Recent Innovations in Speech Technologies

Paolo Baggia

Director of International Standards

Loquendo

GRUPPO TELECOM ITALIA

Outline

- Loquendo Today
- Have There Been Any Recent Innovations?
- Speech Technologies review:
  - Automatic Speech Recognition
  - Text To Speech
  - Speaker Verification and Identification
  - Speech Classifiers
- Standards for Speech Technologies
  - First Generation of Standards
  - Evolutions to a New Generation of Standards
- Conclusions

Company Profile

- **Privately held company (fully owned by Telecom Italia), founded in 2001 as spin-off from Telecom Italia Labs, capitalizing on 30yrs experience and expertise in voice processing.**
- Global Company, leader in Europe and South America for award-winning, high quality voice technologies (synthesis, recognition, authentication and identification) available in 28 languages and 70 voices.
- Multilingual, proprietary technologies protected 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.
- Offices in New York. Headquarters in Torino, local representative sales offices in Rome, Madrid, Paris, London, Munich
- Flexible: About 100 employees, plus a vibrant ecosystem of local freelancers.

International Awards

Loquendo
VOCAL TECHNOLOGY AND SERVICES

Winner of "Market leader-Best Speech Engine" Speech Industry Award 2007, 2008 and 2009

"2008 Frost & Sullivan European Telematics and Infotainment Emerging Company of the Year" Award

Loquendo MRCP Server: Winner of 2008 IP Contact Center Technology Pioneer Award

"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

Have There Been Any Recent Innovations?

Loquendo
VOCAL TECHNOLOGY AND SERVICES

Some areas to be explored:

- **Algorithmic/technological changes**
  - **Research, More data available**
- **Ease of use**
  - **Simplify the use of speech technologies**
- **Exploitation in new sectors**
  - **New devices, new architectures, new applications**
- **Impact of standards on breaking barriers**
  - **10-years of W3C Voice Browsing, IETF protocols, etc.**

Speech Technologies

Loquendo
VOCAL TECHNOLOGY AND SERVICES

Technologies under review will be:

- **Automatic Speech Recognition** – *ASR*
- **Text-To-Speech** – *TTS*
- **Speaker Identification & Verification** – *SIV*
- **Annotation / Classification**

Automatic Speech Recognition

Current ASR Trends

Loquendo
VOCAL TECHNOLOGY AND SERVICES

- **Same core techniques with variations**
  - HMM or hybrid HMM-NN
  - Availability of larger annotated training DB
  - Huge availability of data acquired in real applications
- **Tuning Required**
- **Larger Tasks**
  - Voice search
  - Dictation or Speech-To-Text
  - Natural Language
- **Architectural Changes**
  - Local on-device ASR
  - Remote on-server ASR
  - ASR as-a-Service

Text To Speech

Current TTS Trends

Loquendo
VOCAL TECHNOLOGY AND SERVICES

- **Success of Unit Selection Concatenative technique**
  - Current TTS voices are a good trade-off between intelligibility and naturalness,
    but limited in the control of prosody and expressiveness

- **Challenges:**
  - Prosodic/Intonation control
  - Multilinguality
  - Text Normalization
  - Emotional Speech Synthesis

- **Emerging New Techniques**
  - HMM-based synthesis:
    highly adaptive to allow more flexible voice transformations
  - Mixture of Concatenative and HMM techniques

➔ **TTS is becoming ubiquitous**
Clean integration is needed, TTS as a common resource shared among different Apps

Speaker Verification
Annotation / Classification

Voice Biometrics

- **Early Applications:**
  - Voice in addition to other authentication techniques
  - Successful results: Non critical tasks (password reset)

- **Challenges:**
  - Robustness to device switch
  - Robustness to channel switch (mobile, fixed, VoIP)
  - Voice quality can change (tiredness, sickness, etc.)

➔ **Promising technology for mobile applications**

Speech Classifiers

Loquendo
VOCAL TECHNOLOGY AND SERVICES

Many characteristics can be determined from speech:

- **Gender**
- **Language**
- **Speaker Identity**
- **Emotional state**

New applications:

- **Enrich dialog information / Speech analytics**
- **Security / Limit human intervention**

Challenges:

- **Extend current architecture to include these classifiers**

➔ **Potentially very interesting in many domains**

10 Years of Standards

Loquendo

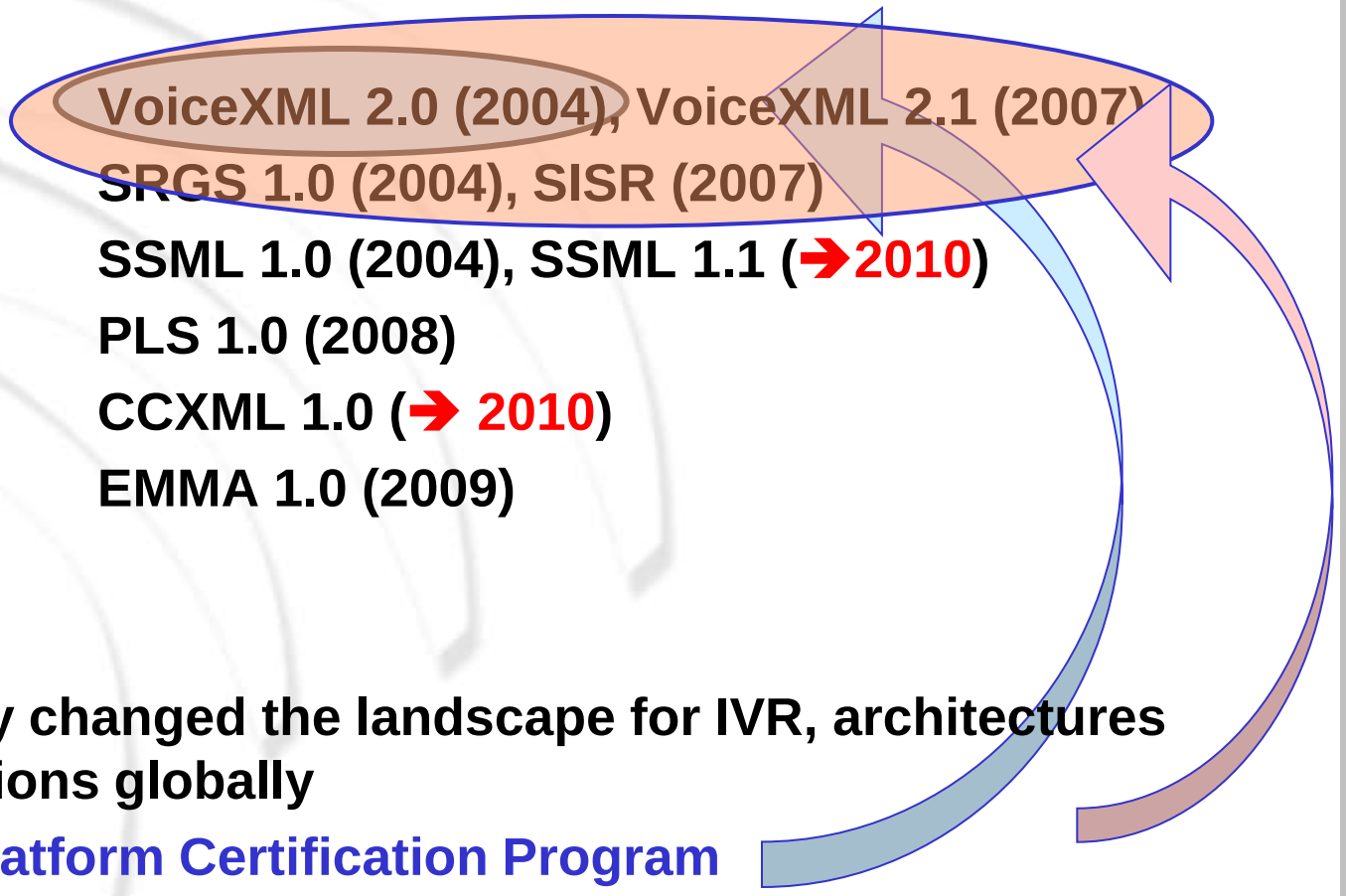
GRUPPO TELECOM ITALIA

First Generation (I)

Loquendo
VOCAL TECHNOLOGY AND SERVICES

Languages (W3C):

- **Dialog:** VoiceXML 2.0 (2004), VoiceXML 2.1 (2007)
- **Grammars:** SRGS 1.0 (2004), SISR (2007)
- **Prompts:** SSML 1.0 (2004), SSML 1.1 (➔2010)
- **Lexicons:** PLS 1.0 (2008)
- **Call Control:** CCXML 1.0 (➔ 2010)
- **MMI Input:** EMMA 1.0 (2009)



➔ **This abruptly changed the landscape for IVR, architectures and applications globally**

➔ **VoiceXML Platform Certification Program by VoiceXML Forum**

First Generation (II)

Loquendo
VOCAL TECHNOLOGY AND SERVICES

Protocols (IETF):

- **MRCPv1, MRCPv2**
- **Other protocols under development for SIP and Voice integration**

Registries (IANA):

- **Language Subtag Registry**
- **(in future) Pronunciation Alphabet Registry**

New Generation of Standards

Loquendo
VOCAL TECHNOLOGY AND SERVICES

Modeling Interaction:

State Charts: **SCXML 1.0 (close to LCWD)**

Voice Browsing:

VoiceXML 3.0

- Detailed semantic description
- Profiling and Modularization
- New dialog resources:
  New media, Speaker Verification and Identification, VCR Controls, etc.
- Integration and flexibility:
  SCXML, Customizable FIA, Event model compatible with DOM3 and browser, MMI Interaction

Conclusions

- Current trends for speech technologies
- Emerging new applications for mobile
- Architectural changes
- Advances in the standards domain

Loquendo
VOCAL TECHNOLOGY AND SERVICES

THANK YOU

for clarifications or questions:

paolo.baggia@loquendo.com