**White Paper**

# Florida Advanced Traveler Information System Project

# FL-ATIS IVR Spanish Recognition Improvement Project – Utilizing the Six Sigma DMAIC Methodology

## Final: Version 1.2

## January 28, 2010

Prepared for:
Florida Department of Transportation
Traffic Engineering and Operations Office
605 Suwannee Street, M.S. 90
Tallahassee, Florida 32399-0450
(850) 410-5600

| DOCUMENT CONTROL PANEL | | |
|---|---|---|
| File Name: | *100128 FLATIS IVR Spanish Recognition Improvement Project_v1_2_*Final.docx | |
| File Location: | Z:\C8I75\Assign 54 - FLATIS\FLATIS IVR Spanish Recognition Improv Proj\100128 FLATIS IVR Spanish Recognition Improvement Project_v1_2_Final.docx | |
| Version Number: | 1.2 | |
| | **Name** | **Date** |
| Created By: | Erik Gaarder, PBS&J | 1/19/2010 |
| | | |
| Reviewed By: | TJ Hapney, PBS&J | 1/25/2010 |
| | TJ Hapney, PBS&J | 1/27/2010 |
| | | |
| | | |
| | | |
| | | |
| | | |
| Modified By: | TJ Hapney, PBS&J | 1/25/2010 |
| | TJ Hapney, PBS&J | 1/27/2010 |
| | TJ Hapney, PBS&J | 1/28/2010 |
| | | |
| | | |
| | | |
| Completed By: | TJ Hapney, PBS&J | 1/28/2010 |

Date Started: October 1, 2009
Date Completed: January 15, 2010
Project Sponsor/Champion: Gene Glotzbach - FDOT
Phone: 850-410-5616
Email: gene.glotzbach@dot.state.fl.us

Additional team members:
Olga Sanchez (LogicTree) – Project Manager
Larry Baldwin (LogicTree) – Director User Experience
Ethan Levine (LogicTree) - FLATIS IVR Sr. UX Designer
Evelyn Perez (PBS&J) - Spanish Transcriber
Lisa (T.J.) Hapney (PBS&J) - QA/QC

## Acronym List

ATIS ..... Advanced Traveler Information System
dB ..... Decibel
DMAIC ..... Define, Measure, Analyze, Improve, Control
DTMF ..... Dual Tone Multi-frequency
FDOT ..... Florida Department of Transportation
FL-ATIS ..... Florida Advanced Traveler Information System
IVR ..... Interactive Voice Response
RTMC ..... Regional Transportation Management Center
SIPOC ..... Suppliers, Inputs, Process, Outputs and Customers

## Table of Contents



## List of Appendices

# 1. Executive Summary

Since launch in June 2009, Florida Advanced Traveler Information System (FL-ATIS) interactive voice response (IVR) recognition has been one of the principal complaints received from the public on the new statewide 511 advanced traveler information system (ATIS). Although tuning was conducted on the system in August 2009, unresolved issues still existed, especially with Spanish recognition. This project was initiated to identify the underlying causes of Spanish voice recognition performance issues.

A cross-functional problem solving team with members from LogicTree and PBS&J was formed to tackle this problem. This team used the Six Sigma Define, Measure, Analyze, Improve, Control (DMAIC) problem solving process to identify opportunities for improvement. First, a qualitative process analysis was conducted to identify potential causes for the IVR recognition issues. Then a detailed analysis of process flows, lessons learned from the previous English tuning report, and brainstorming resulted in identifying the critical dialog states to study and the key input variables. Spanish utterances were analyzed over a two month period (October and- November, 2009).

A major finding of this project was that the voice recognition low confidence level needs to be reset. This rejection confidence level is set too low, causing a false acceptance rate ranging from 15.6 to 18.1 percent in the three dialog states analyzed. This project has recommended that the confidence level be moved from 7 to 18 which will decrease the false acceptance rate in the three dialog states by over 50 percent.

The project was also able to identify grammars in each dialog state that need to be updated in the system. The result of these actions will be improved customer satisfaction, fewer feedback complaints, and the elimination of the need to continue the customer technical support call center. The project also recommends that a “p chart of defectives” be used to monitor the process going forward.

## 2. Define Phase (Improvement Opportunity)

Voice recognition performance for the Florida Department of Transportation's (FDOT) next generation bilingual statewide 511 ATIS is one of the top feedback complaints from users. Over 730 complaints about voice recognition have been received in the 6 month period following the system launch in June 2009 (see Appendix A for the Pareto chart of FL-ATIS User Feedback Issues). Reviewing this feedback and analyzing a high level process map (see Appendix B) one can state the impact of poor voice recognition performance quality on the 511 customer as:

- Inability to receive requested traveler information when the voice recognition process does not understand what was said (i.e., "What was that?").
- Delay in receiving traveler information when the voice recognition process asks for confirmation.
- Receiving the wrong traveler information when the voice recognition system misunderstands what was said.

Voice recognition issues incur costs due to the administrative time needed to respond to dissatisfied customers, and they are also the principal reason the FDOT extended the 511 customer technical support call center at a cost of over $620,000 per year. The team limited its scope to improving the Spanish voice recognition process because it was recognized that this problem was not easily or quickly solvable and would benefit from the application of the DMAIC methodology. The cost of poor quality of the Spanish recognition process is estimated to be $127,000 per year[1]. There is the potential for this cost to increase, if Spanish recognition is not improved, since overall Spanish usage of the 511 system is expected to increase going forward[2].

A Critical to Quality Tree (see Appendix C) was created to identify, organize, and display parts of the process according to areas of critical importance. This project will focus on identifying and correcting the issues related to the accuracy of recognition for the Spanish voice recognition process. For October and November 2009, the non-adjusted recognition rate for Spanish in the Main Menu dialog state was 47.2 percent compared to 59.6 percent for English.[3] The project goal is to get Spanish recognition up to par with English recognition. This will result in improved customer satisfaction (fewer complaints) and improved FDOT confidence in the voice recognition system so that the customer technical support call center does not need to be extended for a second time.

[1] $127,000 per year is based on $10,000 per year for Spanish administrative costs (estimate) and $117,000 as the portion of the call center costs attributable (by population) to Spanish recognition (18.9 percent of $620,000).
[2] Spanish 511 usage is currently around 2 percent but it is expected to increase substantially because 18.9 percent of the state of Florida is Spanish speaking. Spanish population data is based on the 2008 American Community Survey, accessed at factfinder.census.gov on December 30, 2009.
[3] The non-adjusted recognition rate (Total number of "Accepted matches"/ Total number of "caller utterances"). Spanish: 1747/3703 = 47.2 percent, English: 245,575/412,316=59.6 percent.

## 3. Measure Phase (Current State of the Process)

The voice recognition software has a database that stores the utterances it receives, in each dialog state, along with all the pertinent input variables. To keep the project manageable the team needed to focus on the most critical dialog states in the voice recognition process and determine the vital few factors that influence the behavior of the process.

**Key Dialog States**

The team analyzed the call flows for the voice recognition system, the number of non-recognition events in each dialog state, previous studies done on English voice recognition, and identified the following three dialog states as being the most critical to customer satisfaction:

- Ask_MainMenu
- Ask_WhichCityCountyOrHighway
- Ask_WhichSegment

The MainMenu state is the key state for the entire application and has the highest number of non-recognition events. The WhichCityCountyOrHighway and WhichSegment states are the most critical states for accessing incident information which is what most travelers want.

**Key Process Input Variables**

The team relied on LogicTree voice recognition expertise and the process maps to identify the following "vital few" variables affecting Spanish voice recognition:

- Confidence level setting[4];
- Grammar [missing, incorrect, or not sufficient: (e.g., pronunciation)];
- Software issue;
- Noise;
- User error (utterance that is not relevant); and,
- User unclear (lack of clarity in utterance).

**Data Collection**

The team decided to measure all Spanish utterances in the three chosen dialog states over a two month period (October and November, 2009). This resulted in a total of 6,244 Spanish utterances being analyzed (3,698 utterances in the MainMenu dialog state, 1,872 utterances in the WhichCityCountyOr Highway state, and 674 utterances in the WhichSegment dialog state). These sample sizes were determined to be reasonable[5] while still being manageable.

[4] If the confidence score of an utterance is above a software set upper threshold it is accepted, when it is below a lower threshold it is rejected, and clarification is often initiated when the confidence score lies between these two thresholds.

[5] Looking at the smallest sample size, 674, we are still able to say that results from this sample size will have at worst a 3.74 confidence interval at a 95 percent confidence level.

With each utterance the following data was provided with the database query:

- Audio file for utterance;
- Caller ID (when available);
- Date and Time of utterance;
- Raw result: what speech recognition system thought was being said;
- Rejected/Accepted status of utterance;
- Acoustic confidence score[6]; and,
- Average Speech and Background Decibel (dB).

**Categorization of Utterances**

To help with the analysis of the utterances a Spanish transcriber was tasked with listening to each utterance (i.e., audio file), transcribing it, comparing it to the raw result and then categorizing it.

For an utterance that the system had classified as Rejected/"No Match" the Spanish transcriber utilized the following classifications[7]:

| Type | Description |
|---|---|
| Unknown | The grammar seemed to be fine, the audio was "clear enough" and pronunciation was good, but it was still not recognized even though the item WAS in the grammar using parse tool. |
| Unsupported | The caller gave a transcription the current grammar does not support. |
| Pronunciation | The caller spoke this entry with a different pronunciation than what the system expected. The item WAS in the grammar using parse-tool. |
| Missing Entry | The grammar should have included this entry but it did not. Caller said "operator" when it should have been expected. The item WAS NOT in the grammar using parse tool. |
| User Unclear | The caller's speech was unclear, even to the human ear. The item WAS NOT in the grammar using parse tool |
| Noise | A lot of background noise or pure noise must have caused this non-recognition. The item WAS NOT in the grammar using parse tool. |
| User Error | The caller was saying something irrelevant to what the grammar was designed to recognize. |
| Hang Up | The caller hung up while the system was still in the process of recognition. |

[6] The software performs a statistical analysis on each utterance and provides an overall acoustic confidence score for each utterance.

[7] Classifications for this project were developed by LogicTree.

For an utterance that the system had classified as Accepted/"Recognition" the Spanish transcriber utilized the following classifications:

| Type | Description |
|---|---|
| Correct | The audio and transcription correctly matches the Raw Text. The item WAS in the grammar using parse tool. |
| False Accept IG | The audio and transcription do not correctly match the Raw Text. However the item WAS in the grammar using parse tool. |
| False Accept OOG | The audio and transcription do not correctly match the Raw Text. However the item WAS NOT in the grammar using parse tool. |
| False Accept Noise | The audio was just clearly noise or side speech and does not match the Raw Text. The item WAS NOT in the grammar using parse tool. |

**Current Status of Dialog States**

The transcription results for the three critical dialog states: MainMenu, WhichCityCountyOrHighway and WhichSegment are found in Appendix D. Due to the two confidence thresholds, there are a number of recognition rates that could be calculated for each dialog state.[8] To keep things simple for this project, calculations were performed only for a non-adjusted recognition rate[9]and an overall false acceptance rate[10] for the three dialog states. Note that the non-adjusted recognition rate contains both correctly accepted and falsely accepted utterances.

*MainMenu dialog state:*

- Non-adjusted Recognition rate: 47.2 percent.
- False Acceptance rate: 18.1 percent.
- The main root causes for non-recognitions are: *Noise* (33.8 percent), *Unknown* (18 percent), *User Error* (13.8 percent), and *Unclear* (12.1 percent).

*WhichCityCountyOrHighway dialog state:*

- Non-adjusted Recognition rate: 56.1 percent.
- False Acceptance rate: 16.4 percent.
- The main root causes for non-recognitions are: *Unknown* (22.3 percent), *Not Supported* (23.0 percent), *Noise* (17.2 percent), and *User Error* (13.8 percent).

WhichSegment dialog state

- Non-adjusted Recognition rate: 56.2 percent.
- False Acceptance rate: 15.6 percent.

[8] Correct Accepted rate, False Accepted rate, Correct rejected rate, False rejected rate, Correct confirmed rate, False confirmed rate.
[9] Non-adjusted recognition rate = accepted utterances / total utterances, note that no allowances have been made for false accepts nor for root causes outside the control of the voice recognition system, such as noise.
[10] Overall false acceptance rate = total falsely accepted utterances / total utterances

- The main root causes for non-recognition are: *Not Supported* (25.8 percent), *Unclear* (19 percent), *User Error* (18.0 percent), *Unknown* (16.6 percent), and *Noise* (16.6 percent).

## Measurement System Analysis

*Categorization:*

A random spot check was conducted to assess the accuracy of the transcriptions and categorizations. Fifty transcriptions/categorizations were reanalyzed and one issue was uncovered (98 percent accurate). This means that one can say with 95 percent confidence that the accuracy of transcriptions/categorizations will be between 94 percent and 100 percent.

*Confidence Score:*

The confidence score is supposed to reflect how confident the speech recognizer is in the recognition of an utterance. Used in conjunction with the confidence level thresholds, this is how utterances are rejected and accepted. Different speech recognizers use difference methods to calculate the confidence score but in essence the confidence score is based on the probabilities of the various hypothesized recognitions. The major issue with confidence scores is that words that are out of vocabulary (i.e., words that are not part of the grammar set) will corrupt the theoretical model utilized by the speech recognizer. In addition many competing hypotheses (i.e., many words in a grammar set) make the estimation of confidence scores much harder.

*Speech and Background dB:*

We do not know the accuracy and precision of the speech and background dB data points. Analyzing the accuracy of these dB data points is beyond the scope of this project.

*Audio files:*

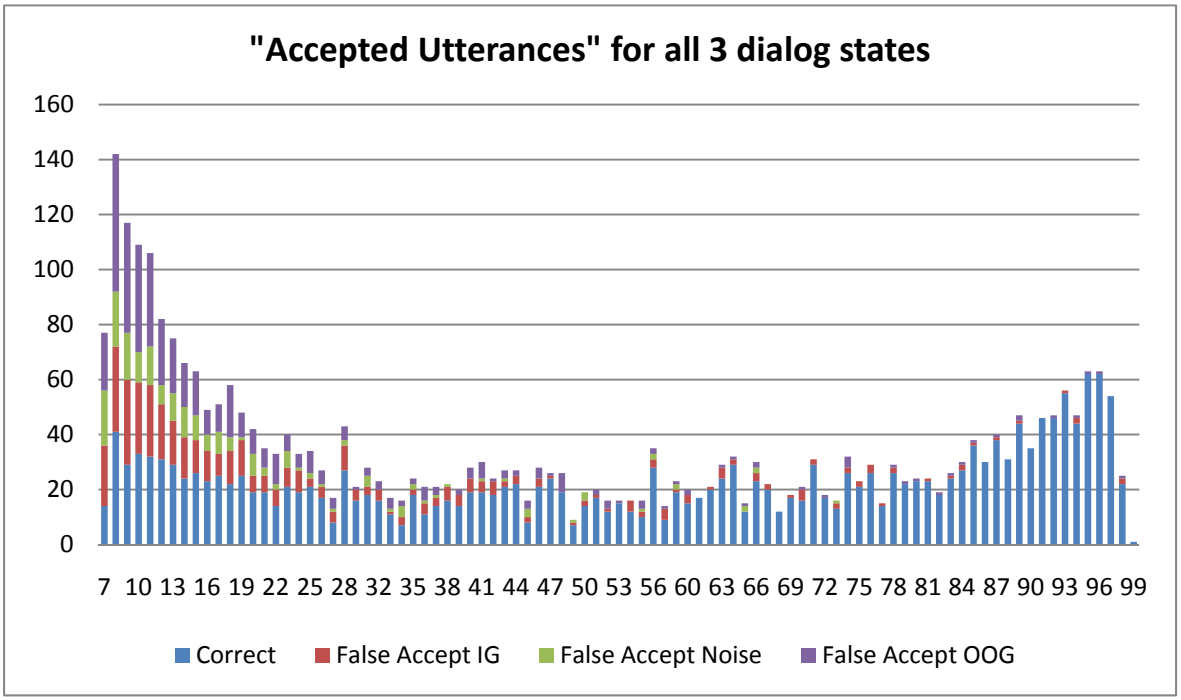
During the transcription/categorization phase a few audio files were encountered that had no audio at all, these were re-examined and some of them actually had small audio files with a small bit of noise, and some audio files were indeed empty. After some research by LogicTree most of these empty audio files were determined to be invalid dual tone multi-frequency (DTMF) entries and categorized in the unsupported category. The few remaining unaccounted for empty audio files (approximately 10) were deleted from the data set before analysis began.

# 4. Analysis Phase (Findings)

## Confidence Level Threshold

The first observation one makes when looking at the data is that the number of false accepts is very high in all three dialog states. The histogram below shows the utterances for the three dialog states that the voice recognition process had "accepted". [11] Note that while the system correctly "accepted" a large number of utterances, shown in blue below, there are quite a few utterances that were falsely accepted, shown in red, green, or purple, below. [12]

There is no historical data on what an acceptable false acceptance rate for this process is, and there is always a trade off between correct acceptances and correct rejections. However, it is quite clear from the histogram that the current rejection confidence level is set too low. Just looking at the utterances at the far left of the histogram, the Spanish voice recognition process is falsely accepting many more utterances than it is correctly accepting.




[11] All three dialog states were analyzed together in this histogram because the confidence threshold is the same across all dialog states.
[12] Currently the rejection confidence level is set at 7.

**MainMenu Dialog State:**

As noted previously the main root causes for non-recognitions in the MainMenu dialog state are *Noise* (33.8 percent), *Unknown* (18 percent), *User Error* (13.8 percent), and *Unclear* (12.1 percent). The main area of improvement leverage here is the "Unknown" category. Noise and Unclear are not under the control of the system and while User Error can be improved through changes in call flows, that is not within the scope of this project. "Unknown" is normally a grammar/pronunciation issue or possibly a software issue of some sort.

The most frequent grammars in the MainMenu dialog state were carefully analyzed and the following was found:

1. "Trafico" (English: "traffic"): 854 utterances, 23% of total utterances in MainMenu[13]
   - "trafico" was misrecognized for a number of other words. Below is a partial list of the words (including frequency) that the system thought were being said, when the user was actually saying "trafico":
     - la cinco (14)
     - transito (11)
     - Brannonville (11)
     - dos cinco (10)
     - Crawfordville (7)
     - Pasco (5)
     - cinco (5)
     - several other words also were misrecognized but with less frequency
   - "trafico" had a moderate percentage of unknown (16.2 percent), leading one to a belief that alternate pronunciations might be needed in the system.
2. "menu principal" (English: "main menu"): 208 utterances, 5.6 percent of total utterances in MainMenu.
   - "menu principal" is somewhat of a User Error. The user is already in the Main Menu and does not need to state "Main Menu". This is an area where further education and possibly call flow changes can be made.
   - "menu principal" had a very high percentage of unknown (33.6 percent), leading to a belief that alternate pronunciations need to be added to the system.
3. "guia informativa" (English: "tutorial"): 114 utterances, 3.1 percent of total utterances in MainMenu.
   - "guia informativa" had a very high percentage of unknown (39.5 percent), leading to a belief that alternate pronunciations need to be added to the system.

[13] Total utterances in MainMenu are based on all utterances, i.e. both accepted and rejected utterances

**WhichCityCountyOrHighway Dialog State:**

The main root causes for non-recognitions in the WhichCityCountyOrHighway dialog state are *Unknown* (22.3 percent), *Not Supported* (23.0 percent), *Noise* (17.2 percent), and *User Error* (13.8 percent). The main areas of improvement leverage here are "Unknown" and "Not Supported". "Not Supported" are words/commands that the system currently does not support.

Many issues that occurred in this dialog state were due to users asking for multiple things at once (e.g., asking for a roadway and a direction at the same time).

The most frequent grammars[14] were carefully analyzed in the WhichCityCountyOrHighway dialog state and the following was found:

1. "miami" (English: "miami"): 115 utterances, 6.1 percent of total utterances in WhichCityCountyOrHighway
   - Careful analysis showed that most issues with this word were due to people asking for multiple things at once.

2. "miami dade" (English: "miami dade"): 75 utterances, 4.0 percent of total utterances in WhichCityCountyOrHighway
   - Careful analysis showed that most issues with this word were due to people asking for multiple things at once.

3. "i noventa y cinco" (English: "I-95"): 51 utterances, 2.7 percent of total utterances in WhichCityCountyOrHighway.
   - "i noventa y cinco" had a very high percentage of unknown (41.2 percent) leading to a belief that alternate pronunciations need to be added to the system.

**WhichSegment Dialog State:**

The main root causes for non-recognitions in the WhichSegment dialog state are *Not Supported* (25.8 percent), *Unclear* (19 percent), *User Error* (18.0 percent), *Unknown* (16.6 percent), and *Noise* (16.6 percent). The main areas of improvement leverage here are "Not Supported" and "Unknown".

The biggest issue in this state was users asking for streets or exits that were not supported.

[14] Turnpike and Main Menu were tied for third place with I-95. There were no notable issues with Turnpike, and Main Menu was already analyzed.

The most frequent grammars were carefully analyzed in the WhichSegment dialog state and the following was found:

1. "todos los informes" (English: "all information"): 145 utterances, 21.5 percent of total utterances in WhichSegment.
   - Analysis showed a small percentage of issues due to unknown (11 percent) and false accepts (9 percent), which might indicate a need to check pronunciation.
   - Analysis showed that many variations of this word were used, "todo los informes" (11 unknown), "todos informes", etc. and some of these variations were not understood, which would indicate the need to add to the grammar set.

2. "todos los reportes" (English: "all reports"): 30 utterances, 4.5 percent of total utterances in WhichSegment.
   - Analysis showed that variations of this word were used; "todo reportes" and "todo los reportes" that were not understood by the system, which would indicate the need to add to the grammar set.

**Average Speech and Background dB**

Scatterplots were created of both the average speech and background dB data with confidence score data. No useful correlations were found between the dB data and any of the potential root causes for rejection.

# 5. Improve Phase (Recommendations)

## Confidence Level

Based on the analysis performed on the Spanish utterances it is recommended that the rejection confidence threshold be moved from 7 to 18. This number should be verified as to its effect on English recognition before it is implemented. Note that raising the rejection confidence level will result in the system rejecting correctly identified utterances along with the falsely accepted utterances.[15] If the non-adjusted recognition rate[16] and a false acceptance rate[17] are recalculated for the three dialog states based on moving the confidence level to 18, the following is shown:

MainMenu Dialog State
Non-adjusted Recognition rate: 32.0 percent
False Acceptance rate: 7.1 percent

WhichCityCountyOrHighway Dialog State
Non-adjusted Recognition rate: 41.4 percent
False Acceptance rate: 7.4 percent

WhichSegment Dialog State
Non-adjusted Recognition rate: 41.5 percent
False Acceptance rate: 7.4 percent

With the new recommended rejection threshold the false acceptance rate was reduced by over 50 percent.

From a literature search on threshold values it was uncovered that change to the system, such as addition of grammars, can affect the validity of the voice recognition confidence thresholds. It is recommended that threshold confidence levels be verified after any major system/grammar change.

Going forward it is recommended that acceptable recognition rates for each dialog state be established. Per the LogicTree voice recognition expert; there are too many factors at play such as grammar complexity and caller environment to come up with an all encompassing standard rate.

From the data there does not seem to be a clear need to move the upper confidence threshold so no recommendations are made to modify it at this time.

[15] At this new confidence level for the MainMenu dialog state 72 percent of the new rejects would be false accepts. At this new confidence level for the WhichCityCountyOrHighway dialog state 61 percent of the new rejects would be false accepts. At this new confidence level for the WhichSegment dialog state 55 percent of the new rejects would be false accepts.
[16] Non-adjusted recognition rate = accepted utterances / total utterances. Note that no allowances have been made for false accepts or for root causes outside the control of the voice recognition system, such as noise.
[17] False Acceptance rate = total false accepted utterances / total utterances.

**Grammars**

It is recommended that the audio files for the grammars identified in the analysis phase be examined in more detail by LogicTree to assist in making the appropriate changes to the system.

It is also recommended that LogicTree take a close look at other, less frequent grammars that were identified as unknown or pronunciation errors in each dialog state to see if further improvements can be found.

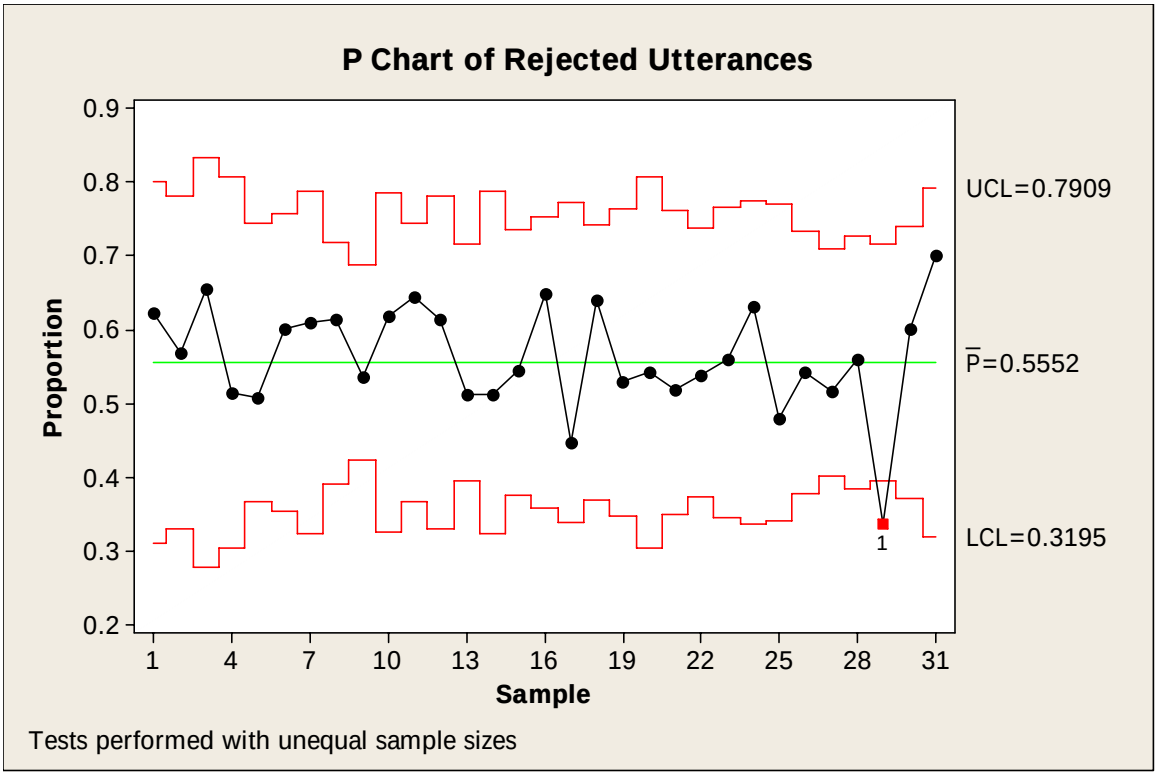
## 6. Control Phase (Monitoring)

Once the recommended changes have been made to the process, ongoing monitoring is required to assist in the day-to-day process management and ensure that improvements are sustained. The first level of monitoring that is recommended consists of customer feedback trend charts. Customer feedback complaints are currently being categorized and their trends should be analyzed on a weekly/monthly basis.

The second level of monitoring recommended is a binomial analysis of the defective items (i.e., Rejects/No Match); in other words a "p chart". This p chart would measure, on a daily basis, the proportion of total "No Match" to the total number of utterances. This can be done for the critical dialog states or, at a minimum, for the entire process.

To assess the viability of utilizing a "p chart", one was created for the month of October using the data collected for this project. The "p chart" below shows the daily proportion of defectives (i.e., number of rejected utterances/ total utterances) in the MainMenu dialog state.

- The center line (green) is the average proportion defective.
- The controls limits (red) are located 3 sigma above and below the center line and provide a visual means for when the process is out of control.




One can see from this chart that one sample, on October 29, 2009, is shown to be out of control, the proportion of defectives was 3 sigma below the average. Upon close

analysis of the utterances on October 29, it was discovered that testing of the Spanish voice recognition process occurred on this day and this caused the higher than normal number of accepted utterances. The above “p chart” highlighted a special cause that disrupted the process and proved that utilizing a “p chart” is a viable way to monitor the process.

# 7. Conclusion

A cross-functional problem solving team with members from LogicTree and PBS&J was formed to improve Spanish recognition. This team used the Six Sigma DMAIC problem solving process to identify opportunities for improvement.

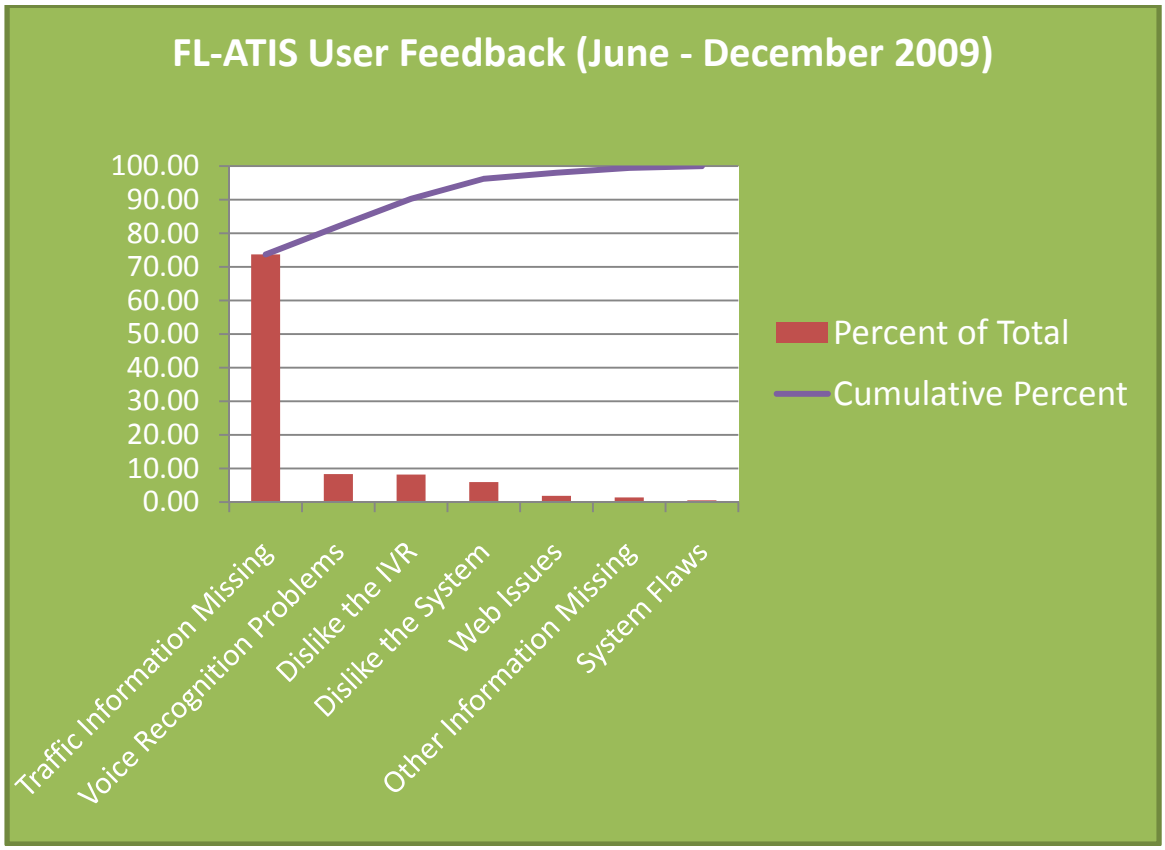
A major finding of this project was that the voice recognition low confidence threshold level needs to be reset. This rejection confidence level is set too low, causing a high false acceptance rate. The project recommended that the confidence level be moved from 7 to 18 which will decrease the false acceptance rate in the three dialog states analyzed by over 50 percent. The project also recommended grammar/pronunciation changes to the system. The result of these improvements will be improved customer satisfaction, fewer feedback complaints, and the elimination of the need to continue the customer technical support call center.

Next steps will consist of a careful review of all recommendations by the entire LogicTree team, including a closer look at the grammar audio files that have been identified as needing review. Further research areas for the team would include call flow analysis and studying other speech recognition parameters, such as Speedvsaccuracy[18], that might help the system performance.

[18] This property specifies the desired balance between speed and accuracy, a value of 0 means quick recognition, a value of 1 means best accuracy, default is .5.

# Appendix A

# Pareto Chart of FL-ATIS User Feedback Issues


FL-ATIS User Feedback (June - December 2009)
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
Traffic Information Missing
Voice Recognition Problems
Dislike the IVR
Dislike the System
Web Issues
Other Information Missing
System Flaws
Percent of Total
Cumulative Percent

# Appendix B

# High Level Process Map for Voice Recognition of Suppliers, Inputs, Process, Outputs and Customers (SIPOC)

*Suppliers:*

- FDOT regional transportation management centers (RTMC) (*Input*: 511 traveler information)
- LogicTree (*Input*: speech recognition software, settings, grammars)
- 511 caller ( *Input*: utterance)

*Process:*

1. Analog waves of a caller's utterances are converted to digital data by sampling the sound. This digital data is divided into small segments and a software program matches these segments to known phonemes[19].

2. The software program examines these phonemes in the context of other phonemes around them and runs a complex statistical analysis comparing them to a large library of known words, phrases, and sentences (grammar). In other words, a caller's utterances are compared to preprogrammed words (grammar), and a software program determines what the user was probably saying with a certain level of confidence (Confidence score between 0-100).

3. If confidence score for a caller's utterance falls below the low confidence threshold (currently set to 7) it is rejected (No Match). If it falls between the rejection threshold and the confirmation threshold (currently set between 7 and 75) it may or may not be accepted depending on the state. If it is not accepted, the user is asked to confirm what the user said. Any confidence score above the confirmation confidence level (currently 75) is automatically accepted.

4. 511 traveler information is provided based on confirmed/accepted utterances per the call flow design.

*Output:* 511 caller receives requested 511 traveler information

*Customer:* A 511 caller wanting to receive traveler information on a covered Florida facility.

[19] A phoneme is the smallest element of a language; there are roughly 40 phonemes in the English language.

# Appendix C

# Critical to Quality Tree

- Critical to Satisfaction
  - Critical to Presentation
    - Types of available 511 information (e.g. Traffic Incidents and Travel Times)
    - Available Languages (e.g. English, Spanish, etc.)
    - Roadway coverage
    - Timeliness of 511 information (entry and removal)
    - Correctness of 511 information
  - Critical to Performance
    - Accuracy of recognition
    - Speed of recognition
    - Ease of access to Information (i.e. number of menus you need to wade thru to get to desired information)

# Appendix D

# Transcription Results for the Critical Dialog States

The overall transcription results for the three critical dialog states: MainMenu, WhichCityCountyOrHighway and WhichSegment are found below.

**MainMenu Dialog State** - Transcription results for utterances the system had recognized/accepted:

| Type | Count | Percent of Total |
|---|---|---|
| Correct | 1077 | 61.7% |
| False Accept IG | 218 | 12.5% |
| False Accept Noise | 183 | 10.5% |
| False Accept OOG | 268 | 15.3% |
| Total | 1746 | |

**MainMenu Dialog State** - Transcription results for utterances the voice recognition process had rejected/"No Match":

| Transcription Result | Total | Percent of Total |
|---|---|---|
| HangUp | 153 | 7.8% |
| Noise | 661 | 33.8% |
| Not Supported | 214 | 11.0% |
| Pronun | 67 | 3.4% |
| Unclear | 237 | 12.1% |
| Unknown | 351 | 18.0% |
| User Error | 269 | 13.8% |
| Total | 1952 | |

**WhichCityCountyOrHighway Dialog State** - Transcription results for utterances the system had recognized/accepted.

| Type | Count | Percent of Total |
|---|---|---|
| Correct | 743 | 70.8% |
| False Accept IG | 154 | 14.7% |
| False Accept Noise | 10 | 1.0% |
| False Accept OOG | 143 | 13.6% |
| Total | 1050 | |

**WhichCityCountyOrHighway_Dialog_State** - Transcription results for utterances the voice recognition process had rejected/"No Match":

| Transcription Result | Total | Percent of Total |
|---|---|---|
| HangUp | 11 | 1.3% |
| Noise | 141 | 17.2% |
| Not Supported | 189 | 23.0% |
| Pronun | 63 | 7.7% |
| Unclear | 99 | 12.0% |
| Unknown | 183 | 22.3% |
| User Error | 136 | 16.5% |
| Total | 822 | |

**WhichSegment_Dialog_State** - Transcription results for utterances the system had recognized/accepted.

| Type | Count | Percent of Total |
|---|---|---|
| Correct | 274 | 72.3% |
| False Accept IG | 47 | 12.4% |
| False Accept Noise | 5 | 1.3% |
| False Accept OOG | 53 | 14.0% |
| Total | 379 | |

**WhichSegment_Dialog_State** - Transcription results for utterances the voice recognition process had rejected/"No Match".

| Transcription Result | Total | Percent of Total |
|---|---|---|
| HangUp | 5 | 1.7% |
| Noise | 49 | 16.6% |
| Not Supported | 76 | 25.8% |
| Pronun | 7 | 2.4% |
| Unclear | 56 | 19.0% |
| Unknown | 49 | 16.6% |
| User Error | 53 | 18.0% |
| Total | 295 | |