Let's build a speakwrite!

Source: http://sci.tech−archive.net/Archive/sci.electronics.design/2005−02/5839.html

---

From: Rich Grise (*richgrise_at_example.net*)
Date: 02/26/05

```
Date: Sat, 26 Feb 2005 04:59:12 GMT
```

This might be a little long and rambling, but it's entirely
on−topic, and in the realm of reality. ;−)

Not too many years ago, I built the hardware half of a voice−input
thing. There was some kind of voice−actuated butler thing on the
market at the time, and PBS had shows about voice−controlled robots.
As it happened, I met a guy who was building robots for home helper
applications, who had heard about the home butler thing, and was
wondering if it was possible to use the outputs from the home butler
thing to control his household robot. It was outputting X10, but
had a serial output, as I remember.

"Cool!" sez I to meself. He had the robot going, by Futaba, and
first, we had to see if I could design a box that could simulate
the selector switch inputs, and transmit control signals via RC.
I did a little control−box−to−Futaba−simulator with an 8X51 −
I think I used an 8035, because an 8035 + 2764 was cheaper than
an 8751 at the time. And I had a 2764 programmer. ;−)

Anyways, during the development of this remote control, the idea
of voice control came up, and it got to the point where we were
actually doing a demo − the voice robot guys had given us a voice
butler that could output serial bytes based on the voice command,
and my interface translated that into robot control signals.

The very first demo worked, for the first voice command, and then
turned into the most embarrasing fiasco I've ever been a party to.

The command was "Up". The robot started to raise its arm, and the
noise from the cheap gearbox/leadscrew actuator thing swamped the
voice−input thing, and we didn't have a panic button. It went
all the way to the stop, the robot broke itself, and we got thrown
out of the voice−butler thing guys' office.

Another problem with that particular voice−input thing was you had
to go through a pre−programmed menu of available commands, and
they wouldn't modify it for anything or anybody − the box is
proprietary, and all that, so it was impossible to implement a

"STOP!" command.

So I set about to design and build my own voice controller.

I had seen another thing on PBS, about human perception, that gave me some ideas about voice input. What it was is, they had some human babies, pre−verbal age, and did a sort of Pavlovian experiment. They had this baby sitting in a baby seat, and they'd offer various stimuli, and rewards, and so on. In this case, the reward was some toy that would light up. They'd have people say certain vowel sounds, and after saying certain ones, the toy would light up, the baby would look over at the toy and laugh and so. It got to the point that on hearing a certain vowel sound, or maybe more accurately, phoneme, they'd anticipate the toy, just like Pavlov's dogs.

The attention−getter for me was that for a particular phoneme, it made absolutely no difference which volunteer spoke the phoneme, the baby was responding to the phoneme. The pitch and timbre of the voice didn't make any difference − it was the relationship between the formant frequencies. Maybe "formant frequencies" is a misnomer − it's more like which harmonics are there − but the point is, the spectrum should have the same _shape_ for a given phoneme, when you bandpass filter it to, say, 300 Hz − 3 KHz, which, according to the 1963 Radio Amateur's Handbook, is where all of the information in voice is anyway.

And it's true.

I built a little interface with eight Don Lancaster active filters spaced across the 300−3K Hz band, and used a 68HC11 to send eight bytes every 10 mSec or so to the computer, where I graphed them.

I was halfway there, right? "Ah" always looks like "ah," no matter who says it. Same with "ee", "eigh", "oh", and so on.

Consonants, however, are another matter. But I figured, if I could dope out the phonemes, I could use some kind of Huffman Soundex table to look up the likely words and stuff.

But I got stopped short at trying to match up shapes of spectra to a look−up table or whatever. This might be a good app for a neural net − what kind of progress are they making there?

If they already have affordable voice input stuff, howcome I've never heard of such a thing?

Thanks,
Rich