Next Generation ASR: from Speech Perception to Speech Recognition

Sorin Dusan

Center for Advanced Information Processing
Rutgers University
Piscataway, NJ

Next Generation ASR Symposium, Oct. 7-8, 2003

Should Speech Recognition Model Speech Perception?

- No clear evidence that humans recognize phonemes first and then words (although very appealing)

- Some experimental evidence show that syllables can be recognized faster than isolated phonemes. Does that mean that humans do not pre-recognize phonemes in words?

- The linear, microscopic approach (successive phonemes make words) might not be the right approach or the only one

- Speech perception is a bi-directional cognitive process (afferent + efferent); Speech recognition should be a bi-directional process (bottom-up + top-down)

- Speech recognition should not be a unitary, simple, linear, pattern recognition process based on frame and phone "classification" but should integrate new knowledge available from the human cognitive process of speech perception

The "Units" of Speech Perception and Recognition

- What are the elementary units of speech perception: phonemes, syllables, words?

- Do these units exist or is speech perception using a simple, non-linguistic time-window for processing?

- In normal reading there are fixations of 200-250 msec corresponding to about 7(+/-3) letters (➔approx. word units)

- In speech perception, syllables and words appear to be better candidates than phonemes as speech perception units

- But, is there a single type of units in speech perception?

- Humans can perceive various phonological units, but are probably using different processes (➔heterogeneous system)

- Should phoneme be dismissed or just joined by other units?

- Speech recognition could be a heterogeneous process using multiple types of phonological units (phonemes, diphones, syllables, words)

ASR: from a Pattern Recognition Approach Towards a Cognitive Process Approach

- Speech recognition alone is an artificial "problem" very rarely used by humans, if ever

- A direction to proceed for a solution to the **ASR=HSR** equation is to extend the pattern recognition approach to a cognitive process approach by using better and more top-down information

Cognitive processing module that help identify and cope with new-, bad-, or unintelligible-speech, and extra-linguistic information