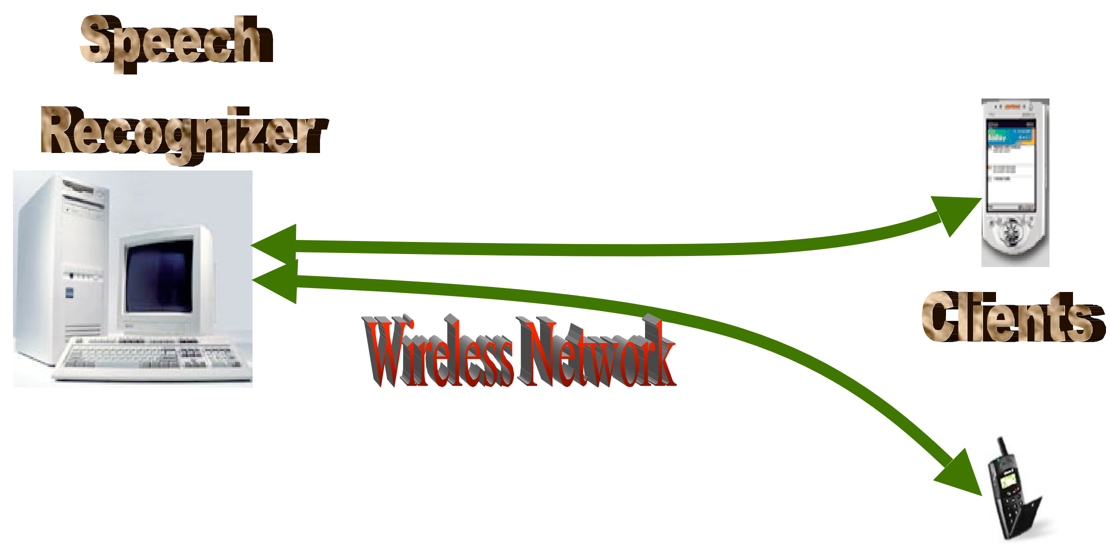
Distributed Speech Recognition

USC | IMSC
UNIVERSITY OF SOUTHERN CALIFORNIA
Integrated Media Systems Center

INTEGRATED MEDIA SYSTEMS CENTER
A National Science Foundation Engineering Research Center at the UNIVERSITY OF SOUTHERN CALIFORNIA

PRINCIPAL INVESTIGATORS

Professor Antonio Ortega
Professor Shrikanth Narayanan



USC STUDENTS, DEGREES

Naveen Srinivasamurthy

BRIEF DESCRIPTION OF DEMONSTRATION

Distributed speech recognition (DSR), enables low complexity (mobile) wireless devices to support speech recognition applications. The acquired speech is transmitted across a wireless network to a remote desktop computer hosting the speech recognizer. To minimize the bandwidth requirement and maximize the battery life of the wireless device, the acquired speech is compressed before transmission. By optimizing the compression performance to the speech recognizer (rather than minimizing perceptual distortion) it is shown that good rate-recognition performance can be achieved. A phoneme recognizer is implemented on the desktop to achieve continuous speech recognition capability.

UNIQUE OR DISTINGUISHING CHARACTERISTICS RELATIVE TO STATE-OF-THE-ART

- Encoding algorithm optimized for speech recognition
- Layered encoding providing a multi-resolution bitstream
- Multi-stage scalable recognition at the server

APPLICATIONS	RECENT HIGHLIGHTS, LEVEL OF DEVELOPMENT, UPCOMING MILESTONES
• Remote recognition • Distributed estimation	• Compression schemes made scalable to provide flexibility in degree of error correction and enabling data driven scalability in the speech recognizer • Mutual information between speech features and phonemes used as a metric for optimizing the encoders for speech recognition applications

UNDERLYING TECHNOLOGIES

- DSR targets the general area of remote speech recognition, in applications where computation/memory requirements precludes light-weight mobile devices from providing state of the art speech recognition services to the user.
- We propose encoding only the features used for recognition instead of the speech waveform to achieve better rate-recognition performance. Furthermore, to handle the problems associated with transmission of data over a wireless channel, the encoders are made scalable. Now the application is able to trade-off between source and channel rates enabling it to adapt to the channel condition and provide best service to the user.
- To handle the computational burden at the server resulting in several clients accessing a server, scalable recognition architectures are adopted. Initial low complexity recognition stages pre-process the data to reduce the search space for later more complex recognizers. To alleviate the overall recognition latency, initial recognition stages make use of the base layers and the later recognition stages make use of the enhancement layers.

LIST OF PUBLICATIONS, REFERENCES, URLs

- http://biron.usc.edu/~snaveen/Publications.html

For additional information, please contact the Principal Investigator listed above via email, or contact

Isaac Maya, Ph.D., P.E. 213-740-2592
Director, Industry and Technology Transfer Programs imaya@imsc.usc.edu

Ann Spurgeon 213-740-4877
Associate Director of Industry Programs aspurgeo@imsc.usc.edu

Integrated Media Systems Center
3740 McClintock Avenue, Suite 131
Los Angeles, CA 90089-2561
213-740-8931 (fax)
For additional information on the Integrated Media Systems Center (IMSC), please visit our Web site at http://imsc.usc.edu