Tuning

Jenny Burr
August 2006
Discussion Topics

• What is tuning?
• What is the process of tuning?
What is tuning?
What is Tuning?

• It is optimizing application performance
  – The caller experience
  – The accuracy of the speech recognizer (ASR engine)
• Optimize based on quantifiable data
• Using statistical significance of real world data
Why is it Important?

• Fine tune the application based on:
  – Caller/Customer experience
  – Goal/transaction completion rates
  – Recognition accuracy rates

• Results:
  – Improved call automation rates
  – Increased user satisfaction
  – Larger return on investment
What are the Goals of Tuning?

• It’s a balancing act
  – Business and caller goals need to be defined at the beginning of the application’s life cycle
    • Goal/Transaction completion rates
    • Caller satisfaction
    • Call Transfer rates
  – Grammar complexity
    • Too many items in the grammars can cause false accepts; too few leads to out-of-grammars
  – Threshold Parameters
Optimizing Recognition Accuracy

• Tuning the Grammars
  – What can be said: words and phrases
  – What is never said: over-generation

• Tuning the Recognition Lexicon
  – How it can be said: the pronunciations
    • Dialect
    • Cross word combination: Phrases often get run together resulting in phonological differences
    • Optimize current default dictionary pronunciations
Optimizing the VUI

• Review the prompts
  – Wording, clarity, volume

• Review the error strategy
  – Are callers able to recover?

• Caller reactions
  • Unexpected language, confusion
Sample Implementation Methodology

**Design Phase**

- Requirements Gathering
- Requirements Analysis
- Functional Requirements Document
  - Creation & Approval
- Sample Call – Conversations
- Voice Talent Selection
- Full Design Specification
  - VUI
  - Business Logic
  - Data Interaction
- Customer Specification Sign-off
- Development
- Testing
- Usability Testing
- Production Tuning
The Tuning Process
The Tuning Process

1. Collect tuning data
   - logging & utterances
2. Transcribe caller utterances
3. Generate statistical reports
4. Analyze performance statistics
   - Recognition, Transaction/Goal Completion
5. Listen to call recordings (if available)
6. Analyze issues and create recommendations
7. Run experiments on recommendations
8. Write report outlining issues and recommendations
9. Review report
10. Implement recommendations
Collect Tuning Data

• Collect from a live production environment
  – Do not collect test calls during User Acceptance Testing
• Need a statistical sample of data
  – Generally 2,000-5,000 calls
  – Or a minimum of 100 utterances per task
  – No changes should be made with fewer than 30 samples
  – Tuning for SLM grammars requires thousands of utterances
• Changes to the application should not occur during this collection period
Discussion

• How can production calls be directed to the application for tuning purposes?
Transcription

- Listen to caller’s utterances
- Determine if the utterance matched the logged recognition result
  - If not, notate what occurred and/or what was spoken
- Notate noise events
  - Background noise, side conversations, coughing, etc
- Spelling in transcriptions should match grammar
## Generate Accuracy Statistics

<table>
<thead>
<tr>
<th>In-Grammar (IG)</th>
<th>Out-of-Grammar (OOG)</th>
<th>Out-of-Coverage (OOC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caller uttered a word or phrase that was “in the grammar”</td>
<td>Utterances that are not covered by any of the active grammars. This category includes both speech utterances and non-voice utterances such as background noise, side speech, and coughing.</td>
<td>A subset of the out-of-grammar statistics that demonstrates the behavior of the system when presented with direct speech utterances that are not covered by any of the active grammars.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correct Acceptance (CA-in)</th>
<th>An utterance in the grammar is correctly recognized.</th>
<th>Correct Rejection (CR-OOG)</th>
<th>An utterance that is not in the grammar is appropriately rejected.</th>
<th>Correct Rejection (CR-OOC)</th>
<th>An utterance that is not in the grammar is appropriately rejected.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Rejection (CR-OOG)</td>
<td>An utterance that is not in the grammar is not in the grammar</td>
<td>False Rejection (FR-in)</td>
<td>An utterance is in the grammar, but is not recognized (it is rejected).</td>
<td>False Rejection (FR-OOG)</td>
<td>An utterance is in the grammar, but is not recognized (it is rejected).</td>
</tr>
<tr>
<td>False Acceptance (FA-OOG)</td>
<td>An utterance in the grammar is mistakenly recognized as another item in the grammar.</td>
<td>False Acceptance (FA-OOG)</td>
<td>An utterance that is not in the grammar is mistakenly recognized as an item in the grammar.</td>
<td>False Acceptance (FA-OOC)</td>
<td>An utterance that is not in the grammar is mistakenly recognized as an item in the grammar.</td>
</tr>
</tbody>
</table>
## Accuracy Statistics

<table>
<thead>
<tr>
<th>Context Name</th>
<th>Number of Transcribed Utterances</th>
<th>IG Count</th>
<th>IG Rate</th>
<th>CA-in</th>
<th>FA-in</th>
<th>FR-in</th>
</tr>
</thead>
<tbody>
<tr>
<td>MainMenu</td>
<td>865</td>
<td>701</td>
<td>81.04%</td>
<td>92.74%</td>
<td>1.57%</td>
<td>5.69%</td>
</tr>
<tr>
<td>Repeat_Details</td>
<td>814</td>
<td>691</td>
<td>84.89%</td>
<td>95.61%</td>
<td>0.87%</td>
<td>3.53%</td>
</tr>
<tr>
<td>Payment_Menu</td>
<td>551</td>
<td>485</td>
<td>88.02%</td>
<td>82.37%</td>
<td>5.82%</td>
<td>11.80%</td>
</tr>
<tr>
<td>YesNoConfirm</td>
<td>475</td>
<td>411</td>
<td>86.53%</td>
<td>97.73%</td>
<td>0.00%</td>
<td>2.27%</td>
</tr>
<tr>
<td>AskPhoneNbr</td>
<td>80</td>
<td>22</td>
<td>27.50%</td>
<td>89.82%</td>
<td><strong>10.18%</strong></td>
<td>0.00%</td>
</tr>
</tbody>
</table>
Reviewing the Statistics

- In-grammar statistics
  - Review for problem areas
  - What is the percentage of false acceptance?
    - Will adding/modifying pronunciations help?
    - Will modifying the grammar help?
  - What is the percentage of false rejections?
    - Is the caller able to recover with the reprompt?
  - Should the confidence thresholds (or other parameters) be changed?
    - How will that affect the false acceptance and false rejection rates?
Exercise

• Review the false accept in-grammar rates and determine ways of improving the recognition.
  – See handout
## Accuracy Statistics

<table>
<thead>
<tr>
<th>Context Name</th>
<th>Number of Transcribed Utterances</th>
<th>OOG Count</th>
<th>OOG Rate</th>
<th>CR-OOG</th>
<th>FA-OOG</th>
<th>OOC Count</th>
<th>OOC Rate</th>
<th>CR-OOC</th>
<th>FA-OOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MainMenu</td>
<td>865</td>
<td>164</td>
<td>18.96%</td>
<td>77.44%</td>
<td>22.56%</td>
<td>108</td>
<td>12.49%</td>
<td>68.52%</td>
<td>31.48%</td>
</tr>
<tr>
<td>Repeat_Details</td>
<td>814</td>
<td>123</td>
<td>15.11%</td>
<td>70.73%</td>
<td>29.27%</td>
<td>102</td>
<td>12.53%</td>
<td>68.63%</td>
<td>31.37%</td>
</tr>
<tr>
<td>Payment_Menu</td>
<td>551</td>
<td>66</td>
<td>11.98%</td>
<td>83.33%</td>
<td>16.67%</td>
<td>42</td>
<td>7.62%</td>
<td>80.95%</td>
<td>19.05%</td>
</tr>
<tr>
<td>YesNoConfirm</td>
<td>475</td>
<td>64</td>
<td>13.47%</td>
<td>70.31%</td>
<td>29.69%</td>
<td>60</td>
<td>12.63%</td>
<td>68.33%</td>
<td>31.67%</td>
</tr>
<tr>
<td>AskPhoneNbr</td>
<td>80</td>
<td>58</td>
<td>72.50%</td>
<td>79.31%</td>
<td>20.69%</td>
<td>26</td>
<td>32.5%</td>
<td>69.23%</td>
<td>30.77%</td>
</tr>
</tbody>
</table>
Reviewing the Statistics

- Out-of-grammar statistics
  - Review problem areas
  - Ensure that the issues are statistically significant
    - Should you add any of the OOG/OOC items to the grammar?
    - Should the prompt be rewritten to provide clearer instructions?
    - Is the system correctly rejecting the OOG/OOC phrases?
Exercise

• Review OOG listing and listen to prompt: determine if the grammar should be modified or if the prompt should be changed
  – See handout
Grammar Over-generation

- Occurs when a grammar contains unlikely and/or unused phrases
- Results in unnecessary complexity
- Could result in nonsensical false acceptances
- Depending on grammar complexity, could result in recognition latency
Exercise

• Review a grammar for over-generation
• Determine items to remove
  – See handouts
Transaction/Goal Completion

• Review the goal/transaction completion statistics
  – Did the calls end with callers completing their goals?
    • If not, where were the problems?
Compile Recommendations

• Determine the problem areas and identify ways to improve the application
  – Prompt changes
  – Error strategy
  – Pronunciations
  – Grammar additions/deletions
  – Weighting grammar items
  – Parameter changes
  – Frequent Timeouts
  – Sensitivity
  – Caller Confusion
Grammar Guidelines

• NEVER “make up” words in the grammar to approximate pronunciations
  – Example: “tee vee” as a grammar option for “t v”
  – Use the dictionary to create any needed pronunciations

• Apostrophes are a good thing - USE THEM!
  – “that’s” not “thats”
  – “i’m” not “im”
  – “i’m all done” not “im all done”

• Use pre-and-post fillers only where appropriate
• Limit synonyms for initial production roll-out
Experiments

• What’s the purpose?
• Test potential grammar and parameter changes against the collected data set
• Regenerate the accuracy statistics to validate the recommendations
• Enumerate the difference the recommended change had on application performance
Tuning Report

• The tuning report outlines
  – Any adjustments:
    • Grammars
    • Pronunciations
    • Parameters
    • Prompts
    • Application logic
Next Steps

- Implement recommended changes
- Gather data for a tuning effort to validate changes (if needed)
Summary

– Data gathering is very important
  • Need enough data per recognition state
  • Data should vary for different caller populations and caller types
– The transcription process is critical
– Ensure that any issues reported are statistically significant
  • Changes to the application should be made based on multiple occurrences of the problem
  • Do not make changes for “one-offs”
Empowering Your customers and employees with speech technologies

SpeechTEKI 2006 Empower
The Voice Solutions Showcase

Thank You