



Speech and Natural Language

Where are we now, and where are we heading?

Ciprian Chelba

`ciprianchelba@google.com`

Case Study: Google Search by Voice

What contributed to success:

- ⑥ clearly set user expectation by existing text app (proverbial “killer-app”)
- ⑥ excellent language model built from query stream
- ⑥ great progress in acoustic modeling using neural networks
- ⑥ clean speech:
 - △ users are motivated to articulate clearly
 - △ smartphones do high quality speech capture
 - △ speech transferred to server error-free over IP
- ⑥ iterations over log (both text and speech) data from users

Challenges and Directions: Speech Recognition

Automatic speech recognition is incredibly complex.
Problem is fundamentally unsolved.

- ⑥ data availability and computing have changed significantly since the mid-90s
 - △ 2-3 orders of magnitude more data and computing are available
 - △ re-visit (**simplify!**) modeling choices made on corpora of modest size
- ⑥ multi-linguality built-in from start, not as an after-thought
- ⑥ managing complexity while delivering the best performance across many languages, applications, etc.

Challenges and Directions: Natural Language Understanding and Dialog

Very hard problem that has been underestimated and somewhat neglected.

- ⑥ develop with the users in the loop to get data, and set/understand user expectation
 - △ data-driven natural language engineering, not hacks
 - △ multi-sensory setup: leverage touch screen, geo-location, perhaps accelerometer
- ⑥ multi-linguality built-in from start, not as an after-thought
- ⑥ managing complexity while delivering the best performance across many languages, applications, etc.

Speech and Natural Language:

Quo Vadis?

- ⑥ Would the technology be the same if we were to restart ASR/NLU research on today's data availability and computing platform?