Handling Unexpected Acoustic Data

Speaker: Jeff A. Bilmes
University of Washington, Seattle
bilmes@ee.washington.edu

ASRU 2007 Panel
organized by Hynek Hermansky

Outline

- Events that are unexpected to a machine.
- By contrast, humans' ability to
  - Detect unexpected events
  - Deal with unexpected events
- How to improve machine's detection of unexpected events.
- How to improve a machine's strategy for dealing with unexpected events when they occur.

Unexpected Events for Machines

- For a statistical model, “expected” things are those that receive a high probability (or likelihood) score. Example: Shannon games.
  - Anything that receives a high score is expected
  - Anything that receives a low score is unexpected
- Thus, since the model is often incorrect, the “correct” hypothesis could be considered unexpected, and the incorrect one could be expected ➔ This is a Problem!

Unexpected events for Humans

- Two things humans are good at include:
  - Knowing when something unexpected occurred.
  - Recovering and continuing on from the potential errors that result from the occurrence of something unexpected.
- Machines are generally poor at both of the above, so both must be improved in ASR systems to help deal with the unexpected.

Unexpected events for Humans

- Humans do temporal and computational entropy coding
  - Much less time and brain power is spent on the expected portions of speech.
  - Backchannels are spread out over a conversation so the speaker (thinks she) knows the person is quickly processing – this encourages speech patterns by the speaker that are optimized for the case when the listener expecting what she is saying.

Unexpected events for Humans

- Humans do temporal and computational entropy coding
- When the unexpected occurs
  - Much more time is spent trying to figure out what was said.
  - Questions, repetitions, speaking rate, speaking style, key phrases (“I said …”) all help to recover from an error.
  - Brain computation increases: This is when human adaptation occurs, at multiple levels.

Knowing the Unexpected

- ASR systems must be able to accurately know when something that it did not expect occurred.
- What are some strategies to improve machine detection of unexpected events?

Improving Machine Detection of Unexpected Events

- One possible solution is score based, e.g.:
  - Low probability score?
    - but sometimes correct hypothesis gets a low score, so this is insufficient
  - Alternate garbage model giving alternate score?
    - Inherent problem in that the garbage model might absorb everything (i.e., how can we get the garbage model to always produce a score that is lower than a non-garbage model scoring non-garbage?).

Improving Machine Detection of Unexpected Events

- Another score solution: Add more training data.
  - This will help, but speech/language evolves, so training data will never fully reflect the the current testing distributions.
- Still another score solution: Adaptation.
  - Crucial to be able to quickly and efficiently adapt to new conditions.
  - Must occur at multiple levels: acoustic, pronunciation, syntactic, semantic.

Knowing the Unexpected

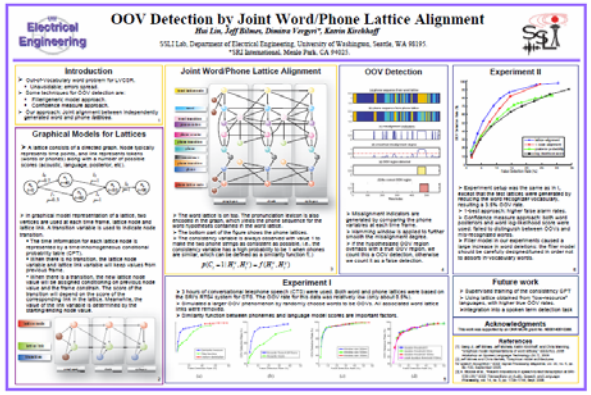
- A new idea: Integration of knowledge from multiple levels of knowledge
  - Background sound event (non speech)
  - Acoustic
  - Phonetic
  - Lexical
  - Syntactic
  - Semantic
- When multiple levels wildly disagree, that time region is suspect, potentially unexpected, and should undergo further inspection.

Knowing the Unexpected: Agree to Disagree

- When multiple levels disagree, candidate region is suspect.
- Approach:
  - consider top N predicted events by each of M multiple levels, total M*N events to consider.
  - If they strongly disagree, the error recovery mechanism kicks in.
  - Error recovery increases compute, slows down recognition, runs adaptation at multiple levels, improves scoring, lexical augmentation, etc.

Knowing the Unexpected: Agree to Disagree

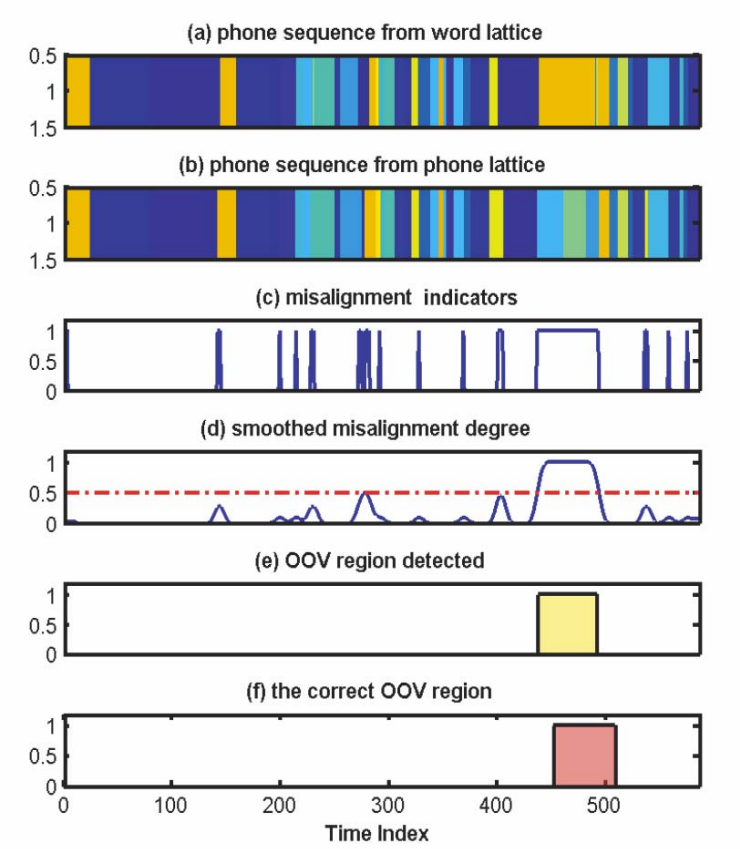
- Very recent examples of this:
  - Hynek Hermansky's recent JHU CLSP workshop 2007 (integration of strongly & weakly constrained classifiers).
  - Hui Lin, Jeff Bilmes, Dimitra Vergyri, Katrin Kirchhoff ASRU2007 (this conference).



OOV Detection by Joint Word/Phone Lattice Alignment

– Hui Lin, Jeff Bilmes, Dimitra Vergyri, Katrin Kirchhoff, ASRU2007.

Poster: This conference, ASRU07!!



Dealing with the Unexpected

- ASR systems must also be able to deal with the case when something unexpected occurred, once it is (accurately) detected.
- This is mostly a case of appropriate dialog strategy, e.g.,
  - Markov Decision Processes (MDPs)
  - Partially observable MDPs (POMDPs)

Dealing with the unexpected

- Does this change the decoding paradigm?
  - Human speech recognition did not evolve in isolation, rather speech evolved in interactive groups, so the approach of just Bayes decision rule (where there is no interactive process that allows one to converge on a solution along with others) is perhaps unnatural.
  - Protocols exist in human conversation for making corrections
    - Backchannels (yea, uh huh, etc.), questions, corrections are expected events.
  - In addition to dialog, this is a case for theoretical game playing strategies.
  - Note, however, that it might be possible to encode a multi-party dialog system using Bayes rule as well…

The End