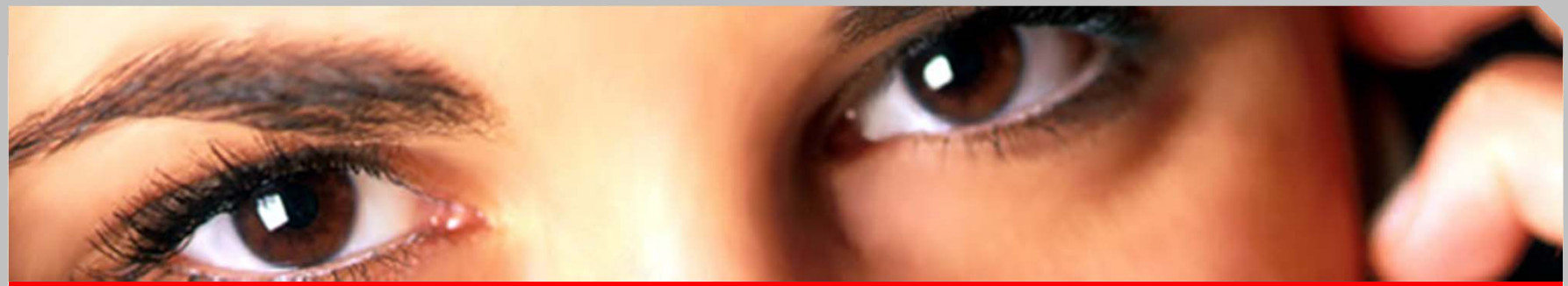


Flexible and Robust ASR Grammars

Paolo Baggia

Loquendo
VOCAL TECHNOLOGY AND SERVICES

Overview

- **Inside the ASR**
- **Language Constraints**
  - Speech Grammars
  - SLMs
  - Pros and Cons
- **More flexible grammars**
  - Garbage Techniques
  - Experimental Results
  - Applications
- **Final Remarks**

Loquendo Today

Loquendo
VOCAL TECHNOLOGY AND SERVICES

- A *Global company* born in 2001 as a spin-off of the Telecom Italia R&D center with over *30 years experience*
- A *Telecom Italia Group* company, which guarantees financial solidity and reliability to Customers
- The *European Leader* in Voice Technologies
- *HQ in Turin (Italy)* with Sales offices in New York (US), Madrid (Spain), Munich (Germany) and Paris (France), as well as a Worldwide Network of Agents.

ISO 9001 CERTIFIED

W3C®

I E T F

- *Multilingual speech technologies* (TTS, ASR, SV)
- *Full support of relevant standards* (VoiceXML, MRCP, VoIP)

"Market leader-Best Speech Engine" Speech Industry Award 2007

Speech Technology magazine's 2007 Market Winner Award

"Best Innovation in Automotive Speech Synthesis" Prize AVIOS-SpeechTEK West 2007

AVIOS Best of AVIOS/SpeechTEK 2007

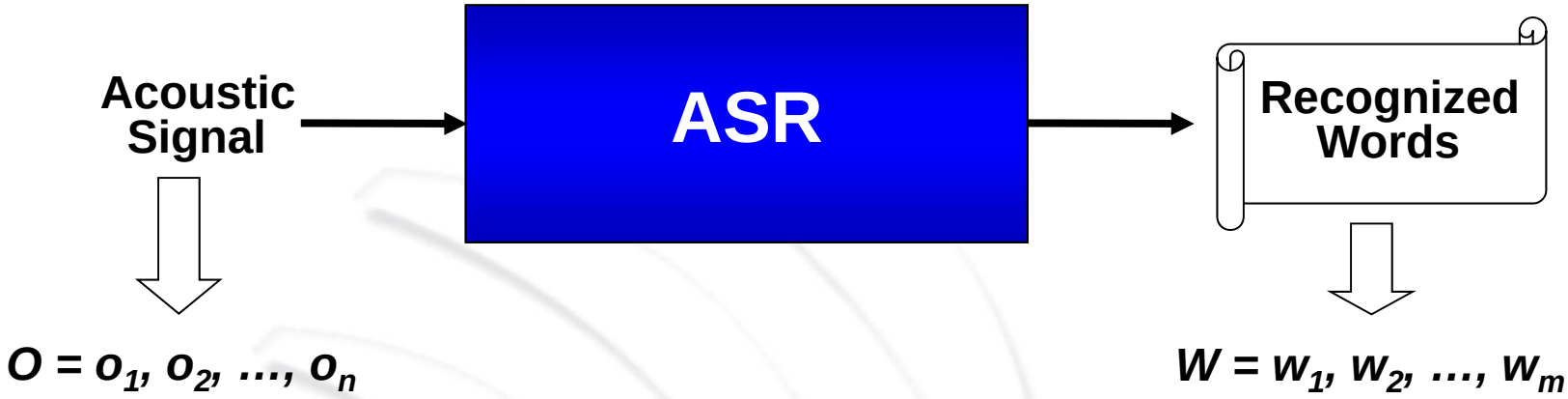
"Best Innovation in Expressive Speech Synthesis" Prize AVIOS-SpeechTEK West 2006

AVIOS Best of AVIOS/SpeechTEK 2006

"Best Innovation in Multi-Lingual Speech Synthesis" Prize AVIOS-SpeechTEK West 2005

AVIOS Best of AVIOS/SpeechTEK 2005

Inside the ASR engine



$$\hat{W} = \arg\max_{W \in L} P(W \mid O) = \arg\max_{W \in L} \frac{P(O \mid W)P(W)}{P(O)}$$

$$\cong \arg\max_{W \in L} P(O \mid W)P(W)$$

Likelihood ***Prior***

ASR Component Probabilities

- $P(O|W)$ **Likelihood**

  Acoustic Models ➔ **ASR engine**

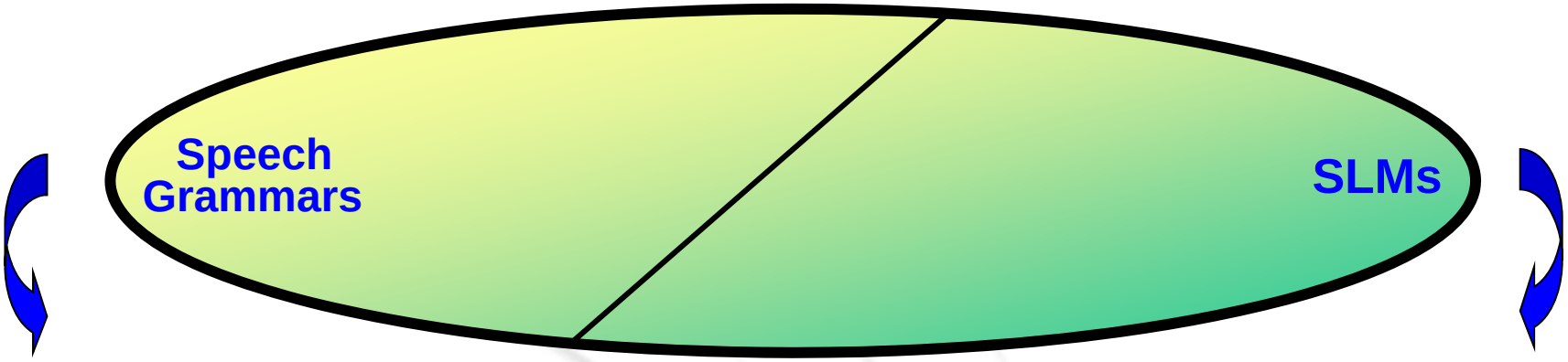
  (possible adaptation)

- $P(W)$ **Prior Probability**

  Language Constraints ➔ **Application Developer**

  COSTLY ACTIVITY

Speech Grammars vs. SLMs



SPEECH GRAMMARS

- **Method:**
  A compact and yet complete description of the user's response
- **Constraints:**
  Rules have to correctly represent sentence construction
- **Issues:**
  - **Standard syntax (W3C SRGS)**
  - **Standard results (Semantic Interpretation – W3C SISR)**
  - **Best performance for in grammar utt.**
  - **Grammar should cover all possible responses**
  - ➔ **Cost of developing a grammar**
  - ➔ **Cost of fine-tuning grammars (need for tuning tools)**

Proposal of new techniques in the middle

STATISTICAL LANGUAGE MODELS

- **Method:**
  Assesses the probability of word occurrence in a sentence
- **Constraints:**
  Probability of 2, 3, …, n preceding words (n-gram)
- **Issues:**
  - **Easy to generate, but only with a specific corpus for each application**
  - **Difficult to assign probabilities to unforeseen events (*smoothing techniques*)**
  - ➔ **Very large corpora are needed**
  - ➔ **Time-consuming transcription of field data required to tune the SLMs**

More Flexible Grammars

- **How can we simplify grammar creation?**
- **We can focus on modeling just the relevant content, and not the rest of the phrase.**

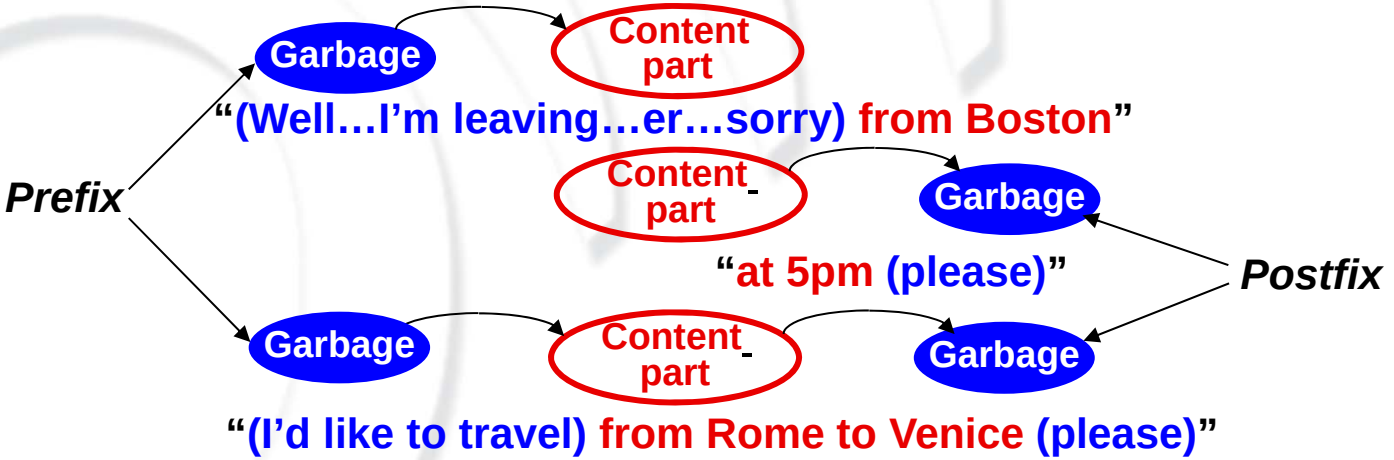
  (I'd like to travel) from Rome to Venice (please)

  (I need to go) from Rome to Venice (as soon as possible)

  (Well…I'm going…er…sorry, a ticket) from Rome to Venice

➔ **Use a special grammar node to discard the rest of the sentence!**

- **Where to put the Garbage rules?**
  - Simplest and most effective:



  - More complex usages are possible, but more tricky!

Example Grammar with Garbage Rules

```
<?xml version="1.0" encoding="UTF-8"?>
<grammar version="1.0" tag-format="semantics/1.0" mode="voice" root="MainRule"
    xml:lang="en-US">

  <rule id="MainRule" scope="public">









  </rule>









</grammar>
```

- SRGS standard grammar

Example Grammar with Garbage Rules

Loquendo
VOCAL TECHNOLOGY AND SERVICES

```
<?xml version="1.0" encoding="UTF-8"?>
<grammar version="1.0" tag-format="semantics/1.0" mode="voice" root="MainRule"
    xml:lang="en-US">

  <rule id="MainRule" scope="public">

    <item>
      from
        <ruleref uri="#city"/>
        <tag>out.from = rules.latest();</tag>
      to
        <ruleref uri="#city"/>
        <tag>out.to = rules.latest();</tag>
    </item>

  </rule>

  <rule id="city">
    <one-of>
      <item> Rome </item><tag>out = "FCO";</tag>
      <item> Venice </item><tag>out = "VCE";</tag>
      <!-- Many other cities here! -->
    </one-of>
  </rule>
</grammar>
```

- SRGS standard grammar
- Content part of the grammar

Example Grammar with Garbage Rules

Loquendo
VOCAL TECHNOLOGY AND SERVICES

```
<?xml version="1.0" encoding="UTF-8"?>
<grammar version="1.0" tag-format="semantics/1.0" mode="voice" root="MainRule"
    xml:lang="en-US">

  <rule id="MainRule" scope="public">
    <item repeat="0-1">
      <ruleref special="GARBAGE"/>
    </item>
    <item>
      from
        <ruleref uri="#city"/>
        <tag>out.from = rules.latest();</tag>
      to
        <ruleref uri="#city"/>
        <tag>out.to = rules.latest();</tag>
    </item>
    <item repeat="0-1">
      <ruleref special="GARBAGE"/>
    </item>
  </rule>

  <rule id="city">
    <one-of>
      <item> Rome </item><tag>out = "FCO";</tag>
      <item> Venice </item><tag>out = "VCE";</tag>
      <!-- Many other cities here! -->
    </one-of>
  </rule>
</grammar>
```

- SRGS standard grammar
- Content part of the grammar
- (Optional) Garbage rules

Input:
"(I'd like to travel) from Rome to Venice (please)"

Semantic results (ECMA obj):
{ from="FCO", to="VCE" }

Garbage Techniques

Loquendo
VOCAL TECHNOLOGY AND SERVICES

a. General Model

- It can be trained at an acoustic level as an average model or as the average of the first N-best activated unit
- A calibration of the general model could be problematic

b. Filler Words

- A vocabulary composed of different words from the speech grammar is inserted as a garbage node
- Efficiency and accuracy performance greatly depends upon the number and choice of the added words

c. Phonetic Models ➔ *Loquendo ASR solution*

- A node containing a chosen subset of the phonetic units of the acoustic model is selected and used as a garbage node
- The "power" of the garbage could be easily modulated through the unit subset selection and the garbage node weights
- Vocabulary size is usually smaller than the list of 'filler words'

Italian	**-1.26%**
Spanish	**-0.10%**
English	**-0.99%**

Average % Accuracy Loss on Built-in Grammars: dates, pin codes, currency amounts, time expressions, etc.

300 Filler Words (random selection)	**90.5%**
300 Filler Words (frequent syllables)	**95.3 %**
41 Filler Words (*Oracle* -> test word)	**98.6 %**
Phonetic garbage	**97.3 %**

Identifying Keywords: months of the year in spontaneous date expressions. The phonetic garbage outperforms the filler words technique except in the *Oracle* case.

Further Considerations

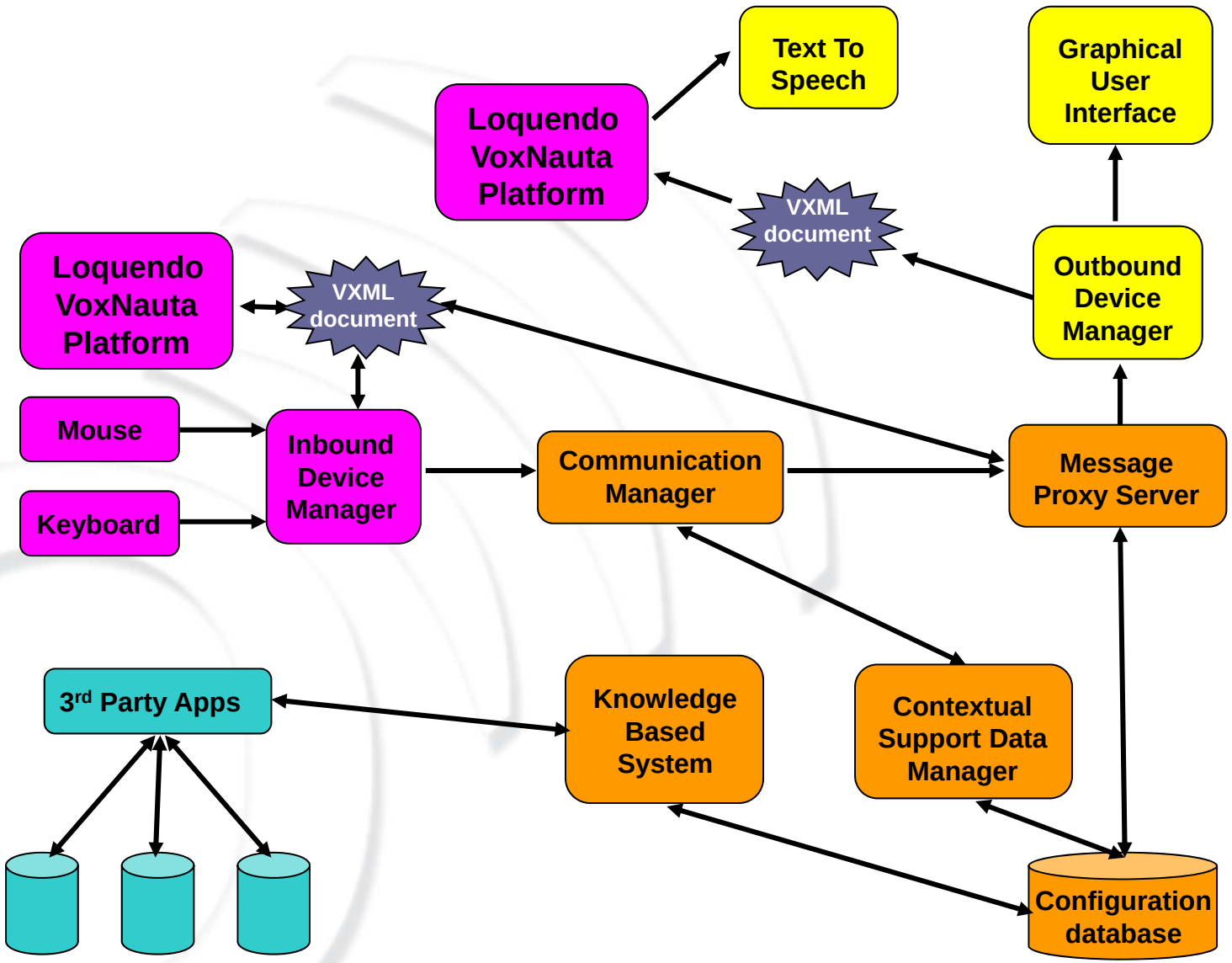
- **Garbage model is a flexible and powerful solution, but …**
  - Difficult to use if the content grammars contains very similar words
  - The number of garbage nodes in a grammar should be limited (4 – 5 max)
  - Garbage nodes could benefit from grammar weights in some circumstances
  - For some tasks the best results are achieved when speech grammars fully cover vocal user formulations
  - Loquendo Phonetic Learner allows analysis of recognition log data to discover frequently used formulations covered by the garbage
- **When the complexity of a recognition task, in terms of user formulations' prediction, is too high…**
  - A Statistical Language Model is better, but an appropriate corpus is needed
  - A grammar with garbage rules could be used as a starting point for the acquisition of the corpus

New Breakthrough with VoxIQ

Loquendo
VOCAL TECHNOLOGY AND SERVICES

- **VoxIQ means:**
  - Massive reduction in speech recognition task
  - Removal of overall vocabulary limitation
  - Exploitation of Loquendo ASR and VoxNauta platform
- **In contact centres, VoxIQ allows:**
  - Complex conversation with support of automatic systems
  - Display of information in line with an ongoing conversation between people, plus automatic form filling
  - Out of context information to be captured and then displayed when that context is being discussed
- **Call centre cost and productivity benefits (estimate):**
  - Agents improve productivity by some 15-20%
  - Complex conversational self service, saving >90% on agent costs
  - Development costs reduced by 80%

VoxIQ Architecture



VoxIQ System Exploits Garbage Rules

Loquendo
VOCAL TECHNOLOGY AND SERVICES

powered by VoxIQ

"(I'd like a) skinny cappuccino (with) 2 sugars"

Results: Milk = Yes
Milk Type = skimmed
Coffee Type = cappuccino
Sugar = Yes
Number of Sugars = 2

Grammars: (4 Coffee Types * 3 Sizes * 3 Milk Types * 2 Sprinkles * 6 Sugars)
432 Combinations
For Just 6 Slots = 2,592 Forms

SLMs: Significant Corpus Collection
Transcription
Training Requirement > 2,592 Utterances
Data Sparsity ➔ Lower Accuracy

Preliminary tests: Accuracy over 90%

Remarks:

- ➔ Loquendo ASR with Garbage rules makes keyword spotting possible.
- ➔ VoxIQ manages the conversation and collects keywords, capturing their semantic value in a form.
- ➔ An example demo was developed in 4 hours with subsequent tuning and refinement in 10 hours

Final Remarks

- **Develop ASR application is a costly activity, even if progresses have been done on standard formats and tools**

- **Garbage rules in grammars:**
  - Greatly simplifies the grammar development
  - May be used to fast prototype systems to be tuned in a second phase
  - Helps in noisy and chatty environments (garbage covers unwanted speech)
  - Promotes a more flexible dialog developments (from system guided dialogs to conversational dialogs)

- **First feedbacks from real applications:**

VoxIQ experience, Carabinieri Virtual Operator (*Italian Armed Force and Police Authority*), applications being done in Spain

Live Demo – Carabinieri Force (in Italian)

Loquendo
VOCAL TECHNOLOGY AND SERVICES

Carabinieri Force – Armed Force and Police Authority in Italy

- Go to www.carabinieri.it web-site
- Click on “Operatore Virtuale” (Virtual Operator)
- Call +39 06 80985232 (It speaks in Italian!)

An automated service for the police force providing info on joining the force and moving up through the service, on upcoming exams, exam results etc. The caller can speak freely and naturally thanks to an extensive use of garbage nodes.

Thank You!

Loquendo
VOCAL TECHNOLOGY AND SERVICES

For more information about Loquendo's products:

- **Visit Loquendo's web-site:** www.loquendo.com
- **Contact us:**

  (for technical issues) paolo.baggia@loquendo.com
  (for business) monica.bisacca@loquendo.com

- **Subscribe Loquendo Newsletter:** **… I'm the editor!**