Speech Technologies on Automotive Embedded Platforms

Alfred Wiesen

April 2010

NUANCE

Speech Technologies on Automotive Embedded Platforms

- Nuance
  - The company
  - The division
- Speech technologies
  - General concepts
  - TTS (principles, challenges)
  - ASR (principles, challenges)
- Automotive market
  - Why speech?
  - Applications
  - Platforms
  - Technological challenges
  - Future trends

NUANCE

2

The Speaker

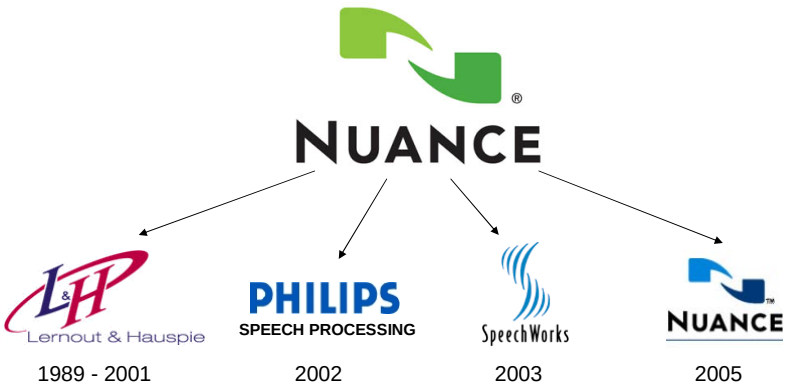
- Alfred Wiesen
  - Civil Engineer ULB 1993
  - End of studies thesis on speech synthesis
  - Lernout & Hauspie (1993 – 2001):
    - Speech compression (CELP, subband coding, …)
    - ASR and TTS SDKs
  - Ubicall Communications (2001 – 2007)
    - Automated attendant systems for hospitals
  - Nuance Communications (2007 – …)
    - Professional Services group in Automotive division
    - String focus on navigation and ASR (voice destination entry, point of interest search, multilinguality, etc…)

NUANCE

3

Nuance: the company

- “Leader in speech and imaging products for businesses and consumers”



NUANCE

4

Nuance: offices

NUANCE

5

Nuance: the divisions

Imaging Solutions

See All Nuance Imaging Solutions

Featured Imaging Solutions:

- eCopy Solutions
- OmniPage OCR Software
- PaperPort Document Scanning Solutions
- PDF Converter Software
- PDF Reader NEW!
- Multi-functional Devices
- Omnipage Capture SDK

- Sign-up for Product Demos
- PDF Converter 6 Try & Buy
- Windows 7 Support

Dragon NaturallySpeaking Solutions

See All Dragon NaturallySpeaking Solutions

Featured Solutions:

- Dragon NaturallySpeaking
- Dragon AudioMining
- Dragon SDK

- Newsletter Sign-up
- I Speak Dragon Contest
- Dragon Blog!
- Windows 7 Support

Healthcare Solutions

See All Nuance Healthcare Solutions

Featured Solutions:

- Dictaphone Enterprise Speech System
- Dragon Medical
- Dragon Medical Mobile Apps NEW!
- eScription
- Focus Infomatics
- Managed Speech Solution
- PowerScribe
- RadCube
- RadPort
- RadWhere
- SpeechMagic
- Veriphy

Nuance Speech / Text Solutions

See All Nuance Speech and Text Solutions

Featured Solutions:

- Authentication
- Automotive
- Customer Care
- Directory Services
- Dragon Mobile Apps NEW!
- Employee Productivity
- Mobile Care
- Mobile Devices
- Mobile Text (T9)
- On Demand Services

Network

- Customer Care
- Telecom
- Directory Services
- Healthcare

Embedded

- Automotive
- Mobile Devices
- Games

NUANCE

6

Speech Technologies

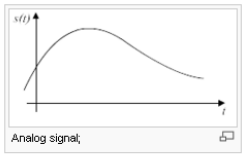
- **Text To Speech (TTS)**
- **Automatic Speech Recognition (ASR)**
- Speaker Verification (SV)
- Dictation

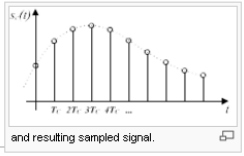
NUANCE

7

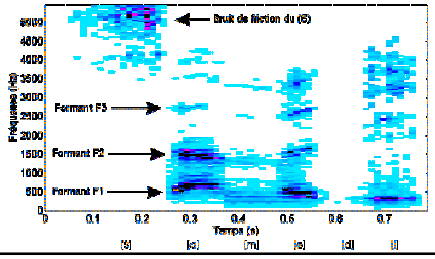
Speech Technologies : Basic concepts

- Signal processing concepts
  - Sampling rate
    (typically 8 kHz for telephony,
    16 kHz / 22 kHz for automotive)





  - Frequency spectrum
    (FFT = Fast Fourier
    transform)



NUANCE

8

Speech Technologies : Basic concepts

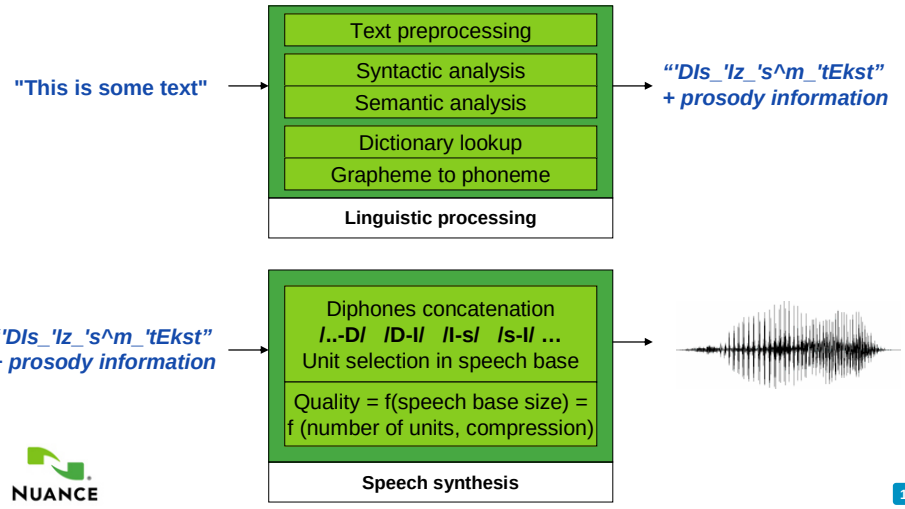
- Linguistic concepts
  - Phoneme : smallest segmental unit of sound employed to form meaningful contrasts between utterances in a given language
    Example of a phoneme is the /k/ sound in the words ***k****it* and *s****k****ill*.

NUANCE

9

Speech Technologies: TTS

- Aim: generate artificial speech from text



NUANCE

10

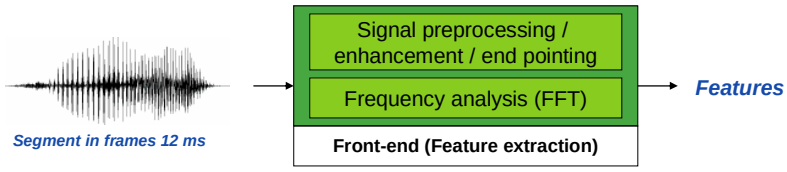
Speech Technologies: TTS

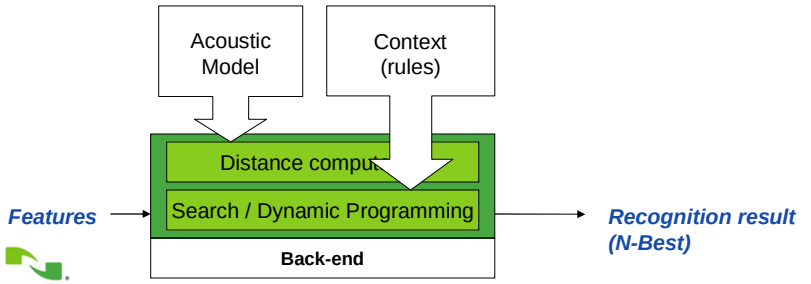
- Typical speech base sizes:
  - **Telephony** 8 kHz:
    Speech base: no compression, all units
    70 → 250 MB for speech base, 8 MB code + dictionaries
  - **Automotive** 22 kHz:
    Speech base: compression (RELP, ADPCM), reduced unit sets
    Standard: 16 MB for speech base, 4 MB code + dictionaries
    Premium light: 64 MB for speech base
- Challenges:
  - Preprocessors and syntactic analysis
  - Multilinguality (foreign words, mixed languages, …)
    « Vous écoutez "Please don't leave me" par Pink »
    « Tournez à droite dans la Bahnhofstraße »
  - Rythm and prosody

NUANCE

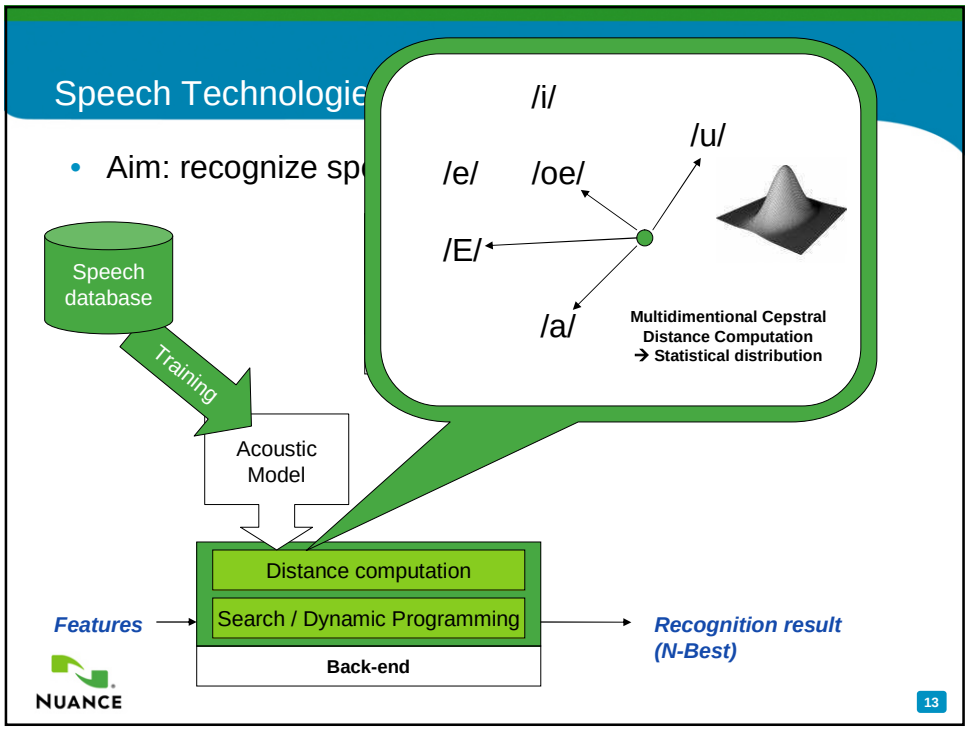
Speech Technologies: ASR

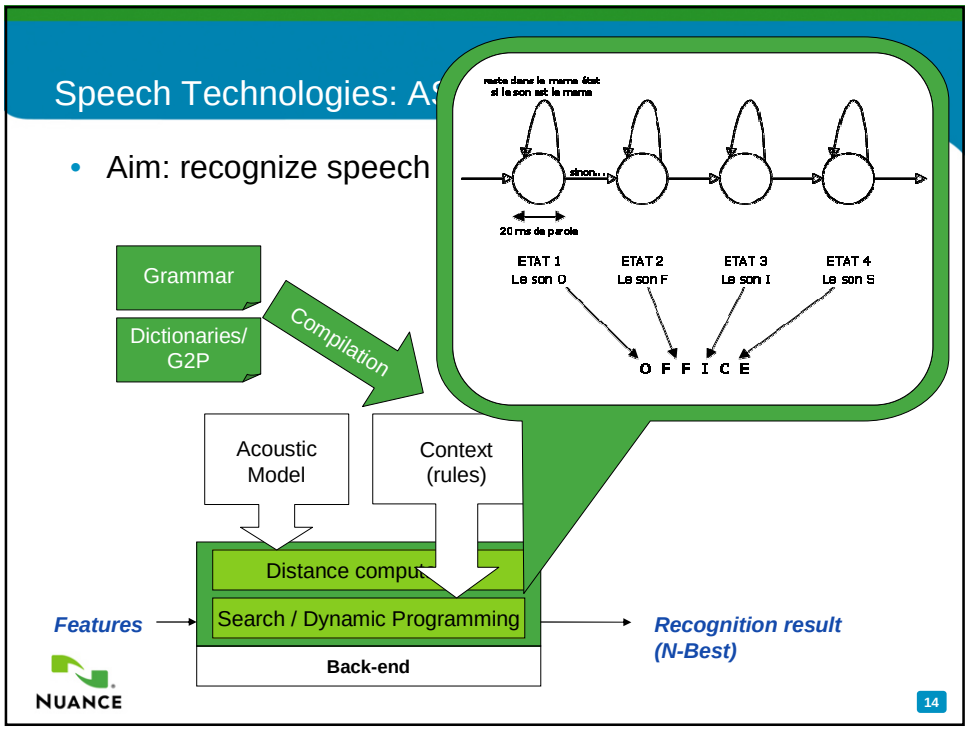
- Aim: recognize speech into text / action





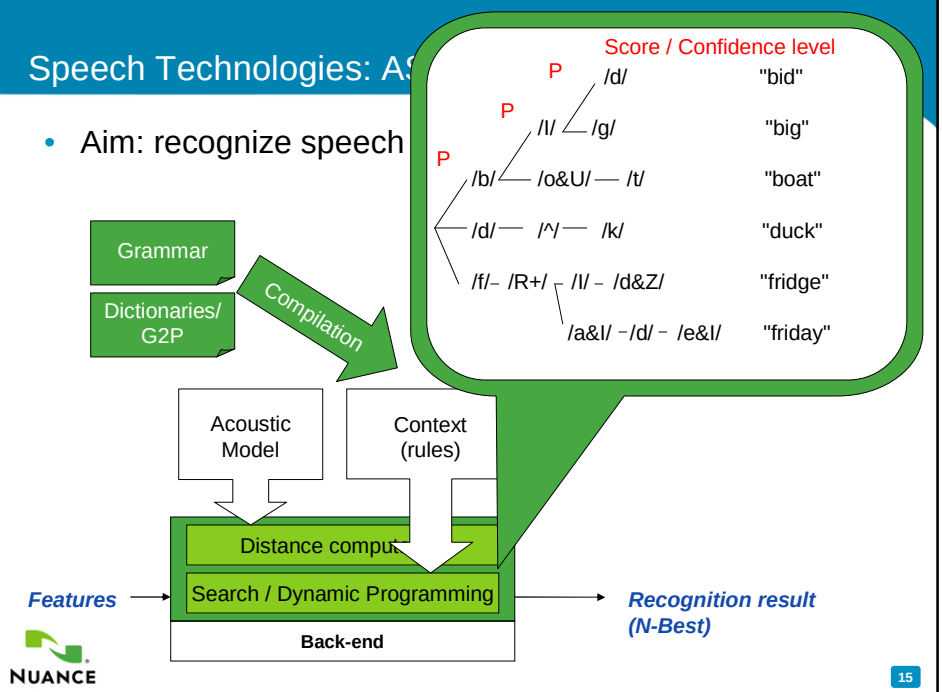
NUANCE


Speech Technologie
• Aim: recognize sp
/i/
/u/
/e/
/oe/
/E/
/a/
Multidimentional Cepstral Distance Computation
→ Statistical distribution
Speech database
Training
Acoustic Model
Distance computation
Search / Dynamic Programming
Back-end
Features
Recognition result (N-Best)
NUANCE
13


Speech Technologies: A
• Aim: recognize speech
reste dans le meme état si le son est le meme
sinon...
20 ms de parole
ETAT 1 Le son O
ETAT 2 Le son F
ETAT 3 Le son I
ETAT 4 Le son S
O F F I C E
Grammar
Dictionaries/ G2P
Compilation
Acoustic Model
Context (rules)
Distance comput
Search / Dynamic Programming
Back-end
Features
Recognition result (N-Best)
NUANCE
14

Speech Technologies: AS

- Aim: recognize speech



NUANCE

15

Speech Technologies: ASR

- Typical data sizes:
  - **Automotive** 16 kHz:
    Acoustic models: from 760 kB to 2.5 MB
    Contexts:
    - Digits, commands : a few kB
    - Navigation (voice destination entry): a few MB for small contexts, 50 MB for full European country, >200 MB full USA
  - **Network** 8 kHz, dictation → hundreds of MB (server based)
- Typical challenges:
  - Audio quality (background noise, etc)
  - Data sizes! RAM and CPU contraints!
  - Accented speech (acoustic model trained for native speakers)
  - Multilinguality (foreign words, mixed languages, …)
    « Ecouter "Think" par Aretha Franklin »
    « Aller à la Bahnhofstraße de München»

NUANCE

16

Automotive Market

- Why speech?
  - Hands free
  - Driver distraction
  - Fast
- Applications:
  - Command and control (ASR) of all car appliances
  - Communication
    - Hands free dialing (ASR), Caller anouncement (TTS)
    - SMS reading (TTS)
  - Navigation
    - Voice destination entry, Point of Interest search (ASR)
    - Reading out street names, traffic messages (TTS)
  - Entertainment
    - Radio station selection, music search (ASR)

NUANCE

17

Automotive Market

- Platforms:
  - Car-kits

  - PNDs = Portable Navigation Devices

  - Built-in Units
    - Car manufacturers
    - Tier-1 suppliers

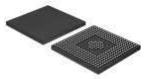
NUANCE

18

Automotive Market Technological Challenges

- Embedded platforms → strict contraints
  - CPU power: ARM9 core, ARM11 core, SH4, PowerPC, MIPS core
    clocked around 300-800 MHz
    → ASR not always real-time
    → Limit latency to a few seconds
  - RAM: depends on platforms
    - Car kits, low-cost PNDs: 32 MB
    - Mid-range PNDs: 64 MB
    - High-end PNDs and in-car units: 128 MB

    NOT ALL RAM AVAILABLE (max 10 – 20 MB for speech)
    → Limit RAM needs at runtime (use techniques like "On demand loading" if storage is fast enough)



NUANCE

19

Automotive Market Technological Challenges

- Embedded platforms → strict contraints
  - Storage limited as well (Flash)
    - Car kits: < 64 MB
    - PNDs: 1 to 4 GB for all data (including maps)
    - In-car units: more space, sometimes even hard disks

    On competitive markets (PNDs cost only 100 €), spending 2 € for more Flash is not acceptable.
    → Limit and compress data as much as possible (e.g. techniques to reuse large context and activate only subparts of it)

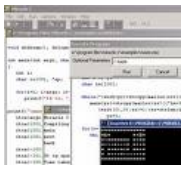
NUANCE

20

Automotive Market Technological Challenges

- Operating systems
  - WinCE, Linux, QNX, Nokia Symbian, Android, Mac OSX / iPhoneOS
- Embedded platforms

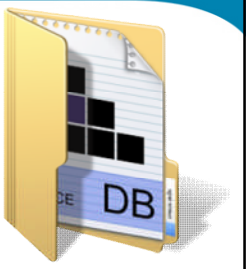
  → porting
  - No assembly
  - Highly portable fixed point C
    - Portable → no dependency on OS, only pure standard C
    - Fixed point → all floating point computation simulated and optimized for speed!



NUANCE

Automotive Market Technological Challenges

- Research challenges on engines:
  - Front-end: better in noise performance, accents (speaker adaptation), …
  - Back-end: faster and more RAM efficient to be able to handle larger data sets
- Typical data sets:
  - Navigation:
    France = 75 k cities (113 k phonetics)
    814 k unique streets (1 M phonetics)
    2 M one-shot entries (4 M phonetics)
    USA = 46 k cities (73 k phonetics)
    27 M one-shot entries (65 M phonetics)
  - Music:
    10 k MP3s → artist, song by artist, album by artist, etc
    → 10 k's combinations



NUANCE

Automotive Market Technological Challenges

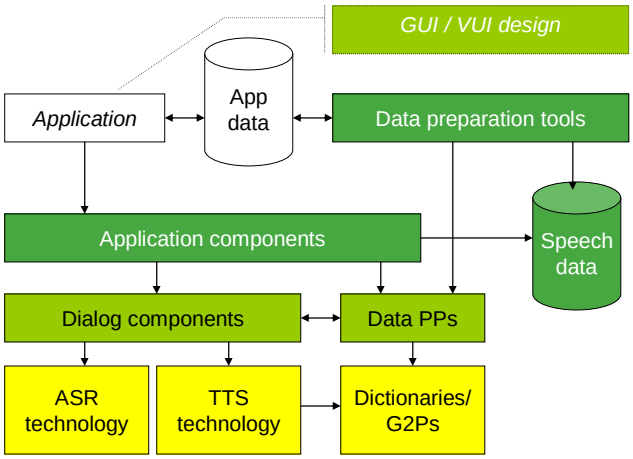
- Data processing
  - Cultural habits
    - Order of fields
    - Optional items
      « [N] Rodeo [Dr] »
      « Avenue du Général Charles de Gaulle »
  - Multilinguality in navigation
    - German driver in France
    - Multilingual countries

Heavy data preprocessing not strictly linked to technology

NUANCE

Automotive Market Future Trends

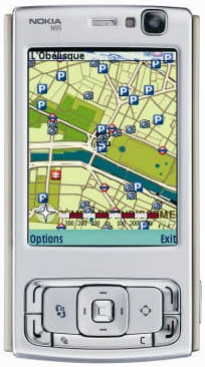
- Need for integrated solutions



NUANCE

Automotive Market Future Trends

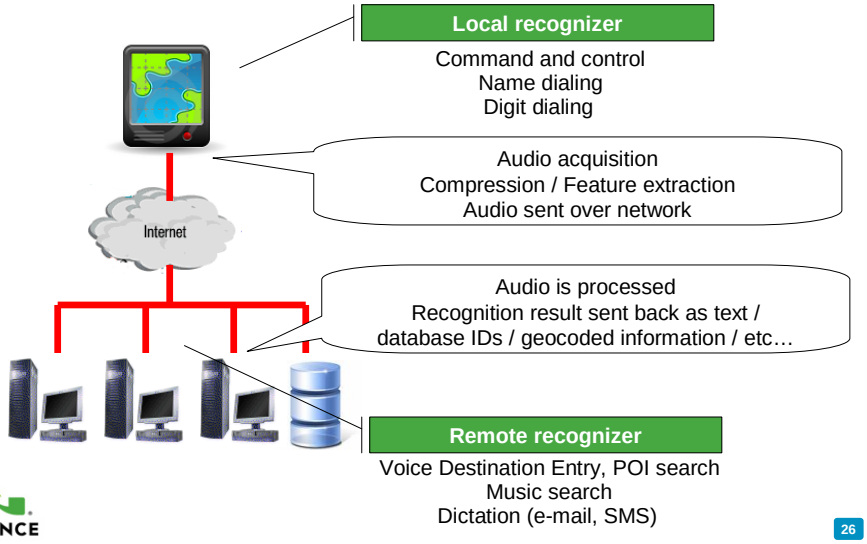
- Navigation services – **Convergence with Mobile**



NUANCE

Automotive Market Future Trends

- Connected services – **Convergence with Network**



NUANCE

Automotive Market Future Trends

- Consequences of these trends
  - Need for more understanding on:
    - Data sets, data preparation
    - GUI work and design
    - Application design
  - Higher level modules introduce OS dependencies
    - TCP/IP and networking
    - Threading, synchronization, …

NUANCE

27

Questions?

NUANCE

28