Challenges in Speech Communication Research

Future areas of interest for speech scientists of the Acoustical Society of America

Dani Byrd

USC Linguistics

Presented at the 75th Anniversary Meeting of the Acoustical Society of America; New York, NY USA May 26, 2004

With support from NIH

A Well-Laid Foundation

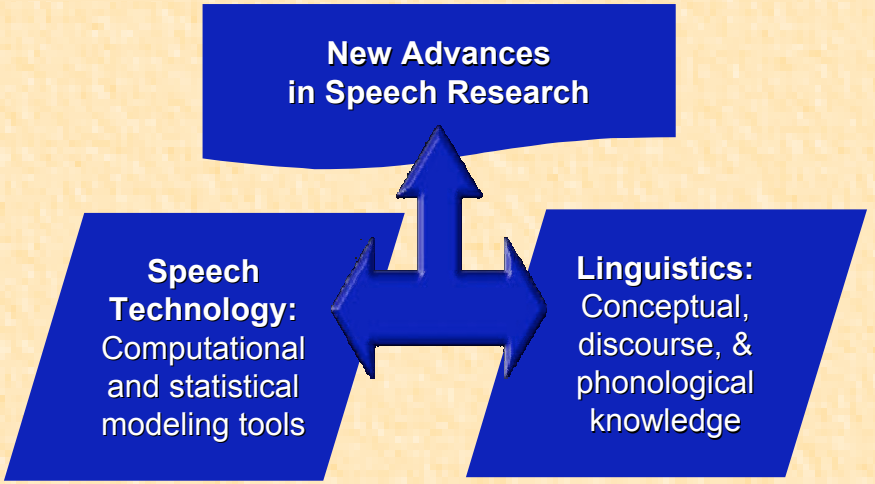
- Acoustical Society research has yielded a detailed understanding of the temporal and spectral properties of the speech signal.
- Strong foundation for pursuing a comparably rich understanding of how this acoustic signal is produced and processed.
  - peripheral: motor, biomechanical, & auditory
  - central: cognitive processing, production, & learning

Future Challenges

Three Promising Areas

1. **Naturalistic spoken language**
2. **Dynamic properties of speech**
3. **A biologically-situated approach**

New Tools Make Advances Possible



First Promising Area

1. **Naturalistic spoken language**
2. **Dynamic properties of speech**
3. **A biologically-situated approach**

Naturalistic Speech

In speech technology	In speech science
• audiovisual	• **Real-world speech variability** in controlled settings
• affectual	– dyadic interactions
• multi-lingual	– spontaneous (non-read) speech
	– multi-modal speech
	– manipulating information structure
	– language exposure in acquisition

Prosody

- "Say _____ again" experiments provided tightly controlled contexts
  - ideal for articulatory & acoustic characteristics of individual speech sounds.
- Now: overt manipulation of **prosodic structure**

Informational groups relevant to cognitive processing (phrasal organization):

When teenagers drive quickly *they get tickets.*

When teenagers drive quickly they get tickets.

Informational prominence within these constituents (focus):

There's no *smoking* on the plane…(but you can drink.)

There's no smoking on the *plane*…(but you can if you take the train.)

Timely Approaches to Prosody

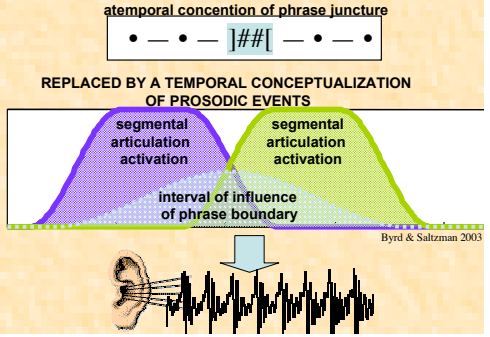
- **How is information structure encoded in articulatory behavior shaping the acoustic signal?**
  - Begin to understand the production and processing of prosodic events, including intonation, as **inherently temporal**.

Phrasal Junctural & Prominence Events:

✓ Interacting with articulatory events

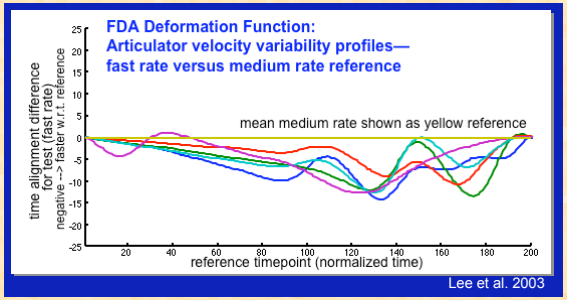
✓ Critically coordinated with those events

✓ Acoustic consequences utilized in real-time cognitive processing



Analyzing Variability

- Statistical tools for quantifying and evaluating **variability**:
- Two examples:
  - connectionist learning models
  - Functional Data Analysis (FDA) (see e.g. Ramsay & Silverman)



Lee et al. 2003

Learning Time-Course & Plasticity

- **Effects of distributional & acoustic variation in the natural language to which learners are exposed**
  - –E.g., rhythmic patterning, input from multiple talkers, patterns obtaining across the lexicon, **distributional frequency & transitional probabilities of phonetic exemplars**

(work by Aslin, Curtin, Maye, Gerken, Holt, Kluender, Kuhl, Lotto, Newport, Saffran, Storkel, Tees, Werker, & others)

Second Promising Area

1. Naturalistic spoken language
2. Dynamic properties of speech
3. A biologically-situated approach

Dynamic Speech

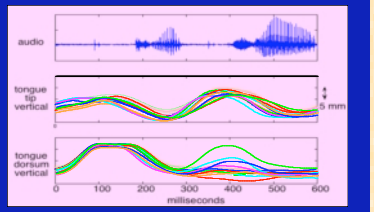
- Production studies will move away from:
  - snap-shot characterizations of vocal tract postures
  - characterizing (static) acoustic cues to phonetic contrasts
- Processing studies will move away from:
  - forced-choice perceptual responses in behavioral studies

An explosion of research in the area of real-time speech production, perception, and processing.

Real-Time Articulation

- Real-time data on articulator patterning and its resulting time-varying aerodynamic and acoustic consequences.

Articulatory point-tracking



Pouplier et al. Subm.

very new advances in
Real-time MRI of the vocal tract

24 frames/sec (no audio)
(Narayanan, Nayak, Lee, Sethy, & Byrd, *JASA* April 2004)

Real-Time Processing

measuring event-related potentials to changes in speech sounds in a 7-month-old infant (by permission, from Kuhl 2004)

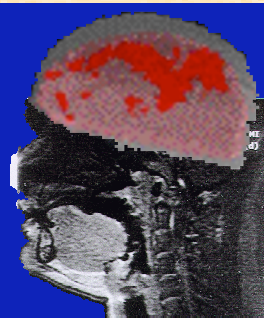
- Real-time cognitive creation and processing of spoken language
  - Time course of speech processing from the auditory to the cognitive
    - For example, MEG, ERP and eye-tracking instruments.
  - Data reduction challenges for high-dimensional data

An Exciting Possibility…

- **Unify studies** of real-time processing using *f*MRI imaging of brain activity during the production of spoken language with studies using real-time high-speed imaging of articulation

Rapidly going back and forth between the **high-speed anatomical imaging and functional brain imaging** during the creation and performance of spoken language

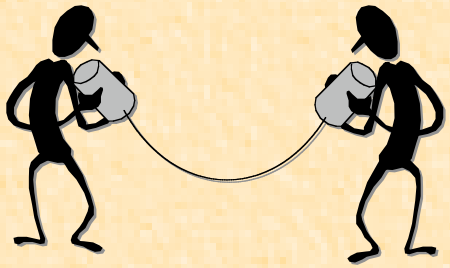
Krishna Nayak (USC): In the future we may be able to pseudo-interleave acquisitions on a 10-50 millisecond basis, with some hardware and *f*MRI SNR improvements.



Third Promising Area

1. Naturalistic spoken language
2. Dynamic properties of speech
3. **A biologically-situated approach**

Biologically-Situated Speech



Language as a process not just shaping but shaped by the coupled interaction among participants in a physical world.

Physiology

Speech acoustics as a key connection allowing improvements in understanding physiologic constraints on spoken language…

Production
Have detailed articulatory-acoustic models. Need **biologically realistic models of the biomechanics of the vocal tract, aerodynamics, & voice production.**

Perception
Have detailed auditory models. Need a better understanding of the **neurophysiology of speech perception.**

Processing
Need knowledge of **cognitive functional organization of the processing of spoken language.**

Complex Systems

- Incorporate knowledge of **complex systems** into mainstream speech research
  - An example: the relevance of the behavior of coupled oscillators to understanding
    - articulatory coordination
    - speaker-listener interactions

NSF's new program solicitation for proposals in the area of Human and Social Dynamics:

> "the dynamics through which individuals and collective entities form, grow, learn, change, and act…; and explorations of the cognitive, computational, linguistic, developmental …[and] biological….processes as dynamic, evolving systems."

Summary

- Future Speech Communication research will take advantage of the detailed knowledge of speech acoustics being provided by ASA scientists to develop a comparably rich understanding of how this acoustic signal is produced and processed.
- Exciting progress can be anticipated in:
  - **naturalistic spoken language**
  - **dynamic properties of speech**
  - **a biologically-situated approach**