Technical challenges in speech recognition in vehicles and in other noisy places

Gabi Ilan, CEO, Command Speech Ltd.

gabi@command-speech.com   **+972 54 453 3168**

Mobile Voice Conference 2012   **San Francisco, March 2012**

Control Mobile Devices by Voice in Noise

About Command Speech Ltd.

- Founded in 2008 to develop advanced speech-recognition software for mobile devices that works in real-life noisy places.

- Speech recognition experts: CEO Gabi Ilan was formerly co-founder and president of Advanced Recognition Technologies, Inc. (ART), pioneer in voice-activated dialing on cell phones.

- Large vocabulary, speaker independent speech-recognition

- Innovative, statistical, compact, text-to-speech (TTS) software

More:

www.command-speech.com

The Challenge

- Reliable voice command also in noisy places (car, crowd)

- Recognition of thousands of words or phrases

- Automated language modeling process

- On-device operation: fast and compact voice recognition and text-to-speech engines for fast apps download

Demo

Why Speech Now?

- **The Need:** Applications like navigation and music selection require the ability to search among thousands of selections in a database, while working in a noisy environment, especially the difficult car environment.

- **The Solution:** A speech-recognition engine that is highly accurate, works on large databases of thousands of entries and is noise-robust.

Accuracy

Noise Robustness

Large Vocabulary
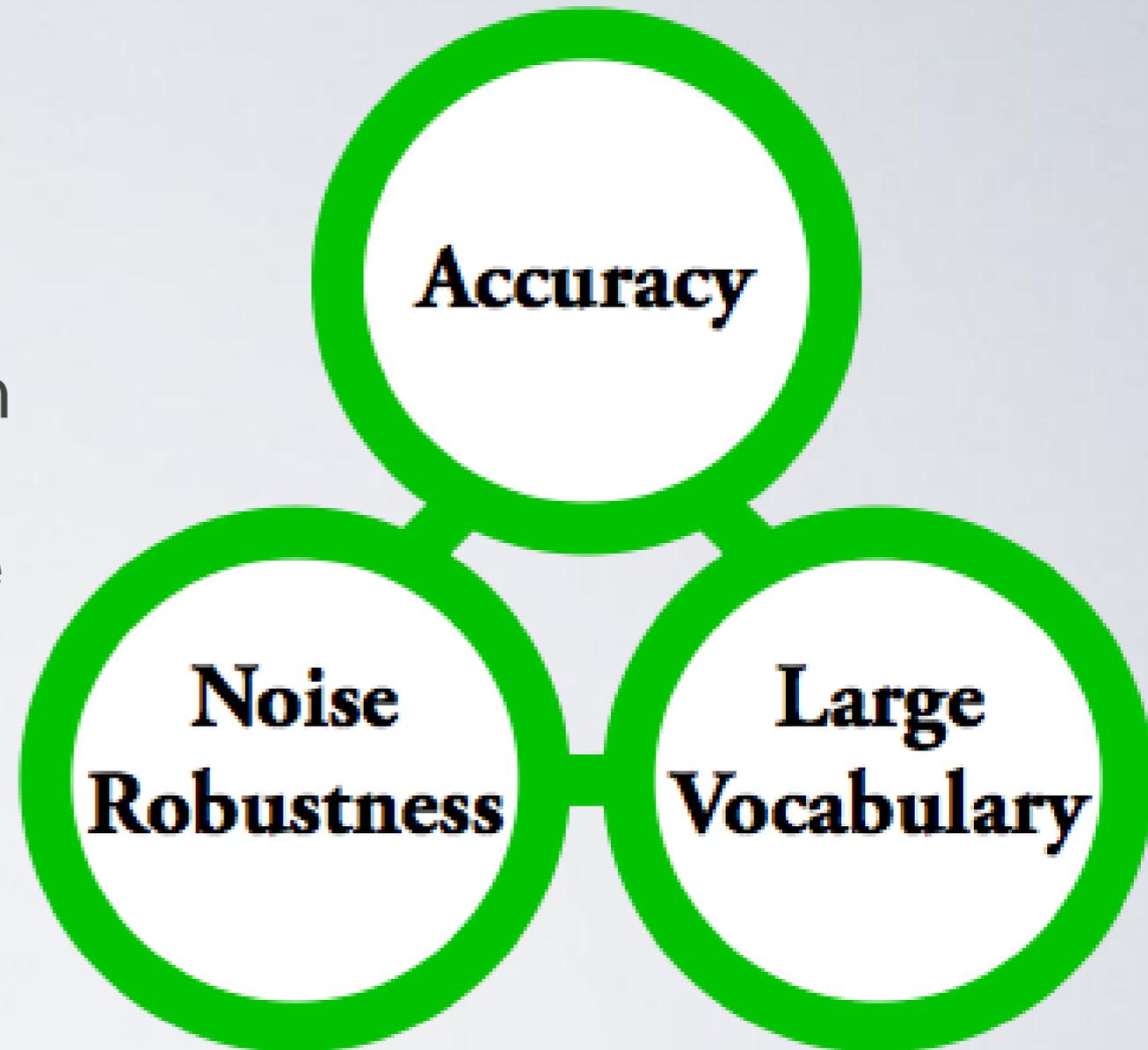
Speech In Cars (1)

- **Voice-Input for Personal Navigation Devices**

  A hot, emerging market

  Solves a real problem: Keying in the destination is complicated, cumbersome and very dangerous while driving

- **Voice Activated Dialing**

  No good solutions available today

  In vehicles – where it is really needed

Speech In Cars (2)

⚫ **Voice Tuning of the Radio**

When more than 6 presets are required

Ideal for Satellite Radio and HD Radio where the number of stations has grown tremendously

⚫ **Voice Selection of Music**

By Song name, Artist name, Genre, etc.

Select from thousands of songs in a device

Smartphones have all these

(and thousands other potential voice command applications)

On-device vs. Remote recognition

| On Device	Over Network
Privacy	Data is shared
Stand Alone. Works anywhere, anytime	Network connection required
Higher SNR – better recognition	Compression can degrade recognition
Immediate response	Delayed reaction due to communications
Free operation	User pays for communications
Limited in size of voice model	No limit on size of voice model on server

Command Speech voice command is an on-device based software

Speech recognition - specifications

Recognizer Type	**Multi-Layer/Multi-Pass proprietary speech recognition software**
User Independency	**User / gender / accent independent**
Training Requirements	**No training required**
Environment	**Optimized for use in noisy environments, such as in moving vehicles or in crowded places**
Number of Entries	**Can support thousands of phrases, enough for very large address books, the number of streets in any city, etc.**
Language Support	**Currently support US English, other languages to follow**
Code Size	**Binary: 1.5 MB Data: 15 MB**
Free Memory Requirement	**5 MB for speech recognition plus text-to-speech engines**
Processor Requirements	**400 MIPS and up**

Technology

1. Multi-layer speech recognition engine structure

    • Time–domain based layer

    • Multi-pass phoneme-based HMM layer

2. Modular and scalable structure to fit size/CPU/functionality

3. Automated, iterative language modeling process

4. Speaker-independent "plug-in" replaceable acoustic models

5. New, efficient and compact, statistical Text-to-Speech engine

Speech-recognition engine structure
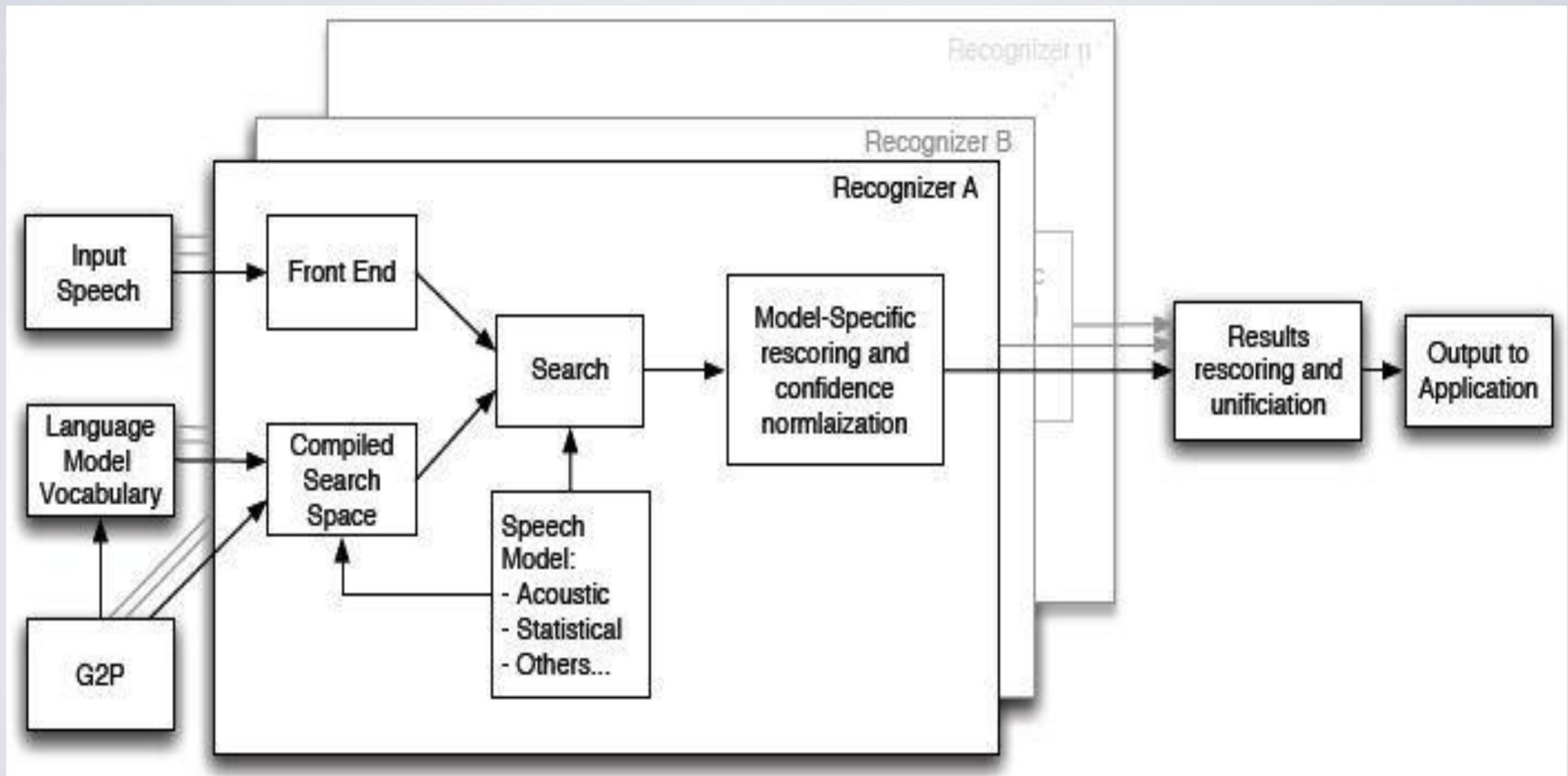
Speech recognition in noisy and in reverberating environment

1. Noise / reverberations tolerance is integral with the recognition

2. One microphone is sufficient

3. Both frequency and time domain analysis and features

4. Careful integration and balance of parallel recognition layers

5. Noisy samples used in creation of acoustic model

6. Using multi-pass, iterative, real-time recognition process

7. More

Thank You