Text to Speech for In-car Navigation Systems

Luisa Cordano
August 8, 2006
TTS in the Navigation Domain

What kind of expectations does the Navigation market have on a TTS system when applying it in its domain?

- Flexibility
- Perfect pronunciation
- Cost savings
- Expressivity
Key aspects of a TTS system

- Unit Selection Technique
- Domain-dependent customization
- Phonetic Input Interface
- Native vs. Foreign Pronunciation
The importance of Unit Selection TTS

- Loquendo TTS is a general purpose (Unrestricted) Text-to-Speech system based on the Unit Selection Speech Synthesis Technique.
- The Unit Selection Technique is at present the only existing technique that enables natural sounding speech.
- Its basic idea is to concatenate long phoneme sequences extracted from large DataBases of speech recorded by a single voice talent.
- Advantages: little speech processing, the natural voice timbre is preserved.
- Drawbacks: scarce control, speech quality depends on the contents of the database.
Unit Selection TTS

Input text

Text Analyzer

Phonemes & prosodic labels (& prosodic values)

Speech Synthesizer

Speech

Domain-representative large text corpus

Statistical Tools for Database Design

Dense text corpus

Tools for:
Speech Acquisition
Phoneme Segmentation
Signal Analysis

Vocal Databases

Language Libraries

ITALIAN
SPANISH
ENGLISH

...
The Synthesis Algorithm

- define a synthesis TARGET
- match phonetic and prosodic labels
- look for similar f0 values at unit junctions
- preferably cut units at diphone boundaries
- concatenate waveforms
- adjust f0 to remove jumps (pitch scaling)
Domain Dependent Customization

- Loquendo TTS can read any text
- In order to improve accuracy and fluency on an application domain it can be customized
  - *Lexical customization*: adapting Text-to-Phoneme conversion via Pronunciation Lexicon, Phonetic Input
  - *Vocal customization*: enhancing the Vocal DataBase with application prompts and recurrent phrases
Domain Dependent Customization

- The navigation domain:
  - A limited number of recurring phrases: Ex. “turn left”, “…”
  - A peculiar and huge lexical domain: addresses, names and toponyms: due to their historical/foreign origin, they do not necessarily follow the standard grapheme-to-phoneme rules of the language

- Customizing Loquendo TTS for Navigation:
  - Navigation Vocal Add-On: the recurring phrases can be recorded in case their synthesis does not sound perfect
  - The exact pronunciation of names and addresses can be specified via the Phonetic Input Interface or via the Pronunciation Lexicon (this does not prevent from possible acoustic defects)
Phonetic Input Interface

- Loquendo TTS let the User control the exact pronunciation of words
- An escape sequence in the input text allows skipping the grapheme-to-phoneme process driven by the Language Library…
- And inputting the desired phoneme sequence directly to the Speech Synthesis process based on the Vocal Database
Phonetic Input Interface

Input Phonetic Transcription (language independent SAMPA Alphabet)

Text Analyzer

phonemes & prosodic labels (& prosodic values)

Speech Synthesizer

Vocal DataBases

Input text

Language Libraries

ITALIAN

SPANISH

ENGLISH

...
Phonetic Input Interface

- The special feature is the *language independence*
- The input phonemes are *Mapped* on the Phonemes of the Voice

Via the Phonetic Interface, an Italian voice can be made to pronounce

\[\text{SAMPA}="\text{santena}\]

for the Italian name “Santena”, overriding the default *san“tena*.

But also

\[\text{SAMPA}=\%\text{le#"granges}\]

For the French “Les Granges”, otherwise transcribed *les#"grandges*.

The foreign *R* and *a~* will be mapped on the Italian *r* and *a*.
Native vs Foreign Pronunciation

- Each Voice has its own native language
- Each Vocal DataBase has its own:
  - Phoneme Set (the db does not contain foreign phonemes)
  - Coverage of Phoneme Sequences (the db contains only sequences that frequent in its native language)

The Foreign Pronunciation feature of Loquendo TTS maps foreign phonemes onto the most similar native phonemes. But it can’t avoid obtaining phoneme sequences not present in the Vocal DB, requiring the concatenation of shorter speech units.

Foreign Pronunciation is:
- Plausible
- Approximated
- Sometimes choppy
Foreign Pronunciation: the Phoneme Mapping Algorithm

The algorithm centers around a Phonetic Similarity Function (PSF) that computes a similarity score of two phonemes that only depends on their phonetic-articulatory features.

Such an approach overlooks:
- Finer language specific aspects of speech perception
- Pragmatic/cultural aspects that may affect foreign pronunciation

But... makes comparison between phonemes free from any language-specific knowledge!
Phonetic Similarity Function

This approach requires, as a first step, the definition of a vector of phonetic-articulatory features for each phoneme.

Each feature has a different weight in the computation of phonetic similarity.

Values of a non-binary feature are placed in a scale of perceptual distance.
Foreign Pronunciation: Demos

English Text:

“Good afternoon everybody! This is a live demo of Loquendo's lifelike text-to-speech technology, using the mixed language feature”

Katrin (German Voice)

Jorge (Spanish Voice)

Bernard (French Voice)

Interactive demos are available on http://actor.loquendo.com/actordemo/
“In 250 yards, bear left, at the roundabout, take the second exit, and follow the signpost indicating, Charles de Gaulle airport.”

**Dave** (American Voice)  🎧 

“In 100 yards, stay in the right lane, and take the main road. After the tunnel, turn left, and follow the signpost indicating, Fiumicino, Leonardo da Vinci airport.”

**Kate** (British Voice)  🎧
Future Works for the Foreign Pronunciation

The problem of the unusual sequences of phones could be partially solved by:

- Augmenting the speech databases with ad-hoc vocal material
- Rewriting the PMM output by reducing the number of unusual phoneme sequences without modifying the plausibility of the reading
- Adding new spectral features to better describe each speech database phone
Empowering Your customers and employees with speech technologies

SpeechTEK 2006 Empower
The Voice Solutions Showcase

Thank You