

# Tuning

Jenny Burr  
August 2006

SpeechTEK | 2006  
The Voice Solutions Showcase



**Empower**

# Discussion Topics

- What is tuning?
- What is the process of tuning?

What is tuning?



**Empower**

# 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

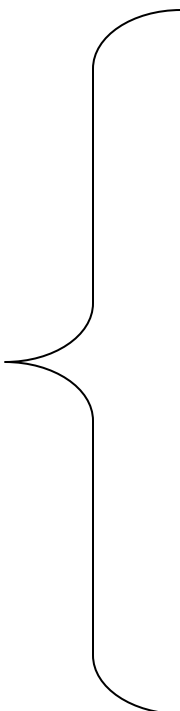
- Tuning the Grammars
  - What can be said: words and phrases
  - What is never said: over-generation
- Tuning the Recognition Lexicon
  - How it can be said: the pronunciations
    - Dialect
    - Cross word combination: Phrases often get run together resulting in phonological differences
    - Optimize 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



**Empower**

# 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

In-Grammar (IG)		Out-of-Grammar (OOG)		Out-of-Coverage (OOC)	
Caller uttered a word or phrase that was "in the grammar"		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.		A subset of the out-of-grammar statistics that demonstrates the behavior of the system when presented with direct <i>speech</i> utterances that are not covered by any of the active grammars.	
<b>Correct Acceptance (CA-in)</b>	An utterance in the grammar is correctly recognized.	<b>Correct Rejection (CR-OOG)</b>	An utterance that is not in the grammar is appropriately rejected.	<b>Correct Rejection (CR-OOC)</b>	An utterance that is not in the grammar is appropriately rejected.
<b>False Rejection (FR-in)</b>	An utterance is in the grammar, but is not recognized (it is rejected).				
<b>False Acceptance (FA-in)</b>	An utterance in the grammar is mistakenly recognized as another item in the grammar.	<b>False Acceptance (FA-OOG)</b>	An utterance that is not in the grammar is mistakenly recognized as an item in the grammar.	<b>False Acceptance (FA-OOC)</b>	An utterance that is not in the grammar is mistakenly recognized as an item in the grammar.

# Accuracy Statistics

Context Name	Number of Transcribed Utterances	IG Count	IG Rate	CA-in	FA-in	FR-in
MainMenu	865	701	81.04%	92.74%	1.57%	5.69%
Repeat_Details	814	691	84.89%	95.61%	0.87%	3.53%
Payment_Menu	551	485	88.02%	82.37%	5.82%	11.80%
YesNoConfirm	475	411	86.53%	97.73%	0.00%	2.27%
AskPhoneNbr	80	22	27.50%	89.82%	10.18%	0.00%

# 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

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

# 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*

**SpeechTEK | 2006**  
The Voice Solutions Showcase

**Empower**

*Thank You*