NASA-TM–2016–219329

# Performance Evaluation of Speech Recognition Systems as a Next-Generation Pilot-Vehicle Interface Technology

*Jarvis (Trey) J. Arthur III, Kevin J. Shelton, Lawrence J. Prinzel III, and*
*Randall E. Bailey*
*Langley Research Center, Hampton, Virginia*

August 2016

## NASA STI Program. . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI Program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI Program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collection of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI Program, see the following:

- Access the NASA STI program home page at `http://www.sti.nasa.gov`

- E-mail your question to `help@sti.nasa.gov`

- Phone the NASA STI Information Desk at 757-864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA-TM–2016–219329

# Performance Evaluation of Speech Recognition Systems as a Next-Generation Pilot-Vehicle Interface Technology

*Jarvis (Trey) J. Arthur III, Kevin J. Shelton, Lawrence J. Prinzel III, and*
*Randall E. Bailey*
*Langley Research Center, Hampton, Virginia*

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

August 2016

## Acknowledgments

The authors would like to acknowledge the contributions of the technical reviewers Vincent Houston and Stella Harrison.

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

# Contents

# List of Tables

# List of Figures

# Nomenclature

| | |
|---|---|
| **API** | Application Programmers Interface |
| **ATC** | Air Traffic Control |
| **AvSP** | Aviation Safety Program |
| **COTS** | Commercial Off-The-Shelf |
| **EP** | Evaluation Pilot |
| **FAA** | Federal Aviation Administration |
| **FMS** | Flight Management System |
| **GVSITE** | Gulfstream-V Synthetic Vision Systems Integrated Technology Evaluation |
| **HUD** | Head-Up Display |
| **ICAO** | International Civil Aviation Organization |
| **IIFD** | Integrated Intelligent Flight Deck |
| **ms** | Millisecond |
| **ND** | Navigation Display |
| **NextGen** | Next Generation Air Transportation System |
| **PFD** | Primary Flight Display |
| **SDK** | Software Development Kit |
| **SSA** | Shared Situation Awareness |
| **SRS** | Speech Recognition System |
| **SV** | Synthetic Vision |
| **SVS** | Synthetic Vision System |
| **TMA** | Airport/Terminal Maneuvering Area |
| **TRACON** | Terminal Radar Approach Control |
| **WAV** | Waveform Audio File |

# Abstract

*During the flight trials known as Gulfstream-V Synthetic Vision Systems Integrated Technology Evaluation (GV-SITE), a Speech Recognition System (SRS) was used by the evaluation pilots. The SRS system was intended to be an intuitive interface for display control (rather than knobs, buttons, etc). This paper describes the performance of the current "state of the art" Speech Recognition System (SRS). The commercially available technology was evaluated as an application for possible inclusion in commercial aircraft flight decks as a crew-to-vehicle interface. Specifically, the technology is to be used as an interface from aircrew to the onboard displays, controls, and flight management tasks. A flight test of a SRS as well as a laboratory test was conducted.*

## 1 Introduction

The Integrated Intelligent Flight Deck (IIFD) project of NASA's Aviation Safety Program (AvSP), was a multi-disciplinary research effort to develop flight deck technologies that mitigate operator-, automation-, and environment- induced hazards. Toward this objective, the IIFD project researched crew-vehicle interface technologies to reduce the propensity for pilot error, minimize the risks associated with pilot error, and proactively overcome aircraft safety barriers that would otherwise constrain the full realization of the Next Generation Air Transportation System (NextGen) [1]. Part of this research effort involved speech recognition systems.

This paper describes two studies evaluating the performance of the current commercial-off-the-shelf (i.e., the commercial "state of the art") Speech Recognition System (SRS) in aviation-domain applications. The speech recognition and text-to-speech technologies are being evaluated as transformative crew-vehicle interface technologies for NextGen.

### 1.1 Background

NextGen doesn't just imply a US airspace transformation to accommodate significant increases in air traffic. NextGen also must consider the tremendous influx of world-wide air carriers and operations, both as they come to US airspace and as US air carriers go world-wide. As such, speech recognition and text-to-speech systems offer potentially "game-changing" capabilities for future aviation operations by: a) enabling NextGen data-link capabilities without a loss of crew situation awareness or modality changes from present-day operations; b) eliminating potential crew or procedures errors or miscommunications; and, c) creating new and unique interface capabilities for safer and more efficient generations of aircraft and aviation operations.

To understand how speech recognition and speech generation system technologies may benefit NextGen flight operations, a historical perspective and understanding of current operations is necessary. Pilot-Air Traffic Control (ATC) aural communications have historically been quite successful yet far from error-free. One content analysis showed that 40% of ATC-pilot communications contained at least one irregularity [2], but the error rate was less than 1% [3]. Fortunately, 60% to 80% of these errors were caught in the read-back process so an estimated rate of only 2.4 communications-related occurrences per million instructions/clearances resulted. However, the proportion of corrected read-back errors varied, where the highest workload sectors had the least corrections: en route controllers corrected 89% of the read-back errors, Terminal Radar Approach Control (TRACON) and local controllers caught only 60% and 63%, respectively, and only 50% of the read-back errors on the ground frequency were corrected. The read-back process provides a critical error-checking mechanism that must be maintained or replicated, or better yet, improved upon for NextGen.

Culture is often discussed in aviation in terms of a "safety culture" for a particularly organization, company, or air-carrier. But more so, regional or national cultural differences must be understood and countermeasures developed as the concept of essentialism - the view that culture is an essential part of every person - predicts that "flight decks must be made sensitive to national culture because people cannot depart from the imprint of their original national culture or be made complexly insensitive to known cultural differences" [4]. This view is shared by the Federal Aviation Administration (FAA) [5] even though no unambiguous historical evidence is available to support (or refute) if regional or national cultural differences have influenced the aviation accident rate [4, 6]. However, the explosion of international travel and airlines and increasing ab-initio training programs abroad suggest that an English-/Western Hemisphere-slanted flight deck will become problematic. This influx of non-Western pilots or operations into aviation operations will only grow in time [7], suggesting that technology which can achieve culture-neutral or culture-tailored interfaces would be advantageous for safety.

Language and culture have historically collided to disrupt or destroy pilot-ATC communications to the point that several accidents have been wholly or partial attributed to language and cultural differences [8] - most notable of which was collision of two B-747 aircraft at Tenerife which resulted in the deadliest civilian aircraft accident ever.

Recently, International Civil Aviation Organization (ICAO) has recognized the importance of language, in particular, and communication, in general [9]. ICAO now has mandated English as the international language of aviation (previously by de facto), but the local language can still be used for domestic operations. More importantly, all signatories must adopt and meet standards for language proficiency, focusing on both speaking and listening skills to promote understanding.

One initiative which could help in reducing the impact of language and culture in pilot-ATC communications is the emergence of data-link communications to replace radio telephony. It was predicted that by the year 2015, 60% of communications would be provided via voice in the most critical phase of flight — in the Airport/Terminal Maneuvering Area (TMA) — reducing to the extent that 85% of Air Traffic Services communications are provided via data-link in the TMA environments by the year 2030 [10]. As presently envisioned, data-link communications [11–17] generally provided positive benefits for pilot-ATC communications:

- The reduction or elimination of message blocking and congestion. Higher efficiency and capacity of communications system resulting in improvement of message delivery time (i.e., when compared to current limited radio-frequency bandwidth, stepped-on messages, and including controller-pilot read-back times).

- The persistence of the message. Unloaded memory burden from lengthy messages, and ability to review later.

- Improved information-processing efficiency and accuracy. Possibility of effective multitasking due to user-pacing communication tasks and elimination of continuous listening workload. Improved information transfer to other ATC and flight deck subsystems.

Current instantiations of the data-link interface change the modality of pilot-ATC communications from aural to visual communications as text read-outs on the Flight Management System (FMS) Control-Display Unit (see Figure 1). For language-/culture-challenged crews, this methodology transforms a "listening" communications task to reading comprehension which is inherently easier and more accurate.

Figure 1. Controller-Pilot Data-Link Capability Demonstration.

While the transition to data-link communications from radio communications would appear to be extremely beneficial, it is far from perfect. Research has shown that:

- Keyboard FMS implementations might be advantageous to non-English-native languages and cultures, but native users find the interaction to be cumbersome and reading is slower compared to aural/oral communication. Wickens, Miller, and Tham [18] found that there was a 3-second delay in pilot requests by controllers. Initiation of message and receipt of acknowledgement is nearly twice as long for a visual-manual data-link system (20 seconds) as for a radiotelephone system (10 seconds) [11, 17, 19]. The FAA [12, 13] reported that this may be, in fact, 15 seconds instead of 20 seconds but is significantly longer for non-routine transmissions and in the final control sector before landing. Lozito, McGann, and Corker [14] reported that pilots took 21.4 compared to 7.9 seconds via voice to acknowledge messages.

- Workload and increased “head-down time” are a major concern [11, 20]
- Users are deprived of urgency emotion when using text instead of voice. Data-link does not adequately convey urgency and other natural voice information [13, 21]
- Users may be deprived of “party-line” information [22]. Pritchett and Hansman [23] reported that 40% of transmissions on party line were considered critical to all pilots on approach.
- Crew coordination, diffusion of responsibility, and reduced cross-check involving automation can occur particularly when data-link is “message-gated” and passed directly into the FMS without the pilot reading the message and entering the data manually [13, 19, 21, 24]. These procedures may induce implicit compliance (i.e., acknowledging the data-link communication as ”WILCO” and executing without thorough checking and verification) and top-down processing [3].
- The read-back process is bypassed and the error-checking mechanism provided by this procedure eliminated [25].

## 1.2 Research Needs

Speech recognition and speech generation (i.e., text-to-speech) research, test, and evaluation for aviation-applications is certainly not unique or new. Numerous efforts have addressed the problem of speech recogntion in aircraft cockpits [26–28], improving pilot-vehicle interaction [29] and applying speech recognition for improved controller-pilot data link interface design [30–36].

These works have shown the potential for speech recognition and text-to-speech system technologies, but their application, to date, within the commercial and general aviation communities has been limited. The military aviation application domain is now seeing its introduction [37, 38].

For success within the commercial and general aviation communities, speech systems technologies must provide the capability to:

- Replicate and improve existing interface modalities
- Enhance current capabilities.
- Create new functionalities.

Research to develop these capabilities is needed.

### 1.2.1 Replicate Existing Interface Modalities

Data-link may improve one source of miscommunications - the inability to get the message from one party to the other - but it does not necessarily address the rest of the communications process - i.e., whether the message was understood and whether it accurately conveyed the speaker's intent [8]. Speech recognition and "text-to-speech" has the potential to create an "equivalent", yet improved radio telephony modality during data-link operations.

The capability for direct "replication" of existing modalities during data-link operations evolves from the concept of using speech recognition of the crew (pilot) to create a "digital" data-link message for transmission and using text-to-speech to decode and broadcast a received data-link message for the crew (pilot).

This direct replication seems simple and straight-forward but research is needed to identify and develop interface technologies which take advantage of data-link where it can best be used, and augment it to overcome its limitations knowing that "failed communications" jeopardize safety when the wrong information is used, situation awareness is lost, or an accurate shared awareness "model" of the present situation is not achieved by both ATC and the flight crew. This work must be done in coordination with methodologies and applications of data-link for trajectory planning, negotiation and execution, during en-route and surface operations.

Just as important, research is needed to proactively apply crew-vehicle interface technologies to identify when the wrong communicative information is being used, generate or enhance situation awareness to the flight crew in a data-link environment while reducing head-down time and workload, and promote the construct of a Shared Situation Awareness (SSA) "model" of the present situation between ATC and the flight crew. This research should create technologies and methods to reduce the propensity for human error, immunity to human error when it occurs, integrity and robustness in non-normal operations, adaptation to non-native English language and non-Western-culture users, capability in transitional and mixed-fleet operations, and realization in general aviation and business aircraft operations as well as air carriers.

### 1.2.2 Enhance current capabilities

One of the key attributes of the present air transportation system is the checks-and-balances which foster safety, such as the use of read-back to verify proper receipt and understanding of instructions. Speech interface technologies can enhance current capabilities to mature the concepts that communications and "understanding" error checking (i.e., surveillance) can be conducted. These concepts have been outlined under previous NASA efforts [39].

These interface capabilities, both of the flight crew and radio telephony, can evaluate the broadcast radio transmissions, read-back and surveillance of the crew awareness and actions. Speech recognition can also be used for stress and workload identification. The goal of this technology is to identify in real-time if the wrong information is being used, if the flight crew situation awareness is being lost or jeopardized, or if a shared awareness of the situation is flawed or absent.

These capabilities can also be used to provide a method for "automatic" route creation, particularly in high-workload environments, and to minimize head-down time in the cockpit. If pilot read-back or radio transmissions for flight and taxi routes were "intercepted" and decoded by a speech recognition system into route instructions for a FMS, the flight crew wouldn't be required to manually enter this information. Obviously, a data-link of this information could be done, but voice could be used to check that the data-link and verbal instructions, if used, were consistent. Again, another error checking process. During taxi operations in particular, highlighted route information [40] has been shown to significantly reduce the potential for runway incursions.

### 1.2.3 Create new functionalities

Finally, the potential for speech as a next-generation pilot-vehicle interface technology is nearly limitless.

Natural language is emerging as the primary future human-computer interface [41]. This analogy - for ease of operation and transition - must be viewed as an emerging capability in aviation, just as the cursor control device has emerged in the past 20 years following this same path. The technology for natural language interfaces to aircraft automation and systems, natural language advisory, assistance, warning, and alerting are needed.

Speech systems technology allow security protocols and concepts via biometrics. Voice authentication would be part of a layered security network to prohibit unauthorized operations. Adaptable automation and biometrics via speech analysis could be used as indicators of stress, workload, or situation awareness.

Even more ground-breaking would be the concept and requirements for a culture-neutral/language-neutral flight deck. Speech interface technologies provide the possibility that the cockpit could be uniquely tailored to the language and culture of the flight crew, yet outside the cockpit, all transmissions are converted into the universal language of digital data-link. With this technology, the potential for cultural-bias or language misunderstanding can be eliminated. The cockpit displays would no longer be hard-wired; but instead, the glass cockpit could transform to the language and culture of choice and natural language interfaces in the native tongue are the primary interface.

## 1.3 Present Study

The examples above highlight the tremendous potential and research challenges for speech interface systems in the aviation domain. Nonetheless, speech interface technology is largely driven by commercial markets since these markets are orders of magnitude larger (in terms of both users and sales). A key facet of aviation speech recognition technology would appear to be how to capitalize on the ready-availability of effective Commercial Off-The-Shelf (COTS) voice synthesis and recognition systems.

A key facet of IIFD research capitalizes on the ready-availability of effective voice synthesis and recognition systems. The development of speech recognition and text-to-speech systems are quite mature and well-addressed elsewhere (e.g., see http://www.speech.cs.cmu.edu/comp.speech/SpeechLinks.html). For IIFD, COTS recognition and synthesis methods are used with application-tailoring to the aviation application. The key research questions are not necessarily tied to the speech recognition and text-to-speech capability, but what can and should be done on the flight deck, given these technologies successfully emerge.

Tailoring of commercial systems for the aviation application brings several unique challenges, such as:

- Extensive use of "spoken" acronyms (e.g.,"HUD" not "H-U-D")
- Extensive use of "names" to associate geographical locations, such as waypoint identifier names (e.g., the waypoint REMTY is pronounced "Remtee"), and flight operations standard operating procedures (e.g., flying the"Tipp Toe Visual").
- Phonetic alphabet ("Alpha", "Bravo", etc.)
- Standardized Radio Communications Phraseology and protocol.
- Extensive list of "company" and "manufacturer" names which imply operational and capability constraints (e.g., "Follow Boeing traffic, at your 3 miles and 12 o'clock")
- Special Use, Special Emphasis Words ("Expedite", "Emergency", "Wilco", "Roger")
- Criticality of the speed and accuracy, i.e., speed and recognition rates for verbal communications (on the order of $99.99^{+}\%$ type accuracy error requirements) in Class B airspace.

In the following, the results of two investigations of system error rates and listener-speaker evaluation for such factors as recognition robustness (speaker independence) and accuracy for COTS speech recognition systems in the aviation-domain are presented. This work represents some initial investigations into these technologies for NextGen which provide path-finder for future IIFD efforts.

## 2 GVSITE SRS Evaluation

A flight test evaluation was jointly conducted (in July and August 2004) by NASA Langley Research Center and an industry partner team under NASA's Aviation Safety and Security, Synthetic Vision System project [42]. A Gulfstream G-V aircraft was flown over a 3-week period in the Reno/Tahoe International Airport (NV) local area and an additional 3- week period in the Wallops Flight Facility (VA) local area. This flight test, known as Gulfstream-V Synthetic Vision Systems Integrated Technology Evaluation (GVSITE), evaluated integrated Synthetic Vision System concepts, critical to the development, and subsequent fielding, of actual Synthetic Vision (SV) Systems. In this context, SV systems include computer-generated terrain presented on the Primary Flight Display (PFD); monochrome textured terrain presented on a Head-Up Display (HUD); plan view or perspective views of computer-generated terrain and obstacles on Navigation Display (ND); and data-link, sensors, and algorithms to provide and verify required information for display. In addition, symbology and algorithms designed to enhance pilot situational awareness during surface operations, and to prevent or alert to potential runway incursions, was also part of the integrated SV system. This paper focuses on the in-flight performance of a SRS that was used as the pilot-vehicle interface for the integrated SV system display concepts.

## 2.1 Flight Test Aircraft

The flight test was conduced using a Gulfstream G-V aircraft (Fig. 2) [42]. The left seat of the G-V was occupied by the Evaluation Pilot (EP) and the right seat was occupied by a Gulfstream Safety Pilot. The left seat included in the installation of two research displays for evaluation of the PFD and ND concepts, an overhead HUD projection unit for evaluation of head-up concepts, and a SRS system for the pilot-vehicle interface to the SV displays (Fig. 2).

Figure 2. G-V aircraft exterior and interior views.

## 2.2 Evaluation Pilots

Ten EPs, representing the airlines, a major transport airport manufacturer, the Federal Aviation Administration, and the Joint Aviation Authority, flew research flights totaling approximately 67 flight test hours. One hundred and forty-five flight test runs were conducted to evaluate the NASA Synthetic Vision System (SVS) concepts in the vicinity of Wallops Island, VA (8 pilots) and Reno/Tahoe International Airport (7 pilots). Five of the ten EPs flew at both test locations.

## 2.3 Speech Recognition System Design for GVSITE

A SRS system was installed in the Gulfstream-V as a pilot-vehicle interface. The SRS was used primarily to facilitate the pilot-vehicle interface to the SVS displays without having to modify hardware or basic ship's systems. It was also used as a way to test the use and utility of a SRS for future commercial and business aircraft flight deck developments. The application used a commercial speech recognition engine to interpret the EPs speech input, an interface application that passed and received information to/from a computer connected via Ethernet, and a synthesis (text-to-speech) module that generated aural messages or played pre-recorded Waveform Audio File (WAV) files (i.e., .wav extension files). The speech recognition product provided a commercial off-the-shelf, speaker-independent speech recognizer with easily-tailored grammar.

The bi-directional SRS allowed the EP to verbally command changes to the SVS displays and provided aural warnings and alerts to the crew when triggered by the SVS research systems. The nominal noise-attenuating David Clark headsets for the G-V aircraft were plugged into a Telex ProCom/2 intercom box which split the pilot's speech input to drive both the nominal G-V intercom input jacks and a SRS function was created using a Microsoft Windows-based application resident on a single computer. The computers audio-in port on the computer accepted the intercom box audio input. A "push-to-listen" function was installed. When the yoke-mounted radio transmit rocker switch was depressed by the EP, a serial input was closed on the SRS computer. The initial closure of the "push-to-listen" discrete triggered the SRS application to start "listening" and release of the "push-to-listen" triggered the SRS application to finish the speech recognition process. This "push-to-listen" implementation was very convenient and easy to use by the EPs since it was essentially analogous to existing radio communications; in this case, however, the EP was communicating with an on-board speech-respondent "assistant."

Rather than establishing a natural language environment, the speech recognition grammar was set-up as a hierarchy to improve recognition rates. The SRS used a 3-word top-level command grammar to issue commands to control the SVS displays (PFD, HUD, and ND) as shown in Figure 3. The first word in the hierarchy was the display device, the second word was the function or display element to be controlled, and the third word was the value or adjective modifier. For example, the command "NAV RANGE 5" would set the navigational display range to 5 nautical miles. Two "exceptions" were also programmed - "cancel" and "repeat." The repeat command (obviously) repeated the previously executed SRS command. This utterance was not used very often. The cancel command was programmed to "undo" the previously executed SRS command. (This command was very well received by the EPs, but it was a more difficult command to execute in terms of having to remember and reload the last display configuration or undoing a previous action.)

Some of the SRS grammar words had alternate pronunciations. For example, the EP could say "HUD" or say each letter as "H-U-D". Similarly, the "NAV" command could have been uttered as "N-D" and the "field-of-view" could be uttered as "F-O-V".

Positive visual feedback of SRS operation was provided. The EP would press a push-to-listen button and speak a command which was interpreted by the SRS. While the push-to-listen button was pressed, a box with plus signs was displayed at the bottom of the PFD (Fig. 4) and HUD. The SRS was set such that if it was at least 40% confident in its interpretation, it would broadcast the command to the

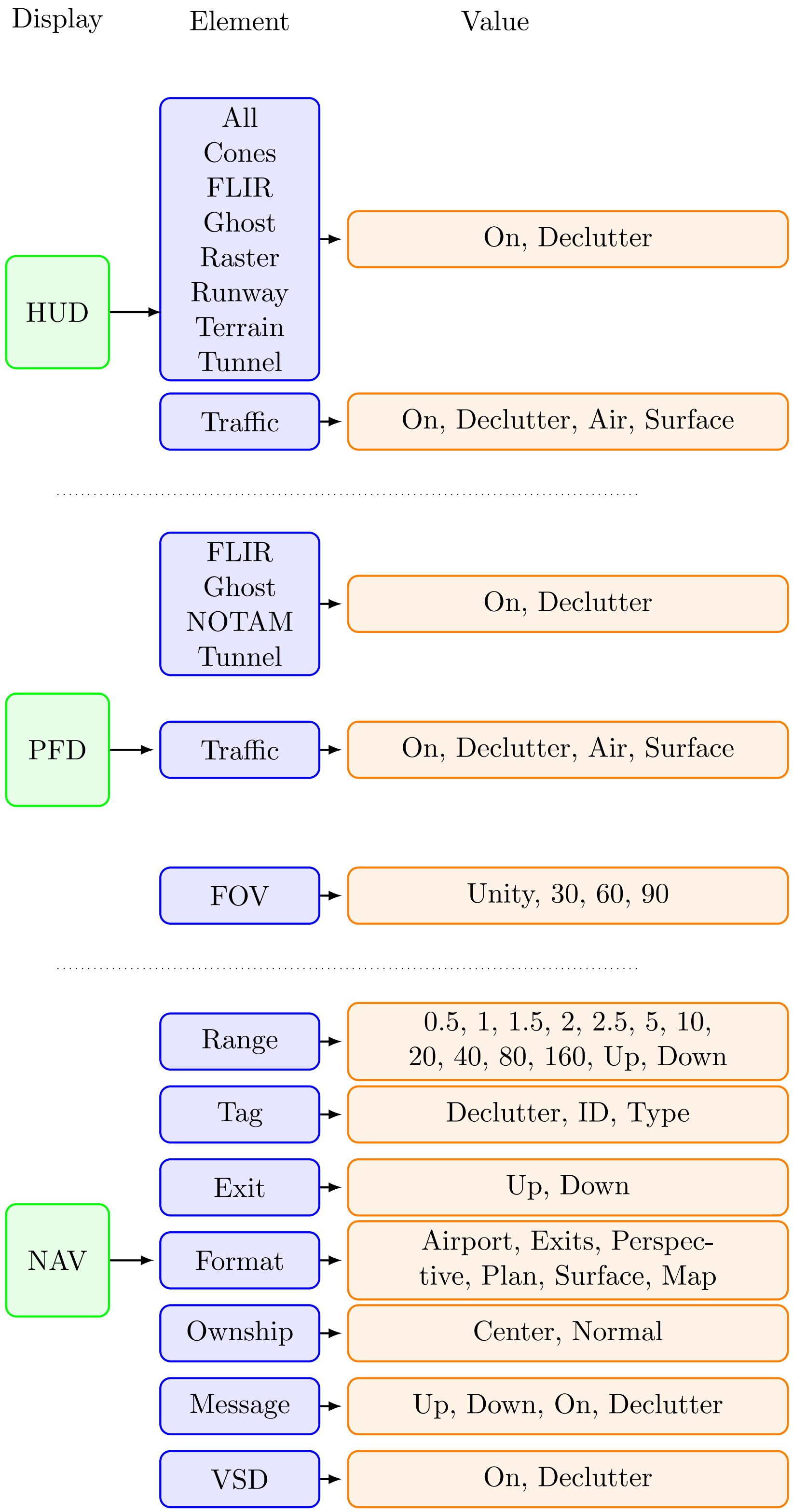


Figure 3. Hierarchical Grammar for GVSITE. The 3 tier grammar structure: 1) Display device, 2) Display element and 3) State.

displays. The interpreted command was then momentarily displayed to the pilot for verification (Fig. 4). If the SRS was less than 40% confident in its interpretation, then a box of minus signs was momentarily displayed at the bottom of the PFD (Fig. 5). The "confidence" value of 40% was selected based on preliminary testing before the evaluation flights began.

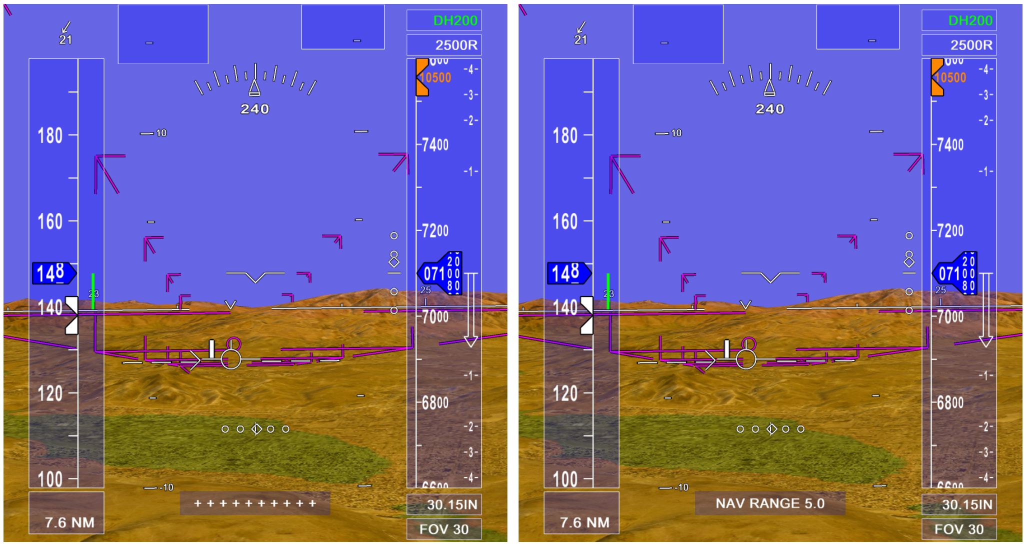


Figure 4. The SRS box awaiting spoken command (left) and displaying the recognized command(right).

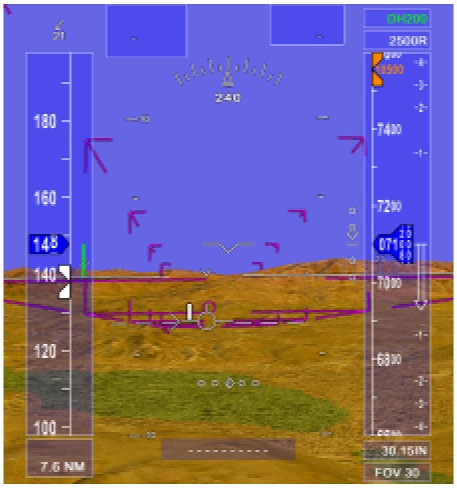


Figure 5. The PFD display when the SRS was not confident in its interpretation.

The complete grammar set-up for GVSITE is given in Appendix A. Some synonyms were also allowed, such as using "NAV" to be the same as "ND." Also, prior to evaluation flying it was determined that the word "off" and "on" was so similar phonetically that very poor recognition rates occurred when both "on" and "off" were included in the grammar. Instead, "declutter" was substituted for "off." In the context of aviation displays, declutter has essentially the same meaning as off; that is, to turn a symbol on a display off is to declutter the display. The use of the word "off" would have been preferred by the EPs since it is most naturally the opposite of "on," but the word "declutter" was acceptable and after training, it became fairly easy to remember and use.

## 2.4 Results

Over the entire flight test, there were 505 total verifiable SRS commands spoken with an overall success rate (accuracy) of 84%. Thus, there were 425 correct recognitions and 80 incorrect recognitions. However, the reported accuracy rate of the SRS software is 96%.

The number of commands spoken as they pertained to a given display is shown in Table 1. The data shows that the PFD and ND were commanded almost identically at a rate four times greater than the HUD. It should be noted that the HUD had hardware symbology controls - a stroke (symbology) and raster (background imagery) declutter controls mounted on the EP's yoke. The SRS commands for the HUD could modulate symbology groups, but the hardware controls toggled the entire stroke or raster HUD components.

Table 1. SRS Commands per Display.

| **Display** | **Commands Spoken** | **Percentage** |
|---|---|---|
| PFD | 222 | 44% |
| ND | 227 | 45% |
| HUD | 56 | 11% |
| **All** | **505** | **100%** |

Each EP averaged 34 SRS commands per test flight, with a maximum of 64 and a minimum of 12. Figure 6 shows the total commands uttered during the flight test for each EP.

There were two types of incorrect recognitions by the SRS:

1. SRS was not confident in matching to any command (i.e., SRS recognition was lower than the threshold level of 40% and therefore, rejected whether the utterance was correctly interpreted or not)

2. SRS incorrectly interpreted a command (i.e., SRS recogntion was greater than the threshold level of 40% but the recognized utterance was not the spoken command; e.g., PFD FOV 60 was interpreted as NAV RANGE 60)

As shown in Figure 6, the errors incurred by EPs varied widely. Two EPs had an error rate of 42% and 37% whereas the remainder were closer to 10% error rates.

Of the commands that the SRS was confident (i.e., exceeded the 40% threshold), the accuracy rate was 96%; that is, only 20 of the 80 incorrect recognitions were

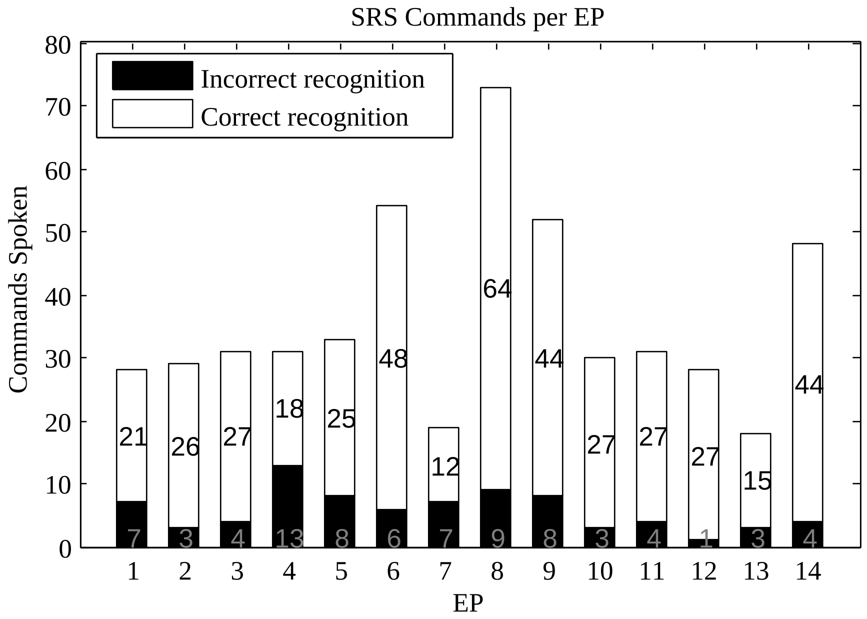


Figure 6. Number of SRS commands spoken by each EP.

misinterpretations. The remaining 60 of the incorrect recognitions were confidence-related where the recognizer performance did not exceed the 40% threshold level and no recognition action was taken (Table 2).

Table 2. Incorrect SRS Commands.

| | **Incorrect Command** | **Percentage** |
|---|---|---|
| Below 40% confidence | 60 | 75% |
| Misinterpretation | 20 | 25% |
| **Total** | **80** | **100%** |

The errors were associated with the display for which the utterance was directed. These data are shown in Figs. 7 - 9. The PFD command accounted for 67 of the total 80 (84%) incorrect recognitions. Given that the ND and PFD were addressed an equal number of times, the PFD accounted for a disproportionate share of the errors.

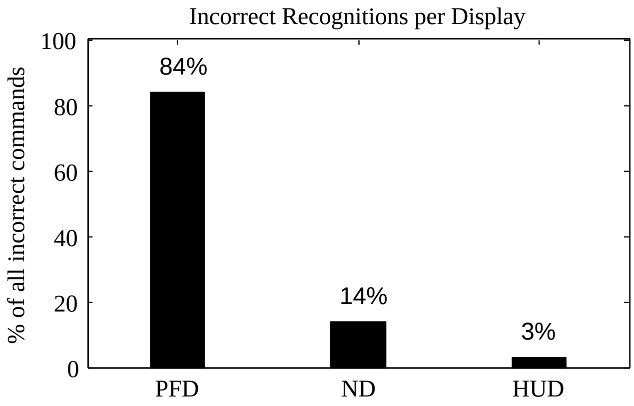


Figure 7. Percentage of Incorrect Recognitions per Display.

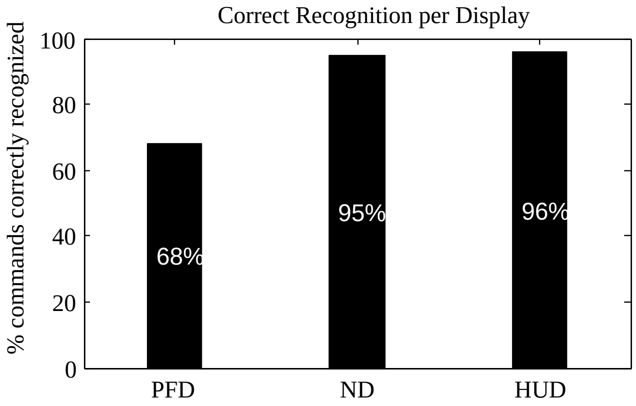


Figure 8. Commands Correctly Recognized per Display.

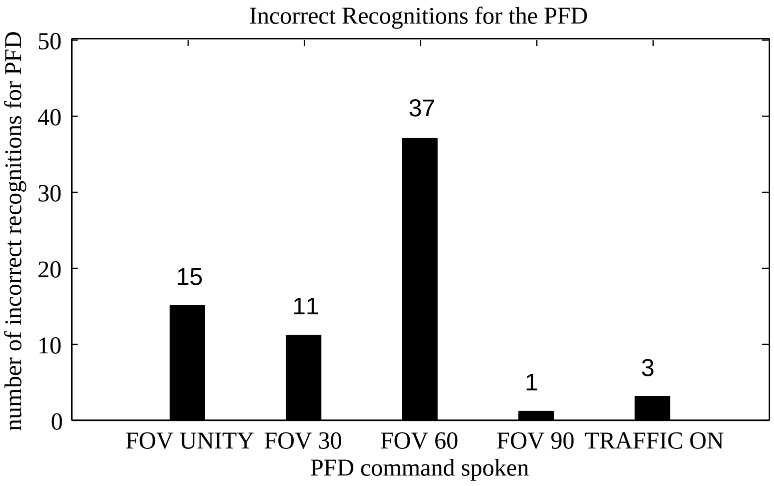


Figure 9. Incorrect Recognitions for the PFD.

## 2.5 Discussion

From all of the GVSITE data flights, the overall success rate of the SRS was 84%. That is, 84% of all SRS commands were interpreted correctly out of a total of 505 SRS commands. Eighty commands were insufficient confidence or an incorrect recognition. For 60 (75%) of the 80 incorrect commands, the SRS was less than 40% confident in the interpretation of that command. For the remaining 20 incorrect commands, the SRS misinterpreted the command (i.e., the pilot uttered "NAV RANGE 5" but the SRS interpreted the command as "NAV RANGE 20")

When the success rate was analyzed on a per display basis, it was found that the ND commands had a 95% success rate and the HUD commands had a 96% success rate, but the PFD had a 68% success rate. The SRS recognition engine has a known practical success rate of 96% which is close to the ND and HUD success rate.

The incorrect recognitions for the PFD were broken-out by command in Figure 9. Clearly, the "PFD FOV" commands contributed to the vast majority of the incorrect recognitions for the PFD commands. The SRS was not confident in the interpretation for these commands most of the time. If the command "PFD FOV" is removed from the analysis, the overall success rate of the SRS would be 96% - the published accuracy rate for the recognizer.

The poor performance of the PFD command was attributable to several factors. A primary factor was the noise-attenuating microphones and head-sets used in the G-V aircraft. Without pilot voice input, the noise-attenuating system canceled out the cockpit ambient noise; however, in doing so, the microphones had a response lag whereby the first utterances into the microphone by the EP had a noticeable volume change (i.e., the volume ramps up once the pilots begin talking, creating an effect whereby the first portion of the word sounds truncated). In this scenario, both the commands "HUD" and "NAV" are nearly phonetically equivalent to their noise-attenuated, truncated versions "UD" and "AV." On the other hand, the command "PFD" is not phonetically equivalent to its noise-attenuated, truncated version "FD." In the cases where the EPs spoke quickly, the noise-attenuating headsets would inadvertently create a difficult recognition problem. This simplistic interpretation was not, but should be, experimentally verified. If confirmed, several changes could be made to ameliorate this effect: a) the grammars may have to be tailored to allow truncated phonetic equivalents; b) the commands could be change to compensate for this effect; or, c) the pilots (users) may be trained to utter something, before depressing the push-to-listen and speaking the desired command.

In addition, another factor in the SRS performance that was potentially significant was the EP audio input volume and quality. During ground checks before each flight, the audio input to the SRS was checked. However, there was not a real-time SRS volume (or quality) monitoring system. This feature was overlooked and should have been installed. The audio volume (and quality) to the SRS is a critical determinate of SRS performance. This was an *uncontrolled* variable in the test since the EPs could and did often modulate their intercom volumes and their boom microphone position which subsequently affected SRS performance. The SRS audio input occasionally was very weak and other times, it was saturated. In any case, audio volume (as a minimum) should be monitored in real-time, preferably by the user using a series of lights to indicate the audio input volume and its operating status (i.e., “high-medium-low”).

## 2.6 EP Recommendations

In addition to these results, pilot comments on the SRS performance and grammar development provide the follow recommendations:

1. Make shortcuts for common commands. For example, allow “RANGE 5” for “NAV RANGE 5” and “VIEW 30” for “PFD FOV 30.” While the methodology for using the hierarchical structure was evident, the “range” command only pertained to the Navigation Display and the “view” command only related to the PFD. Therefore, having to use “NAV” and “PFD” represented extraneous steps to the majority of EPs.
2. Change the cadence of some commands to create a consistent cadence for all, if possible. For example, modify the command “PFD FOV” to a cadence like “NAV RANGE.” This might be done by changing the command “PFD” to a single word (perhaps “Primary”) though a one syllable word is preferred.
3. Change “FOV” to a single word (perhaps “View”). Using “Field-of-View” or even “F-O-V” was verbally cumbersome compared to a single word like “View.”
4. Be consistent with the use of “UP/DOWN” and “INCREASE/DECREASE.” The use of increase/up and decrease/down commands was not always intuitively obvious nor were they programmed in all cases to be synonymous.
5. Essentially all EPs wanted higher accuracy of the SRS on the order of a $99^{+}\%$ recognition performance.

# 3 Laboratory SRS Experiment

Following the GVSITE flight test, a laboratory study was designed to determine a baseline performance of the speaker-independent speech recognition technology. Participants were asked to speak words and phrases from common standard aviation dialog. The study was divided into three segments: single word utterance, short command phrase utterances, and longer ATC clearance phrases. The laboratory study gathered recognition data, measured the basic accuracy of the recognition, and recorded a confidence factor in the recognition output by the SRS. The laboratory study used a different SRS recognizer than the GVSITE flight test.

## 3.1 Participants

A total of 25 (18 male and 7 female) native US English-speaking people participated in the laboratory study. No other information was taken about the participants. The study required approximately 10 minutes for each participant.

## 3.2 Apparatus

A laptop computer with a conventional microphone and earphone connections was utilized for the study. The headset was an Andrea model ANC-700 with an active noise canceling microphone optimized for speech recognition.

Andrea ANC-700 Microphone specifications:

- Noise Cancellation 6 dB/octave
- Frequency Range 100-10,000 Hz
- Impedance at 1 kHz SoundBlaster Interface 300 ohm
- Electrical Signal-to-Noise Ratio 60 dB
- Sensitivity at 1 kHz (0 dB = 1 V/Pa) SoundBlaster Interface -36 dB
- Current Consumption SoundBlaster Interface 0.500 mA

## 3.3 Method

Participants were directed to read words aloud on a screen in three different segments of the study; single word phonetics, short commands, and ATC clearances. The utterances were evaluated for accurate recognition and a confidence factor.

Each participant spoke a total of 71 utterances (Table 3). There were 26 single word utterances (i.e., the aviation phonetic alphabet, Alpha through Zulu) and 45 phrase utterances. The phrase utterances were 39 short phrases and 6 long phrases. The short command phrases were typical flight deck and display management commands used in the previous GVSITE flight test, for example, “NAV RANGE 20.” The longer phrase utterances were taxi clearances, the longest of which was 14 words (19 syllables).

Table 3. All of the 71 utterances each participant spoke.

| | |
|---|---|
| 1 | Alpha |
| 2 | Bravo |
| 3 | Charlie |
| 4 | Delta |
| 5 | Echo |
| 6 | Foxtrot |
| 7 | Golf |
| 8 | Hotel |
| 9 | India |
| 10 | Juliet |
| 11 | Kilo |
| 12 | Lima |
| 13 | Mike |
| 14 | November |
| 15 | Oscar |
| 16 | Papa |
| 17 | Quebec |
| 18 | Romeo |
| 19 | Sierra |
| 20 | Tango |
| 21 | Uniform |
| 22 | Victor |
| 23 | Whiskey |
| 24 | X-Ray |
| 25 | Yankee |
| 26 | Zulu |
| 27 | NAV range 1 |
| 28 | NAV range 2 |
| 29 | NAV range 5 |
| 30 | NAV range 10 |
| 31 | NAV range 20 |
| 32 | NAV range 50 |
| 33 | NAV range 100 |
| 34 | NAV range 200 |
| 35 | NAV zoom out |
| 36 | NAV zoom in |
| 37 | NAV range back |
| 38 | P F D Field of view unity |
| 39 | P F D Field of view 30 |
| 40 | P F D Field of view 60 |
| 41 | P F D Field of view 90 |
| 42 | P F D F O V unity |
| 43 | P F D F O V 30 |
| 44 | P F D F O V 60 |
| 45 | P F D F O V 90 |
| 46 | P F D declutter |
| 47 | P F D traffic on |
| 48 | P F D traffic off |
| 49 | HUD declutter |
| 50 | HUD traffic on |
| 51 | HUD traffic off |
| 52 | Checklist Before Takeoff |
| 53 | Checklist Takeoff |
| 54 | Checklist Climb |
| 55 | Checklist Cruise |
| 56 | Checklist Descent |
| 57 | Checklist Landing |
| 58 | Checklist After Landing |
| 59 | Before Takeoff Checklist |
| 60 | Takeoff Checklist |
| 61 | Climb Checklist |
| 62 | Cruise Checklist |
| 63 | Descent Checklist |
| 64 | Landing Checklist |
| 65 | After Landing Checklist |
| 66 | NASA 557 Taxi To Runway 23 via D F T L |
| 67 | United 231 Taxi At Concourse D via E B A |
| 68 | NASA 557 Hold Short Of Runway 14 Rt at D |
| 69 | United 231 Taxi To Runway 14 Lt via T O B W |
| 70 | NASA 557 Hold At Gate K |
| 71 | United 231 Hold At Concourse J |

In all three segments of the study, the independent variable was the utterance. The dependent variables were the accuracy (correct or incorrect) and the confidence factor (0-100). In the first segment, utterance numbers 1 through 26 were used (Alpha to Zulu). In the second segment, the utterances were approximately 3-5 words long. Utterance numbers 27 through 65 were used in the second segment. Additionally, in the second segment, two different command sets were compared and evaluated for accuracy. The command set starting with "PFD Field-of-View *view angle*" was compared to the set starting with "PFD FOV *view angle*." The second set of commands compared were the set starting with "Checklist *checklist name*" versus "*checklist name* Checklist." The third segment, utterance numbers 65 through 71, were modeled after typical ATC ground control clearances.

## 3.4 Procedure

Each participant was fitted with the headset and the microphone set to a distance proportional to the normal speaking volume. A volume level meter in the software was utilized to consistently set the appropriate position and microphone input level. The height of the microphone was also set below the "Puff line" to reduce the wind noise while speaking "P" sounds.

Each participant was instructed to speak the directed word or phrase. The speech recognition software acquired the audio and applied its recognition algorithms. The recognized utterance was displayed to each participant, and the participant recorded if the utterance was correctly recognized. This continued until all utterances were completed in each segment of the study.

## 3.5 Results

The overall recognition rate for all 71 utterances by all 25 participants (1775 utterances) was 95.5% correct. Per participant, the median was 96%, the maximum was 100%, the minimum was 77% and the standard deviation was 5.52.

### 3.5.1 Segment 1: Single Word Utterance

The recognition rate for all aviation phonetic utterances by all 25 participants (650 utterances) was 94.8% correct, as detailed in Table 4. Performance per participant ranged from 100% to 80% for the phonetic values of 'A' and 'P.'

The recognizer's confidence level (see Appendix E) for the aviation phonetic alphabet utterances (single words) by all 25 participants (650 utterances) is broken down according to each phonetic in Table 5. The data shows that the mean standard deviation for confidence was approximately 8.0. The utterance "Tango" exhibited the greatest variability (standard deviation).

Finally, the percentage correct by participant for the single word utterances is given in Table 6. Eight participants obtained perfect recognition score. One participant only had 77% recognition performance.

Table 4. Mean Percentage of Correct Recognition, All Participants (N=25).

| **Utterance** | **Mean %** |
|---|---|
| Foxtrot, India, Juliet, Quebec, Sierra, Uniform, Victor, X-Ray, Yankee | 100 |
| Charlie, Delta, Echo, Hotel, Kilo, Lima, Mike, Oscar, Romeo, Whiskey | 96 |
| November, Tango | 92 |
| Bravo, Zulu | 88 |
| Golf | 84 |
| Alpha, Papa | 80 |

### 3.5.2 Segment 2: Short Phrase Utterance

Within the second segment, two different command sets were evaluated to determine which set to use. The confidence level for these phrases is tabulated in Table 7, as well as the mean correct recognition rate.

The "PFD Field of View *number*" versus "PFD FOV *number*" set both were recognized 100% of the time. Similarly, the "Checklist *checklist name*" versus "*checklist name* Checklist" command set was only different by 1%. Since the accuracy data revealed no clear advantage, the more natural speech data sets will be used; "PFD Field of View *number*", and "*checklist name* Checklist".

Finally, the percentage correct by participant for the short phrase word utterances is given in Table 8. Fourteen participants obtained perfect recognition score. One participant only had 85% recognition performance, whereas they had 92% performance in the single word utterance test. The participant with the worst performance in the single word utterances, scored 95% in the short phrase utterances.

### 3.5.3 Segment 3: ATC Long Phrase Utterance

Segment 3 was added as a first-look towards future studies of SRS applications in the cockpit as an interface to ATC utterances.

The confidence levels for the long phrase utterances are given in Table 9. The percentage correct by participant for the long phrase word utterances is shown in Table 10.

The percentage correct by participant shows that 8 participants obtained perfect recognition score; however, 4 participants only had 67% recognition performance with a mean recognition rate for all participates of 86%. Most of the ATC phrase utterances were correct, with only one word being incorrect. In the case of the utterance "United 231 Taxi to Concourse Delta via Echo Bravo Alpha", 'Alpha' was mis-recognized 5 times, which correlates to the error rate found with 'Alpha' from the first phonetic segment. Also, it was noted that short syllable words (at, and, to) were dropped many times.

The summary of correct recognitions for all segments is shown in Table 11.

Table 5. Segment 1: Confidence of Phonetic, All Participants (N=25).

| | **Confidence** | | | | | **Recognized** |
|---|---|---|---|---|---|---|
| **Utterance** | **Mean** | **Median** | **Max** | **Min** | **SD** | **Mean Correct** |
| Alpha | 64.76 | 67 | 79 | 36 | 12.0 | 80 |
| Bravo | 68.56 | 71 | 80 | 45 | 8.4 | 88 |
| Charlie | 66.32 | 67 | 87 | 42 | 8.9 | 96 |
| Delta | 80.04 | 82 | 89 | 48 | 8.4 | 96 |
| Echo | 75.12 | 77 | 85 | 57 | 6.6 | 96 |
| Foxtrot | 66.96 | 69 | 80 | 48 | 8.0 | 100 |
| Golf | 65.76 | 69 | 80 | 47 | 9.3 | 84 |
| Hotel | 72.76 | 75 | 88 | 34 | 12.3 | 96 |
| India | 80.60 | 80 | 88 | 67 | 5.0 | 100 |
| Juliet | 73.20 | 74 | 85 | 56 | 6.4 | 100 |
| Kilo | 66.68 | 69 | 82 | 44 | 8.9 | 96 |
| Lima | 75.08 | 75 | 89 | 55 | 9.4 | 96 |
| Mike | 77.56 | 78 | 88 | 49 | 8.4 | 96 |
| November | 73.20 | 77 | 86 | 34 | 12.5 | 92 |
| Oscar | 70.20 | 72 | 80 | 45 | 7.5 | 96 |
| Papa | 66.44 | 69 | 77 | 50 | 7.2 | 80 |
| Quebec | 62.40 | 62 | 76 | 46 | 7.3 | 100 |
| Romeo | 73.32 | 75 | 87 | 45 | 9.8 | 96 |
| Sierra | 61.12 | 60 | 72 | 47 | 6.7 | 100 |
| Tango | 70.64 | 75 | 85 | 0 | 16.9 | 92 |
| Uniform | 68.72 | 68 | 83 | 51 | 7.2 | 100 |
| Victor | 71.04 | 70 | 85 | 49 | 8.3 | 100 |
| Whiskey | 79.44 | 79 | 88 | 71 | 3.5 | 96 |
| X-Ray | 72.56 | 73 | 87 | 54 | 8.4 | 100 |
| Yankee | 69.44 | 73 | 85 | 52 | 10.3 | 100 |
| Zulu | 60.36 | 63 | 72 | 31 | 9.2 | 88 |

Table 6. Segment 1: Phonetics Percent Correct per Participant Sorted by Incorrect Recognitions.

| **Participant** | **Correct** | **Incorrect** | **% Correct** | **Dev from avg** | **Misrecognized** |
|---|---|---|---|---|---|
| 1 | 19 | 7 | 76.9 | -18.0 | B, C, H, N, O, T, Y |
| 2 | 22 | 4 | 84.6 | -10.3 | E, K, T, Z |
| 3 | 23 | 3 | 88.5 | -6.5 | A, P, Z |
| 4 | 24 | 2 | 92.3 | -2.6 | B, O |
| 5 | 24 | 2 | 92.3 | -2.6 | A, G |
| 6 | 24 | 2 | 92.3 | -2.6 | G, M |
| 7 | 24 | 2 | 92.3 | -2.6 | P, R |
| 8 | 24 | 2 | 92.3 | -2.6 | A, L |
| 9 | 24 | 2 | 92.3 | -2.6 | A, P |
| 10 | 25 | 1 | 96.2 | 1.2 | A |
| 11 | 25 | 1 | 96.2 | 1.2 | D |
| 12 | 25 | 1 | 96.2 | 1.2 | G |
| 13 | 25 | 1 | 96.2 | 1.2 | P |
| 14 | 25 | 1 | 96.2 | 1.2 | P |
| 15 | 25 | 1 | 96.2 | 1.2 | B |
| 16 | 25 | 1 | 96.2 | 1.2 | G |
| 17 | 25 | 1 | 96.2 | 1.2 | Z |
| 18 | 26 | 0 | 100.0 | 5.1 | |
| 19 | 26 | 0 | 100.0 | 5.1 | |
| 20 | 26 | 0 | 100.0 | 5.1 | |
| 21 | 26 | 0 | 100.0 | 5.1 | |
| 22 | 26 | 0 | 100.0 | 5.1 | |
| 23 | 26 | 0 | 100.0 | 5.1 | |
| 24 | 26 | 0 | 100.0 | 5.1 | |
| 25 | 26 | 0 | 100.0 | 5.1 | |
| | | **Mean** | **94.92** | | |
| | | **Median** | **96.15** | | |
| | | **SD** | **5.52** | | |
| | | **Max** | **100.00** | | |
| | | **Min** | **76.92** | | |

Table 7. Segment 2: Confidence of Command, All Participants (N=25).

| | **Confidence** | | | | | **Recognized** |
|---|---|---|---|---|---|---|
| **Utterance** | **Mean** | **Median** | **Max** | **Min** | **SD** | **Mean Correct** |
| NAV range 1 | 72.56 | 75 | 81 | 44 | 7.4838 | 96 |
| NAV range 2 | 79.56 | 80 | 86 | 67 | 4.4355 | 100 |
| NAV range 5 | 72.68 | 75 | 82 | 37 | 9.5904 | 96 |
| NAV range 10 | 77.36 | 78 | 83 | 71 | 3.8824 | 100 |
| NAV range 20 | 71.76 | 71 | 78 | 59 | 4.684 | 100 |
| NAV range 50 | 72.76 | 73 | 84 | 66 | 4.6123 | 100 |
| NAV range 100 | 68.68 | 69 | 79 | 60 | 4.58 | 100 |
| NAV range 200 | 70.60 | 72 | 79 | 35 | 8.5147 | 96 |
| NAV zoom out | 70.44 | 73 | 78 | 35 | 8.3869 | 96 |
| NAV zoom in | 72.16 | 72 | 80 | 64 | 4.5797 | 96 |
| NAV range back | 66.44 | 73 | 84 | 0 | 19.929 | 84 |
| P F D Field of view unity | 70.40 | 71 | 79 | 57 | 5.7591 | 100 |
| P F D Field of view 30 | 68.40 | 71 | 82 | 40 | 9.3986 | 100 |
| P F D Field of view 60 | 67.76 | 70 | 81 | 36 | 9.2298 | 100 |
| P F D Field of view 90 | 69.16 | 71 | 76 | 57 | 5.735 | 100 |
| P F D F O V unity | 71.96 | 72 | 81 | 63 | 4.8087 | 100 |
| P F D F O V 30 | 73.92 | 74 | 85 | 60 | 5.7076 | 100 |
| P F D F O V 60 | 72.76 | 73 | 82 | 43 | 7.5899 | 100 |
| P F D F O V 90 | 72.84 | 73 | 82 | 57 | 5.5579 | 100 |
| P F D declutter | 66.76 | 69 | 78 | 37 | 7.7421 | 96 |
| P F D traffic on | 71.20 | 73 | 81 | 56 | 6.7144 | 100 |
| P F D traffic off | 69.28 | 70 | 80 | 56 | 6.4841 | 88 |
| HUD declutter | 63.96 | 64 | 74 | 56 | 4.8346 | 100 |
| HUD traffic on | 69.92 | 70 | 79 | 58 | 5.4077 | 100 |
| HUD traffic off | 67.52 | 69 | 78 | 52 | 6.7769 | 88 |
| Checklist Before Takeoff | 70.92 | 73 | 78 | 48 | 6.1841 | 100 |
| Checklist Takeoff | 69.72 | 72 | 78 | 54 | 7.3116 | 100 |
| Checklist Climb | 73.64 | 74 | 82 | 60 | 5.322 | 100 |
| Checklist Cruise | 73.08 | 75 | 79 | 53 | 6.4026 | 100 |
| Checklist Descent | 71.28 | 73 | 85 | 46 | 7.8396 | 96 |
| Checklist Landing | 69.84 | 73 | 81 | 0 | 15.184 | 96 |
| Checklist After Landing | 67.92 | 71 | 81 | 36 | 9.1511 | 96 |
| Before Takeoff Checklist | 72.56 | 74 | 82 | 42 | 8.1705 | 96 |
| Takeoff Checklist | 72.24 | 75 | 82 | 57 | 6.6538 | 96 |
| Climb Checklist | 72.08 | 72 | 79 | 64 | 3.9887 | 100 |
| Cruise Checklist | 67.72 | 68 | 78 | 57 | 5.712 | 96 |
| Descent Checklist | 71.60 | 73 | 82 | 56 | 6.1779 | 100 |
| Landing Checklist | 73.44 | 75 | 83 | 61 | 5.6648 | 92 |
| After Landing Checklist | 70.72 | 71 | 81 | 59 | 5.8489 | 100 |

Table 8. Segment 2: Commands, Percent Correct per Participant Sorted by Incorrect Recognitions.

| **Participant** | **Correct** | **Incorrect** | **% correct** | **Dev from avg** |
|---|---|---|---|---|
| 1 | 33 | 6 | 84.6 | -12.9 |
| 2 | 35 | 4 | 89.7 | -7.8 |
| 3 | 36 | 3 | 92.3 | -5.2 |
| 4 | 36 | 3 | 92.3 | -5.2 |
| 5 | 37 | 2 | 94.9 | -2.7 |
| 6 | 38 | 1 | 97.4 | -0.1 |
| 7 | 38 | 1 | 97.4 | -0.1 |
| 8 | 38 | 1 | 97.4 | -0.1 |
| 9 | 38 | 1 | 97.4 | -0.1 |
| 10 | 38 | 1 | 97.4 | -0.1 |
| 11 | 38 | 1 | 97.4 | -0.1 |
| 12 | 39 | 0 | 100.0 | 2.5 |
| 13 | 39 | 0 | 100.0 | 2.5 |
| 14 | 39 | 0 | 100.0 | 2.5 |
| 15 | 39 | 0 | 100.0 | 2.5 |
| 16 | 39 | 0 | 100.0 | 2.5 |
| 17 | 39 | 0 | 100.0 | 2.5 |
| 18 | 39 | 0 | 100.0 | 2.5 |
| 19 | 39 | 0 | 100.0 | 2.5 |
| 20 | 39 | 0 | 100.0 | 2.5 |
| 21 | 39 | 0 | 100.0 | 2.5 |
| 22 | 39 | 0 | 100.0 | 2.5 |
| 23 | 39 | 0 | 100.0 | 2.5 |
| 24 | 39 | 0 | 100.0 | 2.5 |
| 25 | 39 | 0 | 100.0 | 2.5 |
|  |  | **Mean** | **97.54** |  |
|  |  | **Median** | **100.00** |  |
|  |  | **SD** | **3.95** |  |
|  |  | **Max** | **100.00** |  |
|  |  | **Min** | **84.62** |  |

Table 9. Segment 3: Confidence by ATC Phrase, All Participants (N=25).

| Utterance | Mean | Median | Max | Min | SD |
|---|---|---|---|---|---|
| NASA 557 Taxi To Runway 23 via D F T L | 100 | 73 | 80 | 65 | 4.32 |
| United 231 Taxi to Concourse D via E B A | 72 | 69 | 79 | 61 | 4.28 |
| NASA 557 Hold Short Of Runway 14R at D | 92 | 67 | 76 | 59 | 4.12 |
| United 231 Taxi to Runway 14L via T O B W | 92 | 72 | 80 | 61 | 4.57 |
| NASA 557 Hold at Gate K | 80 | 65 | 70 | 63 | 2.33 |
| United 231 Hold at Concourse J | 80 | 65 | 78 | 57 | 5.10 |

### 3.6 Optimization

An utterance was marked correct if the participant marked the recognizer guess as correct. The recognizer had an internal algorithm to determine recognition was correct based on an "utterance score." This utterance score was equal to or greater than the utterance threshold setting of 50. In addition to the participant's correct score, the recognizer's score, based on the utterance score, was recorded as well. This data was analyzed to determine an optimized utterance score threshold setting to achieve better recognition rates using the recognizer's utterance score.

Of all the single word phonetic utterances (650), there were 7 occurrences (1.1%) when the SRS was marked a correct recognition but the confidence threshold was less than 50 and, thus, was recorded as incorrect. Conversely, there were 15 occurrences (2.3%) when the utterance was actually incorrect but was determined by the SRS to be correct.

Often in digital avionics design, priorities are set to achieve error detection first then error correction follows. In other words, it is better to get no data than data that is erroneous. For instance, ARINC 429 digital data bus has no error correction capability, but transmits data (error detection) to determine if a data packet was receive correctly. Mirroring this theme, SRS optimization may be set to achieve a lower false positive rate than overall recognition rate.

Changing the threshold setting facilitates some recognition optimization. A "false positive" is where the utterance score was greater than the threshold and deemed correct, but was actually incorrect. To optimize for a minimum false positive recognitions, the threshold could be set to a higher value. Resetting the threshold setting to 52 reduces the false positive rate 0.5%, but there would be a decrease of 0.6% in the overall recognition rate (Table 12). Resetting the threshold setting to 48 would increase the correct recognition rate 1.2%, but also increases the false positive rate by 0.6%. Depending upon priority, SRS optimization by threshold setting is possible within a small range.

Table 10. Segment 3: ATC Phrase, Percent Correct per Participant sorted by Incorrect Recognitions.

| **Participant** | **Correct** | **Incorrect** | **% Correct** | **Dev from avg** |
|---|---|---|---|---|
| 1 | 4 | 2 | 66.7 | -19.3 |
| 2 | 4 | 2 | 66.7 | -19.3 |
| 3 | 4 | 2 | 66.7 | -19.3 |
| 4 | 4 | 2 | 66.7 | -19.3 |
| 5 | 5 | 1 | 83.3 | -2.7 |
| 6 | 5 | 1 | 83.3 | -2.7 |
| 7 | 5 | 1 | 83.3 | -2.7 |
| 8 | 5 | 1 | 83.3 | -2.7 |
| 9 | 5 | 1 | 83.3 | -2.7 |
| 10 | 5 | 1 | 83.3 | -2.7 |
| 11 | 5 | 1 | 83.3 | -2.7 |
| 12 | 5 | 1 | 83.3 | -2.7 |
| 13 | 5 | 1 | 83.3 | -2.7 |
| 14 | 5 | 1 | 83.3 | -2.7 |
| 15 | 5 | 1 | 83.3 | -2.7 |
| 16 | 5 | 1 | 83.3 | -2.7 |
| 17 | 5 | 1 | 83.3 | -2.7 |
| 18 | 6 | 0 | 100.0 | 14.0 |
| 19 | 6 | 0 | 100.0 | 14.0 |
| 20 | 6 | 0 | 100.0 | 14.0 |
| 21 | 6 | 0 | 100.0 | 14.0 |
| 22 | 6 | 0 | 100.0 | 14.0 |
| 23 | 6 | 0 | 100.0 | 14.0 |
| 24 | 6 | 0 | 100.0 | 14.0 |
| 25 | 6 | 0 | 100.0 | 14.0 |
|  |  | **Mean** | **86.00** |  |
|  |  | **Median** | **83.33** |  |
|  |  | **SD** | **11.47** |  |
|  |  | **Max** | **100.00** |  |
|  |  | **Min** | **66.67** |  |

Table 11. Total Correct Recognition for All Participants (N=25).

| **Segment** | **Phonetic (26)** | **Short Phrase (39)** | **ATC Phrase (6)** | **Total (71)** |
|---|---|---|---|---|
| % Correct | 94.9 | 97.5 | 86.0 | 95.5 |

Table 12. Optimization analysis of Confidence Threshold setting.

| **Confidence Threshold Setting** | **Marked Correct** | **Marked Incorrect** | **% Correct** | **Correct but marked incorrect** | **Incorrect but marked correct** | **Total** |
|---|---|---|---|---|---|---|
| 52 | 618 | 32 | 95.1 | 11 (1.7%) | 12 (1.8%) | 23 |
| 51 | 621 | 29 | 95.5 | 9 (1.4%) | 13 (2.0%) | 22 |
| 50 | 625 | 25 | 96.2 | 7 (1.1%) | 15 (2.3%) | 22 |
| 49 | 629 | 21 | 96.7 | 5 (0.7%) | 17 (2.6%) | 22 |
| 48 | 633 | 17 | 97.4 | 3 (0.5%) | 19 (2.9%) | 22 |
| 47 | 636 | 14 | 97.8 | 2 (0.3%) | 21 (3.2%) | 23 |

### 3.7 Discussion

The voice independent and natural continuous speech requirements were successfully demonstrated by the SRS engine. The SRS works well, with better than a 95% recognition rate of 1775 utterances by 25 different participants.

For flight, improved microphone/noise canceling is critical for the input audio signal. In the laboratory test, the SRS performed at the known recognition rate while the aircraft SRS performed well below the known recognition rate. Two major differences between the flight test and the laboratory was: 1) more ambient noise in the flight test compared to the laboratory environment, and 2) the lack of a volume display for the flight test. Fixing these limitations would still yield a best recognition of 96% yet pilots have indicated the SRS would need closer to $99.99^{+}$% correct recognition rate.

## 4 Conclusions

The data highlights some of the issues and challenges of creating a speech recognition system for the aviation domain and identifies some of the specific issues such as the use of the aviation phonetic alphabet.

The data shows that significant research and development is still required. In general, the recognition rate requirements for commercial speech recognition systems are significantly below the recognition rates that are designed for or required by commercial applications. Even as SRSs continue to gain acceptability in consumer electronics (Siri$^{\text{TM}}$, Cortana$^{\text{TM}}$, Amazon Echo$^{\text{TM}}$), aviation communications are not “natural language,” thus it is critical that the recognizers for aviation use be tailored to this unique environment. To increase recognition rates, structured or limited grammars and hierarchical structures, speaker-dependence, and geo-reference or contextual tailoring of the SRS is acceptable for the aviation domain and should be pursued. For example, a database of waypoint names and their pronunciation should be correlated in real-time with aircraft position to increase recognition rates (i.e., a pilot flying in Virginia will probably not be uttering a waypoint name located in California).

The future need for speech recognition systems in aviation applications is growing more critical every day. This push is driven primarily from the emphasis toward and pending criticality of increased data communication between operators in the National Air Space (NAS), especially digital communications by such systems as Aircraft Communications Addressing and Reporting System (ACARS), controller-pilot data-link communication (CPDLC), and future operational paradigms characterized as "Net-Centric Operations" where new operations are enabled by passing status, intent, and performance data between all users for cooperative, coordinated flight operations.

In these operations, human oversight, awareness, and possible intervention is still required even as the machine-to-machine collaboration is growing in volumes and will soon overwhelm the humans-in-the-loop in both information volume and clutter. To handle these data, Increasingly Autonomous Systems (IAS) are needed to effectively inform humans of relevant information (traffic, intent, messaging) being passed and enable interaction or intervention if necessary. The IAS is an autonomous system that understands these communications (and path planning/intent/state data from all aircraft within reception) and parses the information to extract - concisely and succinctly  only that relevant to pilot. The IAS is adaptive – learning from user input and contextual data – via machine learning algorithms, and to be maximally effective, employs human-centered design principals, first and foremost of which is the attribute of bi-directional communication. Speech (text-to-speech and speech-to-text) as the IAS interface method, therefore, becomes critical. Research shows that natural interaction – aural communication – is a prerequisite to create a low workload, intuitive IAS interface.

Significant commercial technology emphasis is on natural language recognition and understanding. As such, future research will include pairing an aviation-specific SRS with, for instance, the IBM Watson technologies [43] to explore possible workload reduction for the commercial flight deck.

IAS are emerging in more aviation applications than just trajectory planning and execution. The technologies - machine learning, cognitive computing, etc. characterized by the IBM Watson  are emerging as viable capabilities to improve the safety and performance within the aviation domain. The technical challenge is to create these Increasingly Autonomous Systems  intelligent machines, using machine learning algorithms, with human involvement and interaction by which the performance of the combined system exceeds that of either system separately. Human-autonomy teaming is critical to the success of the IAS, and as such, speech as a natural, intuitive interface becomes an enabling technology for autonomous systems.

Future research will include pairing an aviation specific SRS with the IBM Watson technologies to explore possible workload reduction for the commercial flight deck.

# References

1. Joint Planning and Development Office. Next-generation air transportation system integrated plan. Technical Report Version 1, U.S. Department of Transportation, December 2004.

2. O.V. Prinzo. An analysis of approach control/pilot voice communications. Technical Report DOT/FAA/AM-96/26, FAA Civil Aeromedical Institute, 1996.

3. K.M. Cardosi. An analysis of en route controller-pilot voice communications. Technical Report DOT/FAA/RD - 93/11, Department of Transportation, Federal Aviation Administration, 1993.

4. E. Hutchins, B.E. Holder, and R.A. Perez. Culture and flight deck operations. Technical Report Sponsored Research Agreement 22-5003, University of California San Diego, January 2002.

5. FAA. The interfaces between flightcrews and modern flight deck systems. Technical report, Federal Aviation Administration Human Factors Team Report, June 1996.

6. J.L. Soeters and P.C. Boer. Culture and flight safety in military aviation. *International Journal of Aviation Psychology*, 10(2):111–133, 2000.

7. J. Croft. Asia-pacific offers much opportunity for expatriate pilots. *Aviation Week & Space Technology*, 9, February 2015.

8. J. Orasanu, J. Davison, and U. Fischer. What did he say? culture and language barriers to efficient communication in global aviation. In *9th International Symposium on Aviation Psychology*, pages 673–678, Columbus, OH, April 1997.

9. Flight Safety Foundation. High stakes in language proficiency. In *Flight Safety Digest*, Jan-Feb 2006.

10. Eurocontrol. Initial communications operating concepts and requirements for the future radio system. Technical report, Eurocontrol/FAA Future Communications Study Operational Concepts and Requirements, January 2005.

11. K. Kerns. Data-link communication between controllers and pilots: A review and synthesis of the simulation literature. *International Journal of Aviation Psychology*, 1(3):181–204, 1991.

12. FAA. User benefits of two-way data link atc communications: Aircraft delay and flight efficiency in congested en route airspace. Technical Report FAA Report FAA/CT-95-4, Federal Aviation Administration Data Link Benefits Study Team, February 1995.

13. FAA. Benefits of controller-pilot data link atc communications in terminal airspace. Technical Report FAA Report FAA/CT-96-3, Federal Aviation Administration Data Link Benefits Study Team, September 1996.

14. S. A. Lozito, S. A. McGann, and K. Corker. Data link air traffic control and flight deck environments: Experiments in flight crew performance. In R. E. Jensen and D. Neumeister, editors, *7th International Symposium on Aviation Psychology*, pages 1009–1015, Columbus, OH, April 1993. The Ohio State University.

15. W.H. Corwin and H. McCauley. Considerations for the retrofit of data link. In *Aerospace Technology Conference and Exposition*, pages 54–59, Long Beach, CA, October 1990. SAE.

16. A.D. Andre, J.M.C. Lins, and J. Wilson. Conveying message criticality via datalink. In *Twelfth International Symposium on Aviation Psychology*, pages 54–59, Dayton, OH, 2003. Wright State University.

17. N. J. Talotta and C. Shingledecker. Controller evaluation of initial data link terminal air traffic control services: Mini-study 3. Technical Report DOT/FAA/CT-92/18, Volume I, U.S. Department of Transportation, Federal Aviation Administration, 1992.

18. C. D. Wickens, S. Miller, and M. Tham. The implications of data-link for representing pilot request information on 2-d and 3-d air traffic control displays. *International Journal of Industrial Ergonomics*, 18:283–293, 1996.

19. M. C. Waller and G. W. Lohr. A piloted simulation of data link atc message exchange. Technical Report NASA TP-2859, NASA Langley Research Center, 1989.

20. J. L. Groce and G. P. Boucek. Air transport crew tasking in an atc data link environment. Technical Report SAE Tech. Paper 871764, SAE International, Warrendale, PA, 1987.

21. R.N.H.W. van Gent. Human factors issues with airborne data link; towards increased crew acceptance for both en-route and terminal flight operations. Technical Report NLR TP 95666, National Aerospace Laboratory, Amsterdam, NL, 1995.

22. A. H. Midkiff and R. J. Hansman Jr. Identification of important "party line" information elements and implications for situational awareness in the datalink environment. *Air Traffic Control Quarterly*, 1(1):5–30, 1993.

23. A.R. Pritchett and R.J. Hansman. Variations in "party line" information importance between pilots of different characteristics. In *8th International Symposium on Aviation Psychology*, pages 673–678, Columbus, OH, April 1995.

24. C. E. Knox and C. H. Scanlon. Flight tests with a data link used for air traffic control information exchange. Technical Report NASA Tech. Paper 3135, NASA Langley Research Center, Hampton, VA, 1991.

25. N. Smith, J. Moses, S. Romahn, P. Polson, J. Brown, M. Dunbar, E. Palmer, and S. Lozito. An assessment of flight crew experiences with fans-1 atc data link. In *Tenth International Symposium on Aviation Psychology*, Columbus, OH, 1999.

26. J.H.L. Hansen. Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition. *Speech Communication*, 20:151–173, November 1996.

27. A.J. South. Some characteristics of speech produced under high g-force and pressure breathing. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 2095–2098, March 1999.

28. E.B. Werkowitz. Speech recognition in the tactical environment: the AFTI/F-16 voice command flight test. In *Speech Tech '84 Voice Input/Output Applications Show and Conference*, pages 103–105, New York, 1984.

29. V. Riley. Developing a pilot-centered autoflight interface. In *2000 World Aviation Conference*, number 2000-01-5598, October 10 - 12 1999.

30. J. Rankin and P. Mattson. Controller interface for controller-pilot data link communications. In *16th DASC*, October 1997.

31. R.A. Faerber and J.L. Garloch. Usability evaluation of speech synthesis and recognition for improving the human interface to next generation data link communication systems. In *The 19th Digital Avionics Systems Conferences, 2000.*, October 2000.

32. A. Lechner, P. Mattson, and K. Ecker. Voice recognition: software solutions in real-time atc workstations. *Aerospace and Electronic Systems Magazine, IEEE*, 17:11–16, November 2002.

33. J.M. Noyes and A.F. Starr. A comparison of speech input and touch screen for executing checklists in an avionics application. *International Journal of Aviation Psychology*, 17:299–315, 2007.

34. S. Damiani, E. Deregibus, and L. Andreone. Driver-vehicle interfaces and interaction: where are they going? *European transport research review*, 1(2):87–96, 2009.

35. H. Zhang and W.L Ng. Speech recognition interface design for in-vehicle system. In *2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, October 2010.

36. M. Draper, G. Calhoun, H. Ruff, D. Williamson, and T. Barry. Manual versus speech input for unmanned aerial vehicle control station operations. In *Human Factors and Ergonomics Society Annual Meeting*, volume 47, pages 109–113, October 2003.

37. R.K. Osgood and Jr. D.R. Chapman. JSF integrated helmet audio-visual system technology demonstration results. In Ronald J. Lewandowski, Loran A. Haworth, and Henry J. Girolamo, editors, *Head-Mounted Displays II*, volume 3058, April 1997.

38. T. Nash. Eurofighter typhoon a record of continuing achievements. *Air & Space Europe*, 1(3):40–45, 1999.

39. T.G. Reynolds, R.J. Hansman, R. Bolczak, and R. Tarakan. Improving surveillance of clearances in future air traffic control systems. Technical Report AIAA 2004-6391, AIAA, September 2004.

40. R. S. McCann, B. L. Hooey, B. Parke, D. C. Foyle, A. D. Andre, and B. Kanki. An evaluation of the taxiway navigation and situation awareness (T-NASA) system in high-fidelity simulation. *SAE Transactions: Journal of Aerospace*, 107:1612–1625, 1998.

41. C. Nass and S. Brave. *Wired for Speech: How Voice Activates and Advances the HumanComputer Relationship*. The MIT Press, 2005.

42. L. J. Kramer, J. J. Arthur III, R. E. Bailey, and L. J. Prinzel III. Flight testing an integrated synthetic vision system. In Jacques G. Verly, editor, *Enhanced and Synthetic Vision Proceedings of SPIE*, volume 5802, Bellingham, WA, 2005. SPIE.

43. D.A. Ferrucci. Introduction to “This is Watson”. *IBM Journal of Research and Development*, 56(3.4):1:1–1:15, 2012.

# Appendix A

## Available commands for the GVSITE flight test

The following tables list the available Speech Recognition System (SRS) commands. In addition, Evaluation Pilots (EPs) could use the short cut commands CANCEL and REPEAT. The CANCEL command would undo the last command. The REPEAT command would repeat the last command spoken.

Table A1. HUD Commands for GVSITE flight test.

| **Display** | **Attribute** | **State** |
|---|---|---|
| HUD | ALL | DECLUTTER |
| HUD | ALL | ON |
| HUD | TUNNEL | DECLUTTER |
| HUD | TUNNEL | ON |
| HUD | TERRAIN | DECLUTTER |
| HUD | TERRAIN | ON |
| HUD | GHOST | DECLUTTER |
| HUD | GHOST | ON |
| HUD | TRAFFIC | DECLUTTER |
| HUD | TRAFFIC | ON |
| HUD | TRAFFIC | AIR |
| HUD | TRAFFIC | SURFACE |
| HUD | CONES | DECLUTTER |
| HUD | CONES | ON |
| HUD | FLIR | DECLUTTER |
| HUD | FLIR | ON |
| HUD | RASTER | DECLUTTER |
| HUD | RASTER | ON |
| HUD | RUNWAY | DECLUTTER |
| HUD | RUNWAY | ON |
| HUD | INSERT | |

Table A2. PFD Commands for GVSITE flight test.

| **Display** | **Attribute** | **State** |
|---|---|---|
| PFD | TUNNEL | DECLUTTER |
| PFD | TUNNEL | ON |
| PFD | GHOST | DECLUTTER |
| PFD | GHOST | ON |
| PFD | TRAFFIC | DECLUTTER |
| PFD | TRAFFIC | ON |
| PFD | TRAFFIC | AIR |
| PFD | TRAFFIC | SURFACE |
| PFD | FOV | UNITY |
| PFD | FOV | 30 |
| PFD | FOV | 60 |
| PFD | FOV | 90 |
| PFD | FLIR | DECLUTTER |
| PFD | FLIR | ON |
| PFD | NOTAM | DECLUTTER |
| PFD | NOTAM | ON |
| PFD | CHANNEL | DECLUTTER |
| PFD | CHANNEL | BOTTOM |
| PFD | CHANNEL | TOP |

Table A3. NAV Commands for GVSITE flight test.

| **Display** | **Attribute** | **State** |
|---|---|---|
| NAV | OWNSHIP | CENTER |
| NAV | OWNSHIP | NORMAL |
| NAV | TAG | DECLUTTER |
| NAV | TAG | ID |
| NAV | TAG | TYPE |
| NAV | RANGE | 0.5 |
| NAV | RANGE | 1 |
| NAV | RANGE | 1.5 |
| NAV | RANGE | 2 |
| NAV | RANGE | 2.5 |
| NAV | RANGE | 5 |
| NAV | RANGE | 10 |
| NAV | RANGE | 20 |
| NAV | RANGE | 40 |
| NAV | RANGE | 60 |
| NAV | RANGE | 80 |
| NAV | RANGE | 160 |
| NAV | RANGE | DOWN |
| NAV | RANGE | UP |
| NAV | MESSAGE | DECLUTTER |
| NAV | MESSAGE | ON |
| NAV | MESSAGE | DOWN |
| NAV | MESSAGE | UP |
| NAV | FORMAT | AIRPORT |
| NAV | FORMAT | PERSPECTIVE |
| NAV | FORMAT | MAP |
| NAV | FORMAT | SURFACE |
| NAV | FORMAT | PLAN |
| NAV | FORMAT | EXITS |
| NAV | FORMAT | ANIMATE |
| NAV | VSD | DECLUTTER |
| NAV | VSD | ON |
| NAV | EXIT | DOWN |
| NAV | EXIT | UP |
| NAV | CLEARANCE | |
| NAV | DIRECTOR | |
| NAV | MAP | |
| NAV | CONFORMAL | |
| NAV | ALIGNMENT | |
| NAV | BORE SIGHT | |

Table A4. EFB Commands for GVSITE flight test.

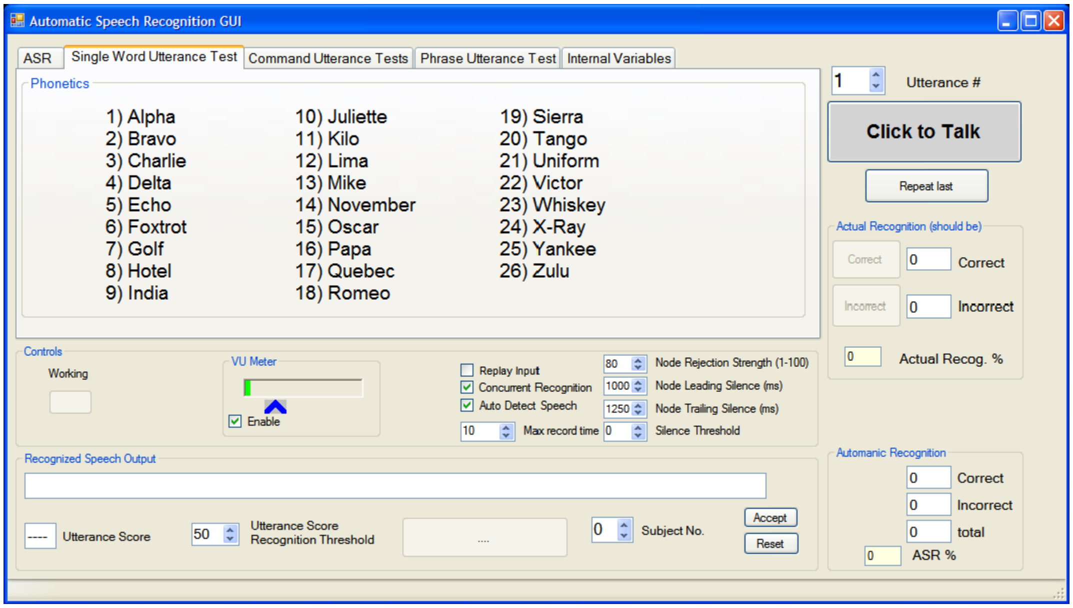
| **Display** | **Attribute** | **State** |
|---|---|---|
| EPAD | OWNSHIP | CENTER |
| EPAD | OWNSHIP | NORMAL |
| EPAD | TAG | DECLUTTER |
| EPAD | TAG | ID |
| EPAD | TAG | TYPE |
| EPAD | RANGE | 0.5 |
| EPAD | RANGE | 1 |
| EPAD | RANGE | 1.5 |
| EPAD | RANGE | 2 |
| EPAD | RANGE | 2.5 |
| EPAD | RANGE | 5 |
| EPAD | RANGE | DOWN |
| EPAD | RANGE | UP |
| EPAD | MESSAGE | DECLUTTER |
| EPAD | MESSAGE | ON |
| EPAD | MESSAGE | DOWN |
| EPAD | MESSAGE | UP |
| EPAD | FORMAT | AIRPORT |
| EPAD | FORMAT | SURFACE |
| EPAD | FORMAT | EXITS |
| EPAD | EXIT | DOWN |
| EPAD | EXIT | UP |

# Appendix B

## Custom Speech Application Software

Evaluation software was developed using Microsoft Visual Studio 2005 development environment using the C# (pronounced 'see sharp') language and the Fonix C# Application Programmers Interface (API). Approximately 2,000 lines of code were written for this study.

Screen shots of the three segments used in the study are shown below.



Phonetics

1) Alpha
2) Bravo
3) Charlie
4) Delta
5) Echo
6) Foxtrot
7) Golf
8) Hotel
9) India
10) Juliette
11) Kilo
12) Lima
13) Mike
14) November
15) Oscar
16) Papa
17) Quebec
18) Romeo
19) Sierra
20) Tango
21) Uniform
22) Victor
23) Whiskey
24) X-Ray
25) Yankee
26) Zulu

1 Utterance #

0 Correct
0 Incorrect
0 Actual Recog. %

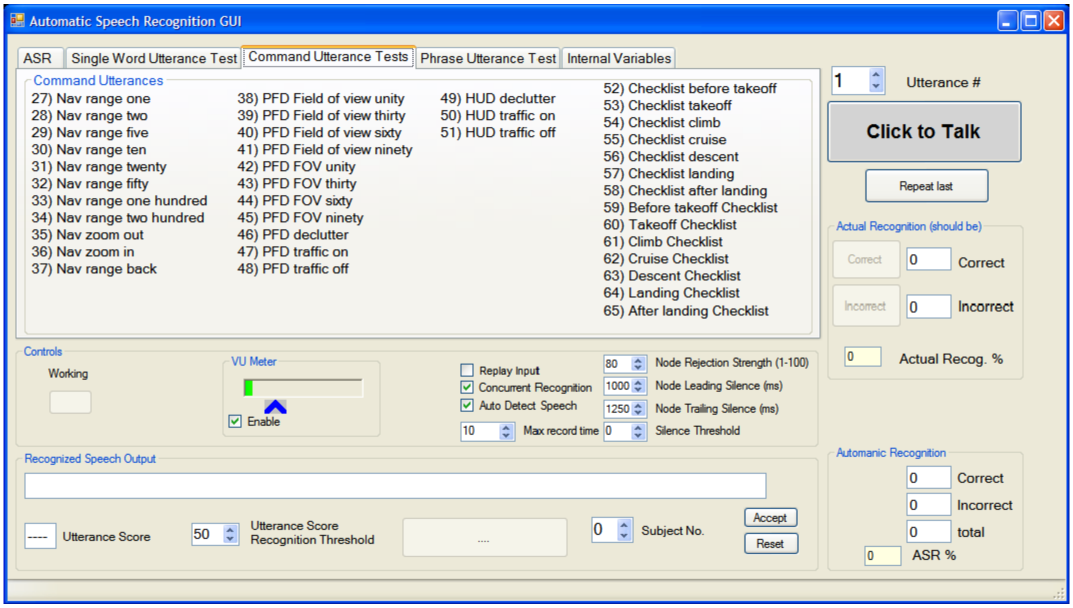
☑ Enable

☐ Replay Input
☑ Concurrent Recognition
☑ Auto Detect Speech
10 Max record time
80 Node Rejection Strength (1-100)
1000 Node Leading Silence (ms)
1250 Node Trailing Silence (ms)
0 Silence Threshold

---- Utterance Score
50 Utterance Score Recognition Threshold
....
0 Subject No.

0 Correct
0 Incorrect
0 total
0 ASR %

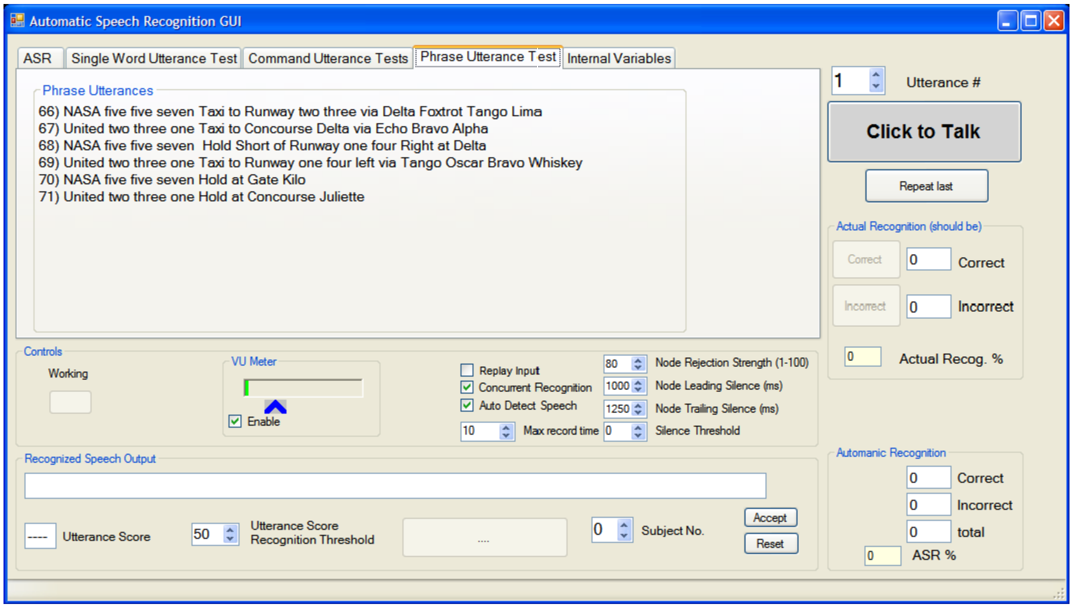
Figure B1. Single word utterance segment screen image.



27) Nav range one
28) Nav range two
29) Nav range five
30) Nav range ten
31) Nav range twenty
32) Nav range fifty
33) Nav range one hundred
34) Nav range two hundred
35) Nav zoom out
36) Nav zoom in
37) Nav range back
38) PFD Field of view unity
39) PFD Field of view thirty
40) PFD Field of view sixty
41) PFD Field of view ninety
42) PFD FOV unity
43) PFD FOV thirty
44) PFD FOV sixty
45) PFD FOV ninety
46) PFD declutter
47) PFD traffic on
48) PFD traffic off
49) HUD declutter
50) HUD traffic on
51) HUD traffic off
52) Checklist before takeoff
53) Checklist takeoff
54) Checklist climb
55) Checklist cruise
56) Checklist descent
57) Checklist landing
58) Checklist after landing
59) Before takeoff Checklist
60) Takeoff Checklist
61) Climb Checklist
62) Cruise Checklist
63) Descent Checklist
64) Landing Checklist
65) After landing Checklist

1 Utterance #
0 Correct
0 Incorrect
0 Actual Recog. %
☐ Replay Input
☑ Concurrent Recognition
☑ Auto Detect Speech
10 Max record time
80 Node Rejection Strength (1-100)
1000 Node Leading Silence (ms)
1250 Node Trailing Silence (ms)
0 Silence Threshold
☑ Enable
---- Utterance Score
50 Utterance Score Recognition Threshold
0 Subject No.
0 Correct
0 Incorrect
0 total
0 ASR %

Figure B2. Short (command) phrase utterance segment screen image.



66) NASA five five seven Taxi to Runway two three via Delta Foxtrot Tango Lima
67) United two three one Taxi to Concourse Delta via Echo Bravo Alpha
68) NASA five five seven Hold Short of Runway one four Right at Delta
69) United two three one Taxi to Runway one four left via Tango Oscar Bravo Whiskey
70) NASA five five seven Hold at Gate Kilo
71) United two three one Hold at Concourse Juliette

1 Utterance #
0 Correct
0 Incorrect
0 Actual Recog. %
☐ Replay Input
☑ Concurrent Recognition
☑ Auto Detect Speech
10 Max record time
80 Node Rejection Strength (1-100)
1000 Node Leading Silence (ms)
1250 Node Trailing Silence (ms)
0 Silence Threshold
☑ Enable
---- Utterance Score
50 Utterance Score Recognition Threshold
0 Subject No.
0 Correct
0 Incorrect
0 total
0 ASR %

Figure B3. Long phrase utterance segment screen image.

# Appendix C

## Speaker-independent engine

A speech application is organized into nodes, which represents the vocabulary and other recognition settings used during speech recognition. The Software Development Kit (SDK) supports word-spotting and grammars nodes. No training is required.

The following diagram contains a speech utterance and identifies the node audio attributes. These node attributes determine how the speech detector frames the utterance before sending it to the recognizer.

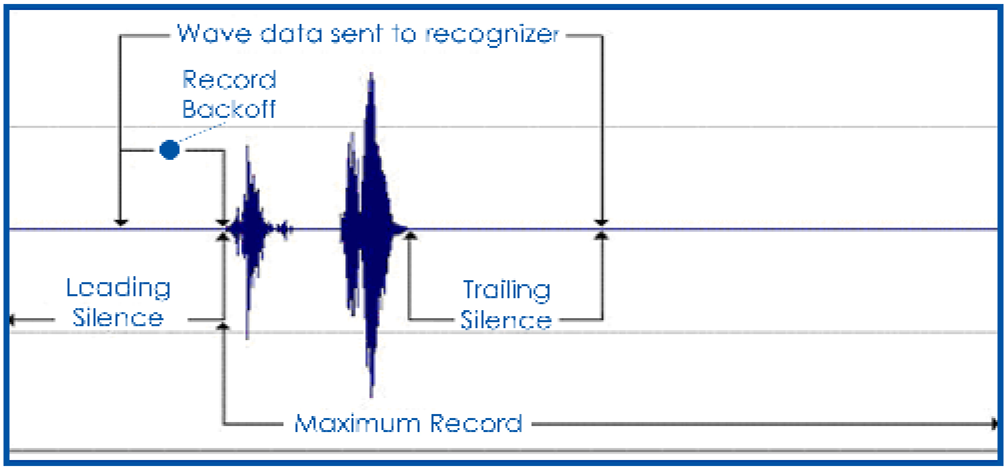


Figure C1. Speech waveform and attributes.

The speaker-independent SRS engine performed the following operations:

1. Audio collection - Raw audio data collected from an input source, such as a microphone. The audio data is sent to the Audio Processing component.
2. Audio processing - “Frames” the audio input using preset values so that the resultant output only includes the audio data necessary for recognition to occur. This data is sent to the Feature Extraction component
3. Feature extraction - Extracts the frequency components every 10 ms from the processed audio data. The collection of frequency components is sent to the Neural Networks component.
4. Neural networks - Extracts phoneme probability estimates from the frequency components, and sends them to the Continuous Word Decoder. Neural networks are key components of the speech recognition technology.
5. Continuous word decoder - Compares the collection of phoneme probabilities against the dictionary and returns a list of all of the word probabilities it finds, in order from highest probability to lowest.

# Appendix D

## Software Implementation

The grammar structure and the dictionary of candidate words were configured differently for each segment of the study. The Application Programmers Interface (API) utilizes a simple script language to implement the grammar structure. Where "vertical line" represents a logical OR, a "Space" represents a logical AND, [ ] represents optional, and "( )" is for grouping.

The first segment dictionary simply contained the 26 phonetics.

```
$phonetics  = (Alpha%A | Bravo%B | Charlie%C | Delta%D | Echo%E |
Foxtrot%F | Golf%G | Hotel%H | India%I | Juliette%J |
Kilo%K | Lima%L | Mike%M | November%N | Oscar%O | Papa%P |
Quebec%Q | Romeo%R | Sierria%S | Tango%T | Uniform%U |
Victor%V | Wiskey%W | X-Ray%X | Yankee%Y | Zulu%Z);

$grammar = $phonetics;
```

The second, (short command utterance) segment contained the following structure:

```
$navcommand = NAV (declutter |
 (zoom (in | out)) |
 ((range ( back |  one%1 | two%2 | five%5| ten%10 |
twenty%20 | fifty%50 | one-hundred%100 |
two-hundred%200)))) ;

$pfdcommand = PFD (declutter |
 (traffic (on | off)) |
 (((Field of view) | (F O V))
 (unity | thirty%30 | sixty%60 | ninety%90)));

$hudcommand = HUD (declutter | (traffic (on | off)));

$chklstcommand = [Checklist] ((Before Takeoff) |
Takeoff | Climb | Cruise | Descent |
Landing | (After Landing)) [Checklist];

$grammar = $pfdcommand | $navcommand | $hudcommand | $chklstcommand;
```

The third segment, (long phrase utterance) contained the following structure:

```
$command = (Hold | (Hold at) | (Hold Short) | (Hold Short Of) | Taxi);

$modifiers  = (To|At);
```

```
$dest = (Ramp | Gate | Concourse | (Runway One%1 Four%4 Left%Lt)|
   (Runway One%1 Four%4 Right%Rt)| (Runway two%2 three%3) |
   (Runway one%1 six%6));
$modifiers2 = (via|at);

$grammar = $callsign $command [$modifiers] $dest [$phonetics]
 [$modifiers2] {$phonetics};
```

# Appendix E

## Software controls definitions

**Utterance Score** an integer (0-100), given by the SRS algorithm representing the confidence in the recognition of the last utterance.

**Utterance Score Recognition Threshold** an integer (0 to 100) that is compared with the Utterance Score. If the Utterance Score is greater or equal to the threshold, the utterance is deemed recognized. This threshold was set to 50 for the entire study.

**Node Rejection Strength** an integer (0-100) setting used in the SRS algorithm to set a threshold for recognizing and/or rejecting out-of-vocabulary words. Raising the value of this attribute makes it harder to recognize words, increasing the number of words that get rejected. Lowering this value makes it easier to recognize words, but increases the likelihood that out-of-vocabulary words might be accepted. This value was set to 80 for the entire study.

**Node Leading Silence** an integer (-1 to 10000 milliseconds) setting used in the SRS algorithm. Leading silence is measured between record start and speech detection. Recording stops if the leading silence time lapse before speech is detected.

**Node Tailing Silence** an integer (-1 to 2000 milliseconds) setting used in the SRS algorithm that represents the maximum length of silence the speech detector waits before determining that speech has ended. This allows for natural pauses in speech, setting a lower value results in a faster return of recognition results. Trailing silence begins after speech detection stops. Recording stops when the trailing silence time is reached. This value was set to 1250 Millisecond (ms) for the entire study.

**Maximum record time** an integer (0 to 120 seconds) that sets max record time after speech is detected. This value was set to 10 seconds for the entire study.

**Silence Threshold** an integer (0 to 500) setting that is designed for high noise environments (speech detected prematurely), but only if recognition is lower than expected. The silence threshold is dynamically set by the program.

**Auto Detect Speech** a discrete setting that turns speech detect on or off during audio collection. If auto speech detect is off, the entire utterance is sent to the recognizer. If speech detect is on, only detected speech (speech, and trailing silence) is sent to the recognizer. This was set to ON for the entire study.

**Concurrent Recognition** a discrete setting that allows concurrent recognition and audio acquisition. The processor must be fast enough to perform speech recognition during audio collection. This was set to ON for the entire study.

**Record Back-off** the time before speech detection. It is included in the data sent to the recognizer. Record back-off can prevent clipping at the beginning of an utterance and was left at the default 250 ms for the entire study.

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704–0188*

**1. REPORT DATE** *(DD-MM-YYYY)*
01-08-2016

**2. REPORT TYPE**
Technical Memorandum

**3. DATES COVERED** *(From - To)*

**4. TITLE AND SUBTITLE**
Performance Evaluation of Speech Recognition Systems as a Next-Generation Pilot-Vehicle Interface Technology

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Arthur III, Jarvis (Trey), J.; Shelton, Kevin J.; Prinzel III, Lawrence J.; Bailey, Randall E.

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**
609866.02.07.07.02

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
NASA Langley Research Center
Hampton, Virginia 23681-2199

**8. PERFORMING ORGANIZATION REPORT NUMBER**
L–20682

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
National Aeronautics and Space Administration
Washington, DC 20546-0001

**10. SPONSOR/MONITOR'S ACRONYM(S)**
NASA

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
NASA-TM–2016–219329

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
Unclassified-Unlimited
Subject Category 03
Availability: NASA STI Program (757) 864-9658

**13. SUPPLEMENTARY NOTES**
An electronic version can be found at http://ntrs.nasa.gov.

**14. ABSTRACT**

During the flight trials known as Gulfstream-V Synthetic Vision Systems Integrated Technology Evaluation (GV-SITE), a Speech Recognition System (SRS) was used by the evaluation pilots. The SRS system was intended to be an intuitive interface for display control (rather than knobs, buttons, etc). This paper describes the performance of the current "state of the art" Speech Recognition System (SRS). The commercially available technology was evaluated as an application for possible inclusion in commercial aircraft flight decks as a crew-to-vehicle interface. Specifically, the technology is to be used as an interface from aircrew to the onboard displays, controls, and flight management tasks. A flight test of a SRS as well as a laboratory test was conducted.

**15. SUBJECT TERMS**
speech recognition, flight test

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | | | STI Information Desk (`help@sti.nasa.gov`) |
| | | | | | **19b. TELEPHONE NUMBER** ***(Include area code)*** |
| U | U | U | UU | 52 | (757) 864-9658 |

**Standard Form 298 (Rev. 8/98)**
Prescribed by ANSI Std. Z39.18