B105 Innovative speech technology

Naturalness and expressivity, efficiency and flexibility in Text To Speech technology

Monica Bisacca, Loquendo
Loquendo Today

- **Global company** formed in 2001 as a spin-off from the Telecom Italia R&D center with over 30 years experience in Speech Technologies
- Complete set of **Multilingual speech technologies** on a wide spectrum of devices
- Full support of international standards (VoiceXML, CCXML, MRCP, VoIP)
- Ready for challenging future scenarios: Multimodality, Security
- A **Telecom Italia Group** company
- Strong and growing presence in Europe, North and Latin America
- HQ in Turin, Offices in US, Spain, Germany and France, and Worldwide Network of Partners

- **“Market leader-Best Speech Engine” Speech Industry Award 2007**
- **“Best Innovation in Automotive Speech Synthesis” Prize AVIOS-SpeechTEK West 2007**
- **“Best Innovation in Expressive Speech Synthesis” Prize AVIOS-SpeechTEK West 2006**
- **“Best Innovation in Multi-Lingual Speech Synthesis” Prize AVIOS-SpeechTEK West 2005**
Loquendo TTS Consolidated Technology

- Supporting Multiple Languages: 22 languages and 53 voices
- Audio prompting tools:
  - Loquendo TTS Director
  - Loquendo Voice Experience
- Mixed Language Capability
- Audio controls & Audio Mixer
- Voice Creator for customized voices
- Lexicon Editor (& plug-in for SMS reading and toponyms)

2007 SpeechTEK Market Winner Award
2007 “Best Innovation in Text To Speech”
2006 “Best Innovation in Speech Synthesis”
2005 “Best Innovation in Multi-Lingual Speech Synthesis”
Loquendo TTS Voice Experience is an extremely user friendly tool for the rapid and hassle-free adjustment of voice parameters: speaking rate, pitch and timbre can be easily toggled up and down by means of clear, no-nonsense controls, as can reverb, echo, sampling rate and stereo balance. **Target**: clients wanting to quickly check the pronunciation of a text or make minor modifications to the speaking rate, audio sampling rate 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 sculpture 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.
Loquendo TTS Director

- Multi-platform Java development tool
- Text is written in the edit box and interactively refined through a "listen & edit" procedure.
- Provides access to Loquendo's TTS Control Tags- i.e. voice, language, prosody, pronunciation, spelling and pausing, audio mixer.
- The Effects menu smoothly guides users to obtain the effect required: "expressive cues" and "plug-in lexicons" (e.g. SMS, toponyms)
- Lexicon editor and SSML.

All TTS features available on TTS Director!!
Command Menus

- File
- Edit
- Control Tags
  - Audio
  - Bookmarks
  - Language
  - Pronunciation
  - Prosody
  - Reading Settings
  - Voice
- Effects
- Configuration
- Tools
- Help
This solution provided an **immediate response** to the market’s requirement for expressivity in speech.

Not yet a fully emotional style for any phrase, but an enhancing repertoire of **Expressive Cues** consisting of a set of pre-recorded formulas (greetings, exclamations, paralinguistic events) which suggest **expressive intention** (to confirm, doubt, exclaim, thank, etc.).

Enriched TTS continues to benefit from the **high quality and natural timbre** achieved with the Unit Selection technique.

---

Good afternoon ladies and gentlemen! Listen to this!

\_Throat I am the American synthetic voice from Loquendo. It gives me great pleasure to be here with you all.

See you later! \_Laugh
Figures of Speech

- These are:
  - **Speech Acts**, divided into intuitive linguistic categories, such as:
    - **Announcements** (dear all! dear customer! ladies and gentlemen! …)
    - **Apologies** (excuse me! I'm so sorry! oh my gosh I'm sorry! …)
    - **Compliments** (good! congratulations! that's great news! …)
    - **Disapproval** (I don't agree! I completely disagree! what a rip off! ..)
    - **Greetings** (all the best! bye bye! hello and welcome! …)
    - **Refusals** (absolutely not! definitely not! I can't! …)
    - **Surprise** (who would have believed it! surprise! …)
    - **Thanks** (thank you! I'm very grateful! thanks for everything! …)

- Non-linguistic interjections
  - (Oh, Aha, Er, Hmm, Doh, Oops…)

- Paralinguistic events
  - (e.g. coughing, laughter, breathing, etc.)
Effects: Speech Acts

- Effects
  - SpeechActs
  - Extras
- Configuration
- Tools
- Help

- .......

- I was like so going to call!
- I'm so sorry!
- I'm sorry!
- apologies!
- excuse me!
- oh my gosh I'm sorry!
- pardon me!
- please forgive me!
- sorry!
Effects: Extra

- File
- Edit
- Control Tags
- Effects
  - SpeechActs
  - Extras
- Configuration
- Tools
- Help
Mixing TTS, Music and Effects in a user-friendly and synchronized way.
Control Tags: Prosody: Reverb, Balance,..
Configuration

- File
- Edit
- Control
- Effects
- Configuration
- Tools
- Help

Variable output frequency coding up to 48 kHz
New custom voices can be created by the adjustment of the prosodic parameters of Loquendo voices:

- Control of the timbre, pitch and speaking rate
- Prosodic settings can be saved to create new custom voices (child, cartoon character, etc.)
- Reading parameters can be saved to create new specialized voices:

Audio file example:

Custom voices

From Susan: her daughter

From Dave: Paul Smith
• File
• Edit
• Control Tags
• Effects
• Configuration
• Tools
  – Voice Creator
  – Lexicon Manager
• Help
Specialized Lexicon

- Reading parameters can be saved to create new **specialized voices**: pronunciation style (e.g. word by word), specialized lexicons (e.g. SMS reading, cities and street names US, UK, German and more coming)

Audio file example:

**SMS reading**

r u don nethng 2mor? would b gr8 2cu.
i'll b @ pub b4 8.
ringl8r. b4n xoxox susan
## Tools: Lexicon Manager

The image shows the interface of the Loquendo Lexicon Manager software. It features a section for English with columns for Original Text, Substitute Text, Phonetic Transcription, and Comment. The software also includes a grid with symbols and letters, likely representing phonetic or linguistic elements.

### Table: English Lexicon

<table>
<thead>
<tr>
<th>Original Text</th>
<th>Substitute Text</th>
<th>Phonetic Transcription</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>asr</td>
<td>Automatic Speech Recognition</td>
<td>\SMPa=;^{s@r/^...^/}</td>
<td></td>
</tr>
<tr>
<td>tts</td>
<td>Text To Speech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>new york</td>
<td>Big Apple</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
<?xml version="1.0" encoding="UTF-8"?>
<speak version="1.0" xml:lang="en">
  <voice name="Paul">
    Good morning. This is Paul Smith. I'd like to speak to Dr. Sunshine, please.
  </voice>
  <voice name="Dave">
    Hold the line a moment, I'll check if she is in her office.
  </voice>
  <voice name="Paul">
    Thank you! 
  </voice>
  <voice name="Dave">
    Yes, Dr. Sunshine is in. I'll put you through.
  </voice>
  <voice name="Susan">
    Hello, this is Dr. Susan Sunshine. May I help you?
  </voice>
  <voice name="Paul">
    Hello, my name is Paul Smith and I'm calling for the position advertised in Sunday's Times of Pizza and spaghetti Express.
  </voice>
  <voice name="Susan">
    Yes, the position is still open. Could I have your name and number please?
  </voice>
  <voice name="Paul">
    Certainly. I'm Paul Smith, and my number is: 415 908455.
  </voice>
</speak>
Loquendo TTS automatically detects each change language and gives the user the option to:

- **Change voice**, choosing from the voices available, based on the language detected.
- **Keep the same voice** as the original language, and apply phonetic mapping between the newly detected language and the original one.

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.

To deal with mixed-language text: web content, movie titles, addresses, e-mail, …
Phonetic Mapping

Applies the foreign language grapheme-to-phoneme transcription-rules to the foreign text, and then maps the transcribed phonemes onto those of the voice's native language in order to access its acoustic units

- Approximate Pronunciation (speaker maintains his native-tongue phonological system when pronouncing foreign words)

German Voice

Italian Voice

French Voice

Spanish Voice
Hear the latest...

| voice=Simon | voice=Dave | voice=Kate | voice=Allison | audio(close) |
| audio(path=C:/VoiceSearch2008) |
| audio(mix=background_pub.wav,loop,name=background_pub.wav,volume=40;fadein=500) |
| pause=1000 |
| audio(mix=02-Lotus.wav,1,name=02-Lotus.wav,volume=10;fadein=1000) |
| pause=4000 |
| voice(name=Allison) Hi! | pause=200 |
| voice(name=Simon) Good morning Allison. | pause=900 |
| voice(name=Allison) Have you started looking for an apartment? | pause=800 |
| voice(name=Simon) Um No, I haven't had time yet. | pause=800 |
| voice(name=Dave) Anyway, it's best to get to know Paris first. |
| voice(name=Simon) Yeah! It's a very big city. |
| audio(mix=footsteps1.wav,1,name=footsteps1.wav,volume=70) |
| audio(mix=wine_glasses.wav,1,name=wine_glasses.wav,volume=80) | pause=2500 |
| voice(name=Allison) Vanguage=French Merci. |
| voice(name=Dave) Vanguage=French Merci. |
| voice(name=Simon) Vanguage=French Merci beaucoup. Vanguage=EnglishGb |
| voice(name=Allison)Vanguage=EnglishUS “Good!” |
| voice(name=Simon) Thank you! That's nearly all the French I know! |
| audio(play=phonRing.wav,1,name=phonRing.wav,volume=20;fadein=3500) |
| audio(play=phonRing.wav,1,name=phonRing.wav,volume=20) |
| voice(name=Dave) Hello? | pause=600 |
| voice(name=Samantha) Hi | pause=1200 |
| voice(name=Dave) Oh | pause=800 |
| voice(name=Samantha) That's great news! | pause=400 |
| voice(name=Dave) Just a minute. |
| audio(mix=chair_sitting_6.wav,1,name=chair_sitting_6.wav,volume=90) |
| pause=100 | volume=40 |
| voice(name=Allison) Throat How do you like the office? |
| voice(name=Simon) It's | pause=10 |
| voice(name=Allison) And the people? |
| voice(name=Simon) Really friendly! I like William a lot! He's amazing with computers. And | language=French |
| voice(name=Jacques Dubois) pause=10 |
| language=EnglishGb is a very nice guy! |
| voice(name=Allison) Vanguage= French Jacques Dubois… | pause=80 |
| language=EnglishUs he's very charming. Everybody likes him. | pause=120 |
| voice(name=Allison) And he has a lovely wife. |
| pause=120 She used to be a pop star when she was young. | pause=240 |
| voice(name=Allison) Have you heard of Susan?
Thank You!

For more information about Loquendo’s products:

- Visit Loquendo’s web-site:
  www.loquendo.com

- Contact me:
  monica.bisacca@loquendo.com

- Subscribe Loquendo Newsletter