A New Kind of Distributed ASR — Phonetic DSR

Melvyn Hunt & Yoon Kim

Novauris Technologies

Overview

- Server-based and local ASR
  - each has its own set of advantages
- "Conventional" distributed ASR ("DSR")
- The Novauris approach to ASR
- This approach permits a new kind of DSR
  - combining many of the advantages of local and server-based ASR
- Why it's particularly good for the *iPhone*
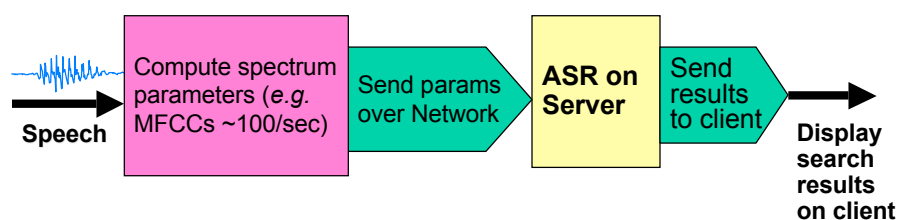- Demo

Server-based & Local ASR

- Most ASR on mobile devices is either:
  - **Server-based**
    - the speech waveform is transmitted over a network to a server and the recognition result is sent back
  - Or **fully local**
    - the ASR is carried out locally on an embedded processor in the mobile device

The different advantages of the two approaches

- **Server-based ASR:**
  - Can be used with very large databases
  - and databases with sensitive information
  - databases can be easily updated
  - users' speech is available for analysis

- **Local ASR:**
  - no transmission delays
  - no transmission costs
  - adaptation to user's voice is easier
  - avoids coding distortion, esp. with noisy speech
  - avoids telephone bandwidth limitations
  - unaffected by loss of network
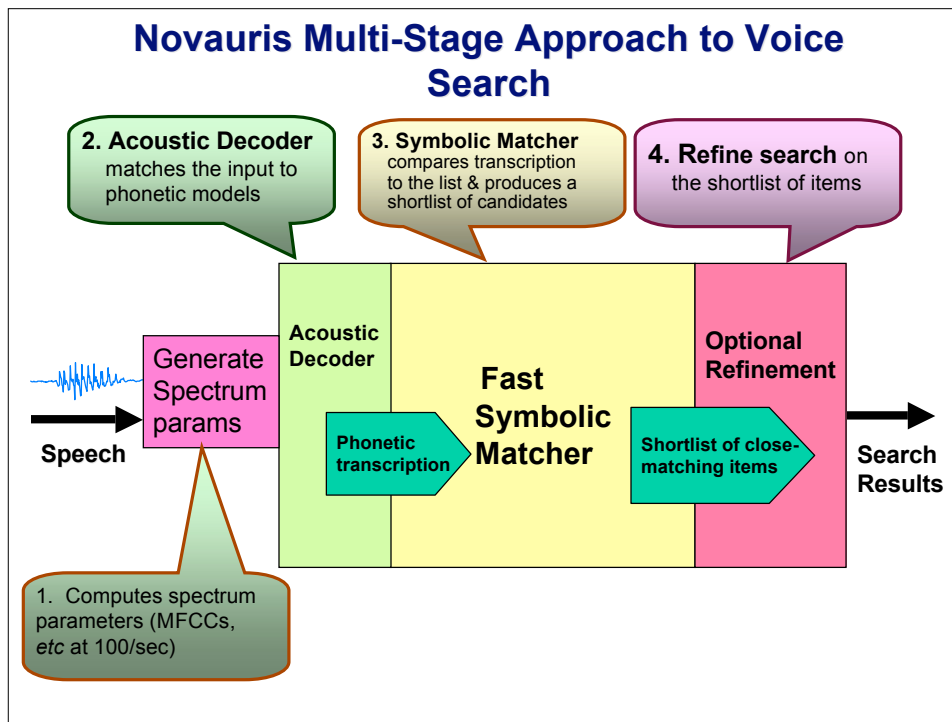
(Conventional) Distributed ASR

Essentially server-based but spectrum parameters are computed locally

Best-known example is the Aurora standard
– but not in widespread use

Distributed ASR

- Distributed ASR computes spectrum parameters (typically MFCCs) locally
- Sends them over the network:
    - 100 frames per second of ~13 parameters
- Avoids coding distortion – esp. in noise
- Database can be kept on the server
- **But** has transmission costs & delays of normal server-based ASR

Novauris Multi-Stage Approach to Voice Search

Phonetic Distributed ASR

- Novauris' *NovaSearch*® approach opens up a new kind of DSR
  - instead of transmitting spectrum parameters we can send the phonetic transcription
- The transmission rate is only **80 bits/sec!**
- Sometimes, data is returned to the client for decision refinement
  - this needs a somewhat higher bandwidth
  - but most networks are asymmetric

How does Phonetic DSR change the picture?

- Like server-based ASR:
  - ✓ gives access to very large databases
  - ✓ gives access to databases with sensitive information
  - ✓ databases can be easily updated
  - ✗ users' speech available for analysis

- Similar to local ASR:
  - ✓ <u>low</u> transmission delays
  - ✓ <u>low</u> transmission costs
  - ✓ adaptation to user's voice is easier
  - ✓ avoids coding distortion, esp. with noisy speech
  - ✓ avoids telephone bandwidth limitations
  - ✗ unaffected by complete loss of network

Why PhDSR is particularly well suited to the *iPhone*

- Unlike most smartphones, the *iPhone*:
  - has no external memory facility
    - flash cards, *etc.*
  - Data has to be stored with the application
  - makes fully local applications, such as recognition of all US addresses, all but impossible
- Phonetic DSR poses no problems

Demonstration

- We chose an application that could definitely not be done fully locally:
  - Name and address recognition
  - From a set of 245 million items
- First ever public demo
- Only just completed
- Lack of 3G in the lecture room no problem!
- Server in our office in England
- Fingers crossed!

Concluding Remarks

- Phonetic distributed ASR is a new development for Novauris, made possible by our *NovaSearch*® approach to ASR
  - combining many of the advantages of fully local and of server-based ASR
- We believe it to have significant potential
- Thanks for your attention