssons* [D-uppsats](#) "Ofelia - kvinnlig syntesröst med västsvensk dialekt". [MBROLA-project](#) tries trying to collect and document synthesis voices from around the world, and here you can listen to the result of Adina Svensson synthesis voice "[sw2 female voice](#)".

[ [Back](#) ]

---

© William Sandqvist [william@kth.se](mailto:william@kth.se)