The Best of Breed in Speech: Loquendo TTS & Loquendo ASR

Rosanna Duce
V.P. International Sales

March 2 2009
Voice Search Conference 2009
Agenda

- Loquendo Today
- The value of Speech in Customer Relationship Management
- Loquendo Speech Technologies for Call Centers:
  - **Text To Speech** to create lifelike prompts
  - **Automatic Speech Recognition** in a natural speech environment
Company Profile

- Privately held company (fully owned by Telecom Italia), founded in 2001 as spin-off from Telecom Italia Labs, capitalizing on 30 yrs experience and expertise in voice processing.

- Global Company, high quality voice technologies (synthesis, recognition, authentication and identification) available in 26 languages and 62 voices.

- Multilingual, proprietary technologies protected by over 100 patents worldwide

- Financially robust, break-even reached in 2004, revenues and earnings growing year on year

- Growth-plan investment approved for the evolution of products and services.

First Impressions Make A Difference

The Key to Success in Voice Applications and CRM:

- A natural, accurate, well-designed speech interface
- And first-class ASR and TTS to power it.

Your speech service is your customer’s frontline!
A natural and user-friendly VUI is the key to an enhanced customer experience.

Are You Ready for the Future?

www.loquendo.com

Create a User Experience Worthy of Your Customers
Loquendo speech technologies allow you to create an enjoyable and consistent user experience through highly effective, reliable and natural interaction.

Streamline Your Business and Prioritise Your Spending
Upgrading a customer contact centre with Loquendo speech technologies makes your service more cost-effective, more responsive and better tailored to your clients’ needs.
Speech in your Call Center to drive profit

- 24h access to information, anywhere, simply by phone
- Easy to use: natural interaction compared to traditional IVR systems

- Effective ‘burst’ management and decreased customer waiting time
- Offering multi-language and personalized service

- Helps reduce the “digital divide”: accessible without PC, including by elderly and differently able people

- Realizes profit:
  - Automating Customer Care Interaction with Speech lowers costs by reducing the need for staffing live agents at a 1-to-1 ratio with consumers
  - Understanding the Caller’s intent and route their request to the correct destination
....And to improve Customer satisfaction

Can you tell me at what time you open???

Please hold, all our agents are busy.

Wait for the next available agent...

Nobody’s answering ....

Waiting and waiting ...

☠☠@#

!! ☣ ☠ ✂
Speech is the more natural interaction in any Call Center…

**Information services:**
- Directory Assistance
- News
- Location Information
- Timetables
- Tracking and tracing
- Entertainment
- Access to DB

**Transactional services:**
- Phone banking, online trading
- Booking services (hotel, theater, cinema, travel...)
- Electronic Payments
- Voice commerce (online shopping malls, music, etc.)

**Personal Communication services:**
- DB Access
- Virtual Operator
- Personal Communication:
  - Address book
  - Agenda
  - Voice Activated Dialing
  - E-mail access by phone
  - Appointment notification
  - Centralized voicemail
Speech Engines in the Value Chain & Product Positioning

Basic Resources to be used in application developments:
- Specialized lexicons, grammars
- Reusable Dialogue Objects
- Customization

Turn-key MRCP (v1 & v2) Server:
For simple interfacing with IVRs and voice platforms

Speech Engines, SW only:
- Text-To-Speech
- Automatic Speech Recognition
- Speaker Verification and Identification
- Language Identification

For telephony server, desktop, embedded & mobile devices

Windows, Linux, WinMobile, Symbian, …

O.S.

Hardware

Speech Engines

MRCP Server

Building Blocks

Applications

Servers

Speech Engines

Loquendo Focus
Voice technologies in Call Center environments

INTERNET

Multi-channel Access (Web, SMS, WAP)

APPLICATION SERVER
CONTENTS

LEGACY SYSTEMS

OPERATORS

Loquendo TTS
Loquendo ASR

Major IVR Platform

Data
XML
HTML
VoiceXML

PBX/ACD

CTI Server

pre-existing IVR

VoiceXML

Voice Search Conference 2009
Rosanna Duce
How Speech technologies can help

Create the voice content for your Call Center with TTS instead of using costly human speakers and recording studios.

Loquendo TTS samples
Multi-language: 26 languages, 62 voices – and more coming!

Truly Natural and Expressive sounding voices for highly…

Emotional pronunciation:

- Commonly used phrases such as “How are you?” or “You’ve got to be kidding!” to confirm, exclaim, thank, express doubt, etc.

Plug-in lexicons: for reading e.g. SMS abbreviation, emails

Audio Mixer: to have complete control over all audio sources (including sampling rates and coding) – audio files can be mixed, looped, faded in and out, or synchronized with synthetic speech

Mixed Language Capability:

- **Language Guesser**: that automatically identifies the language of any text so that by means of…

- **Phonetic Mapping**: any of Loquendo’s voices can correctly pronounce any foreign word (e.g. Spanish words in English text)
Multilanguage Capabilities

Many recent TTS applications increase need for dealing with mixed language input

Given an input text with multi-lingual content, the expectation is

- **Automatically detect** each language change, or **control** a specific language change with appropriate commands inserted in the input text
- **Change voice** - choosing from the available voices, based on the language
- Alternatively, **keep the same voice** - speaking in a different language so that a single voice can speak several languages

A couple of strategies have been proposed:

- The **polyglot** approach: multilingual databases are recorded from a polyglot speaker
- The **phonetic mapping** approach

Many movies have been produced and filmed in Mexico, even though in many instances you would never know it. We can quote, for example: "Y tu mamá también", or "La última noche", that are examples of films shot in Mexico.

**Navigation systems**

**Entertainment (movies, songs, etc.), news, interviews**
Expressive Synthesis

Expressive TTS means

- the use of sophisticated algorithms for audio signal manipulation that, based on prosodic models for different expressive styles (angry, happy, sad,…), is able to generate expressive speech (hot topic for research)

Enriching TTS with Expressive Cues

- This is an **immediate response** to the market’s requirement for expressivity in speech
- Not yet a fully emotional style for any sentence, but an enhancing repertoire of **Expressive Cues**, such as apologies, greetings, exclamations, etc.
- Enriched TTS continues to benefit from the **high quality and natural timbre** achieved with the Unit Selection technique

- More informal applications (IVR, gaming, virtual operator, etc.)
- Applications using alerts, directions
Audio Controls and Innovative Features

Mixing synthetic speech and Music
- Insertion of special effects applied to the audio signal (reverb, echo, fade-in, fade-out, balance, graphic equalizer, ...)
- Different output codings and frequencies can be chosen

Customizable voices
- Change in timbre - save and reuse a customized voice

Specialized application contexts
- SMS reading
- Navigation systems (street names, ...)

Enrichment for multimedia presentations
- Reading of synchronized subtitles
- Dubbing

SMS reading
r u don nethng 2mor? would b gr8 2cu.
i'll b @ pub b4 8.
ringl8r. b4n xoxox susan
Loquendo TTS Tools: Lexicon Manager
New custom voices can be created by the adjustment of the prosodic parameters of Loquendo voices:

Control over the timbre, pitch and speaking rate

Prosodic settings can be saved to create new custom voices (child, cartoon character, etc.)
Loquendo TTS Director is a complete development environment for creating your own voice prompts, and for designing your own personalised voices.

**Target:** clients wanting to edit their prompts at a more complex level and adjust parameters with far more precision, as well as to add pauses, phonetic transcriptions, and tailor-made lexicons for atypical pronunciation.
And now the ASR engine .....  

**Automatic Speech Recognition**

Loquendo ASR benefits from the input of Loquendo TTS which inherits multi-language capability, language guesser and natural speaking features
Loquendo ASR

- **Speaker-Independent**

- **Highly Accurate Speech Recognition** – thanks to integration of neural networks and hidden Markov models, Loquendo ASR has high performances and increases efficiency with large vocabularies (up to hundreds of thousands of words).

- **Broad Vocabulary & Flexible Recognition** – recognizes up to 1,000,000 words; supports isolated and continuous speech.

- **Extended Standards Support** – optimized for VoiceXML applications; MRCP and AURORA compliance; complete grammar standards’ support, both W3C SRGS and SISR.

- **High Efficiency** – low-computational power requirements enable a large number of recognition channels to run simultaneously, both with small and large vocabularies.

- **Rapidly Extensible to new languages**

- **Powers Loquendo Speaker Verification technology**
A complete set of simple and powerful features guarantees truly robust speech technology, enabling:

- Advanced **barge-in capability** to guarantee high reactivity and robustness to noise and background speech.
- A patented speech enhancement method for **improved recognition performances in noisy conditions**.
- A **flexible rejection mechanism** which identifies any linguistic expressions that are not acceptable within a specific domain.
- **Dialog-flow management** which is achieved through confidence values provided for all the Nbest hypotheses returned – on a sentence-by-sentence & word-by-word basis.
- **Garbage rules** definition to match arbitrary spoken sequences not modeled by the grammar.
Loquendo ASR: Enhanced Transcriptions

Loquendo ASR grammars are transcribed into phonemes, by means of the Loquendo TTS transcriber. This allows:

- **Highly Accurate Phonetic Transcriber** – specialized for each supported language
- **Seamless import** of Loquendo TTS enhancements and customizations
- Usage of the same **phonetic lexicons** between Loquendo ASR & TTS.
- Loquendo TTS **Lexicon Manager tool** can be used to add:
  - Alternative pronunciations
  - Acronym expansions
  - Transliterations
- **Multilanguage capabilities** are a very powerful feature for Loquendo ASR
TTS Multilanguage Capabilities is a very effective feature to be used in Loquendo ASR to create international grammars:

- Even if a single acoustic model is active, Loquendo ASR grammars can contain words of different languages.
- The foreign words are transcribed in the different language using the phonetic mapping features of the Loquendo TTS transcriber.
- Field results (e.g. air traffic control) showed the effectiveness of this technique.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<grammar version="1.0" xmlns="http://www.w3.org/2001/06/grammar"
    xml:lang="en-US"
    tag-format="semantics/1.0-literals"
    mode="voice" root="love">
  <rule id="love" scope="public">
    <one-of>
      <item> I love you! </item>
      <item xml:lang="it"> Ti amo! </item>
      <item xml:lang="fr"> Je t’aime! </item>
      <item xml:lang="es"> Te quiero! </item>
      <item xml:lang="de"> Ich liebe dich! </item>
      <item xml:lang="ru"> Я люблю Вас </item>
    </one-of>
  </rule>
</grammar>
```

Transcription language, then mapped into English.
Loquendo ASR: Tool Suite

- **LogCollector Tool** to select recorded audio material
- **Acoustic Model Adaptation Tool** to efficiently adapt the recognition engine to difficult conditions, such as:
  - **different audio channels** (wireless, multimedia microphone, PDA, etc)
  - **different environments** (in-car, public areas, factory)
  - **application-dependent vocabulary** (specific jargon, such as aeronautical terms)
  - **ways of speaking** (regional accents, fast speech)
- **Phonetic Learning Tool** to improve performance using data collected in the field to deal with
  - **additional linguistic formulations**
  - **complex speech recognition applications**
  - **non-native speakers or regional accents**
How can we simplify grammar creation?

We can focus on modeling only the relevant content, and not the rest of the phrase.

E.g.
(I’d like to travel) from New York to Chicago (please)
(I need to go) from New York to Chicago (as soon as possible)
(Well…I’m going…er…sorry, a ticket) from New York to Chicago

Use a special grammar node: to discard the rest of the sentence!
Where in the phrase do we position the garbage?

Simple usage

More complex usages are possible and the user can speak freely almost without rules ....
With Loquendo TTS:

- You do not need to use professional speakers to create your prompts
- You can create your “testimonial” voice
- You do not need anymore costly professional recording studio
- You can work in a multi-language environment

And in combination with the ASR:

- You can use the same speech DB already tuned for the TTS
- You can “adapt” the environment to the different noises
- You can allow the Caller to speak almost freely
- You can be 7*24 reliable without idling time

And your Call Center can really drive profit improving, in the meantime, the Customer satisfaction.
For more information please:
Keep an eye on: www.loquendo.com
Contact: rosanna.duce@loquendo.com

- Keep in touch with Loquendo news, subscribe to the Loquendo Newsletter
- Try our interactive TTS demo: insert your text, choose a language, and listen
- The latest News at a click
- Consult the Loquendo Newsletter online
- Keep up to date on events and initiatives
- For further information, fill in our Contacts Form