Biologically Inspired Signal Processing

Breakout Session V

Symposium on Next-Generation ASR

Georgia Tech, 10/7/2003

Why aren't all speech recognizers biologically inspired?

- Too many auditory models were naively applied --- one needs to couple inspiration with good engineering practice.

- Computers in the 1980s were too slow.

- Auditory model (n.): The stuff from physiology & perception that we haven't put into our recognizers yet.

- They are – we just don't know it yet. Methods developed for engineering reasons are often discovered, after the fact, to be similar to biological systems.

Why should we be inspired by biology?

- Speech is the problem, hearing is the solution (Allen, 2003).

- Machine learning methods good at learning parameters, less good at learning structures – biology can suggest structures, e.g., build FSMs to model structures in dendritic connections.

- Speech recognition = mathematical model of speech perception. Collaborative NIH proposals may be possible, e.g., some psychologists are tired of simple stimuli, and want to know what psychophysical results would be of use to engineers building ASR.

Biologically inspired ideas currently under-utilized by ASR

- Localized time-frequency event detection, inspired by cortical revcor maps, phonetics (Stevens' landmark model), and speech perception (e.g. Miller's results).

- Binaural hearing --- making use of fine temporal information in the auditory representation that is currently discarded.

- Pitch --- pitch may be less important for transcribing speech into text, but it is important for tasks important to the organism, e.g. emotion, source localization & separation, semantics & syntax (prosody).

How can we use biological inspiration?

- Needed: a Journal of Negative Results. Biologically inspired ideas are very different from state of the art --- only way to know if they are better or worse is by trying them.

- Needed: micro-tasks on which crazy biomimetic ideas can be tested. Example: Nelson Morgan's innovation ladder, in which novel ideas are successively refined using the numbers task, then a 500-word conversational speech task, then a full evaluation with limited training.