SPEECH RECOGNITION

Defining System Requirements

Abstract

This document compares traditional measures of speech recognition accuracy with additional measures that will better indicate the usability of a speech system. The document also provides guidelines as to how the reader may best define system requirements.

Gary Pearson

Gary.pearson@adacel.com

1 Background 3
2 Purpose of the Document 3
3 What is the Biggest Issue with SR Usability? 3
4 In-Grammar and Out-of-Grammar Metrics 4
4.1 In-Grammar Metrics 4
4.2 Out-of-Grammar Performance 4
5 In Grammar Metrics for Speech Recognition 4
5.1 Traditional Measures of Speech Recognition Accuracy 5
5.1.1 What is a Speech Recognition System 5
5.1.2 Word Accuracy 6
5.1.3 Sentence accuracy 6
5.1.4 Command Accuracy 7
5.1.5 When to Use Word and Sentence Accuracy 8
5.1.6 Practical Command Accuracy 8
6 Out-of-Grammar Performance 9
6.1 The Impact of OOG Utterances 9
6.1.1 Good Match 9
6.1.2 Failed Utterance 9
6.1.3 False Positive 10
6.2 How to Handle OOG 10
6.3 Defining Supported Phrases 11
7 Using Statistical Language Models 12
8 Example Requirements 12
8.1 General Requirements 12
8.2 Supported Phraseology 13
8.3 System Synthetic Voice Response 13
8.4 Accuracy and Testing (In-Grammar) 14
8.5 Recognition Results Processing (OOG) 14
8.6 Responses to Command Processing 14
8.7 Measuring Results Processing Performance 15
8.8 Speech Recognition Tools 15
8.9 Data Logging 16

Defining Requirements for Speech Recognition Systems

1 Background

Adacel has been successfully deploying speech recognition (SR) applications for more than 20 years. Adacel's expertise in the use of speech recognition technology, ranges from large and highly complex air traffic simulation for controller and pilot training, to advanced command and control systems for aircraft cockpits. With thousands of speech positions installed, Adacel has a wealth of knowledge gained from the analysis of empirical data. This analysis has led to a clear understanding of the difference between a statistically accurate speech system and one that is more focused on practical usability.

We have analyzed thousands of hours of speech recognition results, from hundreds of users from an install base that is significantly larger than all of our competitors combined.

The fundamental takeaway from this experience is that word recognition accuracy is NOT a measure of the practical usability of a speech recognition enabled application. While standard and adapted metrics are important, their primary use is as a measure of baseline accuracy of a specific ASR (Automatic Speech Recognition) engine and performance of the ASR when the speaker uses only phrases supported by the application.

2 Purpose of the Document

This document will discuss the application of standard accuracy measures and suggest additional measures that will provide a clearer indication of the usability of speech recognition in simulation applications. The intent of the document is to provide the reader with a number of suggested measures that could be used in defining formal requirements.

3 What is the Biggest Issue with SR Usability?

Adacel has a long and successful record of delivery of speech recognition enabled systems that have achieved certified word accuracy rates as high as 99%. We have even demonstrated 95% or more accuracy with non-native English speakers with extremely difficult accents. Even with this impressive track record we still receive requests for system improvements. Given that our word accuracy is higher than real world pilot and controller communications, why do we strive to improve this industry leading capability?

The biggest issue with speech recognition as an input, is not the accuracy against supported phrases, it is in fact the performance of the system when provided with unsupported phraseology. A speech recognition system will only match against sentences it is program to recognize. What happens when you use phrases that are unsupported (Out-of-Grammar)? Not only should the system handle Out-of-Grammar (OOG) phrases, it must be possible for the purchaser to define and test how this should be measured?

4 In-Grammar and Out-of-Grammar Metrics

It is Adacel's experience that the requirements specified in request for proposal documents tend to be designed to measure in-grammar accuracy. There has to date been little or no focus on a more important issue and that is the measure of Out-of-Grammar (OOG) performance.

4.1 In-Grammar Metrics

In-grammar is defined as an utterance that matches exactly, a phrase defined in the application grammar. The typical metrics for In-grammar utterances are word and sentence accuracy.

In the early days of speech recognition and air traffic control training, word and sentence accuracy were acceptable measures, as training systems were primarily deployed in school houses for new control training. In this environment students have no prior knowledge of ATC phraseology and they are expected to learn and comply with simple and clearly defined grammar. If a student fails to comply with official grammar and the recognizer failed, it seemed acceptable to forgive the ASR and assign blame to the student.

Traditional metrics as defined in section 5.1 should continue to be used in formal requirements documents. They give an excellent indication of baseline speech recognition accuracy of an ASR. However, traditional in-grammar metrics need to be considered alongside OOG performance metrics.

4.2 Out-of-Grammar Performance

OOG is defined as a spoken utterance where one or more of the words in the utterance (or the order in which the words are spoken), do not exactly match one of the systems supported phrases. In typical testing, any OOG utterance is eliminated from the results when calculating word and sentence errors. In real world use, OOG utterances are very common, especially in systems that are now being more widely deployed at operational units and used by experienced and trainee controllers using phraseology not specified in FAA7110.65. OOG performance is discussed in section 6

5 In Grammar Metrics for Speech Recognition

There are numerous challenges to be overcome when developing speech recognition for training and simulation. Two of the most prominent questions to be answered when specifying functional requirements and performance accuracy criteria are "How do you state the requirement for measuring SRS accuracy?" and "How do the accuracy and speech requirements translate into real world use?" By clearly defining and understanding these issues, evaluation of proposals from multiple vendors can be based upon like-for-like submissions.

Typical request for proposals (RFPs) often include very sparse requirements. For example:

> The system shall support phraseology as defined in FAA 7110.65 and locally adopted terminology
>
> The system shall have a word accuracy of at least 95% and a sentence accuracy of at least 90%.

As written, these requirements when reported compliant, provide no indication of the usability of the product. They are also written in a manner that allows pretty much any vendor to report compliant.

This lack of detail almost always results in a SRS that performs poorly within the training and mission rehearsal environment; creating user frustration and wariness of speech recognition applications. Furthermore, it will lead to a false understanding of the true cost of ownership of a speech enabled simulator.

5.1 Traditional Measures of Speech Recognition Accuracy

This section describes three traditional methods for defining system accuracy and proposes and additional metric (practical command accuracy) that is seldom used but would provide more meaningful data if adopted. The accuracy of an ASR, is often reported as “Word Accuracy” (WA) and/or “Sentence Accuracy” (SA). Before describing each of these measures and also the application of “Command Accuracy” and “Practical Command Accuracy”, we should understand what these numbers really measure.

5.1.1 What is a Speech Recognition System

A speech recognition system is comprised of a number of key components

1. The ASR – This software executes the actual task of recognition. The input is audio and the output is one or more (Called N-Best) recognition hypothesis and an interpretation of the actions to be executed by the attached simulation device.
2. The grammar and dictionary – The grammar contains a list of all phrases the ASR is expected to recognize. The dictionary is a phonetic representation of the words that make up the supported phrases.
3. The acoustic model – This component provides the ASR with a model of the speaking characteristics of the target user, e.g., US English. A system that uses a US English acoustic model will not perform as well for a French accented English speaker.
4. Post Recognition Processor – Any effective SRS will include a post processor that will determine how to action the returned hypothesis.
5. Synthetic response or Text-to-Speech (TTS) – The TTS generates responses from the simulated entities.

Standard metrics as described in this section simply provide an indication of the performance of the ASR. While it is possible under ideal conditions to achieve a 100%-word accuracy, it is not possible to achieve this level of performance in deployed conditions. There are simply too many variations that impact the ASR performance including accents, incorrect or unsupported grammar, noisy environments and many more.

5.1.2 Word Accuracy

Word accuracy is calculated by comparing what was said against what was recognized, determining if the recognized phrase required any of the spoken words to be deleted or substituted, and if any words had to be added to achieve the recognized result. This is better understood with the following ATC example:

> During the simulation the controller speaks the phrase, Eagle one, ***tower***, cleared to land runway two seven left.
>
> The SRS returns the phrase – "Eagle One, cleared to land runway two seven left"

You can clearly see that although what was spoken was not fully recognized, the actual result works perfectly in that the intent of the spoken command was achieved. However, using traditional word accuracy we would be led to believe otherwise.

Analysis to determine word accuracy would show that the word "tower" "spoken by the controller was not part of the SRS result. This represents one (1) word deletion. Ten words were spoken and the result contained only nine words.

The speech recognition Word Accuracy (Wacc) performance is then evaluated using the following formula:

$$W_{acc} = 100\left(1 - \left(\frac{W_s + W_I + W_D}{W}\right)\right)\%$$

Where: W is the total number of words to be recognized and WS, WI and WD are the number of words which were substituted, inserted, and deleted respectively.

The word accuracy result of the above example would be 90%. This is significantly lower than the typical +95% accuracy called out in system requirements documents.

So is word accuracy a valid measurement of SRS usability? Although Adacel can comfortably achieve high word accuracy requirements, experience has shown this approach should not be used as the sole measurement of the expected performance of a speech recognition system.

5.1.3 Sentence accuracy

Sentence accuracy is often seen as an additional accuracy requirement, but again it gives no indication of expected practical performance. Sentence accuracy is simply the percentage of sentences spoken that contained no word errors. In the previous description of word accuracy, the sentence accuracy would be 0% because the result contained one or more errors.

It would seem reasonable to assume that a system with a high word accuracy would have a high sentence accuracy, but that is almost always a false assumption. Take the following examples.

1. System 1 has 100 test phrases with 10 words in each phrase and a word accuracy of 90% (100 word errors).
2. Test 1 shows that every second sentence contains two word errors for a total of 100 word errors, leading to a calculated sentence accuracy of 50%.
3. System 2 has 100 test phrases with 10 words in each phrase and a word accuracy of 80% (200 word errors).
4. Test 2 shows that the first 20 sentences each contained 10 word errors for a total of 200 word errors and a calculated sentence accuracy of 80%.

Although system 1 had a higher word accuracy, the resulting sentence accuracy was significantly lower than system 2. This simply demonstrates that word and sentence accuracies cannot just be taken at face value when trying to determine the usability of a speech recognition system. A simulation system that produces a meaningful error every second phrase would likely be unusable.

5.1.4 Command Accuracy

Command accuracy has been shown to be far more important in determining the usability of a system than word or sentence accuracy, i.e., was a command recognized and processed correctly and was the appropriate response initiated. This time we will use a flight simulation example to demonstrate the concept:

> SRS grammar supports the phrase – “Tower, Eagle 1, request left downwind runway two one left”
>
> However, the pilot uses the phrase – “Tower, Eagle 1 ***looking*** for left downwind runway two one left.”

And

> The SRS recognizes - “Tower, Eagle 1, request left downwind runway two one left.”

In this example the SRS has achieved the desired semantic recognition result providing a command accuracy of 100%.

Command accuracy also cannot be used as a sole measure of usability, as demonstrated in the following example:

> Synthetic controller states, “Eagle one fly heading two seven zero.”
>
> Pilot replies, “Fly heading two seven zero, Eagle one.”
>
> SRS recognizes –“Fly heading two ***six*** zero, Eagle one.”

For this SRS error the system would produce a response from the synthetic controller reissuing the heading instruction; much as a real controller would if the pilot had incorrectly heard the controller instruction. From a requirements statement and accuracy measurement perspective, this creates difficulties as the practical outcome is arguable more realistic than a system that consistently produces 100%-word accuracy.

5.1.5 When to Use Word and Sentence Accuracy

As has been demonstrated here, word and sentence accuracy are not suitable measurements of system usability. However, they are excellent measures for initial baseline assessment of competing ASR systems. As long as each system is tested using the same audio test set in batch recognition mode, the results will provide useful information from which to gauge potential of the speech engine being used by the vendors. For best results, in addition to the use of a single test set, the test set itself should comply with the following guidelines

1. The audio should contain only phrases that are fully compliant with the supported test grammar. Out-of-Grammar (OOG) testing shall be covered elsewhere in the document.
2. The grammar should only contain words that are included in the system dictionary.
3. The test audio should be designed to cover a large percentage of the words contained in the grammar, i.e., a typical application may need to recognized 2000+ words, the test would provide minimal value if the test audio contained less than 50% of those words.
4. The test audio should contain a broad range of supported phrases
5. The test audio should be spoken by a large cross section of speakers and a representative split between male and female speakers.
6. The test audio should cover a significant number of words that could not be expected to be found in an English dictionary (beacon names, callsigns, and geographic features

5.1.6 Practical Command Accuracy

A third approach to speech recognition accuracy that is more appropriate to the ATC environment, Adacel terms “Practical Command Accuracy.” Practical command accuracy is simply a combination of digit accuracy (word accuracy for digits only) and command accuracy. Accuracy of a single command is either 100% or 0%; 100% if the recognized result does not change the semantic meaning or 0% if the result does change the semantic meaning of the spoken phrase. This definition of command accuracy does not allow any measurement for accuracy of components such as digits;

these require the application of digit word accuracy. Digit accuracy (Dacc) is calculated by the following formula:

$$D_{acc} = 100\left(\frac{D_t - D_i}{D_t}\right)\%$$

Where: Dt is the total number of digits spoken to be recognized and Di are the number of incorrectly recognized digits.

So to expand the example above so that the pilot repeats backs the corrected heading:

Synthetic Controller states, “Eagle one fly heading two seven zero.”

Pilot replies, “Fly heading two seven zero, Eagle one.”

SRS recognizes –“Fly heading two ***six*** zero, Eagle one.”

Synthetic Controller replies, “Eagle one, negative, fly heading two seven zero.”

Pilot replies again, “Fly heading two seven zero, Eagle one.”

SRS recognizes – “Eagle one fly heading two seven zero.”

In this example 100% command accuracy and a digit accuracy of 87.5% have been achieved (the total digit count for this example is eight (8) including the call sign digit). Although the digit word accuracy number seems low, when measured over a larger representative sample of the supported phraseology, digit word accuracy greater than 95% and a very high command accuracy can be achieved.

6 Out-of-Grammar Performance

In-grammar metrics are clearly valuable in the assessment of system accuracy but as previously explained they do little to measure practical usability. It is fine to understand how accurate the speech recognition is when we stick to supported phrases but how do we define and measure the performance in the all too common OOG situations.

6.1 The Impact of OOG Utterances

An OOG utterance will result in one of the following outcomes:

1. The utterance is closer to the intended phrase than any other supported phrase, that a good match will be made
2. The utterance is far enough out of grammar that the utterance will fail
3. The utterance is closer to a supported phrase than to the intended phrase, resulting in a false positive

The impact of each of these outcomes are as follows

6.1.1 Good Match

All is well, we got lucky. No further discussion required.

6.1.2 Failed Utterance

When an utterance fails, nothing happens. This is an acceptable given that in the real world if a pilot failed to understand an instruction, he will simply ask for the controller to repeat the instruction. Given that a failed utterance has no effect on the physical behavior of the simulated aircraft, the requirements should specify the acceptable response from the speech system, e.g., if a valid callsign is used with an OOG utterance, you would expect the callsign to ask for clarification. If an invalid callsign is used with an OOG utterance, then it is reasonable to expect no response.

6.1.3 False Positive

A false positive has undoubtedly the biggest impact on a simulation exercise.

False positive: A false positive error, commonly called a "false alarm" is a result that indicates a given condition has been fulfilled, when it actually has not been fulfilled.

False positives if processed by the simulator will lead to the synthetic pilot executing commands that are not reasonable within the context of the scenario often leading to an out of control traffic situation that was no fault of the student. Note that false positives can occur with in-grammar and OOG utterances. They are both equally damaging.

6.2 How to Handle OOG

The SRS will only recognize what it has been designed to recognize, so any test that involves free speech will lead to unquantifiable and invalid accuracy results. Although mathematical accuracy is a valuable and necessary starting point, recognition accuracy does not represent the performance of the system in use outside controlled test conditions.

Adacel's primary technical goal for speech recognition and synthetic response system is:

1. Regardless of the content of a transmission, the speech recognition system shall at all times produce a response that appears appropriate and at no time shall the simulator process phrases that would not be executed in a real world situation, e.g., A pilot on short final is not likely to accept an instruction to report left base, without questioning the instruction.

This challenge, although difficult, can be met. With each iteration of Adacel's LEXIX SRS and as more systems are deployed, more data is collected from the user community. This data is invaluable in moving closer to Adacel's stated goal. To determine how well this goal is being achieved, Adacel has developed a set of criteria to increase system robustness, usability, and user satisfaction. These criteria are:

1. The SRS shall attempt to eliminate false positive recognition. A false positive occurs when the speech engine accepts a match against a phrase other than the one that was spoken by the user.
2. The SRS shall not permit the processing of false positives by the simulator if such action has an unrealistic and negative impact on the traffic situation.

3. Some false positives may be acceptable if a similar situation also happens in real world communications. However, a false positive that happens consistently for the same situation is not acceptable. Acceptable false positives should be considered a rare exception and not the rule.
4. The synthetic entities shall respond in a manner that seems natural to the user should an invalid request be recognized, whether the invalid request is the user or SRS recognition mistake.
5. The system shall be capable of processing partial recognition results where an utterance contains more than one instruction, i.e., it is very important for usability that the system can handle the recognition of the callsign separate for the issued phrases. See section 8.6

6.3 Defining Supported Phrases

As discussed, a requirements document often contains very loose definitions for supported phrases (grammar), e.g., the system shall support all appropriate phrases defined in 7110.65 and locally adopted terminology. This definition raises the following issues

Locally supported terminology is an unconstrained requirement. No vendor can truthfully report compliant and furthermore no vendor can develop estimates for level of effort to support this requirement leading to apples and oranges pricing comparisons.

If a purchaser selects a vendor based upon these requirements and the vendor turns out to be non-compliant, the purchaser is in a no win situation. Even if it can be contractually argued that the vendor is responsible for all costs to become compliant, the system can be unusable until the necessary changes are made. The purchaser has no mechanism to determine how long it will take the vendor to become compliant.

The size of the phraseology support issue can be almost insurmountable. In 2001 prior to the USAF Tower Simulation System (TSS) program award, Adacel analyzed 7110.65 to determine how many supported phrases were necessary to meet the speech recognition requirements of the TSS program and deduced that estimating for 150 phrases would easily suffice. After 13 years of USAF program experience the supported phraseology is now counted in the high tens of thousands with numerous requests for new additions seeming to be endless.

When defining requirements, we recommend that the focus be placed on the functions that you wish the simulator to support, e.g., Taxi an aircraft, land an aircraft etc. These functions are almost always the cost drivers as they require the support or addition of sometimes very complex simulation functionality.

Adacel terms these actions as atomic commands. Our current system has support for over 600 atomic commands, with many of the atomic commands being driven by thousands of phrase variations, e.g., the traffic information command supports over 8000 spoken phrases.

When issuing requirements, a vendor can provide a like for like pricing comparison without knowing specifically what phrases may be needed. Example

1. The system shall support the following commands

    1. Taxi
    2. Taxi Via
    3. Taxi and Hold Short
    4. Follow
    5. Take-Off
    6. Closed Traffic

    etc.

2. The vendor shall support 1000 spoken phrases

Section 8.1shall expand on these ideas.

7 Using Statistical Language Models

This document has focused on the use of context free grammar speech (CFG) recognition. CFG has proven to be the most usable speech technology to date. It provides the benefit of high accuracy but at the cost of having to determine what phrases need to be supported. Statistical Language Models (SLM) operate differently. They do not require phrases to be defined in advanced but the do require the creation of a statistical model that the ASR uses to determine what was likely spoken. Initially SLM would seem more appropriate but unfortunately although the upfront work is minimized, the downside is a significant drop in accuracy. At the time of writing this document, SLM is not considered a viable solution. Having said that, the choice of technology should be left to the vendor. If the vendor can comply with the performance measures specified in the requirements, then the choice of CFG or SLM becomes a vendor implementation issue and not a program issue.

8 Example Speech System Requirements

This section of the document will offer some suggestions for the definition of speech recognition requirements for simulation and information on possible evaluation processes that should be used prior to contract award. The requirements assume that the speech recognition system will be part of the acquisition of a simulation system. This section is not intended to be a comprehensive list of requirements.

8.1 General Requirements

The system shall support a speech recognition interface at the following positions

1. Local

2. Ground
3. Flight Data
4. [position]
5. [position]

The speech system shall support continuous speech that does not require pauses between words or phrases or for the user to adapt the speed at which they talk.

The latency of the speech system shall be less than 300ms

The speech recognition system shall support Male and Female US English speakers

The speech system shall support the following simulation actions

1. Engine start
2. Taxi
3. Pushback
4. Hold short
5. [Action5]
6. [action x]

The vendor shall report compliance with the required simulation actions and provide details of additional actions supported by their simulator

8.2 Supported Phraseology

The system shall support a minimum of 50,000 [Enter Appropriate Value] spoken phrases for the defined simulation commands.

The vendor shall provide details of the number of supported phrases in their system.

It shall be possible to chain together 8 phrases in a single transmission

The system shall support the use of the break command

The system shall support the correction of phrases in mid utterance

8.3 System Synthetic Voice Response

The system shall include [x] synthetic voices to generate responses to issued phrases

Voices shall include [x] female and [y] male

The system shall include the following accents

1. US English
2. UK English
3. Italian English
4. [Accent 4]

5. [Accent 5]

The vendor shall provide the means for a local user to modify the standard responses with locally appropriate alternatives

8.4 Accuracy (In-Grammar)

It is recommended that the testing be carried out prior to the selection of a vendor

The system shall provide a minimum word accuracy of 95%

The system shall provide a minimum command accuracy of 95%

The system shall provide a minimum digit accuracy of 95%

The system shall provide a minimum callsign accuracy of 95%

8.5 Testing (In-Grammar)

Testing shall be performed in batch recognition mode against a purchaser provided standard test set of [insert number] utterances. Utterances will be accurately transcribed by the purchaser and provided for test purposes only. Utterances will be “In-Grammar” only

> Note that although this requires the purchaser to provide test utterances, this is the most effective method by which to make accuracy comparisons between competing systems. If a vendor has prior access to the test set, it can be used to “cheat” the test. Batch recognition is more appropriate than live measure in this case as it is consistent. It can also be done without a connected simulator at a location of the purchaser’s choice. The vendor should be provided with a transcript of the audio to be tested. These transcripts will be used to determine if the test system has support for the test phrases and to determine if additional words need be added to the dictionary.
>
> Although it is ultimately up to the purchaser to determine the number of test utterances, it should be noted that choosing a large number of utterances that cover a broad range of commands, spoken by multiple speakers, will provide the most meaningful results.

8.6 Recognition Results Processing (OOG)

The system shall not permit the processing of results that contain sequences of incompatible commands, e.g., American One Two Three cleared to land, report level five thousand

The system shall not permit processing of commands that are inappropriate with the flight state of the aircraft, e.g., cleared to land when aircraft is taxiing for departure

The system shall support partial command processing, i.e., if an utterance contains more than 1 command it will be capable of processing those commands that are correctly recognized even although one or more of the commands may have failed recognition.

8.7 Responses to Command Processing

In the case where a valid aircraft callsign is recognized but the entire utterance was rejected, the system shall generate a synthetic response from the callsign for the controller to repeat the instruction.

In the case where the callsign was not recognized, there shall be no response from the simulator

In the case where a callsign is recognized that is not active on the frequency but is similar to an active callsign, the active callsign will state “was that for me”. E.g., Air Canada One Two Three is recognized but Air Canada Two Two Three is on frequency, “Tower Air Canada Two Two Three was that for me”.

In the case where a partial recognition of the result occurred the aircraft will acknowledge what was understood and ask for clarification for the remainder. E.g., “American Four Five Zero, climb and maintain eight thousand, report over [Fix Name]. Roger climb and maintain eight thousand, say again remainder American Four Five Zero”.

8.8 Measuring Results Processing Performance

The system shall achieve less than 5% false positive processing (processing means passing the result to the simulated entity for action) at the atomic command level. False positives are only considered in error if they result in the synthetic entity performing a physical action (Turn, climb, land). A false positive is not considered an error if it results in the passing of information only, e.g., aircraft instructed to land, receives wind direction and strength.

The system shall achieve a less than 5% incompatible command processing at the sentence level. e.g., American Four Five taxi to parking, report level five thousand. The result is binary, fail or pass. In this example it is a fail.

The system shall achieve less than 5% inappropriate command processing. E.g., cleared to land when aircraft is taxiing for departure

8.9 Speech Recognition Tools

The vendor shall provide tools to permit the local user to modify or add to the supported phrases to better support local terminology.

The vendor shall provide tools to permit the addition of local spoken names without the need for software changes, e.g., aircraft callsigns, navigation beacons, geographic features

The vendor shall provide a means for the user to analyze recorded utterances. The tools should at a minimum provide feedback on the improper use of the press to talk switch (clipping), saturation of the audio feed, low audio volume and a means to determine if the utterance was In-Grammar or OOG

The provided tools shall not require the user to have specialized speech recognition knowledge

It shall be possible to teach an ATC specialist the fundamental basics of the tools in one working day (8 hours).

8.10 Services

The vendor shall offer a service to improve the performance of the system for users with non-standard accents. The vendor shall provide details of the methods, process, purchaser responsibilities and cost for such a service.

8.11 Data Logging

The system shall provide logging tools available to the end user.

The logging tools shall store data that includes the following information

1. Login Name of speaker
2. Name of scenario
3. Date and time of execution of scenario
4. Number of rejected utterances including time the rejection took place
5. Report on poor press to talk execution (clipping of audio)
6. Report on volume of audio settings (saturation of microphone)
7. Confidence scores of each utterance