NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

TESTING TEMPLATE AND TESTING CONCEPT OF OPERATIONS FOR SPEAKER AUTHENTICATION TECHNOLOGY

by

Marek M. Sipko

September 2006

Thesis Advisor: James F. Ehlert
Second Reader: Pat Sankar

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** September 2006	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE: Testing Template and Testing Concept of Operations for Speaker Authentication Technology			**5. FUNDING NUMBERS**
6. AUTHOR(S) Marek M. Sipko			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

This thesis documents the findings of developing a generic testing template and supporting concept of operations for speaker verification technology as part of the Iraqi Enrollment via Voice Authentication Project (IEVAP). The IEVAP is an Office of the Secretary of Defense sponsored research project commissioned to study the feasibility of speaker verification technology in support of the Global War on Terrorism security requirements. The intent of this project is to contribute toward the future employment of speech technologies in a variety of coalition military operations by developing a pilot proof-of-concept system that integrates speaker verification and automated speech recognition technology into a mobile platform to enhance warfighting capabilities. In this phase of the IEVAP, NPS developed a generic testing template and supporting concept of operations for speaker authentication technology. The intent of this project was to contribute toward the future employment of speech technologies in a variety of coalition military operations by developing a testing template along with a concept of operations to conduct such testing.

14. SUBJECT TERMS voice recognition, speaker authentication, speaker verification, automated speech recognition technology, voice recognition testing template			**15. NUMBER OF PAGES** 119
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

TESTING TEMPLATE AND TESTING CONCEPT OF OPERATIONS FOR SPEAKER AUTHENTICATION TECHNOLOGY

Marek M. Sipko
Major, United States Marine Corps
MBA, National University, 1993

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
September 2006

Author: Marek M. Sipko

Approved by: James F. Ehlert
Thesis Advisor

Pat Sankar
Second Reader

Dan Boger
Chairman
Department of Information Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This thesis documents the findings of developing a generic testing template and supporting concept of operations for speaker verification technology as part of the Iraqi Enrollment via Voice Authentication Project (IEVAP). The IEVAP is an Office of the Secretary of Defense sponsored research project commissioned to study the feasibility of speaker verification technology in support of the Global War on Terrorism security requirements. In this phase of the IEVAP, NPS developed a generic testing template and testing concept of operations for speaker authentication technology. The intent of this project was to contribute toward the future employment of speech technologies in a variety of coalition military operations by developing a testing template along with a concept of operations to conduct such testing.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to thank my dear wife Dorota for her encouragement and help to achieve all my goals at the Naval Postgraduate School.

I would also like to thank my thesis advisors, Mr. Jim Ehlert and Dr. Pat Sankar, for their guidance and encouragement throughout the past year while I worked on this project.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ASR	Automated Speech Recognition
BCCF	Baghdad Central Correctional Facility
BFC	Biometric Fusion Center
CONOPS	Concept of Operations
COTS	Commercial Off The Shelf
CTI	Computer Telephony Integration
DET	Detection Error Tradeoff
DOD	Department of Defense
DTMF	Detecting Dual Tone Modulation Frequency
EER	Equal Error Rate
EIS	Enterprise Information Systems
FAR	False Accept Rate
FMR	False Match Rate
FNMR	False Non-Match Rate
FRR	False Reject Rate
GUI	Graphical User Interface
GWOT	Global War on Terrorism
IEVAP	Iraqi Enrollment via Voice Authentication Project
ISN	Internment Serial Number
IZ	International Zone
JVM	Java Virtual Machine
MIT	Massachusetts Institute of Technology
NAE	Nuance Application Environment
NCA	Nuance Caller Authentication
NCS	Nuance Call Steering
NL	Natural Language
NPS	Naval Postgraduate School
NVP	Nuance Voice Platform
OSD	Office of the Secretary of Defense
PIN	Personal Identification Number
POC	Proof of Concept
ROC	Receiver Operating Characteristics
ROI	Return on Investment
SIP	Session Initiation Protocol
SLM	Statistical Language Model
SOP	Standard Operations Procedures
TTS	Text to Speech
VA	Voice Authentication
VLV	Variable Length Verification
VOIP	Voice over Internet Protocol
VUI	Voice User Interface
VV	Voice Verification

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. OVERVIEW

This research documents the findings of developing a generic testing template and supporting concept of operations for speaker verification technology as part of the Iraqi Enrollment via Voice Authentication Project (IEVAP). The IEVAP is an Office of the Secretary of Defense (OSD) sponsored research project that studies the feasibility of speaker verification and speech recognition technology in support of the Global War on Terrorism (GWOT) security requirements. The IEVAP is organized into several project phases that are intended to take the POC system from concept development to operational testing in Iraq [6]:

- **Phase 1.—**Pilot menu-driven laptop system and demonstration that voice authentication technology can work with sufficient accuracy.
  - **Phase 1A.—**Develop and demonstrate a bilingual voice-activated, menu-driven phone system in English and Arabic.
  - **Phase 1B.—**Test and demonstrate speaker verification technology in English.
  - **Phase 1C.—**Test and demonstrate speaker verification technology in Iraqi-Arabic.
- **Phase 2.—**Detailed development of enrollment applications.
- **Phase 3.—**Preparation of systems/applications for deployment.
- **Phase 4.—**Deployment.
- **Phase 5.—**Operational testing in Iraq.
- **Phase 6.—**Broader deployment decision.

In the spring of 2005, the Naval Postgraduate School (NPS) developed and successfully tested Phases 1A and 1B of the IEVAP. In Phase 1A, NPS developed a bilingual (English and Jordanian-Arabic) speech application that demonstrates the

viability of speaker verification technology [6]. During Phase 1B, NPS conducted a test to assess the accuracy claim of Nuance's package speaker verification system application, Nuance Caller Authentication 1.0 (for North American English). The NPS test consisted of 68 speaker enrollments and 411 speaker verification attempts. Upon completion of the test, NPS conducted a single data-point analysis yielding a system accuracy of 95.87% [6]. This thesis expands prior Phases 1A and 1B findings by discussing specific areas of voice recognition technology to include discussion on Markov chains. Additionally, the Resource Provisioning Guidelines section discusses estimation criteria used for resource determination needed for a voice recognition system deployment to include discussion on the Erlang-B formula. The Erlang-B formula gives the probability of blocking in a system where a large population makes use of a finite number of resources. Testing and concept of operations for testing templates are this thesis's specific deliverables providing ready made references for future voice authentication system performance tests.

B. RESEARCH QUESTIONS

- Is it possible to successfully develop a generic testing template in support of a reliable and user friendly speaker verification technology?
- Is it possible to develop and demonstrate a generic concept of operations (CONOPS) for testing a reliable and user friendly speaker verification technology?

C. SCOPE OF THESIS

This thesis focuses on developing a template for a generic voice authentication test plan and a concept of operations for such a testing. Additional research and development will be required to transition this speaker verification technology to an operational system.

The value of this research includes:

- Selecting the most appropriate hardware, software, and peripherals for a mobile demonstration kit (laptop, voice input devices, etc.) for implementing speaker verification and ASR technologies.
- Having available a generic voice authentication test plan for any future testing.
- Having available a generic concept of operations (CONOPS) that could be utilized prior and during any future testing.

D. RESEARCH METHODOLOGY

This research will use the qualitative approach for data collection and analysis. This research will consist of an analysis of the speaker verification technology and associated suite of equipment and devices through literature reviews, interviews, prior and concurrently conducted NPS tests and demonstrations.

E. THESIS ORGANIZATION

Chapter II contains the requisite background information that supports this research. This information includes a description of basics of voice recognition technology to include discussion of capabilities and limitations of the voice recognition component technologies. Chapter III describes provisioning guidelines used when designing voice recognition and authentication systems. Chapter IV provides the description of the template for a generic voice authentication test plan. Chapter V contains the description of the template for a generic system concept of operations. Finally, Chapter VI describes the conclusion drawn from the results of the research and provides recommendations and suggestions for further study.

THIS PAGE INTENTIONALLY LEFT BLANK

II. VOICE RECOGNITION TECHNOLOGY

A. PRIOR TECHNOLOGY

Interactive voice response, called IVR, is in wide use today and has been around for a number of years. With this technology, a caller enters information in response to prompts by pressing their touchtone keypad. The tones created by this action (called dial tone modulated frequency or DTMF) are recognized by the system, allowing callers to interact with a computer application over the phone. IVR, unfortunately, has a number of limitations. While fine for simple tasks with few menus of choices, IVR can be extremely inefficient with anything more complex, requiring tedious levels of hierarchical menus. Furthermore, IVR runs into severe complications when dealing with long lists of choices (such as city or stock lists). Also, applications that require hands-free operation do not lend themselves to IVR. For instance, a person driving a vehicle cannot safely navigate a cell phone keypad while weaving through rush hour traffic [10].

Another technology created early in the quest for automated speech recognition (ASR) was template matching. This was an extremely inefficient and limited way of getting a computer to recognize speech. One or more templates representing the waveform of a spoken word is created for each word meant to be recognized. Both the storage requirements and the template matching algorithms restrict the system to small vocabularies. Furthermore, continuous speech recognition was not possible because the comparison mechanism only allowed recognition of individual words and not phrases. This type of recognition was not speaker-independent, as the stored templates reflect only a single way to utter the word [10].

Over the past decade, speaker recognition technology has made its debut in several commercial products. The specific recognition task addressed in commercial systems is that of verification or detection (determining whether an unknown voice is from a particular enrolled speaker) rather than identification (associating an unknown voice with one from a set of enrolled speakers). Most deployed applications are based on scenarios with cooperative users speaking fixed digit string passwords or repeating prompted phrases from a small vocabulary. These generally employ what is known as

text-dependent, or text-constrained, systems. Such constraints are quite reasonable and can greatly improve the accuracy of a system; however, there are cases when such constraints can be cumbersome or impossible to enforce. An example of this is background verification where a speaker is verified behind the scene as he/she conducts some other speech interactions. For cases like this, a more flexible recognition system is needed, one that is able to operate without explicit user cooperation and independent of the spoken utterance (called text-independent mode). This thesis focuses on the technologies behind these text-independent speaker verification systems. A speaker verification system is composed of two distinct phases—a training phase and a test phase. Each phase can be seen as a succession of independent modules [10]. Figure 1 shows a modular representation of the training phase of a speaker verification system.

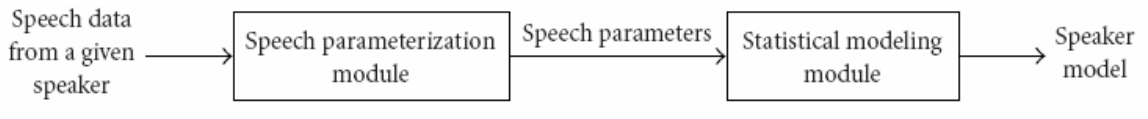

Figure 1. Modular Representation of the Training Phase of a Speaker Verification System [From Ref. 10]

The first step consists of extracting parameters from the speech signal to obtain a representation suitable for statistical modeling, as such models are extensively used in most state-of-the-art speaker verification systems. The second step consists of obtaining a statistical model from the parameters. This training scheme is also applied to the training of a background model. Figure 2 shows a modular representation of the test phase of a speaker verification system.

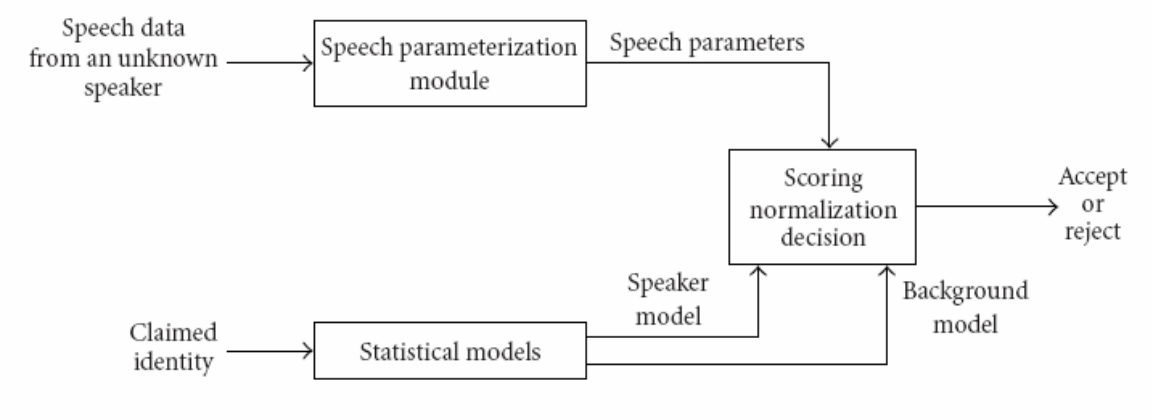

Figure 2. Modular Representation of the Test Phase of a Speaker Verification System [From Ref. 10]

The entries of the system are a claimed identity and the speech samples are pronounced by an unknown speaker. The purpose of a speaker verification system is to verify whether the speech samples correspond to the claimed identity. First, speech parameters are extracted from the speech signal using exactly the same module as for the training phase. Then, the speaker model corresponding to the claimed identity and a background model are extracted from the set of statistical models calculated during the training phase. Lastly, using the speech parameters extracted and the two statistical models, the last module computes some scores, normalizes them, and makes an acceptance or a rejection decision. This normalization step requires some score distributions to be estimated during the training phase or/and the test phase. A speaker verification system can be text-dependent or text-independent. In the former case, there is some constraint on the type of utterance that users of the system can pronounce (for instance, a fixed password or certain words in any order). In the latter case, users can say whatever they want. This thesis describes state-of-the-art text-independent speaker verification systems [10].

B. NATURAL LANGUAGE ASR

Natural language (NL) ASRs have become a standard for speech recognition. Natural language gives developers the freedom to work with large, extensible grammars and is not limited to a few words or tones. Callers can also speak continuously to systems, and they do not have to stop after each spoken word for recognition. An

additional benefit of today's ASRs is greater speaker independence. For instance, systems from Nuance Corporation, a world leader in the deployment of voice interfaces, can recognize and distinguish between the widely different ways in which people speak. Best of all, natural language ASR allows users to speak in complete sentences and phrases and still be recognized [10].

There are many ways in which ASR applications can improve upon interacting with live operators. People are often hesitant to request "live" human operators to do repetitive tasks. For example, when requesting stock quotes, very few people are willing to call a human operator repeatedly just to get a few current prices. But, when they know they are interacting with a computer, they are much more inclined to make multiple queries. Machines are also well suited to recognizing long alpha (letter) strings. People find it difficult to remember these strings, while machines can quickly recognize and use them. An example of this can be found in the UPS, or US Postal Service package tracking system [10].

ASR voice technology enables the applications that are driving growth in three major markets: enterprise, telecommunications and the Internet. Enterprises like brokerages, banks, airlines and retailers have applications in stock quotes and trading, travel planning and shopping. Wireless and wireline telecommunications carriers have voice-enabled applications like dialing, directory assistance, and access to voicemail and email messages. Internet applications are also emerging with voice-commerce applications, information delivery through voice portals (e.g., movie directories, driving directions and traffic updates) and web content access over the phone. Additionally, both the government and the military offer a plethora of ASR employment opportunities. Voice verification and voice authentication in support of force protection and security operations are examples of such applications.

C. SPEECH DEFINITION

Sound concerns variations of air pressure over time. These "waves" of air pressure impinge on ears, causing information to be transmitted to brains. Typically, the

sound waves from speech will have characteristic signatures, called waveforms, as seen below in the graph of air pressure vs. time for the utterance, “May 14th, 1998.”

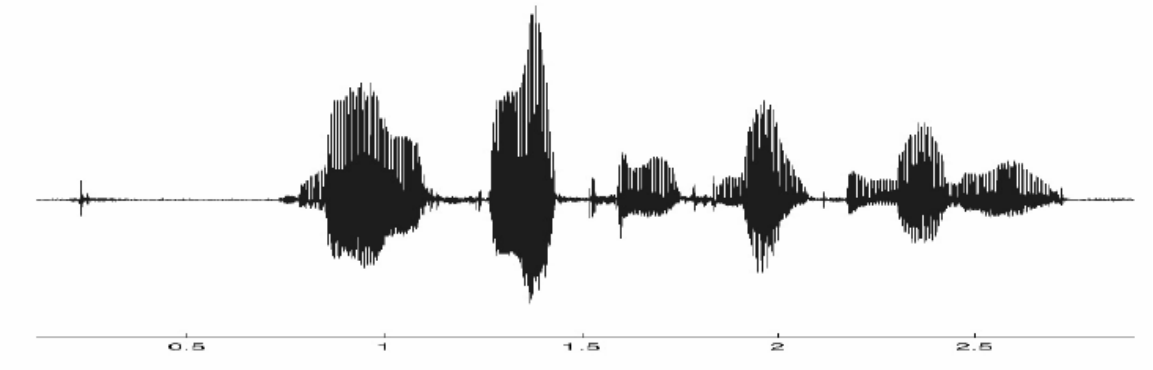

Figure 3. Air Pressure vs. Time Example [After Ref. 10]

Voice recognition/authentication software performs a mathematical transformation on these waveforms to give information on the sound intensity in a set of frequency bands for a particular moment in time (usually in 10 msec samples). Within that 10 msec sample, the energy levels in the specific bands will be representative of a particular part of speech called a “feature.” For instance, in Figure 4, one can see what the ‘m’ looks like in “May,” the ‘f’ in “for” and so on. The computer can then use these signatures to recognize the words when comparing them to predefined mathematical models and then draw out the meaning, or semantic content, of what was said [10].

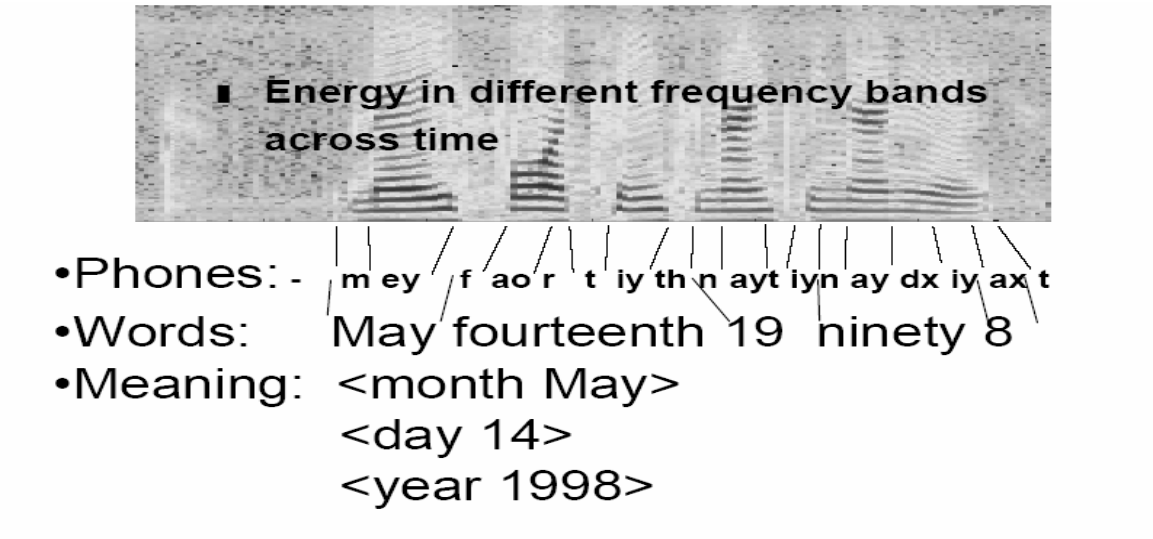

Figure 4. Speech Sample [From Ref. 10]

D. THE GRAMMAR

The first element making up the recognition package is the grammar. In the grammar, a file is created defining all of the allowable user utterances permitted for a given task in a given application. For example, an application calls for a grammar defining basic greetings functions. In Nuance's Grammar Specification Language (GSL), the grammar would look something like Figure 5. The grammar has a name which appears at the top and starts with a capital letter (in this case, "Sentence" is the grammar name). The allowed phrases are then specified under this name and can be stored in a simple text file for compiling.

Another important aspect of grammar is the ability of a developer to specify how the recognizer is to associate meaning (semantic interpretation) to the recognized speech. This is done by assigning a value to a variable, called a "slot," when a given phrase is recognized. In the sample below, the grammar slot "greet" is filled with the value "hello" or "goodbye" depending on which phrase is recognized [10].

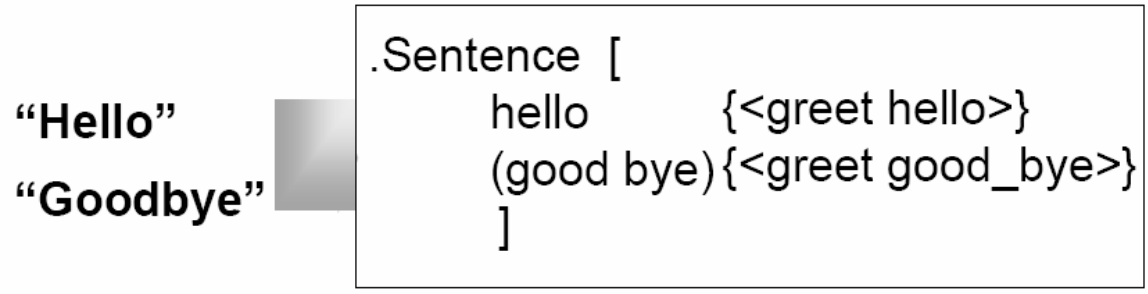

Figure 5. Grammar Specification Language (GSL): Speech Representation [From Ref. 10]

E. THE DICTIONARY

The second major component needed for recognition is the dictionary. The dictionary defines the phonetic pronunciation of the words contained in the grammar. This is done by assigning one or more strings of the appropriate phonetic units, called "phonemes," to each word found in the grammar. Every language has a finite number of sounds represented by phonemes (English has roughly 41) [10].

The phoneme can be defined as "the smallest meaningful psychological unit of sound." The phoneme has mental, physiological, and physical substance: human brains process the sounds; the sounds are produced by the human speech organs; and the sounds are physical entities that can be recorded and measured.

For an example of phonemes, consider the English words *pat* and *sat*, which appear to differ only in their initial consonants. This difference, known as contrastiveness or opposition, is sufficient to distinguish these words, and therefore, the "P" and "S" sounds are said to be different phonemes in English.

F. ACOUSTIC MODELS

The recognizer contains mathematical "acoustic" models for each spoken phoneme taken in the context of the phonemes that directly precede and follow it. It takes the supplied grammar and dictionary and forms entire models for each possible phrase based on these acoustic models. So, in comparison with the old ASR technique, in which a waveform model for an entire word was used for matching the utterance, a much more flexible approach is now used where the individual triphone building blocks are matched to the numerical templates of the acoustic model [10].

G. THE SEARCH SPACE

A search space is created containing each possible set of phrases and pronunciations as allowed by the grammar, dictionary and acoustic models. When a caller speaks, the incoming waveform is transformed into a string of features that are matched against the available paths in the recognizer search space. The recognizer picks the most probable path and returns the corresponding phrase as the recognition result. For instance, there is a simple grammar containing three possible phrases: "one," "two," and "three." The diagram below, courtesy of Nuance Corporation, displays some of the possible probabilistic paths that are allowed, called "Markov Chains." "Markov chains" are used when a caller speaks; the incoming waveform is transformed into a string of features that are matched against the available paths in the recognizer search space. The recognizer picks the most probable path and returns the corresponding phrase as the

recognition result. These probabilistic paths are the "Markov chains" that are used extensively in voice recognition technology. Shown on the bottom are the various sequence and duration of phonemes that can be arranged to correspond to speaking the word "one." The best match of the utterance to a path allowed by the grammar determines the recognition of the utterance [10].

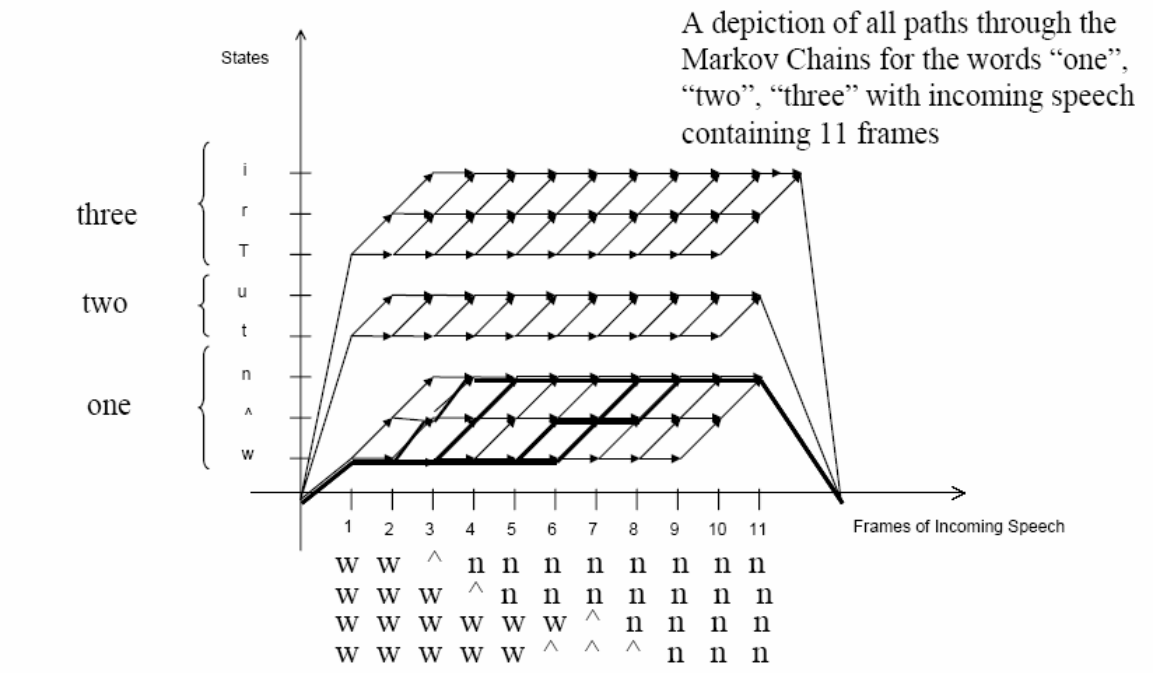

Figure 6. Recognition Search Space Example [From Ref. 10]

H. THE FLOW OF RECOGNITION

The recognition process can be recapped as follows: a caller speaks an utterance which is captured in a waveform that goes through the front-end processing, outputting the speech features (a vector of numbers representing samples of the waveform). The recognizer receives the speech features, along with three other inputs [10]:

- The dictionary has phonetic pronunciations for words in the grammar files (no meaning is drawn from the dictionary).
- The acoustic model set provides a linguistic representation of the expected caller base (British English, American English, Australian English,

Canadian French, German, Mandarin, Latin American Spanish, Japanese, Jordanian Arabic, etc.).

- The recognizer also takes input from the grammar file. From all the input, the recognizer comes up with the word string it hypothesizes with some level of confidence that the speaker uttered.
- The recognizer then chooses the best match of the feature string to a path in the search space (an allowed phrase in the grammar), after which, in the interpretation phase of recognition, the semantic interpretation from the corresponding phrase in the grammar is returned to the application.

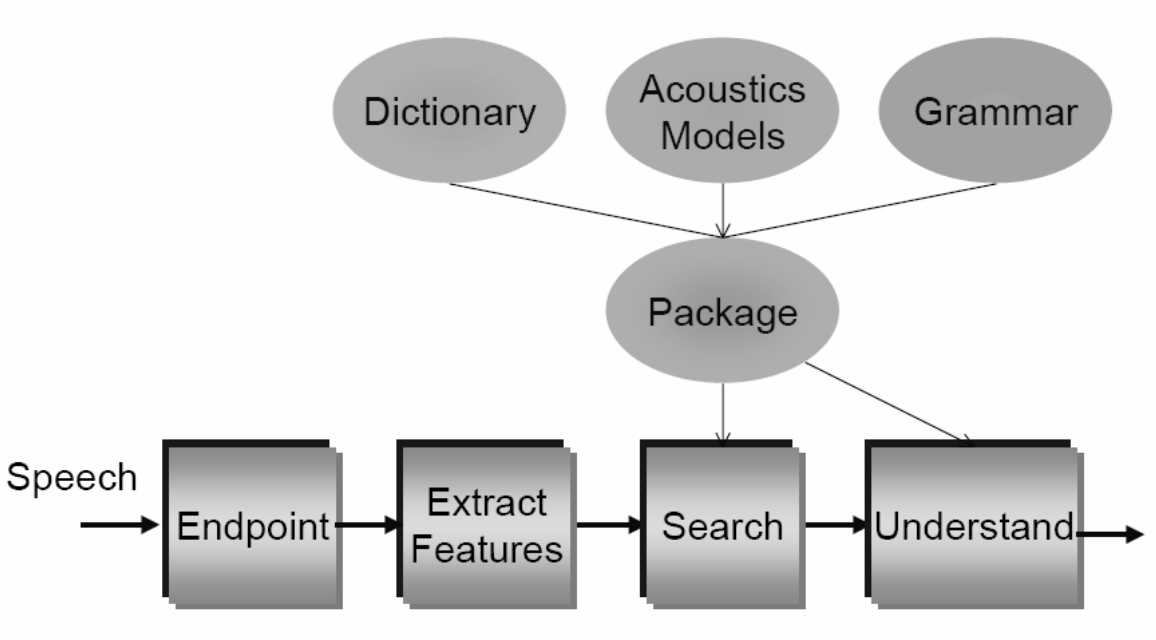

Figure 7. The Flow of Recognition Process [From Ref. 10]

I. RECOGNITION PERFORMANCE

Recognition accuracy is a partial measurement of the performance that a given application is experiencing with customer interactions. Accuracy rates are measured in terms of the number of successful recognitions made by the system for callers speaking phrases allowed by the grammar. This rate is called "CA-in" or "correct accept in-grammar."

There are several benefits from achieving higher levels of accuracy. First, it improves the experience for the caller and makes the system more usable. With higher accuracy, there is less need for callers to repeat themselves (which can be frustrating) or correct errors in recognition. This will have the additional bonus of increasing transaction success rates and lowering the number of "operator" requests. Fewer operators directly translates into higher cost savings. Second, the system design is simplified because developers do not have to "design-in" extra application logic for poor recognition. This reduces the development time and reduces the effort needed to maintain and tune the system. Accurate systems are also more efficient; without the need for corrections or repetitions of information, the average transaction time will decrease. This allows for higher call volumes on a given system, lowering the hardware requirements as compared to a less efficient system (another form of cost savings) [10].

J. CLASSIFYING RECOGNITION EVENTS

Recognition events can be separated into three basic categories: invalid, out-of-grammar (OOG) and in-grammar (IG) [10].

- **Invalid:** Errors in which the recognizer is not sent the proper segment of speech)
- **Out-of-Grammar:** OOG events can be either correctly rejected or falsely accepted (an error). According to Nuance internal testing, typically, a properly designed system should not experience OOG rates exceeding 10%. If so, the tuner should look to the prompts to more carefully direct users into making in-grammar utterances.
- **In-grammar:** IG events can be either correctly accepted (good), falsely accepted as a different allowed in-grammar utterance (an error), or falsely rejected by the recognizer (another error). Accuracy involves minimizing these errors and maximizing the number of correct accepts (CA-in).

K. FACTORS AFFECTING RECOGNITION PERFORMANCE

The quality of speech sent to the recognizer has a large effect on the recognition performance. Speech heard in an environment with lots of background noise reduces accuracy by confusing the recognizer and may cause problems with barge-in (the ability to interrupt prompts in an ASR system). Channel, or transmission path, differences can cause large variations in the quality of the speech sent to the recognizer. For instance, listening to the same sound or sentence coming across a cell phone vs. a regular handset vs. a speaker phone can indicate is the variety of sound quality. The defined pronunciations and the acoustic models may not know what to do with a strongly accented utterance resulting in False Reject (FR) type errors. FR type errors occur when a user is truly trying to pass verification under his/her identity and is rejected as being an imposter.

Grammar capability also directly affects the recognition performance—how broad is the grammar; does it provide good coverage as directed by the prompts; are dynamic lists employed? It is extremely important that developers take great care to produce a broad grammar because the broad grammar will provide the widest possible coverage, thus minimizing Out of Grammar errors. Grammar ambiguity occurs when similar sounding utterances are contained in the grammar leading to recognition. If the recognizer has a large search space (e.g., overly large grammars) intensive load may be placed on the CPU of the machine performing recognition. This may have the effect of increasing latency (the amount of time to return a recognition result after the caller has finished speaking). More complex acoustic models require more processing time during recognition. Furthermore, a less than fully developed acoustic model may result in a lowered ability to match the various pronunciations of the calling population, increasing the error rate [10].

L. THE TUNING PROCESS

The process of optimizing recognition performance is called “tuning.” Typically, to tune an application, the developer needs to launch a pilot phase. In this phase, the system is opened up to a limited population of callers solely for the purpose of gathering data on the system performance [10].

1. Logging Data

The first step in the tuning process is getting data from real callers on the system. This is accomplished by recording user speech made into the system and getting a file containing the system's response, called a "log file." The recorded speech is then sent to a transcription service and transcribed according to a given proprietary software transcription convention.

2. Updating the System

The improved grammars, prompts and parameters are then fed back into the application, where the performance may be monitored to ensure the changes were positive. The tuning process may be performed iteratively if additional cycles prove necessary.

M. PERFORMANCE REPORTING

Typically, when analyzing system performance, a tuner works with performance reports generated from comparing the transcriptions to the logged utterances. In this case, a standard accuracy report has been generated to show the application tuner what the various standard error levels are and IG vs. OOG rates. The tuner may then adjust the grammars, prompts and parameters, and observe the effectiveness via additional reports. The marked improvement in the error rates after tuning should be noted. According to Nuance, the untuned system is actually performing well (error rates from 10-15%). However, tuning brings down the error rates to only a few percent [10].

N. SPEAKER AUTHENTICATION BASICS

There are two parts of voice authentication—Enrollment (or training) and Verification [10].

- **Enrollment—**Whenever new users connect to a verification system, they are required to make an identity claim by providing an identifying phrase (e.g., an account number). The users first must go through a one-time voice enrollment process that provides enough speech samples to allow

the system to "learn" their voice. From these utterances, certain features of the voice and physiology are extracted, creating a "voiceprint" that is stored on a database for later reference during the verification phase. The voiceprint created is not a set of audio samples, but a matrix of numbers that represent the characteristics of the user's voice and vocal tract, and is quite independent of the specific utterances used to create the model.

- **Verification—**For the verification phase, users call the system subsequent to the initial Enrollment and speak an identifying phrase. The recognizer recognizes the utterance and uses it to bring the corresponding caller voiceprint from the database into memory. Next, the verifier generates scores from comparing one or more utterances to the voiceprint and to a "background model" or composite imposter model (made from a combination of many other speakers). Based on the scores and configuration thresholds included in the recognition package, the Verifier determines whether the speaker is who they claim to be.

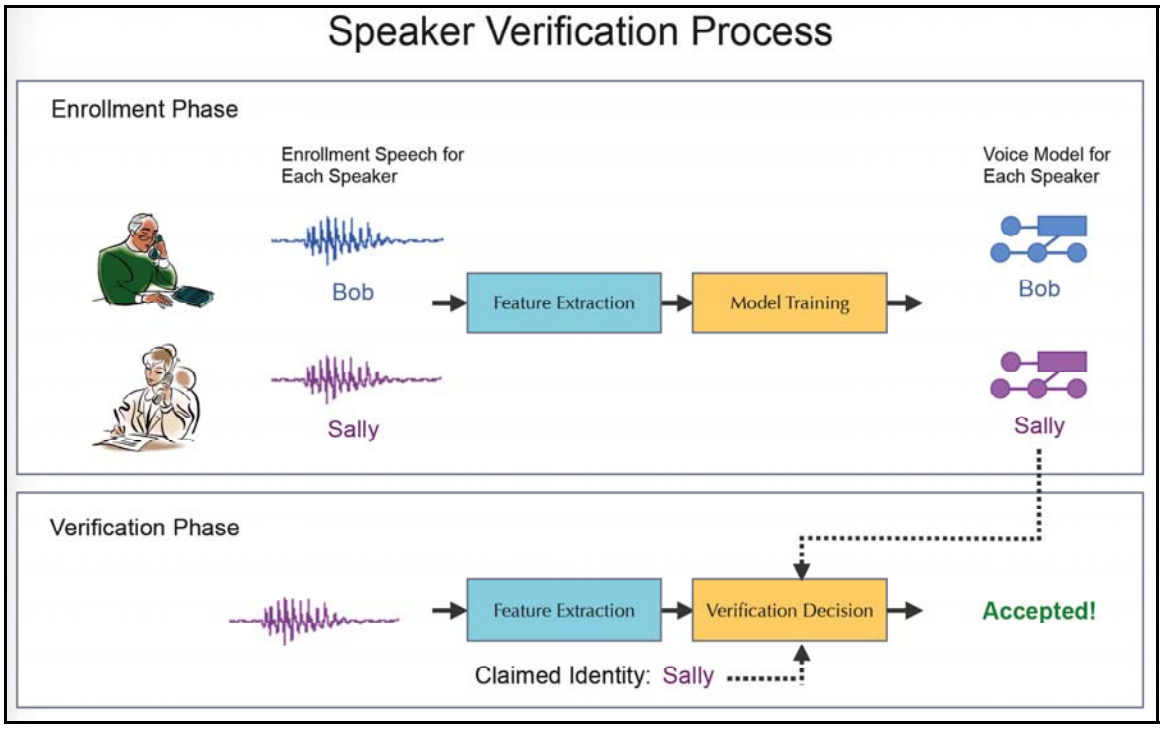

Figure 8. Speaker Authentication Basics Process [From Ref. 10]

O. TUNING THE VERIFIER SECURITY LEVEL

Authentication accuracy uses different metrics than those used to determine recognizer performance. Typically, verifier performance is measured relative to two types of errors: False Accepts (FA) and False Rejects (FR) [6], [10,], [13].

- **False Accepts—**FAs occur when a user tries to break through the verification while claiming to be another user and the system accepts this as true. The user is falsely accepted into the system.
- **False Rejects—**FRs occur when a user is truly trying to pass verification under his/her identity and is rejected as being an imposter.

A tradeoff occurs when the system can be adjusted to allow less FR, causing more FA, and vice versa. This tradeoff can best be seen from the receiver operation curve (ROC), which plots the FR rate vs. the FA rate as shown in Figure 9. When tuning verification performance, one seeks the appropriate FA/FR trade-off that is optimal for the type of system. A high security rate is achieved by adjusting the system to have a lower FA rate. A system with higher convenience can be created by lowering the FR rate. Typically, one can balance between the two and achieve equal error rates (EER) of roughly 0.9% for FR and FA [6], [10], [13].

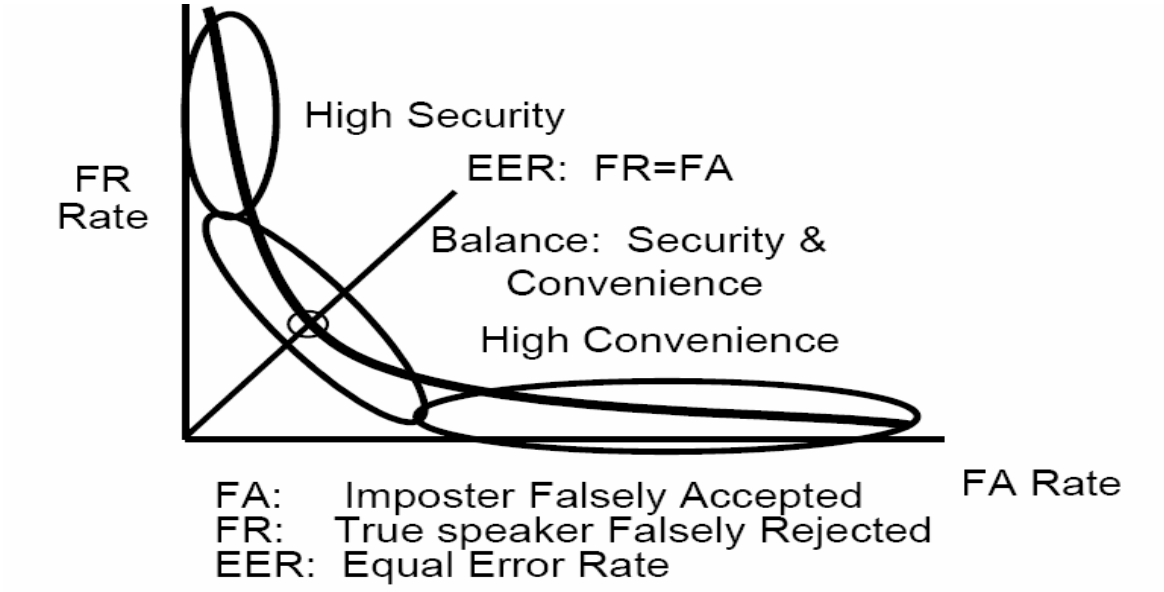

Figure 9. Security/Convenience Tradeoff [From Ref. 10]

P. DATABASES AND VOICEPRINTS

A voiceprint is not a recording of the user's voice; it is a binary file containing a matrix of numbers that reflect physical characteristics of the person's vocal tract as well as behavioral characteristics of the way the person speaks. Having someone's voiceprint would not allow a malicious user to break into their account. When the user enrolls, the verifier calculates voice and physiology features that are built into a voiceprint model. The process cannot be reversed to produce utterances to break into the system, because the impostor does not know the algorithm and parameters used to produce the model (assumes no insider information), and because the model does not encapsulate specific waveforms. The voiceprints are contained in 20KB files that do not grow if they are adapted or improved over time. For instance, Nuance supports Oracle and ODBC-compliant databases for storing the voiceprints. The system developer can write custom database providers if voiceprints need to be stored on some other database. For prototyping and tuning, one may also use file system databases [10].

Q. THE FIVE STEPS PROJECT METHOD OF SPEECH RECOGNITION SYSTEM DEVELOPMENT

As with any new system development project, a well-planned and consistent methodology should be utilized to achieve the best results. With speech-enabled systems, certain aspects of the project development process are specific to speech alone. There are five distinct phases: 1) the requirements analysis, 2) the design, 3) the development, 4) the testing, and 5) the tuning and monitoring [10].

- **Requirements**: The developer must analyze the business, user and application requirements to scope the project.
- **Design**: The developer will lay out the persona and audio design, the dialog design and the application design from the requirements gathered in phase one.
- **Development**: The actual application development takes place. This necessitates producing the audio for the system, developing the grammars, and putting together the code and hardware for the application.

- **Testing**: The developer would thoroughly and entirely run the application. The developer would test and retest all facets of the application to ensure that each feature of the application is working per the requirements' specifications. During the testing phase, it is important that iterative usability testing be carried out to ensure that the application meets the user's needs.
- **Tuning and Monitoring**: When the pilot tuning is carried out, the in-depth recognition performance is analyzed and optimized. After this has been completed, the application is deployed and the system is monitored for additional tuning and to ensure that the performance does not degrade over time. Many of the testing cycles may be performed iteratively to get the most out of the system.

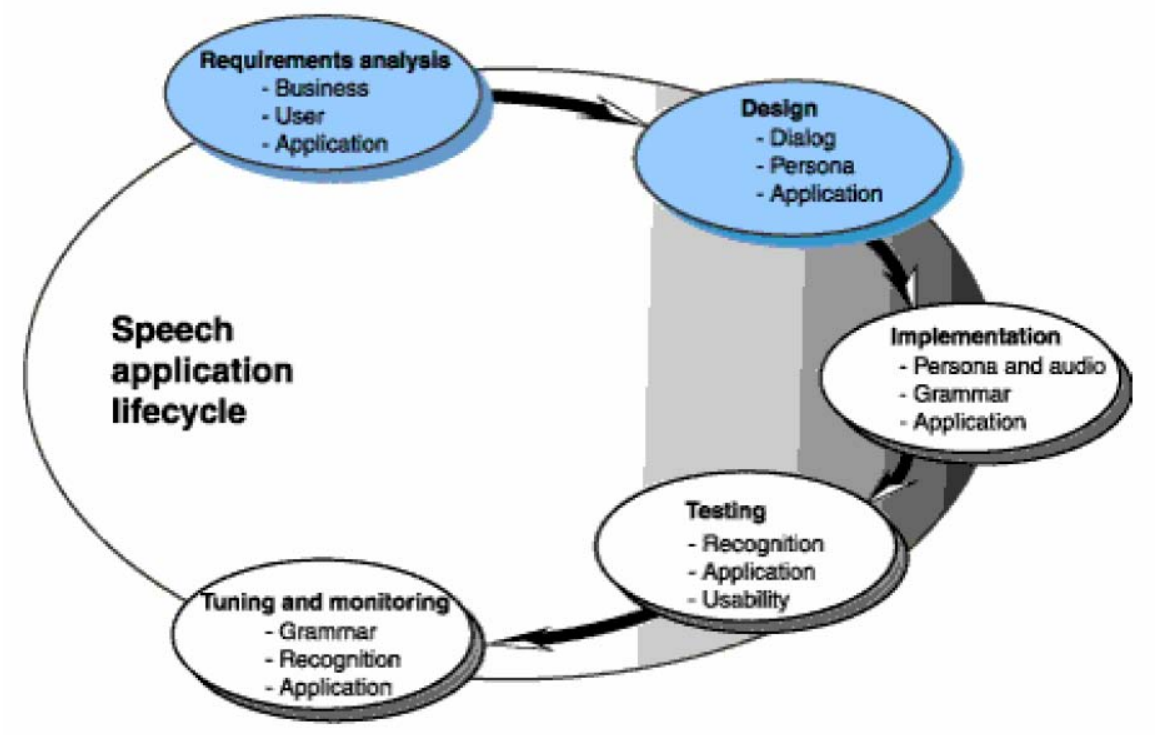

Figure 10. The Five Steps Project Method of Speech Application Lifecycle [From Ref. 10]

R. DESIGN PRINCIPLES

Speech Recognition systems require extraordinarily good design. Dialog design is part science, part evolving art form. Nothing is straightforward, and developers must contend with many trade-offs. The art of dealing with human communication is a very complex field in and of itself. In the end, the user's satisfaction and overall experience is strongly affected by the Dialog Design. With this in mind, a good system should get the job done efficiently and must not confuse or frustrate the caller [10].

The concept of persona is crucial when designing speech recognition systems. The persona defines the caller's relationship with the system. It should be appropriate to the business model of the application it serves.

The Dialog Specification is the key document for a speech system. It contains the complete definition of the dialog from the standpoint of the caller and is the primary tool for communicating with the customer about the application details. It also becomes the basis for the developing and testing the application. The dialog design specification defines every detail of the application speech interface, including [10]:

- The call flow and system logic
- Error handling (incorrect speech, no speech, too much speech, repeated errors, etc.)
- Universals (opting out to an operator, canceling a request, etc.)
- Prompts
- Recognition grammars and return values
- All details for each recognition state.

Ideally, system developers should follow a user-centered design methodology that is incorporated in the Five Steps Project Method. Each aspect of the voice-user interface (VUI) design process focuses on the user's experience. The various steps that form the foundation of proper VUI design are [10]:

- Laying out the requirements
- Performing high-level design of the application

- Detailed dialog/grammar design
- Validation and tuning of the application

S. THE CORE PRINCIPLES OF VUI DESIGN

There are four key principles to adhere to when designing a voice-user interface [10]:

- **Accuracy:** Achieving a high degree of recognition accuracy is a key component for a successful system. A system with a good recognition performance increases user satisfaction.
- **Graceful Error Recovery:** When errors occur, it is important to recover quickly and efficiently. The system must be aware of when there is a problem, and quickly reestablish clarity through effective prompting.
- **Efficiency:** The system needs to deliver just enough information to the caller, at just the right time. Overly long prompts need to be avoided, and barge-ins need to be enabled to speed the user's interaction with the system.
- **Low Cognitive Load:** The end user should not be overloaded with too much information or too complex a task. The developers should remember that a voice interface is significantly different from a graphical one. Users can remember only so much at a time.

All of these principles lead to clarity, making the application enjoyable and easily understandable to the caller.

III. RESOURCE PROVISIONING GUIDELINES

A. BACKGROUND

This chapter outlines how to provision for a minimum hardware and software architecture to ensure the expected level of service, whether the system is running in normal mode or under abnormal circumstances—such as during a software upgrade or while a telephony session service is being automatically restarted. Additionally, a discussion on determining the number of telephone ports (T1 provisioning) needed to handle the traffic, along with VoIP provisioning, is included as appropriate [11].

Availability of a system refers to a system or process being ready for use. High availability refers to the quality of a system that is up most of the time. Normally, availability is determined by two measurements [11]:

- Mean Time between Failures (MTBF) measures equipment reliability. MTBF is equal to the total equipment uptime in a given time period, divided by the number of failures in that period.
- Mean Time to Repair or Replace (MTTR) measures maintainability and indicates how quickly a system can be restored to service. MTTR is equal to the total equipment downtime for a given time period, divided by the number of failures in that period.

To improve reliability and ensure that downtime is minimal in case of failure, the system should include these features [11]:

- **No single point of failure**: All components of the system have some form of redundancy.
- **Standby capability**: The system can switch to a standby component when failure is detected on a primary component.
- **Error self-detection capability**: The system can detect something is going wrong before failure.

B. T1 PROVISIONING

T1 provisioning is based on the evaluation of expected call traffic during the busiest hour and the expected average call time. Using these values and the Erlang-B formula, developers can determine the number of channels needed for a given probability of blocking.

The Erlang-B formula gives the probability of blocking in a system where a large population makes use of a finite number of resources. The blocking probability represents the chance of all channels being busy when a call comes in. The only way to prevent blocking entirely is to configure as many channels as there are potential callers into the system, which is obviously unrealistic. Normal levels of blocking are from 1 to 10 percent. The Erlang-B formula assumes that all calls are placed independently and that the number of potential callers is large [11].

The total demand for channels is measured in Erlang units. Traffic is determined by multiplying the total expected number of calls per time unit during the busiest hour by the average time a channel is in use. For example, if the call rate is two calls per second and the channel's average holding time is twenty seconds, then the traffic is forty Erlang. The following table (based on data obtained from Nuance) presents three typical cases for T1 provisioning. Based on hypothetical traffic and an acceptable level of blocking—2 percent is usually an acceptable level of blocking for telephony services—the Erlang-B formula is used to calculate the required number of channels (assuming no fractional T1). The last two columns show the level of blocking in the occurrence of T1 failures (1 and 3 lost T1s) [11].

Traffic (Erlang)	Minimal number of channels	Number of PRI T1s (23 channels)	Configured number of channels	Blocking with 1 lost T1	Blocking with 3 lost T1s
200	214	10	230	3.5%	21.2%
1000	1008	44	1012	3.2%	6.9%
5000	4939	215	4945	2.3%	3.0%

Table 1. T1 Provisioning Examples [From Ref. 11]

To ensure the provisioned number of channels continues to match the needs of the service, the traffic patterns should be monitored (calls per second and average call duration) throughout the life of the service.

C. VOIP PROVISIONING

When using Voice-over-IP (VoIP), developers or network administrators must provision a data link to carry the Session Initiation Protocols (SIP) messages along with voice data. Most of the bandwidth is used by the voice data. Voice data is carried using Real-Time Transport Protocol (RTP) with a payload type of G.711 and packets of 20 milliseconds. Each channel used takes a maximum of 92.2 Kbps at the link level (per Ethernet 802.3 standards). SIP signaling for setting up and tearing down a call amounts to roughly 2 KB in each direction. If call duration averages 20 seconds, signaling takes less than 1 percent of the bandwidth. Thus, to account for all traffic, it can be safely assumed that a minimum 100 Kbps per channel is needed on the link with the VoIP service provider. Blind call transfers require only signaling bandwidth. Bridged transfers double the voice path and require 200 Kbps at the link level[11].

D. HARDWARE

This section describes host and telephony hardware that can be used with the voice recognition architecture package. This thesis does not endorse any particular vendor; however, the following vendor hardware examples were either used by Nuance or NPS while running their various voice recognition applications. Specifically, in June 2005, NPS completed Phases 1A and 1B of the IEVAP. The objective of Phase 1A was to develop and demonstrate a bilingual voice-activated menu-driven phone system in English and Arabic. The objective of Phase 1B was to test and demonstrate speaker verification technology in English [6].

1. SPARC Solaris Hosts

The Sun Sunfire 280R server has been recommended by Nuance and it was used by Nuance for internal testing. This server is configured to have a high degree of availability [11], [14]:

- Dual hot swappable redundant power supplies
- Two hot swappable hard drives
- Dual CPUs
- Optional dual redundant network interfaces

The following lists specifications concerning the Sun Sunfire 280R server [14]:

- **Processor**: Powered by up to two high-performance 1.2 GHz UltraSPARC III Cu processors, Sun's binary compatible next-generation CPU module.
- **Rack Assembly**: Fits standard 19-inch rack with sliders for easy servicing and upgrading of CPUs, PCI cards, memory and power supplies (up to eight systems per 72-inch rack).
- **Front Accessible**: Front-accessible power supplies and software-mirrored disk drives.
- **Hot-Swap**: Redundant hot-swap power supplies, independent power cords and hot-plug disk drives.
- **Remote Management**: Sun Remote System Control (RSC) for remote monitoring of key components or remote power on/off. With the battery backup in the RSC board, the remote management functions will be available for up to 40 minutes after a complete power failure.

Figure 11. Sun Fire 280R Server [From Ref. 14]

2. Microsoft Windows Hosts

Phase 1B NPS tests used the following two Microsoft Windows laptop servers for internal testing. These laptops were configured to have a high degree of availability [6].

- Dell Latitude 15.4" D810 Intel Pentium M770 Processor (2.13 GHz), 2GB DDR2-533 SDRAM, 80GB Hard Drive, Intel Pro/Wireless 2915 (802.11 a/b/g, 54 Mbps) and integrated Bluetooth
- Dell Latitude 12" D410 Intel Pentium M755 Processor (2.00 GHz), 2 GBDDR2-533 SDRAM, 80GB Hard Drive, Intel Pro/Wireless 2915 (802.11/b/g, 54 Mbps) and integrated Bluetooth
- Sony F-V420 Unidirectional Natural Sound Vocal Microphone

These two computers (host) are required to demonstrate the bilingual application per Nuance technology requirements. The laptop computers listed above were chosen for their processing power, memory capability and mobility. The input device (microphone) was selected based on its ease of use in developing and testing the speech application. The below figure depicts the testbed setup [6]:

Figure 12. NPS Testbed Hardware Setup [From Ref. 6]

3. Telephony Hardware

The reference architecture supports the CG 6000 telephony board from Natural Microsystems, which has an estimated MTBF of 132,403 hours and supports four T1s (96 channels). The following lists some of the features associated with this board [8], [11]:

- Up to 120 universal IVR/VoIP/fax ports
- Low-latency media streaming
- On-board RTP/RTCP
- Dual 10/100Base-T interface
- Both single-slot 6U CompactPCI and PCI solutions
- T1/E1 digital trunk PSTN interfaces
- Natural Access software environment
- Full-speed H.100/H.110 bus with 4,096 timeslots to support interoperability with other boards in open-architecture, high-capacity systems

Figure 13. Natural Microsystems CG 6000 Telephony Board [From Ref. 8]

For VoIP deployments, Cisco gateways, such as the AS5300, are recommended, as these devices support up to four T1s or E1s and PRI Q.931. The AS5300/Voice Gateway is NEBS Level 3 compliant. The Cisco® AS5350 Universal Gateway is the only one-rack-unit (1RU) gateway supporting 2-, 4- or 8-port T1/7-port E1 configurations that provides universal port data, voice and fax services on any port at any time (Figure 20). The Cisco AS5350 Universal Gateway offers high performance and high reliability in a compact, modular design. This cost-effective platform is ideally suited for Internet service providers (ISPs) and enterprise companies that require innovative universal services [1], [11].

Figure 14. Cisco AS5350 Universal Gateway [From Ref. 1]

4. LAN Switch

The Extreme Networks Summit48 switch was used by Nuance for internal testing. This switch offers many fault tolerance features including multiple load-sharing trunks, multiple spanning trees, Extreme Standby Router Protocol, and redundant, load-sharing power supplies. Summit48si has a non-blocking architecture with 17.5 gigabits of throughput with wire-speed performance on every port. Bidirectional rate shaping allows users to manage bandwidth on Layer 2 and Layer 3 traffic flowing both to and from the switch. DiffServ and 802.1p deliver varied levels of service for time-sensitive, demanding applications for voice, video and data, and ensure efficient bandwidth usage. Eight hardware queues provide granularity for multiple applications, and guarantee low latency/low jitter for time sensitive applications (voice and multimedia) with support for advanced scheduling algorithms [4], [11].

Figure 15. Extreme Networks Summit48 Switch [From Ref. 4]

5. IP Load Balancer

The BIG-IP 1000 IP Application Switch was used for Nuance's internal testing. This IP load balancer comes with high-availability features that provide the required functionality and redundancy to avoid single points of failure within networks. The BIG-IP 1000 provides the power of a 1 (Gb) X 8 (10/100) switch, with a platform that features fewer ports, with integrated SSL [7], [11].

Figure 16. BIG-IP 1000 IP Application Switch [From Ref. 7]

THIS PAGE INTENTIONALLY LEFT BLANK

IV. VOICE VERIFICATION TESTING TEMPLATE

A. PERFORMANCE MEASURES

Performance of a speaker-verification system is based on the measure between two types of errors present in biometric systems, specifically False Match Rate (FMR) and False Non-Match Rate (FNMR). FMR and FNMR are more commonly referred to as False Accept Rate (FAR) and False Reject Rate (FRR) respectively. The following definitions are provided [6], [10], [13]:

- False Accept is the false acceptance of an invalid user, such as in the case of an impostor breaking into a system (also known as a Type-I error).
- False Reject is the false rejection of a valid user, such as in the case of rejecting a true speaker (also known as a Type-II error).

The tradeoff between FAR and FRR exists in every biometric system. For instance, if a system's threshold is set to allow for greater user convenience, the probability of false rejections decreases, and the likelihood that an imposter can break into the system increases. Likewise, the opposite would hold true; if a system's threshold is set to allow for greater user security, the probability of false rejections increases (FRR rises) while the likelihood that an imposter can break into the system decreases (FAR diminishes). System performance at all the operating points (thresholds) can be depicted in the form of a receiver operating characteristic (ROC) curve. A ROC curve is a plot of FAR against FRR for various threshold values for a given application. An example of an ROC curve is shown in Figure 17, in which the desired area for a given application is at the lower left of the plot, where both types of errors are minimized [13].

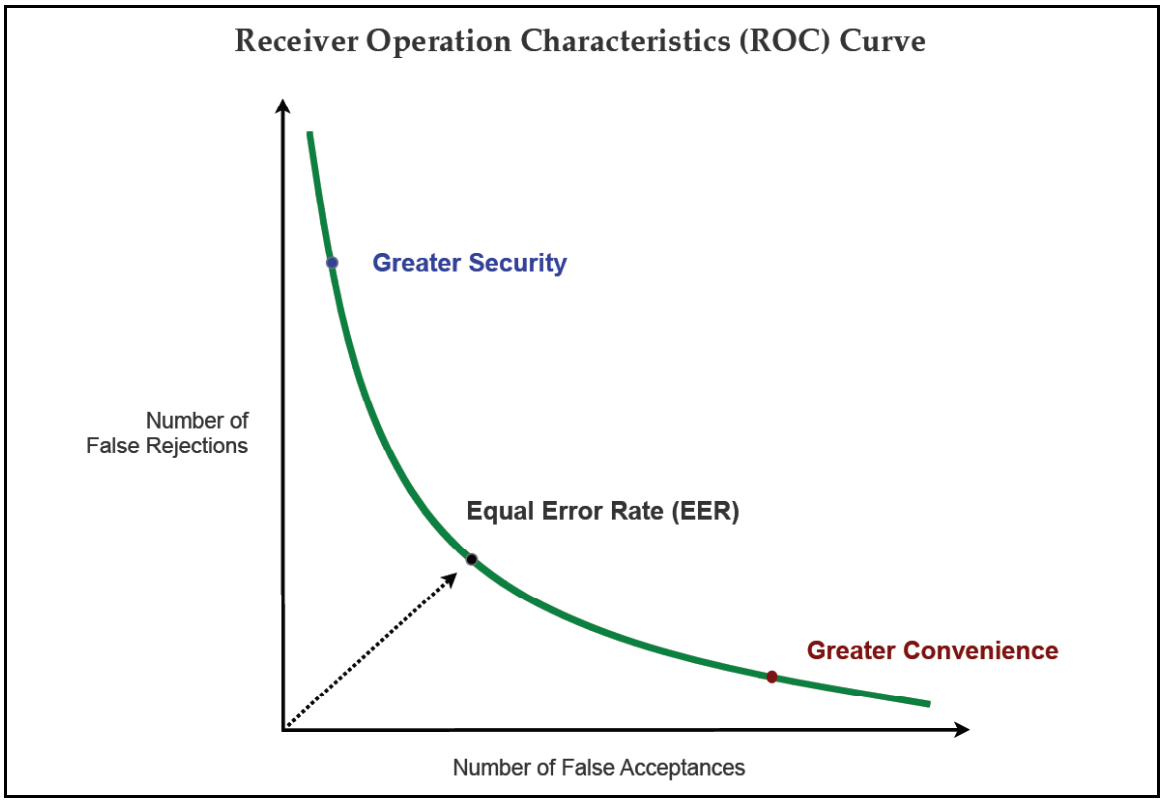

Figure 17. Receiver Operating Characteristics (ROC) Curve [From Ref. 13]

In Figure 17, the point on the curve at which the FRR and FAR is equal is called the equal error rate (EER). Often, the EER is used as the single summary number to gauge the performance of a speaker verification application. The green line shown in Figure 18 represents the various thresholds to which a given application can be set. For instance, in applications that required greater security, one would set the threshold of an application to the left of the ERR along the green curve, reflecting a lower probability of false accept but, at the same time, accepting a higher probability of false reject.

More recently, a variant of an ROC curve, called the detection error tradeoff (DET) curve has been employed, especially in the academic and national research institutions. The DET curve plots the same tradeoff shown in a ROC curve using a normal deviate scale. This has the effect of moving the curves away from the lower left corner when performance is high and producing linear curves. The advantage of a DET

curve over a ROC curve is that it allows easier comparisons of multiple data sets. Figure 18 shows the comparison of a data set plotted on two different curves—the DET curve and the ROC curve.

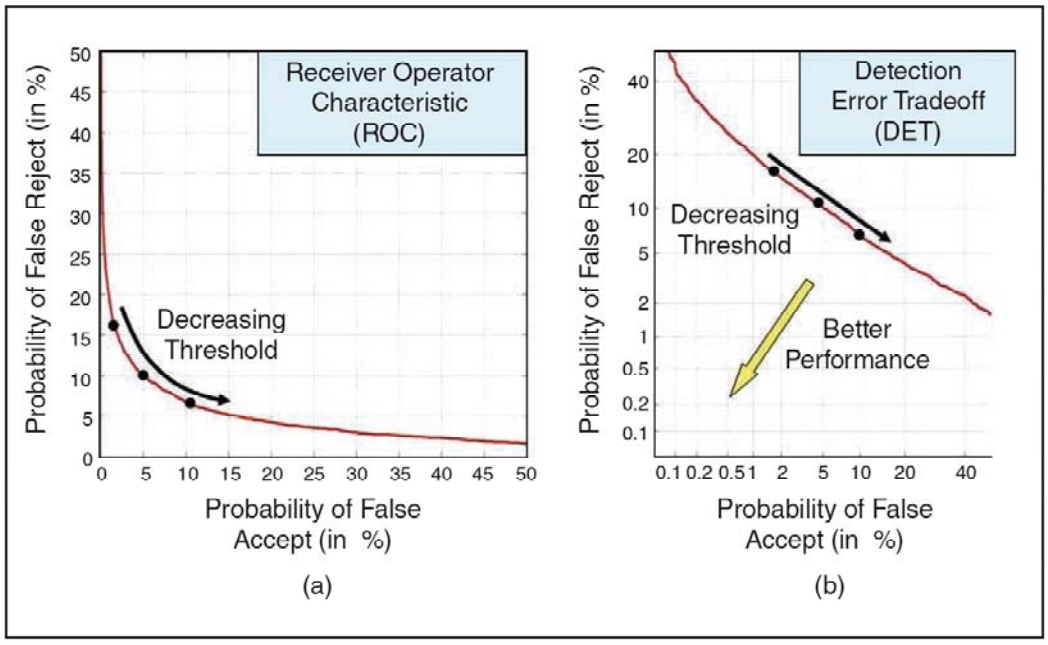

Figure 18. ROC Curve and DET Curve [From Ref. 13]

B. CONFIDENCE INTERVALS

Estimating statistical parameters, such as mean or variance from a set of samples, can result in "point estimates." Point estimates are single number estimates of the parameters in question. While very useful in many applications, one limitation of a point estimate is the fact that it conveys no idea of the uncertainty associated with it. If many such point estimates are used in the same analysis, it can become challenging to decipher which estimate is the best/most accurate [13].

On the other hand, a confidence interval provides a range of numbers (between a lower limit and an upper limit) with a certain degree of probability as to the possible interval of the respective point estimate. Thus it is easier to conclude that the point estimate with the shortest confidence interval is the most robust and reliable [13].

C. STATISTICAL BASIS CRITERIA

In the spring of 2005, NPS conducted a voice verification test as part of Phase IB of the IEVAP. The statistical analysis in the design of the NPS voice verification test was based on the following simplified scenario [6], [13]:

Assume that N speakers, taken at random from the envisaged user population, provide data for the trial. For simplicity, assume also that, for any given trial condition, each speaker makes one verification bid, whose result is either correct or incorrect, and that the results of different speakers' bids are independent. Let the probability of an incorrect verification result for any one bid — that is, the underlying population error rate — be p. Then the observed number of errors, r, is binomially distributed with mean Np, variance $Np(1-p)$; and the observed error rate r/N has mean p and variance $p(1-p)/N$.

Using the normal approximation to the binomial distribution, the 95% confidence limits on the observed error rate is expressed as $p \pm 1.96*\text{sqrt}((p(1-p)/N))$.

This equation was computed by measuring 95% of the area, i.e., a 95% probability on the normal distribution curve, which corresponds to a value of 1.96σ, where σ is the standard deviation.

When $p = 0.01$ (or when the population error rate is 1%), the confidence limits are as follows: $\pm 1.96*\text{sqrt}((0.0099/N)) = 0.01 \pm 0.195/\text{sqrt}(N)$

Setting N equal to 1000 gives confidence limits of: 0.01 ± 0.00617 (i.e. $1\% \pm 0.617\%$) on the observed error rate.

D. SYSTEM ACCURACY

Within the context of speaker verification technology, system accuracy is dependent on numerous factors, such as the level of user cooperation and the type of system constraints levied on a given application. In speaker verification applications, speech used for system enrollment and verification can span from text-dependent to text-independent, resulting in different levels of system performance. In a text-dependent application, a speaker states the same text during enrollment and verification and the speaker-verification system has prior knowledge of this text. Whereas in a text-

independent application, the system has no prior knowledge of the text to be spoken, which makes it more complex for the system to process [6], [13].

The System accuracy is defined in the following algebraic equations:

Accuracy of the System = (NT - (NFRR + NFAR)) / NT
= (NTAR + NTFR) / NT

where
False Reject Rate (FRR) = NFRR / NT
False Accept Rate (FAR) = NFAR / NT
NT = NTAR + NFRR + NFAR + NTFR.
where,

NT	The total number of valid verification attempts
NTAR	The total number of true accepts
NFRR	The total number of false rejects
NFAR	The total number of false accepts
NTFR	The total number of true failures.

E. BRIEF SUMMARY OF PHASE 1B ACCURACY TEST OF NORTH AMERICAN ENGLISH

In Phase 1B of this project, NPS successfully conducted a speaker verification test to assess Nuance's speaker verification technology based on the performance measures of the FRR and FAR. During the test, NPS did not impose any restrictions on the callers in terms of the type of phone used or from where the calls originated. The Nuance ROC analysis yielded an equal error rate of 3% (FRR based on 411 trials, FAR based on 4,300 trials) and a system accuracy of 94%, while the NPS analysis yielded a FRR of 2.91% and a FAR of 1.2% (based on 411 verification attempts) and a system accuracy of 95.87%. The ROC analysis equal error estimates of the NPS test were in the same range as the average estimates of the equal error rate (FRR = 3.4%, FAR = 3.4% and system accuracy = 93.2%) by Nuance based on other similar datasets. This validated the NPS test in spite of the smaller number of enrollments and speaker verification attempts [6], [13].

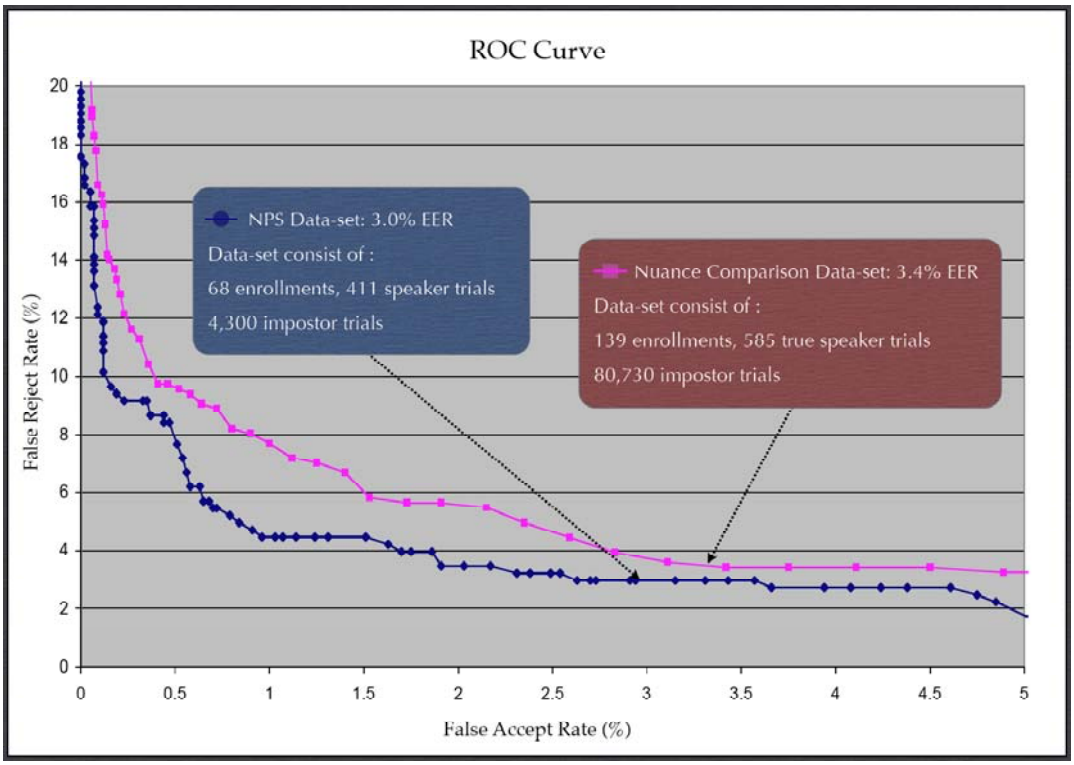

Figure 19. NPS Phase 1B Test Results [From Ref. 6]

F. ESTIMATES OF CONFIDENCE INTERVALS FOR THE NPS TEST

The NPS test had 68 speakers. The confidence interval computed using Normal Approximation for the various test data sets are given in Table 2 below [6]:

Analysis Type	**Confidence Interval for False Reject Rate using Normal Distribution**	**Confidence Interval for False Accept Rate using Normal Distribution**
NPS Single Point Data Analysis on the NPS data set	$N = 68$ $p = 2.91$ **2.91% ± 2.3647%**	$N = 68$ $p = 1.22$ **1.22% ± 2.3647%**
Nuance ROC analysis of the NPS data set	$N = 68$ $p = 3.0$ **3.0% ± 2.3647%**	$N = 68$ $p = 3.0$ **3.0% ± 2.3647%**
Nuance ROC analysis of an independent data set	$N = 139$ $p = 3.4$ **3.4% ± 1.654%**	$N = 139$ p=3.4 **3.4% ± 1.654%**

Table 2. Confidence Intervals for the NPS Voice Verification Test [From Ref. 6]

G. TEST TEMPLATE SCOPE AND OBJECTIVES

1. Test Scope

In this paragraph, the experimenter should describe the scope of the test plan. This write up should describe the test focus, e.g., focus on testing and demonstrating speaker verification technology in the Iraqi-Arabic language.

2. Test Objectives

The test objectives paragraph should include [6], [13]:

- Data to be collected, e.g., collect enrolled voice samples of approximately 1,000 Iraqi Arabic callers and store these samples in a database.
- Data to be extracted, e.g., extract voice features and store them as templates in a database.
- Data to be collected and subsequently stored in a database in terms of specific verification voice samples. Additionally, known imposter attempts will be recorded and tracked to measure the system false alarm rate accomplished by computing how many times the user can successfully break into the system; e.g., collect 10,000 verification voice samples, ten from each of the thousand Iraqi Arabic callers, and store them in the database (either temporarily or permanently depending on the protocol). These known imposter attempts are recorded and tracked to measure the system false alarm rate, done by calculating how many times the user can successfully break into the system.
- Indicate how many, out of the total number of verification attempts, will be imposter attempts; e.g., out of the 10,000 verification attempts at least 500 attempts should be imposter attempts in which an Iraqi Arabic caller will try to break into the system by pretending to be the owner of another account, i.e., an account that was not setup by the person himself/herself
- Indicate if tracking of the type of phone used is accomplished to estimate cross channel effects on the tested language accuracy; e.g., keep track of the type of phone used—such as cell phone, land line or voice over IP

(VoIP)—so that cross channel effects on the Iraqi Arabic Language accuracy can be estimated and reported.

- Collect a significant number of voice verification samples in the presence of pronounced background noise; e.g., collect a minimum of 500 voice verification samples in the presence of pronounced background noise such as automobile noise, loud music, airport or public places noise, household equipment noise, military vehicles noise, military explosion noise and multiple human speakers in the background.
- Keep track of the gender of callers to study the effect of male and female voice; e.g., keep track of the gender of callers to study the effect of male and female voice as it affects the overall Iraqi Arabic voice verification accuracy.
- Match the collected feature set during the verification phase with the template of the person whom the calling person claims to be.
- Set a decision threshold that results in different accept/reject criteria.
- Compute the miss rate and false alarm rate based on the previously collected data to measure the performance characteristics of the system.
- Report the measured system accuracy as a single point data analysis after eliminating enrollment and verification attempts that are true user failures (due to improper use of the system by not following directions) and not system failures.
- Have the associated vendor perform an automated imposter analysis in terms of an ROC curve by considering every voice verification attempt as a possible imposter against every enrolled account, after eliminating enrollment and verification attempts that are true user failures (due to improper use of the system by not following directions) and not system failures. NPS will measure the FRR and FAR from this ROC analysis.
- Provide match and non-match score distributions.
- Provide Automatic Speech Recognition (ASR) error rates by testing the tested language verifier on a pre-selected set of specific language words

and phrases typically used in a given application, or some such similar application; e.g., provide Automatic Speech Recognition (ASR) error rates by testing the Iraqi Arabic language verifier on a pre-selected set of Iraqi Arabic words and phrases typically used in the BCCF application, or some such similar application.

- Provide Text-to-Speech (TTS) error rates based on testing the tested language synthesizer module on a pre-selected set of the tested language words and phrases that might be used in that specific application; e.g., provide Text-to-Speech (TTS) error rates based on testing the Iraqi Arabic Language synthesizer module on a pre-selected set of Iraqi Arabic words and phrases that will be used in the BCCF application.
- Draft a detailed report on the tested language verification test summarizing all of the results and the details of the test protocol and procedures; e.g., draft a detailed report on the Iraqi Arabic Voice Verification Test summarizing all of the results and the details of the test protocol and procedures.

3. Test Procedures

- **Source and Nature of Test Subjects**

Sources and nature of test subjects will have to be identified. For instance, Phase 1B of IAEVP utilized NPS students as test subjects. Phase 1C needs Iraqi Arabic speakers, so NPS has identified four possible avenues to recruit the 1,000 Iraqi Arabic speakers required to support this test [6], [13]:

(1) Utilize Defense Language Institute (DLI) faculty and students
(2) Outsource to Nuance
(3) Partner with educational/linguistic institutions
(4) Utilize Iraqi Arabic immigrants.

Nevertheless, it is important to identify test subjects as soon as possible to ensure smooth recruiting efforts, enabling a sufficient number of test subjects. NPS is required,

by law, to receive a clearance from the Protection of Human Subjects Committee. Each test subject is required to sign a consent form, privacy act statement, and a debriefing form [9]. (Appendix B contains the appropriate forms necessary for the approval to use human subjects).

4. Training of Test Subjects

The test plan should identify training for the prospective test subjects to ensure they understand what will be required of them to successfully complete the actual testing phase. For instance, during Phase 1C of the IAEVP, NPS is planning to offer limited training for the prospective callers as described below [13]:

- Web-based Training: NPS will request Nuance to host a website of their Iraqi Arabic Caller Authentication system at least one month prior to the actual test start, so that prospective users can learn how the Iraqi Arabic Caller Authentication system can be used for enrollment as well as verification.
- In the event NPS utilizes DLI students, or any other institution teaching Iraqi Arabic, NPS would need to perform an actual demonstration of the Nuance Iraqi Arabic Caller Authentication system at least two weeks prior to the start of the test phase.

5. Test Phases

The test plan should identify a network diagram required for the actual testing. Also, each distinct test phase needs also be described. For instance, in Phase 1C testing, NPS identified a network diagram for the Nuance Caller Authentication System, as shown below in Figure 20. The test is performed in two phases, with the first phase for enrollment and verification, and the second phase for verification only [13].

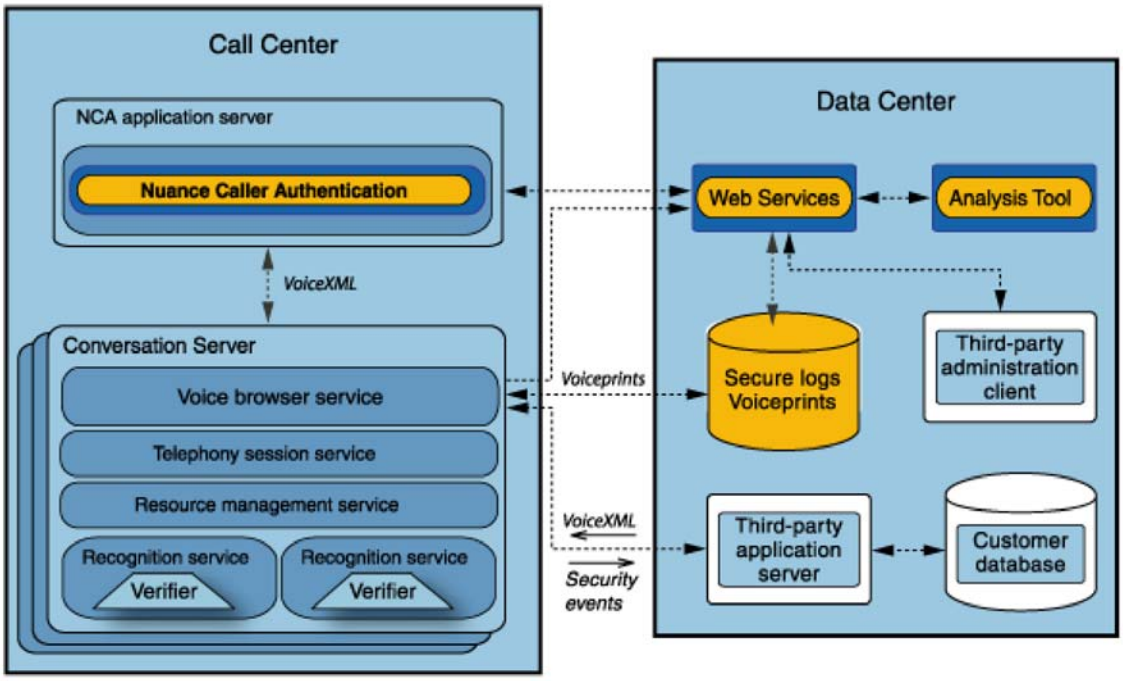

Figure 20. Nuance Caller Authentication System Network Diagram [From Ref. 13]

6. Time Needed for Enrollment and Verification Attempts

The test plan should include a thorough discussion of the time needed for enrollment and verification attempts. Normally and very conservatively, the enrollment takes about five minutes and each verification attempt takes two minutes. So, depending on how many callers there are, developers will be able to deduce how long the enrollment verification takes place. Knowing that information, the developers should understand how many days will be required to complete the whole testing scheme. For instance, for Phase 1C testing, the information below describes the steps required to provide the needed time to complete that testing project [13]:

Let us assume that the enrollment takes five minutes and each verification attempt takes two minutes. Hence, the verification of a thousand callers will take 5,000 minutes and 10,000 verifications will take 20,000 minutes. Assuming that the phone lines are available twenty-four hours a day, this requires fourteen days to complete the entire enrollment and verification. Since we plan to have three simultaneous lines, at best this whole experiment could be completed in a total of five to six days.

However, assuming that the system will not be used twenty-four hours a day, and the calling pattern will be at random, we are allocating a total of thirty days—fifteen days for the enrollment and verification phase, and another fifteen days for the verification phase only. If there is a surge of more than three simultaneous callers at a time, the system will prompt a message for the user to wait on line and will give an approximate expected time of availability, just as is done in many hotel and airline reservation systems. This will give the opportunity for the user to decide to stay on line or hang up and call at a later time [13].

7. Enrollment and Verification Phase of Iraqi Arabic Voice Samples and Initial Verification

The test plan should specify how and where callers should call in to test the system. There should be several (at least two to three) dedicated phone numbers to which test subjects can call in to enroll and verify. There should be a carefully crafted enrollment and verification dialog that would guide subjects through the enrollment process. It should be simple, clear and to the point. An example of an excellent dialog process is the one used by NPS during Phases 1B and 1C. During those phases, callers would dial one of three specified toll-free numbers (provided by NPS) and then would enroll once and verify at least four times [13].

The enrollment dialog was as follows [6], [13]:

When calling the system, the caller is initially greeted with a bilingual welcome message:

"Hi, Welcome to Baghdad Central Correctional Facility's Visitor Center (same prompt repeated in Iraqi-Arabic)."

After the initial greeting, the caller is prompted to select a language:

"To continue in English, say 'English'. To continue in Arabic, say 'Arabic' (Arabic welcome prompt spoken in Iraqi-Arabic)."

When the user chooses the option to schedule a visitation to the prison, the system plays back an initial prompt asking the caller if he or she has enrolled in the system. If

the caller replies “yes,” then the system will proceed to the speaker verification dialog. If the caller replies “no,” the system will proceed to the speaker enrollment dialog.

System: “In order to use our automated scheduling system, you must be an enrolled user. Are you an enrolled user? If you are, say ‘yes.’ If you’re not, say ‘no’ and I’ll help you to enroll in our system.”

Caller: “Yes.”

System: “To get started, go ahead and say, or key-in, your 10-digit account number.”

Caller: “No.”

System: “To get started on the voice enrollment process, I need your 10-digit account number. If you don’t have an account number, or if you’ve lost it, please go to your nearest police station to register for a new account. If you have the account number, go ahead and say or key it in now.”

(Note: for the purposes of this test NPS will assume that the user’s ten digit phone number will act as the ten-digit account number.)

If the caller provides a 10-digit account number, the system asks the caller to confirm his or her answer. If the caller confirms his or her answer, the system then checks to see if the account number is valid or not. Once, the system has validated the account number, the system then asks the caller to repeat from one to nine in order to authenticate the caller’s voice biometric. If the system authenticates the caller, the system proceeds to the next dialog. If the system does not authenticate the caller, the system repeats the authentication process. If on the third attempt the system cannot authenticate the caller, then the system plays back a prompt informing the caller to re-register or will connect the caller to a live agent (if available).

System: “Thanks, I heard 8005551212 is that right?”

Caller: “Yes.”

If the account number is already not enrolled, the system will come back and ask the caller:

System: “You are not enrolled in the system. Will you please enroll?

System: “But before you enroll, please indicate whether you are using a land line, cell phone or voice over IP. Say “land” if you are using land line, “cell” if you are using the cell phone or “IP” f you are using voice over IP.”

Caller: “Cell.”

System: “In order to enroll you need to count the digits 123456789 ”

Caller: “123456789.”

System: “Can you please repeat it once more?”

Caller: “123456789.”

System: “Can you please repeat it once more for the last time?”

Caller: “123456789.”

System: “Thank you. You are now enrolled in the system.”

The verification dialog will be very similar to the enrollment dialog, except that the system will already know the user has a valid registration number. In this instantiation, the system will request that each caller repeat the digits 123456789 only once.

Each specific dialog is different depending on a situation and what needs to be accomplished. System owners and system users need to be consulted throughout the design process to make sure the proposed dialogs address project needs and requirements.

8. Imposter Trials

The test plan should specify how imposter trials will be conducted. For instance, before the actual test, a group of pair subjects should be identified to request permission to break into the account. The pairs would know each other’s account numbers to allow them to make several attempts into the other person’s account. The system tester would

keep track of these imposter calls so that, based on the system performance, the false match rate can be computed and reported. Below is an imposter trial plan used during Phase 1C testing [13]:

NPS plans to collect at least 500 imposter trials. This will be accomplished as follows:

a) NPS will identify 50 pairs of callers who will be requested to break into the account of their partner.

b) The pairs will know each other's account number and hence will be asked to make at least five verification attempts into the other person's account, in addition to the valid verification attempts into their accounts.

c) NPS will keep track of these imposter calls so that, based on the system performance, the false match rate can be computed and reported.

9. Processing of Consent Forms

According to NPS rules and regulations, consent forms need to be signed for the use of human subjects in any experiment. Certainly, any use of "live" callers will require consent forms, and human subjects review and use. Hence all subjects will be required to sign and fax a consent form prior to participating in this test. This consent form will not request demographic or personal information except for minimum requirements such as the account number (phone number), gender and the type of phone line (land line, cell phone, voice over IP, etc.). All information provided should be treated as personal and confidential. The actual participation should be voluntary in nature and subjects should be given an opportunity to withdraw from tests at any time for any reason [6], [9], [13].

10. Test Facilities/Environment

The test plan should also identify the actual location of the system. Equipment and software lists should be provided, along with how it will be connected to the phone system. As an example, the test facilities/environment plan utilized for the Phase 1C testing is described below [13]:

The Nuance Iraqi Arabic Caller System Test equipment will be located in the in the Glasgow building at NPS. The computer equipment will be comprised of the hardware and software listed in sections 5.2 and 5.3. The caller authentication system will be hooked up to at least three lines that are toll-free numbers for the callers.

11. System Test Schedule

The test plan should also list a realistic timeline of the test, identifying each important milestone or critical point. Below is an example timeline considered for the Phase 1C testing [13]:

Nuance Application Development (Phase 1C)	01 Apr 06
Enrollment and Verification Phase	01 Apr – 14 Apr 06
Verification Only Phase	01 May – 14 May 06
Analyze/Interpret Data	31 May 06
Initial Report on the Analysis	30 Jun 06
Conduct Formal Demonstration	01 Aug 06
Draft Thesis	15 Aug 06
Final Thesis Submission	31 Aug 06

The bottom line is that each timeline should be different for each test but, nevertheless, has to be included in the plan.

12. Resources

A test plan should also list resources available. The first list should include human resources involved with the testing, specifically listing NPS personnel, vendor representatives and, finally, the project sponsor. Next, a detailed list of hardware and software should be included, listing every major end items for each respective category. Below is an example of resources listed for the Phase 1C test plan [13]:

NPS

Mr. Jim Ehlert, Program Manager
Dr. Pat Sankar, Subject Matter expert
Major Marek Sipko, Master's student

Nuance (personnel to be assigned)

Project Manager
Support Manager
Director of R&D

OSD (for final approval)
Gerard Christman
Brian Fila

Hardware

- (2) Dell Latitude 15.4" D810 Intel Pentium M770 Processor (2.13 GHz), 2GB
- (2 sets) DDR2-533 SDRAM, 80GB Hard Drive, Intel Pro/Wireless 2915 (802.11 a/b/g, 54 Mbps) and integrated Bluetooth.
- Dell D/Dock Expansion Station for Dell Latitude.
- Plantronics SupraPlus Binaural w/ Voice Tube Headset.
- Plantronics MX10 Headset Switcher Multimedia Amplifier.
- Plantronics DA60 USB to Headset Adapter.
- Sony F-V420 Unidirectional Natural Sound Vocal Microphone.
- Intel Netstructure PBX-IP Media Gateway 8 port.
- LaCie 100GB P2 Mobile Hard Drive USB/FW.

Software

- Windows 2000 and or 2003
- Norton Antivirus Client Edition.
- Nuance Voice Platform 3.0.
- Microsoft Office 2003.
- Xten X-Pro SIP Softphone for Windows.

13. Roles and Responsibilities

A test plan should also list participating organizations in the project. Each organization should have a detailed list of responsibilities directly assigned to that particular organization. Below is an example of roles and responsibilities listing for Phase 1C test plan [13]:

Participating Organizations

The primary organization for the Phase 1C project is NPS. Other participating organizations are as follows, with their expected roles and responsibilities as perceived by NPS:

- **Office of the Secretary of Defense (OSD)**
  - Project sponsor and the overall authority for the validation of the successful completion and delivery of this product.

- Provide IEVAP project Concept of Operations (CONOPS).

- **Defense Language Institute (DLI)**
  - Provide language expertise to help acquire the Iraqi Arabic language spoken samples with the participation of students and faculty.
- **Biometrics Fusion Center**
  - Provide Subject Matter Expertise on an "as needed" basis or as directed by OSD.
- **Nuance**
  - Provide the Iraqi Arabic core voice recognition engines.
  - Provide technical support to NPS.
- **Naval Postgraduate School**
  - Provide the Program Manager for Phase 1C.
  - Prepare the statements of requirements of all Nuance modules and Iraqi Arabic Voice Verification test plan.
  - Ensure delivery of the Iraqi Arabic Language Model, Iraqi Arabic Verification Model and the Nuance Iraqi Arabic Caller Authentication System.
  - Conduct the Iraqi Arabic voice verification/authentication test.
  - Analyze the test results and provide summary.
  - Provide final report (interim final report and final thesis).

14. Reviews and Status Reports

The test plan should include a description of reviews and status reports. Examples of these are report on the ROC/DET performance curves of the entire database and report of any limitations/shortfalls discovered in the Phase 1C software and associated history of fixes. Additionally, final deliverables should be The Iraqi Arabic Voice Verification Accuracy Test Report to include the identification of a set of equipment (hardware, software and peripherals) that will be able to demonstrate the feasibility of the concept. If there is a master's thesis involved with the project, such a

thesis should also be listed as a final deliverable. Below is an example of Reviews and Status Reports listing for Phase 1C test plan [13]:

Test Deliverables

The following files and reports will be derived from the testing data:

- Report on the ROC/DET performance curves of the entire database
- Report of any limitations/shortfalls discovered in the Phase 1C software and associated history of fixes.

Equipment

Listed below are the initial cost estimates for the equipment expense.

(a) Hardware:
(1) Laptops and Backup Hard Drives: $8,190.00
(2) Telecommunication Equipment and Accessories, 1-800 number: $2,500.00.

(b) Software:
(1) Nuance: $450,000.00 (Iraqi Arabic)
- NVP 3.0 SP4, Vocalizer 4.0, NAE 3.0 SP4 (V-Builder)
- Microsoft Speech Server 2004
- Microsoft Windows 2000 Pro
- SIP Foundry's SipXphone

(c) Nuance technology courses: $3,900.00

Initial Research Cost Estimate

(a) Nuance: $450,000
(b) NPS Equipment: $10,190.00
(c) NPS Travel: $17,000
(d) DLI Support: $20,000
(e) NPS Faculty Labor: $85,500.00

Final Deliverables

(a) The Iraqi Arabic Voice Verification Accuracy Test Report to include the identification of a set of equipment (hardware, software and peripherals) that will be able to demonstrate the feasibility of the concept.

(b) NPS master's thesis on "Testing and demonstrating speaker verification technology in Iraqi-Arabic as part of the Iraqi Enrollment via Voice Authentication Project (IEVAP) in support of War on Terrorism (WOT) security requirements," by Major Marek M. Sipko.

15. Benefits of the Study to the Sponsor

The test plan should include statements related to benefits of the subject testing and study to the sponsor. These would include statements related to the origination of the project, value added estimation, and recommendations for further studies/research as appropriate. Below is an example of Benefits of the Study to the Sponsor statements for Phase 1C test plan [13]:

The testing and demonstration of the speaker verification technology in Iraqi-Arabic was specifically requested by OSD. Subsequent NPS research is intended to contribute toward the future employment of voice authentication technologies in a variety of coalition military operations. The value added from this research includes:

(a) The selection of the most appropriate hardware, software, and peripherals for a mobile demonstration kit (laptop, voice input devices, etc) suitable for this technology.

(b) The integration of existing voice authentication technology (Nuance) into a hardware and software suite that utilizes the output of the voice authentication process and performs other functions depending upon the output of the authentication process.

The following is a preliminary listing of further studies for additional NPS students in support of this research project:

(a) Communication architecture research to identify the optimal medium to employ voice authentication technology in Iraq, e.g., comparison of 802.11, 802.16, cellular, and POTS technologies.

(b) Concept of Operations (CONOPS) research on the employment of voice authentication technology in support of other military applications and domains.

(c) Costs benefit analysis on the deployment of voice authentication technology.

(d) Statistical comparison of the success and failure rates of voice authentication technology versus other biometric technologies.

(e) Additional voice authentication proof-of-concept research in support of other voice authentication technologies or in support of other critical low-density foreign languages, e.g., Pashto, Dari and Farsi.

16. Issues/Risks/Assumptions

The test plan should include statements and discussion related to issues, risks and assumptions. The successful completion of any project is critically dependent on a variety of factors and influences; a delay in any one area could significantly affect the timely completion of this project. Each project is different, with its own peculiarities and nuances. All of these issues will have to be accounted for and presented in this section. Below is an example of Issues/Risks/Assumptions discussion for the Phase 1C test plan [13]:

- NPS assumes that part of this project will be funded by NPS prior to 30 November 2006, such that effective planning and implementation by Nuance for the Iraqi Arabic Language model and verification model can commence in December 2006.
- NPS assumes that the remainder of this project will be funded by OSD prior to 31 December 2006, such that effective planning and implementation of the Nuance Caller Authentication Iraqi Arabic Localization module can commence during the first week of January 2006.
- NPS assumes that collaboration with NPS and DLI will result in collection and storing of the actual voice data samples to support testing no later than 30 April 2007.

THIS PAGE INTENTIONALLY LEFT BLANK

V. SYSTEM CONCEPT OF OPERATIONS TEMPLATE

A. EXPERIMENTS

The formulation of experimentation is centered on the evidence that is required to develop or test a theory or assess an application—evidence that can provide answers to some question or questions at hand. The word *evidence* is defined as "something that furnishes (or tends to furnish) proof [15]." Based on this definition, evidence is not equated with proof, which is a logical product of analysis or a conclusion, but with the inputs to an analytical or thought process. Both observation and testimony can constitute evidence; however, not all evidence is equally relevant, valid, replicable, or credible. Properly designed and conducted experiments or tests greatly increase the likelihood that the data collected, the observations made, or the testimony that is also known as expert opinion elicited will have these desirable properties. Multiple experiments and analyses are required to establish relevance, validity, repeatability, and, ultimately, credibility [2]. Thus, the conduct of properly designed and sequenced experiments is integral to any test, including voice recognition tests.

B. ANALYSIS

Analysis takes the data provided by tests, combines it with previously collected data, and develops findings that serve as the basis for drawing conclusions related to the issues or questions at hand. Statistical theory forms the scientific basis for determining the probability that the observed data have a given property with a given level of confidence, or in other words, that there is little likelihood that the result occurred by chance. Increasingly, this analysis extends into areas of complexity, where analysis is more challenging and requires new approaches and tools intended to identify emergent behaviors and system properties [3].

Analysis needs to take place before, during, and after the conduct of each test. The conceptual model provides a framework and point of departure. There are many analytical techniques that can be utilized and care must be taken to employ the

appropriate method or tool. The findings developed in each of the analyses that are conducted should be used to update the conceptual model reflecting what needs to be accomplished to make it better [2].

C. HUMAN SUBJECTS

Human subjects are normally part of the experiments, and usually they are part of voice recognition tests. They must be identified and arrangements must be made for their participation, since voice recognition tests and experiments require a large pool of them. Human subject recruiting is a very difficult task. However, any human subject recruiting should be conducted very carefully and diligently because the results of the test can only be applied with confidence to the population represented by the subjects [2].

Subjects definitely need to be unbiased in that they have no stake in the outcome of the experiment. For example, if the experiment is assessing the military utility of an innovation (e.g., Iraqi Arabic voice recognition), people responsible for the success of that innovation (e.g., program manager, chief scientist, or master's thesis student) are inappropriate subjects. Even if they make every effort to be unbiased, there is ample evidence that they will find that almost impossible. Moreover, using such "contaminated" personnel will raise questions about the experiment results. In demonstration experiments, of course, such advocates will often be at the core of those applying the innovation. In this situation, where the utility of the innovation has already been established, their extra motivation to ensure that the innovation is successfully implemented becomes an asset to the effort. This is especially true in situations when future funding directly depends upon success of the demonstration [2], [6], [13] .

Subjects must also be available for the entire time required. Very often, the type of people needed as subjects are professionally very busy. The more skilled they are, the more demand there will be for their time. Hence, the experimentation team must make the minimum necessary demands on their time. At the same time, requesting insufficient preparation time for briefing, training, learning required training techniques, and working in teams, as well as insufficient time to debrief them and gather insights and knowledge developed during their participation undermines future tests. Failure to employ the

subjects for adequate lengths of time could, in most likelihood, compromise the experiment and even may make it impossible to achieve its goals [2].

D. TRAINING

Subjects need to be trained as part of the pretest activities. Training should precisely address the skills to be used in the test. Ideally, the subjects being trained should be the same people who will participate in the experiment. Full trials should be run (although they may well lack the rigor of those to be made during the experiment) in order to enable dialogue between the members of the experimentation team, the subjects, and the technical personnel supporting the rehearsal [2]. Voice recognition and authentication tests are inherently technical and interactively complex. Therefore, they require a rigorous subject training to ensure future test successes.

Most of the results of the pretest will be obvious, but others may require some reflection and study. Time should be built into the schedule to provide an opportunity to reflect on the results of the pretest and to take corrective actions as appropriate. Because elements of an experiment are heavily interconnected, changes needed in one area could affect or depend upon changes in other aspects of the experiment. For example, learning that the subjects need more training on the systems they will use must have an impact on the training schedule and may require development of an improved human-computer interface as well as new training material. Voice recognition and authentication tests may require subjects possessing a foreign language capability. This means that test subjects could be dispersed geographically throughout the United States or even abroad. If this is the case, then a web-enabled training application might be the only way to adequately brief and train the subjects before the actual test commencement. Nevertheless, it is critical that human subjects are trained and understand what will be required of them during the actual test [2], [13].

E. SPEAKER VERIFICATION PERFORMANCE MEASURES

There are four significant performance measures for speaker verification [12]:

- **False acceptance (FA) rate**—the probability that an imposter is accepted into the application. The FA rate is not the percentage of calls that result in a false acceptance, since this assumes that a large majority of callers are true speakers. The FA rate is the chance of being accepted given that there is an imposter. For example, a 1.0% FA rate does not mean that 1.0% of the total calls will be falsely accepted; it means that 1.0% of the imposters will be falsely accepted by the application. The total percentage of calls that result in a false acceptance is therefore equal to the FA rate multiplied by the probability that a caller is an imposter.
- **False rejection (FR) rate**—the probability that a true speaker is rejected by the application. It is assumed that almost all callers are true speakers; therefore, the FR rate should be close to the percentage of all calls that result in a false rejection.
- **Reprompt rate**—the probability that a caller is prompted for additional utterances, when variable-length verification is turned on.
- **Knowledge rate**—the probability that a caller experiences knowledge verification.

Verification accuracy is measured along a curve, called the receiver operation curve (ROC) that maps the FA rate and the FR rate pairs that can be achievable for an application (see Figure 21). It is critical to understand that verification performance can only be specified by noting the FA rate and the corresponding FR rate at the same threshold [12].

The application can operate anywhere on the ROC curve. The location of the operating point on the curve is dictated by the verification thresholds required for a given application. Developers can modify the verification performance thresholds as needed by a given application by choosing a different operating point (a different FA rate/FR rate combination). As the FA rate is being decreased, it is more difficult to get into the

application, however, the FR rate increases. This relationship between FA and FR is definitely developer specific and should be set according to needs and wants of the development team [12].

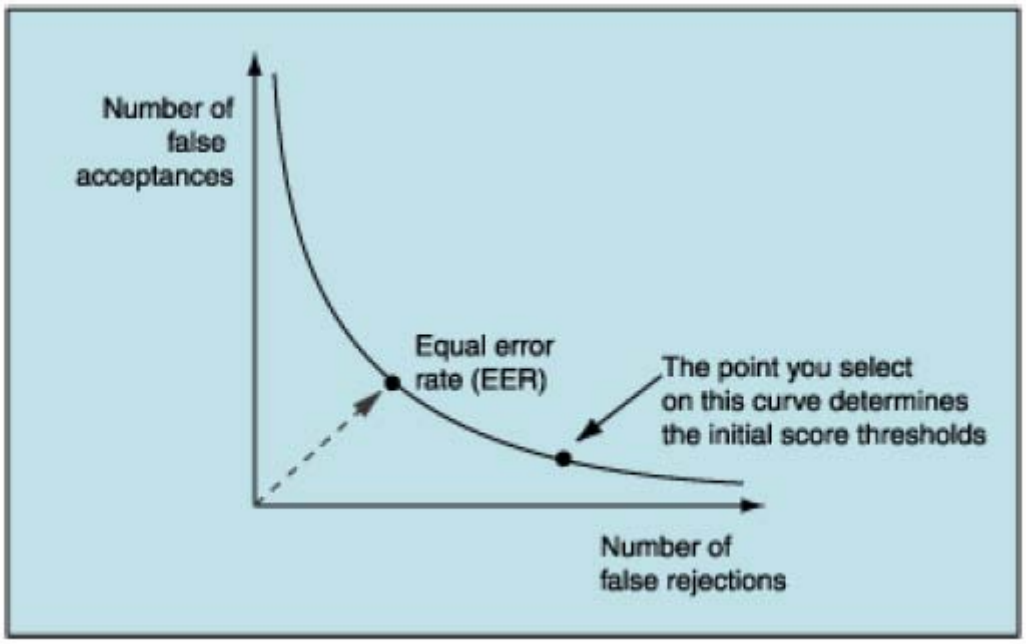

Figure 21. ROC Performance Curve [From Ref. 12]

F. SPEAKER IDENTIFICATION PERFORMANCE MEASURES

There are three significant performance measures for speaker identification [12]:

- **False acceptance (FA) rate**—The probability that an imposter is accepted into the application. The FA rate is not the percentage of calls that result in a false acceptance, since this assumes that a large majority of callers are true speakers. The FA rate is the chance of being accepted given that there is an imposter. For example, a 1.0% FA rate does not mean that 1.0% of the total calls will be falsely accepted; it means that 1.0% of the imposters will be falsely accepted by the application. The total percentage of calls that result in a false acceptance is therefore equal to the FA rate multiplied by the probability that a caller is an imposter. For speaker identification, an imposter is defined as a speaker who is actively trying to break into the system but who is not part of the group that is tested. For example, there is a family account with two family members: Martha

Smith and Robert Smith. If Robert Smith tries to break into the application using Martha Smith's identity, he is not considered an imposter.

- **False identification (FID) rate**—the probability that a speaker is incorrectly identified in a group. The FID rate for a group can be determined as follows:

$$\text{FID rate for a group} = \frac{\text{Number of calls falsely identified in a group}}{\text{Total number of calls to that group}}$$

To determine the overall FID rate for an application, the following equation applies:

$$\text{FID total} = \sum_i \text{FID group}_i \times P\langle \text{group}_i \rangle$$

- **False rejection (FR) rate**—the probability that a true speaker is rejected by the application. It is assumed that almost all callers are true speakers; therefore, the FR rate should be close to the percentage of all calls that result in a false rejection. This measure is calculated independently from the FID rate. Therefore, even if the true speaker is incorrectly identified, the FR rate is calculated for the true speaker and not the falsely identified one.

G. COLLECTING DATA TO MEASURE THE PERFORMANCE

To generate a ROC curve for a specific application and then select the operating point, application developers must collect appropriate data. The following describes the type of data that developers must collect to measure the FR, FID, and FA rates [12].

- **Measuring the FR rate—**To measure the FR rate, developers perform true speaker trials where true speakers attempt to access the application. Such applications can compare utterances by true speakers against their own voiceprints and measure the rejection rates at various thresholds. This data typically comes from verification sessions during data

collections, limited deployments, or rolled-out applications. The number of true speaker trials needed for statistically meaningful results depends on the target FR rate.

- **Measuring the FA rate—**To measure the FA rate, developers perform imposter trials, where imposters attempt to access the application. There are two possible methods for measuring the FA rate:
  - Developers can utilize a Nuance application called "batchrec" to simulate imposters by verifying utterances from a user against voiceprints from other users. Developers can use true speaker trials in a round-robin fashion to simulate impostor attempts, so long as all simulated impostor attempts are from different speakers than the voiceprints tested against.
  - Developers can have a large number of live imposters try to access the application. If knowledge verification is used, these imposters should be informed imposters: they should know the correct knowledge information so that they will be accepted by the knowledge verification component of the application. Developers will then be able to measure the voiceprint verification performance and the knowledge verification performance separately.
- **Measuring the FID rate—**To measure the FID rate, developers perform closed-set identification trials where true speakers attempt to access the application. This application must test identification utterances from each enrolled speaker of a group against the voiceprints of all the members in the group and measure the rejection rates at various thresholds. This data (speakers, groups) typically comes from verification/identification sessions during data collections, limited deployments, or rolled-out applications. To assess the performance accurately, developers must have a representation of the groups (size, composition, number) close to the deployed application. The FID rate applies to speaker identification only.

- **Statistical significance of performance measures—**How much data to collect is a difficult question when trying to produce accurate performance measures. When collecting the initial round of data, very few training and verification calls are required to set a reasonable performance threshold, ensuring that the limited deployment will perform well. This is not the case when collecting data during limited deployments and rolled-out applications. This section describes how much data is necessary for performance evaluations with limited deployments and rolled-out applications like NPS's Phases 1A and 1B tests. The goals of data collection during limited deployments and rolled-out applications are:
  - Evaluate the performance and see how the application is operating.
  - Tune the application performance by setting the operating point or by making changes to the training and/or verification dialogs. To attain these goals, developers have to get an accurate measure of the application performance. All performance measurements, whether for recognition or verification, include a certain amount of noise since the events described (for example, false acceptance, false rejection) are probabilistic. This type of noise adds an error to the measurement. The more data on which the performance measure is based on, the lower the error is. Statistical significance is when enough data has been collected so that the measurement noise is low compared to the quantity being measured. The voice recognition industry "rule of thumb" is that, for statistical significance, developers need to get at least 30 examples of each type of error that developers are interested in. For instance, developers need to see at least 30 false acceptances and 30 false rejections. Using this rule of thumb, developers can then determine the number of true speaker trials and imposter trials using the following formulas:

$$\text{Number of true speaker trials} = \frac{30}{\text{FR rate}}$$

$$\text{Number of imposter trials} = \frac{30}{\text{FA rate}}$$

For example, suppose that the desired FR rate of the application is 10%. How many true speaker trials are necessary for statistical significance? In order to see 30 FRs, 30 has to be divided by .10; therefore, 300 true speaker trials are required. The lower the FR rate, the higher this number is. Again, suppose that the desired FA rate is 1.0%. How many imposter trials are necessary for statistical significance? To see 30 FAs, 30 has be divided by 0.01; therefore, 3,000 imposter attempts are required.

H. TESTING PROTOCOL

To test the performance of an application to the level of certainty described in the previous section, and given a target of 1% FA and 5% FR, Nuance recommends collecting at least 600 true speaker trials and 3,000 independent impostor trials [12].

To collect this amount of data, Nuance recommends the following procedures [12]:

- 150 enrollees are chosen. The enrollees should have the same gender distribution as would be expected from the application user population.
- Two periods of time are chosen: an enrollment period and a verification period.
- During the enrollment period, each enrollee enrolls in the application. No verification trials should be conducted during the enrollment period.
- During the verification period, each enrollee will access his or her own account four times. If users would normally access the application from the same phone, the enrollees should use the same phone that was used for enrollment. No enrollments should be conducted during the verification period.

If only common verification utterances are used for enrollment and verification, the data collection process can stop immediately. Impostor trials can be created from the verification calls using the round-robin method described earlier. Using this method, developers can simulate 3,000 impostor trials by using "batchrec" (Nuance specific applications only) to run each of the 1,000 verification calls against three additional voiceprints [12].

If group-specific utterances, such as an account number, were used for enrollment and verification, live impostor tests must be collected as follows [12]:

- 200 impostor speakers are chosen. Enrollees can be impostors.
- Using the 150 accounts that have been enrolled, 300 lists are generated, each list with 10 randomly chosen account numbers. The 150 accounts are distributed to the lists such that each account appears approximately an equal number of times in the lists. The lists are randomly distributed to the impostors.
- An impostor data collection period is chosen. It is critical to not have any true speaker trial attempts during the impostor collection period, or the results will be inaccurate unless significant post-processing of the data is performed.
- During the impostor period, each impostor attempts to access each of the 10 accounts on each list only once.

Analysis is performed to measure the False Reject Rate from the true speaker trials and the False Accept Rate from the impostor trials. The False Reject Rate will be defined as the percentage of true speaker calls that are rejected by the application. Some calls during the true speaker period might be classified as impostors based on manual listening. The False Accept Rate will be defined as the percentage of impostor trials that are accepted during the off-line “batchrec” tests or the live impostors [12].

I. SYSTEM CONCEPT OF OPERATIONS TEMPLATE

The following describes a generic System Concept of Operations. It could be utilized when making considerations and plans for voice recognition and authentication applications testing. This proposed system ConOps is not designed to be an answer for all facets of a system development and testing, but certainly it is a good starting point. This template assumes that there is an existing system that will be replaced by a proposed system. The following is the proposed template [5]:

1. Overview

The first section of the Concept of Operations (ConOps) document provides four basic elements: system identification, an overview of the document, a high-level overview of the proposed system, and a brief description of the scope of effort required to take the system from the current state to the final future state of deployment that will be achieved at the conclusion of the proposed deployment. The following paragraphs describe these in further detail.

1.1 Identification

This section contains the proper title, identification number, and abbreviation, if applicable, of the system or subsystem that the ConOps applies to. If a system's related ConOps documentation has been developed in a hierarchical manner, the position of this document relative to other ConOps documents should be described.

1.2 Document Overview

This section summarizes and expands on the purpose for the ConOps document. The intended audience for the document should also be described. The audience can be a variety of people with various levels of technical knowledge and backgrounds. Therefore, it is important that document be clearly written to clearly define technical terms and utilize layman English parts of the document. The purposes of a ConOps document will, in most cases, be:

- To communicate user needs and the proposed system testing expectations
- To communicate the system developer's understanding of the user needs and how the system testing will verify such needs

1.3 System Overview

This section briefly states the purpose of the proposed system testing to which the ConOps applies. It describes the general nature of the system, and identifies the project sponsors, user agencies or departments; system developers; maintenance and support entities; and the operating centers or sites that will run the system. It also identifies other documentation that is relevant to the present or proposed system and its pertinent testing efforts [5].

A high-level graphical overview of the system is strongly recommended. This can be in the form of a physical layout diagram, a top-level functional block diagram, or some other type of diagram that depicts the system and its environment. Documentation that might be cited includes, but is not limited to, project authorizations, relevant technical documentation, significant correspondence, documentation concerning related projects, risk analysis reports, and any feasibility studies [5].

2. Referenced Documentation

This section lists the publisher, document identification number, title, revision, and date of all documentation referenced in the ConOps document. This section should also identify a point of contact for all documents not available through normal channels [5].

3. Current System Situation

This section of the ConOps describes the objectives to be tested, and the system or situation as it currently exists. The Current System Situation basically answers the following questions [5]:

- What is the system?
- What is the system supposed to do?
- Who owns, operates, and maintains the system?
- How well does the system perform?
- When is the system used?
- How does the system operate?
- What other systems does it talk to?

If there is no current system, this section will describe the reasons and motivations for developing the new system. In addition, this section will introduce the problems, needs, issues, and objectives that need to be addressed by the proposed system and pertinent verification tests. This enables the reader to understand better the reasons for the desired changes and improvements. Specific elements that may be documented in this section are outlined in the sections below. If there is no current system, this section will be described as non-applicable [5].

3.1 Background, Objectives, and Scope

This section should basically provide an overview of the current system or situation, including the background, mission, objectives, and scope of the current system, as applicable [5].

3.2 Operational Constraints

This section should include a description of limitations on the operational characteristics of the system. This could include limits on hours of operation, hardware limitations, or resource limitation [5].

3.3 Description of the Current System or Situation

This section should provide a thorough description of the current system, including operational characteristics; major system components; component interconnections; external system interfaces; current system functions; diagrams illustrating inputs, outputs, data flows; system costs; and performance statistics.

Additionally, this section should include a brief description of user classes and other people who interact with the system. A user class is distinguished by the way users interact with the system, and is classified according to common responsibilities, skill levels, work activities, and the ways they interact with the system [5].

3.4 User Profiles

This section should include a description of how the users interact with the system and the scenarios when they interact with the system. The section should also discuss how the users interact with each other. For example, a supervisor user class may have certain capabilities that an operator class may not have with the system, and the ConOps should describe when, why, and how such an interaction takes place to achieve a system objective or function [5].

3.5 Support Environment

This section should describe how the system is supported and maintained, including the maintaining department or agency; facilities; equipment; support software or hardware; and repair or replacement criteria. The section should also identify whether the system owners will maintain the system or a vendor will be contracted to maintain the system according to a contractual agreement [5].

4. Justification and Nature of the Changes

This section describes the shortcomings of the current system or situation that motivate development of a new system or modification of an existing system, and also describes the nature of the desired changes and assumptions for the proposed system. Specifically, the following information is detailed in the following paragraphs [5].

4.1 Justification for Changes

This section should include the reasons for changes introduced by the proposed system, including [5]:

- New or modified user needs, missions, or objectives
- Dependencies or limitations of the current system

4.2 Description of the Desired Changes

This section should include a summary of the new or modified capabilities, functions, processes, interfaces, and other changes needed to respond to the justifications previously identified. This should include [5]:

- Capability changes (i.e., functions and features to be added, deleted, or modified)
- System processing changes (i.e. changes in the process or processes of transforming data that will result in new output with the same data, the same output with new data, or both)
- Interface changes (i.e., changes in the system that will cause changes in the interfaces that will cause changes in the system)
- Personnel changes (i.e., changes in personnel caused by new requirements)
- Environmental changes (i.e., changes in the operational environment)
- Operational changes (i.e., changes to the user's operational policies, procedures, or methods)
- Support changes (i.e., changes in the support or maintenance requirements)
- Other changes (i.e., a description of other changes that will impact the users)

4.3 Change Priorities

This section should include any prioritization or ranking regarding the proposed changes. The section should define what features are essential, what features are desirable, and what features are optional [5].

4.4 Changes Considered but Not Included

This section describes assumptions or constraints applicable to the changes and new features in this section. This should include all assumptions and constraints that will affect users during development and operation of the new or modified system [5].

5. Concepts for the Proposed System

This section describes the proposed system that results from the desired changes specified in the fourth section of the ConOps document. The format follows the format of the third section to make it easy to understand the role of the proposed system in solving the problem stated in the beginning of the document. This includes a high-level description of the proposed system that indicates the operational features to be provided without specifying design details. Methods of description to be used and the level of detail in the description will depend on the situation. The level of detail should be sufficient to explain how the Proposed System is envisioned to operate in fulfilling user needs and requirements [5].

In some cases it may be necessary to provide some level of design detail in the ConOps. The ConOps should not contain design specifications, but it may contain some examples of typical design strategies for the purpose of clarifying the proposed system's operational details. In the event that actual design constraints need to be included in the description of the proposed system, they should be explicitly identified as requirements to avoid possible misunderstandings [5].

Specifically, the fifth section should include information on the:

- Proposed system's background, objectives, and scope of the system itself and the corresponding performance verification tests
- Operational policies or constraints imposed on the proposed system and the corresponding performance verification tests
- Description of the proposed system and the corresponding performance verification tests
- Modes of operation
- User involvement and interaction
- Support environment

5.1 Background, Objectives, and Scope

An overview of the new or modified system, including the background, mission, objectives, and scope, should be provided, as applicable. In addition to providing the proposed system's background, a brief summary of the system's motivation should be provided. The goals for the new or modified system should also be defined, as well as the strategies, solutions, tactics, methods, and techniques proposed to achieve these goals. The goals should also describe in detail strategies, solutions, tactics, methods, and techniques needed for thorough performance verification tests [5].

5.2 Operational Policies and Constraints

The operational policies and constraints that apply to the proposed system should be also described. This includes, but is not limited to, such elements as hours of operation, staffing constraints, space constraints, and hardware constraints [5].

5.3 Description of the Proposed System

A thorough description of the proposed system should be provided that includes [5]:

- Operational environment and its characteristics

- Major system components and the interconnections among these components
- Capabilities or functions of the proposed system
- Relationship to other system

- Charts and accompanying descriptions that depict inputs; outputs; data flows; and manual and automated processes so the proposed system or situation is sufficiently understood from the user's point of view.
- Cost of system operations
- Deployment and operational risk factors
- Performance characteristics
- Quality attributes, such as reliability, accuracy, availability, expandability, flexibility, interoperability, maintainability, portability, reusability, supportability, survivability, and usability
- Provisions for safety, security, privacy, integrity, and continuity of operations in emergencies

Since the purpose of this section is to describe the proposed system and how it should operate, it is important that the description of the system be simple and clear enough that all intended readers can fully understand it. It is important to keep in mind that the ConOps should be written in the user's language. Graphics and pictorial tools should be used wherever possible. Useful graphical tools include, but are not limited to, the work breakdown schedule (WBS); sequence or activity charts; functional block diagrams; and relationship diagrams.

The description of the operational environment should identify the facilities, equipment, computing hardware, software, personnel, and operational procedures needed to operate the proposed system. This description should be as detailed as necessary to give the readers an understanding of the numbers, versions, capacity, etc., of the operational equipment to be used.

The author or authors of a ConOps should organize the information in this section as appropriate to the proposed system, as long as a clear description of the proposed system is achieved. If parts of the description are lengthy by nature, they can be included in an appendix or incorporated by reference. An example of material to be included by reference might be detailed operations or policy manual [5].

5.4 Modes of Operation

This section should describe the proposed system's various modes of operation. Examples of modes operation include standard, after-hours, maintenance, emergency, training or backup as applicable [5].

5.5 Support Environment

The support and maintenance concepts and environment for the proposed system should be documented. This section should include the support agency or agencies; facilities; equipment; support software; repair or replacement criteria; maintenance levels and cycles; any other areas concerning support environment [5].

6. Operational Scenarios

A scenario is a step-by-step description of how the proposed system should operate and interact with its users and its external interfaces under a given set of circumstances. Scenarios are written in layman's language and should be no technical as much as possible. Scenarios should be described in a way that will allow readers to walk through them and gain an understanding of how all the various parts of the proposed system function and interact. The scenarios tie together all parts of the system, users, and other entities by describing how they interact. Scenarios may also be used to describe what the system should not do.

Pre-deployment system performance verification tests should reflect these operational scenarios in order to verify if the proposed system would support such operational scenarios [5].

Scenarios should be structured so that each describes a specific operational sequence that illustrates the role of the system, and its interactions with users and other systems. Operational scenarios should be described for all operational modes of the proposed system. Each scenario should include events, actions, inputs, information, and interactions as appropriate to provide a comprehensive understanding of the operational aspects of the proposed system [5].

Scenarios are an important component of the ConOps and, therefore, should receive substantial emphasis. Fully presented scenarios should result in an enhanced reader understanding of all benefits resulted by the proposed system. The number of scenarios and level of detail specified will be proportional to the complexity and criticality of the project.

7. Summary of Impacts

This section describes and summarizes the operational impacts of the proposed system from the user's perspective. This section can also include a description of the temporary impacts that can be realized during the development, installation, or training periods. This information is provided to allow all affected end-users (both individuals and agencies) to prepare for the changes that will be brought about the new system, and allow them to plan for any possible future implications and impacts. Implications and impacts can be characterized into several areas, including operational impacts, organizational impacts, and developmental impacts to include any testing issues and anticipated challenges and difficulties [5].

8. Analysis of the Proposed System

This section provides a summary of the benefits, limitations, advantages, disadvantages, alternatives, and trade-offs considered for the proposed system.

Improvements to the system should be documented. This includes a qualitative and quantitative summary of the benefits to be provided by the proposed system, and can include new capabilities, enhanced capabilities, deleted capabilities, and improved performance. In addition, any disadvantages or limitations should also be provided [5].

The major alternatives considered, the trade-offs among them, and the rationale for the decisions reached should be summarized in this section. In the context of a ConOps document, alternatives are operational alternatives and not design alternatives, except to the extent that design alternatives may be limited by the operational capabilities desired in the new system. This information can be useful in determining, now and later, whether a given approach was analyzed and evaluated, or why a particular approach or solution was rejected. This section should describe the proposed system's costs based on assumptions that are clearly stated. Additionally, an approximate schedule for the development should also be included [5].

9. Notes

This section should contain any additional information that will aid in understanding of the ConOps document. If there are not any notes, this section should still be included with a notation that there are not any notes at this time. Subsequent revisions of the ConOps usually require that notes be added [5].

10. Appendices

To facilitate the ConOps' ease of use and maintenance, some information may be placed in appendices to the document. Each appendix should be referenced in the main body of the document where that information would normally have been provided. Appendices may be bound as separate documents for easier handling [5].

11. Glossary

The inclusion of a clear and concise compilation of the definitions and terms used in the ConOps document that may be unfamiliar to readers is important. A glossary should be maintained and updated during the ConOps' concept analysis and development processes. To avoid unnecessary work due to misinterpretations, all definitions should be reviewed and agreed upon by all involved parties [5].

VI. CONCLUSIONS

A. SUMMARY DISCUSSION

This thesis documented the findings of developing a generic testing template and supporting a generic concept of operations for speaker verification technology as part of the Iraqi Enrollment via Voice Authentication Project (IEVAP). In this phase of the IEVAP, NPS developed a generic testing template and testing concept of operations for speaker authentication technology. The intent of this project was to contribute to the future employment of speech technologies in a variety of possible military applications by developing a voice authentication testing template, along with a concept of operations to conduct such testing.

Additionally, this thesis provided information concerning basics of voice recognition technology to include discussions on a number of key voice recognition concepts and definitions. Also, resource provisioning guidelines were included to provide information on how to provision for a minimum hardware and software architecture to ensure the expected quality of service (QoS). Speaker verification and speaker identification performance measures were also provided in relation to a proposed testing template and a generic concept of operations for such testing. Moreover, a high level discussion on concepts and actions related to experiments, analysis, human subjects and their training was also included for thesis depth purposes. Finally, testing protocol discussion delivered some additional information concerning other key items related to testing work.

B. RECOMMENDATIONS FOR FURTHER RESEARCH

Voice recognition offers a plethora of research opportunities for students and industry. The following is a list of recommended further studies for NPS students in support of voice recognition technology.

- Develop a test to assess the performance of the Iraqi-Arabic speaker verification and speech recognition language modules for Phase 1C of the IEVAP.

- Conduct a cost-benefit analysis on the deployment of speaker verification technology in military applications.
- Conduct a comparative analysis of 802.11, 802.16, cellular, and landline technologies in support of the employment of speaker verification technology in military applications.
- Conduct a review and comparative analysis of possible military applications concerning voice recognition technology.

APPENDIX A: TERMS

Acoustic Adaptation: A feature that analyzes task-specific data like recorded utterances and recognition results, and adapts acoustic models accordingly.

Acoustic Confusability: Refers to the closeness of words in the way they sound. Example: 'Newark' and 'New York' are acoustically confusable.

Acoustic Model: Mathematical models representing the various contextual triphones present in speech. Different models are used for different languages (e.g., US English, UK English, Swedish) and/or special environments such as hands-free phones.

Adaptation: A process of enhancing an existing voiceprint during training, using new data.

Algorithm: A sequence of instructions/steps that instructs a computer system what operations to perform.

Ambiguity: In the context of voice application refers to the case where a recognized utterance maps to more than one natural language result in the current grammar.

Barge-in: The ability for callers to interrupt a prompt by speaking.

Batchrec: *batchrec* is a Nuance tool that performs offline recognition on a set of recorded audio files, prints the results, and, if a file is supplied containing transcriptions or nl-transcriptions of the data files, scores the results. *batchrec* also lets developers test dynamic grammar and speaker verification features. *batchrec* is useful for:

- Establishing recognizer accuracy on a known set of audio files.
- Measuring recognizer speed.
- Estimating performance on a live task, in advance.

- Tuning the recognizer configuration while holding the recognition task constant.

Built-in Grammar: A VoiceXML grammar element representing a grammar that is provided directly by the platform.

Correct-Accept In-Grammar (CA-in): An utterance which is covered by the grammar and was accepted by the recognizer.

Call flow: The logical flow of a speech application, including various dialog states, primary paths of informational exchanges, transactional denials, and decision logic, outlined in a flow chart.

Call log: The text file that records all recognition activity performed during a single call session.

Cluster: Two or more hosts running the entire set of Nuance Voice Platform (NVP) services, with each host configured to perform a specific role. Each cluster includes a primary Management Station and one or more other hosts including browser hosts, recognition hosts, resource hosts, audio output hosts, CTI gateway hosts, and V-Server hosts.

Conditional Transfer: Call transfer method similar to a blind transfer, except that the application waits before disconnecting from the two parties until either the third-party line is ringing (without far-end dialog) or they are successfully connected (with far-end dialog).

Configuration Information: Information specified by the application developer when creating a recognition package for verification. The information includes the minimum

and maximum number of utterances that will be processed by the Verifier, as well as the accuracy threshold that will help the Verifier make verification decisions.

Cognitive Load: The informational burden placed on a caller's memory. Often referenced in regards to the limitation present for auditory information delivery. For example, users cannot remember long list of items or listen to long prompts. Similarly, the last instructions provided by the prompts are the ones often remembered.

Confidence Rejection Threshold: Sets the limit in confidence score below which all recognitions are rejected.

Context Sensitive Help: A dialog technique by which help prompts are designed based on the context of the transaction.

Conversation Server: Nuance Voice Platform component including services for voice applications, including VoiceXML interpretation, recognition, verification, and text-to-speech.

Core: Grammar portion containing the most important meaning-bearing words.

Coverage: Refers to all utterances that a grammar contains.

Correct-Reject Out-Of-Grammar (**CR-out)**: An utterance which is not covered by the grammar and was rejected by the recognizer.

Delayed Help: A dialog technique by which help is delivered without the user having to ask for it, usually after they have been given the opportunity to talk.

Diagnostic Log: Text file containing message output generated by a specific Nuance Voice Platform service. Each message in the log represents an event.

Dialog: Interaction between a user and a voice application. A single unit of interaction or single transaction is often referred to as a *dialog state*.

Dictionary: Refers to the file which contains pronunciations for given words specified in phonetic units.

Directed Dialog (system initiative): A dialog technique that prompts the user for each separate piece of information in order to complete a transaction.

Dynamic Grammars: A *dynamic grammar* is a grammar that can be compiled at run time. This is necessary if the complete application grammar cannot be determined until runtime or if the grammar needs to change at runtime. Examples include a personal contact list in a voice-activated dialing application or a database search result.

Echo Cancellation: When an application plays a prompt, a portion of the outgoing energy is reflected in the input channel as an *echo*. This effect is more pronounced with analog telephone lines, but still exists even with digital lines such as T1/E1 or ISDN-PRI since the overall telephone network itself provides a path for reflecting the prompt. *Echo cancellation* improves the quality of a speech signal by diminishing any echo that might have been introduced by the telephone line.

Endpointing: A process used for recognition accuracy and efficiency, it is critical that the system distinguish leading or trailing background noise or silence from the utterance itself before sending it to the recognizer.

Error Handling: Refers to the dialog techniques used to handle errors whether they are emanating from the users or the system.

Escalated Help: Dialog technique to provide more help as the user shows signs of difficulty. Escalation is usually based on the number of errors the user is experiencing in a given state.

External Rule Reference: A means of accessing grammars stored in a file system or on a web server.

False Accept (FA): A mis-recognition in the instance of when a caller makes an utterance that gets incorrectly recognized as something else. Technically, it means that the interpretation returned by the recognizer does not match the one for the transcribed utterance.

False-Accept In-Grammar (**FA-in)**: An utterance which is covered by the grammar but was mis-recognized (i.e. accepted) by the recognizer. This is also referred to as a substitution. Example: The sequence “eight two three” is in-grammar but mis-recognized for “a two three” which is also in-grammar.

False-Accept Out-Of-Grammar (FA-out): An utterance which is not covered by the grammar but was recognized (i.e. accepted) by the recognizer. Example: The word “pizza” is Out-Of-Grammar (OOG) but mis-recognized by “plaza” which is in the grammar.

False Acceptance: A type of verification error occurring when:

- An imposter says the correct information and is recognized
- A true speaker or an imposter does not say the correct information, but the recognizer mistakenly reports that the user spoke the correct information

Fillers: A term which refers to:

- Verbiage added around the important pieces of information
- A mechanism to explicitly exclude certain subgrammars from the confidence scoring mechanism

Flattened/Un-flattened: A parameter that tells *nuance-compile* to create a smaller binary representation of the grammars (un-flattened), which runs slightly more slowly. By

default, grammars are fully expanded, or flattened, during compilation, meaning that whenever a subgrammar is referenced, it is expanded in that location, regardless of where else it might be referenced. In grammars that reference a subgrammar more than once, this leads to binary files that are larger than necessary but that provide optimum recognition performance.

Flexible (user initiative) dialog: Utilizes natural language and allows filling multiple semantic slots in a single utterance while accepting a wide variety of responses. Example: "How can I help you?"

False-Reject In-Grammar (FR-in): An utterance which is covered by the grammar but was rejected by the recognizer. This is sometimes an indicator of threshold sensitivity.

Generate: This Nuance program examines a compiled grammar and traverses possible paths, generating sentences. The program can work with either a top-level grammar or a subgrammar. This tool should be used to help to determine whether the existing grammars provide the correct level of sentence coverage. The Generate program can also be use to generate scripts for data collection experiments, or to test whether particular sentences can be recognized by a grammar.

Graceful error recovery: Refers to the dialog techniques used to recover from errors in a natural and friendly way. These techniques advocate that the system is at fault and then builds the error recovery strategies around that premise. Example: The user is not recognized. The system should say: "I'm sorry I didn't understand" as opposed to "Please speak more clearly."

Grade of Service: The Nuance Grade of Service target is to provide a response to the caller within 2 seconds 95% of the time. Of course, one can configure the response time to be even less and with higher probability.

Grammar: Users' responses must be included in a recognition grammar, which is the collection of all possible utterances at a given point in the call flow. Otherwise, these utterances will be rejected, as they are *out of grammar*. It is absolutely critical that prompts be designed carefully and concurrently with the grammar.

Hypothesis: In the context of Nbest list, this term refers to an item returned in the list.

In-Grammar (IG): The utterance is found in the grammar.

Implicit Confirmation: A dialog technique by which pieces of information are played back to the user without asking them to confirm.

Interpretation: Refers to the meaning associated to a certain expression (entry) in the grammar. Interpretation is both the text and the slots/values pairs returned.

Interactive Voice Response (IVR): This is a somewhat misleading term that typically refers to a platform for creating DTMF-based (touch tone) applications. Today all major IVR vendors have speech recognition integrations, but the majority of IVR applications are still DTMF-based.

Just-in-time Instructions: A dialog technique to deliver help or information based on what the user said. Example: The user asks to go to the trading menu from the main menu. The system informs the users that next time they want a quote, the users can say it directly from the main menu.

Latency: The amount of time between two events. In the specific case of this research there are several aspects of latency:

- Nuance recognition latency is the time between the end-of-speech and the recognition server returning the result to the application.

- Application latency is the time between the application getting the result from Nuance and then performing an external request (database, mainframe host or some other device such as a tape silo) which performs a lookup function and upon return of the data, then act upon it to the caller.
- Network latency is the time for the network to respond and transmit the data whether it is via LAN or WAN etc.

Mixed initiative dialog: A dialog which enables a caller to fill multiple semantic slots with a single utterance. Any slot which has not been filled by the initial utterance will then be filled individually in a directed fashion.

Multi-slot: Refers to a recognition result or grammar entry containing more than one natural language (NL) slot.

Natural Language (NL): Refers to the possibility for the users to use normal sentences and for the recognizer to be able to extract the meaning (i.e. one or more slots filled) from such input.

N-best: Refers to the N responses that can be returned by the recognizer when configured to do so.

Nuance Grammar Builder (NGB): Nuance's GUI Development environment for grammars.

NL Interpretation: The actual values returned by one or more slots.

NL Slot: An entity returned by the recognizer which is a placeholder for the actual values recognized (i.e. interpreted by the grammar). The NL Slots are used by the application to execute the dialog flow and logic.

NL Structure: A construct used in grammars used to return data structure. For example, the date slot is comprised of $date.day and $date.month.

nl-tool: *nl-tool* is a Nuance tool that performs natural language interpretation on sentences. *nl-tool* takes sentences from standard input; developers or administrators can enter sentences one at a time and see the resulting interpretation(s). The developers can also use *nl-tool* in batch mode by creating a file with a list of sentences and using input redirection. *nl-tool* prints out the interpretation(s) for the sentence, as well as the number or words used in creating each interpretation.

Nuance Standard Grammars: Standard grammars that are delivered with the Nuance software and tools.

Nuance-resources file: A file in which parameters and/or contexts are defined for a specific grammar package.

Nuance-resources.site file: A file in which parameters are defined for all applications running for a specific nuance installation.

Open Development/Deployment Platform (ODP): Also known as the Nuance reference platform. This is an alternative to IVR platforms, consisting of either a Windows NT or Unix system with Dialogic or Natural Microsystems telephony cards.

Out-Of Coverage (OOC): Refers to utterances that technically cannot be covered by a grammar. Examples are noise, silence etc.

Parse-tool: A Nuance tool that tests whether a sentence can be parsed correctly by a grammar. By default, *parse-tool* tries to parse the sentence with any grammar in the specified package, whether that grammar is top-level or not. One can also explicitly specify the grammar to parse against.

Path: Refers to a specific location in a file system tree.

Persona: The consistent character that is captured by the voice and audio environment of a voice-enabled application. It is the "face" of the experience for the user.

Probability: In the Nuance context, refers to the weight assigned to utterances or groups of utterances. Setting probabilities is a task adaptation technique used to improve recognition accuracy.

Prompts: A system's dialog design is ultimately governed by carefully worded prompts and users' responses to them. Prompts consist of pre-recorded or synthesized text-to-speech messages played to either elicit a response from the caller or to deliver information to them. The intimate connection between the wording of prompts and users' responses cannot be underestimated. If, during any part of the application development process, a developer changes a prompt, a corresponding change to the grammar is required since this change will likely cause users to respond differently.

Provisioning: The calculation of how many particular computers are required to perform the speech processing in the customer's proposed system (recognition, verification, TTS)

Recognition Client: Also known as the RecClient, this is the process that handles the interaction between an application and the Nuance System. The RecClient manages audio input and output (typically over telephone lines). The RecClient supports limited call control capabilities and provides the interfaces that you call to invoke Nuance recognition services. Speech application developers use one of the available APIs that access the RecClient.

Recognition Engine: Also known as "recognizer" this term refers to the recognition algorithms in general.

Recognition Package: All of the compiled grammars for a specific acoustic model.

Recognition Parameters: Those parameters that affect the behavior of the recognizer, RecClient, etc.

Recognition Search Space: The set of all possible phrases and pronunciation specified by the current grammar and dictionary. In examining these possibilities, the recognizer uses a hierarchy of search mechanisms that allow it to select the most likely hypothesis for recognizing incoming speech from this set of possible hypotheses.

Recognition Server: Also known as the RecServer, this is the process that performs recognition and natural language interpretation of utterances, as requested by an application via a RecClient. Speech application developers will not access the RecServer directly; instead, they use one of the Nuance APIs to the RecClient which, in turn, requests services from the RecServer. Alternatively, the developer can use an IVR interface, which will then access the Nuance System. In most cases, integration developers use one of the RecClient interfaces to indirectly access the RecServer.

Recognition State: A state in which the recognizer is active and usually consists of a prompt, followed by recognition.

Resource Manager: The Nuance Resource Manager performs real-time load balancing. It ensures that recognition and verification tasks are distributed evenly across the available RecServers, thus reducing hardware requirements and improving the quality of service. The Resource Manager is also the key component for fault tolerance. If a RecServer becomes disabled, the Resource Manager will stop sending recognition requests to it. All RecClients and RecServers connect to the Resource Manager. The Resource Manager keeps track of the recognition packages supported by each server, monitors the load on each server, and allocates an appropriate server for each recognition request.

Sample dialog: An excerpt of dialog based on the dialog specification or call flows.

Shortcuts: A mechanism by which the user can bypass certain dialog states to get a transaction completed. For example: Main menu commands accessible in sub branches of the dialog, ability to trade from the main menu even though this functionality is available in a separate state.

Skip List: Refers to the algorithms or strategies used to eliminate choices from a list based on a decision criterion.

Slot Definitions File: A file used by the Nuance compiler that lists all slots used in the package and can be returned by the recognizer.

Slot Name: The name of the slot defined in the grammar package. This is the key that the application uses to associate meaning to language (i.e. retrieve values from the recognition result).

SpeechObjects: These are open, reusable and customizable application components, which facilitate application development to encapsulate discrete pieces of conversational dialog to allow users to focus on the user interface. That is, the dialog for a speech application rather than the underlying interactions with the recognition engine. SpeechObjects use a SpeechChannel object to access the recognition client functionality.

Subgrammar: A grammar construct that can be referred by other top-level or sub-grammars.

Task Adaptation: Refers to the various advanced tuning techniques enhancing recognition performance based on an application specific caller population, channel, utterances distribution etc. Some task adaptation techniques are the usage of probabilities in grammars, acoustic model tuning, and statistical language modeling. These techniques all require large amounts of field data to be statistically valid.

Telephony Control: The IVR System provides basic telephony functionality when used with a Dialogic, Natural Microsystems (NMS), Aculab card or some other proprietary card. For telephony support, a wide variety of IVR toolkits and systems are available commercially as well Nuance can provide telephony functionality which includes:

- Placing a call
- Answering the phone
- Detecting hang-up
- Detecting Dual Tone Modulation Frequency (DTMF) tones
- Transferring calls
- Setting up trombone calls (a limited form of conferencing)

Text-to-Speech (TTS): The ability of a computer to play back written text as audible speech.

Top-level Grammar: In static grammars, this term refers to a grammar preceded by a dot. Example: .GetAccountNumber […]

Transcription: The act of listening to the recording of the spoken words and keying their Phonetic representations into a text file. These transcriptions are then used to help tune the application and to help determine errors.

Universals (Globals): Commands that the user can say throughout the application. Common examples of such commands are: help, repeat, and operator.

Utterance: The sounds that make up words. There are approximately 44 different phonemes (English language) that when strung together in different sequences make up the words of the language. A phrase uttered by a caller at a given state in a speech application. These phrases are recorded and saved as audio files (or “wav files”) on recognition client machines during the tuning phases of a project.

Voice Portal: A single access point via the telephone providing access to an aggregated set of services (information, commerce, communication, etc.).

Voice Site: A node on the voice web that contains voice-enabled content accessible via the telephone.

VoiceXML: An emerging standard markup language for creating voice applications.

VoiceXML Interpreter: Software that interprets VoiceXML markup language and generates a voiced dialog.

Voice Web: A Network of voice portals and voice sites that people can access from any telephone.

Voice Web Server: A Nuance software bundle to support VoiceXML based voice sites or voice portals (includes Nuance's VoiceXML Interpreter, Nuance 7.0 ASR and optional Nuance Verifier and Vocalizer)

Voice Service Provider (VSP): Analogous to ASPs or ISPs, these are companies that host a range of voice applications.

APPENDIX B: NPS SAMPLE CONSENT FORMS, LETTERS, AND PRIVACY ACT STATEMENT

To: Protection of Human Subjects Committee

Subject: APPLICATION FOR HUMAN SUBJECTS REVIEW FOR THE IRAQI ARABIC INTERACTIVE VOICE RESPONSE SYSTEM.

1. Attached is a set of documents outlining a proposed experiment to be conducted over the next nine months in support of the Office of Secretary of Defense (OSD) sponsored project.

2. We are requesting approval of the described experimental protocol. An experimental outline is included for your reference that describes the methods and measures we plan to use.

3. We include the consent forms, privacy act statements, and debriefing forms we will be using in the experiment.

4. We understand that any modifications to the protocol or instruments/measures will require submission of updated IRB paperwork and possible re-review. Similarly, we understand that any untoward event or injury that involves a research participant will be reported immediately to the IRB Chair and NPS Dean of Research.

Very Respectfully,
James Ehlert

APPLICATION FOR HUMAN SUBJECTS REVIEW (HSR)	**HSR NUMBER *(to be assigned)***
PRINCIPAL INVESTIGATOR(S): Mr. James Ehlert, Research Associate, Information Science Department, (831) 656-3002	
APPROVAL REQUESTED [X] New [ ] Renewal	
LEVEL OF RISK [ ] Exempt [ ] Minimal [X] More than Minimal Justification: Human subjects will be required to demonstrate the use of an Arabic language-based voice-activated menu-driven phone system and provide feedback on the functionality of the phone system.	
WORK WILL BE DONE IN:	ESTIMATED NUMBER OF DAYS TO COMPLETE:

DLI Foreign Language Center Bldg 420 Defense Language Institute Monterey, CA 93943	273 Days (01 Jan – 30 Sep 06)
MAXIMUM NUMBER OF SUBJECTS: 700 (DLI Arabic Language Instructors and Arabic students)	ESTIMATED LENGTH OF EACH SUBJECT'S PARTICIPATION: 20-30 minutes
SPECIAL POPULATIONS THAT WILL BE USED AS SUBJECTS [ ] Subordinates [ ] Minors [ ] NPS Students [X] Special Needs Arabic Linguists	
OUTSIDE COOPERATING INVESTIGATORS AND AGENCIES [ ] A copy of the cooperating institution's HSR decision is attached.	

TITLE OF EXPERIMENT AND DESCRIPTION OF RESEARCH.

1. The title of the research is "Testing and demonstrating speaker verification technology in Iraqi-Arabic as part of the Iraqi Enrollment via Voice Authentication Project (IEVAP) in support of War on Terrorism (WOT) security requirements."

2. The purpose of this research is to create a pilot system using existing commercial off the shelf (COTS) technologies in order to help manage detention visitation at the Baghdad Central Correction Facility (BCCF).

This system will serve as a proof-of-concept (POC) system in the demonstration and pilot evaluation of an Arabic voice-activated menu-driven phone system using existing COTS interactive voice response (IVR) technology in order to expedite a visitor's entry to a controlled facility/secure space (Baghdad Central Correction Facility).

This research is a continuation of series of investigations into the application of voice technologies by developing a POC system that integrates IVR technology into a mobile platform in order to demonstrate a functionality that does not exist in order to meet war-fighter requirements.

Specifically, this research is intended to contribute toward the future employment of voice authentication technologies in a variety of coalition military operations. The value added from this research includes:

- Demonstrating the viability of this technology for subsequent research and development.
- Selecting the most appropriate hardware, software, and peripherals for a mobile demonstration kit (laptop, voice input devices, etc) for implementing IVR technology.

3. The demonstration and testing of the POC system will be conducted in coordination with linguistic support from the Defense Language Institute (DLI). The subject population of this research will consist of Arabic Language Instructors and Arabic language students at the DLI. The test and evaluation of this POC system will be in completed in two phases. Use of human subjects will be required in both phases of this research. Human subjects will be required to demonstrate the use of the Arabic voice-activated menu-driven phone system and provide feedback on the functionality of the phone system. Listed below are the testing milestones.

a. Test Phase 1: Test the voice-activated menu-driven phone system in Iraqi Arabic (01 Mar – 05 Mar 06). Utilizing Nuance IVR software, and with Arabic language support from the DLI, the goal is to test the voice-activated menu-driven phone system in Iraqi Arabic by having human subjects enroll (one time) and verify their identity (four times) via a telephone.

b. Test Phase 2: Test the voice-activated menu-driven phone system in Iraqi Arabic (01 – 05 Apr 06). Utilizing Nuance IVR software, and with Arabic language support from the DLI, the goal is to test the voice-activated menu-driven phone system in Iraqi Arabic by having human subjects verify their identity (five times) via a telephone.

I have read and understand NPS Notice on the Protection of Human Subjects. If there are any changes in any of the above information or any changes to the attached Protocol, Consent Form, or Debriefing Statement, I will suspend the experiment until I obtain new Committee approval.

SIGNATURE_____ DATE_____

PARTICIPANT CONSENT FORM

1. **Introduction.** You are invited to participate in a study on the demonstration of the Iraqi Arabic Interactive Voice Response System. With information gathered from you and other participants, we hope to demonstrate the use of an Iraqi Arabic voice-activated menu-driven phone system using existing COTS interactive voice response (IVR) technology in order to expedite a visitor's entry to a controlled facility/secure space. We ask you to read and sign this form indicating that you agree to be in the study. Please ask any questions you may have before signing.

2. **Background Information.** The Naval Postgraduate School's Voice Authentication Technology Research Group is conducting this study.

3. **Procedures.** If you agree to participate in this study, the researcher will explain the tasks in detail. There will be 10 required sessions with each session lasting 1-2 minutes: User will test and evaluate the proof-of concept system by calling into IVR phone system during which you will be expected to accomplish a number of tasks related to appointment scheduling using your Arabic language capabilities.

4. **Risks and Benefits.** This research involves no risks. The benefits to the participants are gaining techniques for the demonstration of this technology for subsequent research and development.

5. **Confidentiality.** The records of this study will be kept confidential. No information will be publicly accessible which could identify you as a participant.

6. **Voluntary Nature of the Study.** If you agree to participate, you are free to withdraw from the study at any time without prejudice. You will be provided a copy of this form for your records.

7. **Points of Contact.** If you have any further questions or comments after the completion of the study, you may contact the research supervisor.
8.
9. **Statement of Consent.** I have read the above information. I have asked all questions and have had my questions answered. I agree to participate in this study.

---------- ----------
Participant's Signature Date

---------- ----------
Researcher's Signature Date

MINIMAL RISK CONSENT STATEMENT

NAVAL POSTGRADUATE SCHOOL, MONTEREY, CA 93943

Participant: VOLUNTARY CONSENT TO BE A RESEARCH PARTICIPANT IN: THE DEMONSTRATION OF AN IRAQI ARABIC INTERACTIVE VOICE RESPONSE SYSTEM.

1. I have read, understand and been provided "Information for Participants" that provides the details of the below acknowledgments.

2. I understand that this project involves research. An explanation of the purposes of the research, a description of procedures to be used, identification of experimental procedures, and the extended duration of my participation have been provided to me.

3. I understand that this project does not involve more than minimal risk. I have been informed of any reasonably foreseeable risks or discomforts to me.

4. I have been informed of any benefits to me or to others that may reasonably be expected from the research.

5. I have signed a statement describing the extent to which confidentiality of records identifying me will be maintained.

6. I have been informed of any compensation and/or medical treatments available if injury occurs and is so, what they consist of, or where further information may be obtained.

7. I understand that my participation in this project is voluntary; refusal to participate will involve no penalty or loss of benefits to which I am otherwise entitled. I also understand that I may discontinue participation at any time without penalty or loss of benefits to which I am otherwise entitled.

8. I understand that the individual to contact should I need answers to pertinent questions about the research is Mr. James Ehlert, Principal Investigator, and about my rights as a research participant or concerning a research related injury is Dr. Jurgen Sottung, Program Manager, Language Science and Technology, Defense Language Institute. A full and responsive discussion of the elements of this project and my consent has taken place.

_____

Signature of Principal Investigator Date

_____

Signature of Volunteer Date

_____

Signature of Witness Date

PRIVACY ACT STATMENT

NAVAL POSTGRADUATE SCHOOL, MONTEREY, CA 93943

PRIVACY ACT STATEMENT

1. Purpose: The purpose of this research is to create a pilot system using existing commercial off the shelf (COTS) technologies in order to help manage detention visitation at the Baghdad Central Correction Facility (BCCF). This system will serve as a proof-of-concept (POC) system in the demonstration and pilot evaluation of an Iraqi Arabic voice-activated menu-driven phone system using existing COTS interactive voice response (IVR) technology in order to expedite a visitor's entry to a controlled facility/secure space.

2. Use: Data collected from this research will be used for statistical analysis by the Departments of the Navy and Defense, and other U.S. Government agencies, provided this use is compatible with the purpose for which the information was collected. Use of the information may be granted to legitimate non-government agencies or individuals by the Naval Postgraduate School in accordance with the provisions of the Freedom of Information Act.

1. Disclosure/Confidentiality:

   a. I have been assured that my privacy will be safeguarded. I will be assigned a control or code number which thereafter will be the only identifying entry on any of the research records. The Principal Investigator will maintain the cross-reference between name and control number. It will be decoded only when beneficial to me or if some circumstances, which is not apparent at this time, would make it clear that decoding would enhance the value of the research data. In all cases, the provisions of the Privacy Act Statement will be honored.

   b. I understand that a record of the information contained in this Consent Statement or derived from the experiment described herein will be retained permanently at the Naval Postgraduate School or by higher authority. I voluntarily agree to its disclosure to agencies or individuals indicated in paragraph 3 and I have been informed that failure to agree to such disclosure may negate the purpose for which the experiment was conducted.

   c. I also understand that disclosure of the requested information, including my Social Security Number, is voluntary.

_____

Name, Grade/Rank (if applicable) DOB SSN
[Please print]

_____

Signature of Volunteer Date

LIST OF REFERENCES

1. Cisco website: www.cisco.com (accessed Mar 2006)
2. Code of Best Practice for Experimentation, David S. Alberts, DOD, 2005
3. Complexity Theory, James Moffat, DOD, 2003
4. Extreme Networks website: www.extremenetworks.com (accessed Mar 2006)
5. Florida's Statewide Systems Engineering Management Plan, State of Florida, Mar 2005
6. Proof of Concept: Iraqi Enrollment Via Voice Authentication Project, NPS Master's Thesis, Samuel Lee, Sep 2005
7. Kemp Technologies website: www.kemptechnologies.com (accessed Mar 2006)
8. Natural Microsystems website: www.nmscommunications.com (accessed Mar 2006)
9. NPS Human Subject Review, NPS Instruction
10. Nuance Speech University, Introduction to Speech Recognition, Nuance Inc., Sep 2005
11. Nuance Voice Platform High Availability Engineering Guide, Nuance Inc., Nov 2004
12. Nuance Voice Platform, Verifier Developer's Guide, Nuance Corporation, Nov 2004
13. Phase 1C Iraqi Arabic Language Voice Verification Accuracy Estimation Test Plan, NPS, Dr. Pat Sankar, Dec 2005
14. Sun Microsystems website: www.sun.com (accessed Mar 2006)
15. Webster Third International Dictionary, 2002

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Marine Corps Representative
   Naval Postgraduate School
   Monterey, California

4. Director, Training and Education
   MCCDC, Code C46
   Quantico, Virginia

5. Director, Marine Corps Research Center
   MCCDC, Code C40RC
   Quantico, Virginia

6. Marine Corps Tactical Systems Support Activity (Attn: Operations Officer)
   Camp Pendleton, California

7. Dan Boger
   Naval Postgraduate School
   Monterey, California

8. James F. Ehlert
   Naval Postgraduate School
   Monterey, California

9. Pat Sankar
   Naval Postgraduate School
   Monterey, California

10. Gerry Christman
    OSD-NII
    Washington, District of Columbia