AFRL-IF-RS-TR-2004-259
Final Technical Report
September 2004

AIR FORCE DUAL-USE SCIENCE & TECHNOLOGY TWO-WAY VOICE-TO-VOICE TRANSLATOR

Integrated Wave Technologies, Incorporated

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

STINFO FINAL REPORT

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2004-259 has been reviewed and is approved for publication

APPROVED: /s/

STEPHEN E. SMITH
Project Engineer

FOR THE DIRECTOR: /s/

JOSEPH CAMERA, Chief
Information & Intelligence Exploitation Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this 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	3. REPORT TYPE AND DATES COVERED
	SEPTEMBER 2004	Final Apr 02 – Oct 03

4. TITLE AND SUBTITLE
AIR FORCE DUAL-USE SCIENCE & TECHNOLOGY TWO-WAY VOICE-TO-VOICE TRANSLATOR

5. FUNDING NUMBERS
C - F30602-02-2-0050
PE - 62805F
PR - 459E
TA - AT
WU - FT

6. AUTHOR(S)
Timothy McCune

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Integrated Wave Technologies, Incorporated
4042 Clipper Court
Fremont California 94538

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/IFEC
525 Brooks Road
Rome New York 13441-4505

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
AFRL-IF-RS-TR-2004-259

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: Stephen E. Smith/IFEC/(315) 330-7894/ Stephen.Smith@rl.af.mil

12a. DISTRIBUTION / AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

This project demonstrated hardware and software technology to produce a miniature two-way voice translator capable of operating in high noise environments. This effort built upon Integrated Wave Technology's work with a miniaturized one-way translator. This involved adding separate audio channels for second microphones and speakers. Separate channels allow the circuit board to process the input from the interviewer's microphone and to play audio directly to the interviewee. The system was designed to be handheld or pocket-mounted (for eyes-free, hands-free operation), to consume a fraction of a watt of power, and to work in high background noise with high accuracy. Speaker independence for the Arabic language was also a goal of this research. There was marginal success due to limited amounts of speech for building universal templates in this language. The accuracy and utility of this device has been demonstrated and documented in this report.

14. SUBJECT TERMS
Speech Recognition, Machine Language Translation, Synthesis, Pattern Recognition, Real-Time Software/Hardware, Computer Interface, Error Correction

15. NUMBER OF PAGES
30

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Table of Contents

List of Figures

Summary

This project was to demonstrate hardware and software technology to produce a miniature two-way voice-to-voice translator capable of operating in high noise environments. This effort built upon IWT's work with a miniaturized one-way translator.

IWT's translators are built upon an optimized pattern recognition system. IWT's one-way voice-to-voice technology has achieved important results, according to testing by the Naval Air Warfare Command's Training Systems Division (NAWC-TSD). These tests were funded by the National Institute of Justice (NIJ) and the Special Operations Command (SOCOM). For field use in high-noise environments, IWT's one-way Voice Response Translator (VRT) has achieved better accuracy and noise immunity than other available devices, according to the NAWC-TSD tests. NAWC-TSD also stated the VRT was also the only device known to have eyes-free, hands-free capability.[1]

IWT's technical approach is to co-develop hardware and software to maximize the key technical objectives of accuracy, noise immunity and low power consumption. Implementation of this technology is on a proprietary circuit board that contains the recognition unit, the sound amplifier and battery charging control. Code is written for relatively small processors, in this case a 33MHz ATMEL ARM. The recognition software is written to allow the processor to run as slowly as possible, reducing power consumption without lowering performance.

Concurrently with this effort, IWT made changes to the one-way translator board as a new production run was beginning. This involved adding separate audio channels for second microphones and speakers. Separate channels allow the circuit board to process the input from the interviewee's microphone and to play audio directly to the interviewee.

The initial milestone for the two-way effort was to produce a one-way, speaker-independent unit retaining IWT's noise immunity and accuracy. The second milestone was operating and demonstrating a dual-handset system with acceptable overall system accuracy and noise immunity. The third milestone was incorporation of an initial Q&A software architecture that allows for speaker-independent, foreign language responses in realistic background noise settings. The fourth milestone was the inclusion of a mission-driven, expanded set of phrases to allow for a useful Q&A. The fifth milestone was operational testing of the developmentally tested device.

[1] User feedback for the one-way device is included as Appendix B.

Development to reach the first milestone was successful. IWT's integrated recognition board provides a potentially effective basis for this limited two-way voice-to-voice translation. Further, the basic recognition algorithm, as implemented in the new two-way application, worked for this purpose.

Discussions with military and law enforcement users determined the dual-handset approach would not be useful in field applications. Users stated this form factor could be used as a weapon against the person conducting the interview. User groups stated IWT's current Voice Response Translator form factor, modified with an additional microphone for two-way translation, was preferable. For that reason, the dual handset form factor was abandoned and a modification of the existing form factor used. This had the additional benefit of allowing for hands-free, one-way voice-to-voice communications by the user.

The third milestone was reached successfully. The application as developed was able to work in high (greater than 95dB) background noise for the short two-way responses.

Reaching the fourth milestone was frustrated by difficulties in obtaining sufficient quantities of Arabic foreign language utterances, particularly for female speakers. Official Arabic sources were in short supply for this research. Language samples from private sources were obtained, but availability and quality was limited. While the samples obtained achieve reasonable accuracy in some tests, overall accuracy was lower than expected.

Our inability to reach Milestone Four prevented our deploying the device operationally and reaching Milestone Five. IWT is continuing this effort with other US Government entities and will provide regular activities reports to AFRL.

Program Discussion:

Introduction

This program centered on developing and demonstrating a capability to provide miniaturized, robust, limited two-way voice-to-voice translation. Several more elaborate systems have been developed by other companies and achieved varying levels of effectiveness, but these rely on laptop or much larger computers.

IWT's system is designed to be handheld or pocket-mounted (for eyes-free, hands-free operation), to consume a fraction of a watt of power, and to work in high background noise with high accuracy. There are also other efforts to build small two-way, voice-to-voice translators, and IWT's efforts are to develop performance levels to make a two-way VRT competitive with these other options.

Background

This project was to build an advanced field translator with two-way capabilities. The proposal for this project responded to a stated military requirement for voice-to-voice portable language translation. The requirement flows both from the increased level of operations among Coalition forces as well as the diverse groups of civilians encountered by US forces. The rapidly changing language requirement, combined with the variety of personnel exposed to non-English-speaking persons, means that traditional language familiarization training will not be able to cope with the US military's language translation requirement.[2]

Some description of IWT's current work with its one-way translator is useful. IWT has developed specialized speech recognition technology that will allow it to develop a device for evaluation. In some tests done by NAWC-TSD, the accuracy of the system exceeds 99% in adverse conditions and in the presence of background noise.[3]

Continuing evaluation work performed by the Naval Air Warfare Command's Training System Division (NAWC-TSD) has shown IWT's

[2] The Defense Language Institute has been working closely with IWT and provides content for its one-way translator as a way of supplementing its traditional language work.

[3] These results were obtained when more effective microphones were used with the VRT in place of ones previously supplied with it. Microphones of these capabilities are now standard with the VRT. NAWC-TSD also replaced voice commands with low sound energy with more effective ones. This also is now standard VRT practice.

technology to be superior to other voice-to-voice translators in key areas.[4] NAWC-TSD is performing these evaluations for the Justice Department's National Institute of Justice and the Special Operations Command.

NAWC-TSD compared the VRT with the DARPA-developed Phraselator and the Ectaco UT-103. Each of the devices demonstrated unique performance characteristics. For the VRT, superior areas were accuracy, performance in high noise environments and near-instant response after a voice command was given.

The 2003 NAWC-TSD report stated that the VRT achieved 98 percent accuracy in 95dB of ambient noise, the highest level tested.

"After considering the results of the test and unit behavior during testing, the VRT seems to be the easiest, least intrusive to use device," it said. "The main advantage of this unit is the fact that no user intervention is required for operation. After turning the unit on, setting the phrase group (Coast Guard in this case), the user simply talked to the unit."

"The other translator devices, with push-to-talk, and the GUI (Phraselator) required operator intervention. The VRT was also the fastest unit, with response times of less than a second," according to the NAWC-TSD report. "Testing for the VRT proceeded faster than with the other two units as it was easier to determine if the phrase was recognized (i.e., the unit responded immediately).

"All VRT responses occurred in about 1 second or less," the report said. Phraselator response time was between four and five seconds, and the UT-103 was just over three seconds on average.

There are various ongoing programs to build laptop and smaller-sized two-way voice-to-voice translators. Some examples are: a system by S-Minds, funded by INSCOM; a system by Language Systems, Inc. (LSI), funded currently by the National Institute of Justice and the Office of Naval Research; a system by Carnegie Melon/Lockheed Martin funded under an Army Chaplains program; and a system by Voxtec, funded by DARPA. The S-Minds, CMU/LM and LSI systems are based on high-end, commercially available laptop computers, while the Voxtec system is based on developed/integrated hardware.

IWT's two-way system would be distinguished from other efforts in a manner similar to the way its one-way system is different from other one-way

[4] NAWC-TSD issues reports periodically on these devices. Specific information cited in this report comes from the most recently issued report, "Support for Prototype Development and Results of Initial Field Testing," Prepared for: Office of Science & Technology, National Institute of Justice, By: Naval Air Systems Command, Training Systems Division – Orlando, 12350 Research Parkway, Orlando, Florida 32826, December 2003." A new report is due in August 2004.

translators. On one hand, IWT's systems are very simple to use. The one-way VRT has only a single button[5], an on/off switch and a microphone for operation. The trade-off in not using a keyboard or screen as a back-up interface limits effectiveness in some complex situations. Field personnel thus might use a variety of available one-way and two-way translators depending on specific mission/task requirements.

The IWT two-way system retains the simplicity and effectiveness of its one-way system. Noise immunity, accuracy and ultra-low (one-tenth of a watt) power consumption are important performance characteristics. Within the operational parameters described below, the system would be effective for limited two-way translation. The system should be more suitable in tactical situations where the size, power consumption and lesser noise immunity are a hindrance to laptop-based devices.

The device described in this report builds upon the successful aspects demonstrated in the one-way translator. Such a device would have the following characteristics:

1) Recognition Accuracy. Near 100 percent accuracy is needed for users to communicate effectively. Substitution errors will degrade operations severely and greatly diminish user confidence in the device. The goal of this effort was to produce a device with near 100 percent accuracy.

2) Noise Immunity. Ambient noise in "quiet" locations such as offices is often 20dB to 50dB because of equipment, ventilation and other occupants. Noise in field areas and cities is often over 100dB. The goal of this effort was to produce a device able to operate in these environments.

3) Miniaturization. The device described in this report would, when developed fully, be carried routinely by users and must be as small as possible to provide this capability without being burdensome. The goal of this effort was to produce a translator with a system weight of less than one pound. Final test system weight was about 12 ounces.

[5] Functions controlled by this button are: erasing all of a user's trainings; playing an "Emergency Phrase" without using a voice command; instructing the unit to go "on standby"; retraining just the initial group of commands; and determining which language/phrase group is active.

4) Ease of Use. The goal of this effort was to build a device users can learn to operate with minimal training, perhaps received by viewing a simple video or reading a very short manual. The degree to which the device is accepted for widespread use is related significantly to its ease of use.

5) Ease of Language Upgrade/Revision. Languages encountered by law enforcement, military, social worker and tourist users will change. Users will need to add languages to the unit and to add to or revise phrase sets already included in the unit. The goal of this effort was to build such a device

6) Low Power Consumption. The unit will be operated by most users away from line power. Consistent with the requirement to miniaturize the device is a need to keep battery size as small as possible, and allowances for this should be no more than several ounces. The goal of this effort was to product a device that uses small batteries and is able to operate for long periods of time free from line power.

7) Ruggedness. The translator will be used routinely in austere environments and will certainly be dropped occasionally and transported roughly. In addition, it will be subjected to moisture and heat. The goal of this effort was to create a device that is militarily and commercially viable by being able to withstand such treatment for a period of years.

8) Low Eventual Production Cost. The ability of military and commercial customers to acquire this translator will be related to its eventual production cost. The goal of this effort was to design a unit that relies on low-cost processors and circuitry that implement elegant, innovative design solutions.

One DUST program factor is the probability of commercial markets for this device. These include:

- Law enforcement officers in the US and other countries;
- Other government personnel dealing with non-English-speaking persons on a routine basis such as emergency medical personnel, social workers, firemen, building inspectors, aid workers, and workplace safety personnel;

- Commercial enterprises dealing on a routine basis with persons not speaking English; and

- Governments and companies overseas dealing with foreign languages.

The primary technical risk in this project is the transfer of the lab-tested speaker independent/noise immune/language adaptable algorithm to a field-useful system. IWT's development work indicated this could be done by making a collective analysis of a template group, and that in operation there would be no lowering of noise immunity or increasing real-time response.

An overall system risk concerns the operational effectiveness of the voice-to-voice translator concept. The English-to-other language phrases must convey the intended meaning from the interviewer to the interviewee. IWT drew upon the experience of persons who have used two-way, voice-to-voice translation to provide an initial phrase architecture.

Possible Benefits of a Two-Way Translator

There are several benefits a two-way voice-to-voice translator – as described in this report – might produce. Operational testing will be needed.

Possible benefits include:

- Higher levels of response enabled

- Enhanced user acceptance

- Quicker, more accurate feedback from interview subjects for simple items

- More effective form-driven questioning

- Greater interaction between interviewer, interview subject

Difficulties in Implementing Two-Way Translation

The one-way VRT is designed for relatively simple tasks. The VRT allows users to: Identify themselves and their mission; Tell subjects what they're doing;

Tell them what they want to do; and To ask simple questions that can be answered non verbally.

The task of the two-way translator is more complicated. Even simple questions such as `How are you?' can produce a large number of formal, informal, structured, unstructured responses. Even within a language, there are wide variations for dialects, slang/idiom and relationship of the interview subject to the interviewer.

Form factors also present user issues. This effort envisioned a dual-handset device joined by a connecting cable.[6] IWT conducted discussions with the following organizations to collect feedback concerning the form factor and overall architecture:

- U.S. Coast Guard
- U.S. Special Operations Command
- U.S. Marine Corps
- Naval Special Warfare Group Four
- Metropolitan Nashville Police Department
- Anaheim Police Department
- Lexington (KY) Police Department
- Arcadia (CA) Police Department
- Naval Air Warfare Center Training Systems Division
- National Institute of Justice

Data were collected through discussions with persons acquiring, testing and/or deploying IWT's one-way translator. No one favored the dual-handset form factor. Several reasons were given:

1. Users feared that the cable connecting the handsets might be used as a weapon. Though releases/quick disconnects could be employed to detach the cables, feedback stated that even a detached cable would be a problem.
2. Users stated that having the dual-handset form factor would be awkward and bulky.
3. Users said most often that they would like the device to be convertible to a one-way translator and have a form-factor similar to the existing one.

[6] Wireless connections are problematic for devices deployed with military forces as frequencies for these connections vary in some countries.

Based on this feedback, we proceeded with a form-factor based upon the current translator design. Either a single microphone can be plugged in the place of the headset[7] or the current headset can be used with a second microphone for the interviewee. The circuit board was adapted to accommodate two microphone channels.

Architecture Design

IWT's basic speech recognition technology is speaker dependent. Building a two-way translator based on this would be awkward at best. The proposal for this project described a method for creating speaker-independent templates for selected foreign-language phrases.

IWT first experimented with this approach 10 years ago in Japan. Native Japanese speakers serving as test subjects trained their voices using the IWT system to record numbers as templates.[8] The system then held these templates – numbers from zero to nine – active. Other native Japanese speakers whose voices were not trained on the system then spoke these words. In limited tests using only three non-trained persons, accuracy was 100 percent.

The next test of speaker-independent capability on IWT's systems was the incorporation of speaker-independent user selection for the one-way VRT. IWT's tests of this feature showed it had nearly 100 percent accuracy when only words within the domain (digits one through eight) were spoken, but out-of-context errors occurred for "eight" and "two", which are characterized by vowel sounds surrounded by weak fricatives.[9]

Our conclusion was that optimal command size for VRT operation is three-to-five syllables, but one-syllable commands can be used if the environment is structured to avoid out-of-context utterances.[10]

A key design feature in voice response systems – ranging from prototype two-way translators to telephone-based dialogue devices built by different companies – is expected/directed response. Many service companies such as Federal Express and United Airlines have created Automatic Speech Recognition

[7] A jack connection would replace the current hard-wire installation.
[8] This was a DOS-based system using a 486-processor laptop.
[9] These specific tests involved 10 new users selecting user numbers 12 times each.
[10] A system design constraint placed on the VRT by NIJ was that it not false trigger and that a push-to-talk switch not be used to accomplish this. NAWC testing and use by Special Operations units and the Defense Language Institute have confirmed that for the one-way version, false triggering is at or near zero.

(ASR) systems that work well for routine tasks such as tracking/rerouting packages and obtaining the status of lost luggage.

These systems rely on users following explicit instructions describing available options and the phrases needed to provide or obtain information. Instruction phrases appear usually to run about seven syllables to provide sufficient discrimination among options, the "yes" and "no" are often presented as a domain of two. Data entry includes numerals and cities, and while these can be as short as a single syllable, domains are short and defined well.

These telephone-based ASR systems have the advantage of being able to use whatever level of computing resources is required, size and power consumption not being design constraints. The results they achieve are impressive considering that a wide variety of input devices and connections are used.

Generally, these systems do not have to work in high levels of ambient noise. Also, accents can cause problems with such systems, as with any speaker-independent software.[11]

Law enforcement-military two-way voice-to-voice systems such as Language Systems, Inc.'s (LSI) "SpeechTrans" and "CopTrans" products use "expected response" to limit domains and phrase construction.

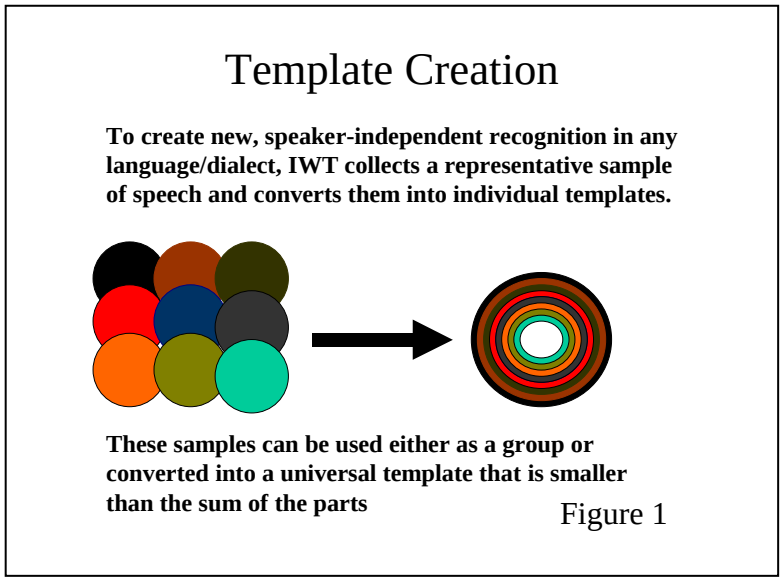

Figure 1

IWT's challenge was to devise a system that provided a level of usefulness similar to these systems while retaining its core qualities of noise immunity, accuracy, small size and low power consumption.

In Figure 1, IWT's basic technique for speaker-independent recognition of foreign language phrases is described. The simple technique of creating an array of templates provides the basis for converting a speaker-dependent recognition

[11] CNN.com Monday, November 17, 2003, SHREVEPORT, Louisiana (AP) -- Southern drawls have thwarted voice recognition equipment used by the Shreveport Police Department to route non-emergency calls.

system into a speaker-independent one. This technique is not novel, but doing this while maintaining IWT's tested core qualities was important for the success of this effort.

The subject language for this effort was Arabic, to cover as many dialects as possible. The original plan was to collect responses in a wide variety of dialects. This proved to complicate planned dialogues, so the plan was improved to use Modern Standard Arabic responses collected from a wide variety of speakers.

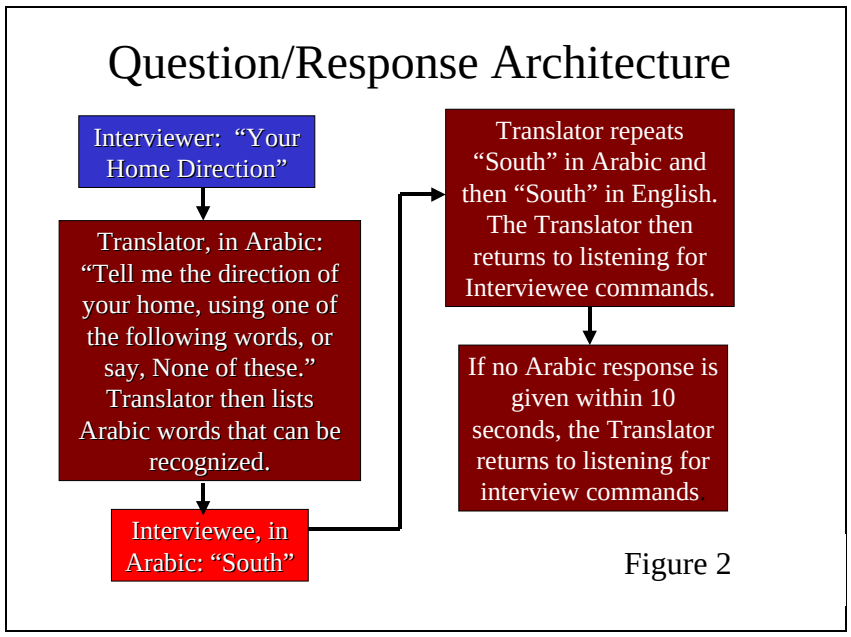

Figure 2

An initial phrase structure was developed prior to the voice template gathering exercise. Per Figure 2, this structure was a simple question to require the interviewee to indicate in which direction – North, South, East or West – that he or she lived. Another phrase structure required an answer in digits, zero through nine.

The dialogue proceeds per the flow chart in Figure 2. There is a verification function that follows the recognition of the initial response. Interviewers can bypass this verification by pushing the reset button. The unit then returns to interview mode.

Phrase Selection/Creation

Once the phrase list was crafted, a native Arabic speaker located 25 subjects to provide templates.[12] To facilitate cooperation, speakers were identified to IWT only by sex and dialect of Arabic spoken. The speakers used were:

[12] IWT as assisted in this effort by Ms. Sanaa Kholfi, a Moroccan operations research doctoral candidate at George Mason University.

- Four Egyptian-Arabic males
- Five Jordanian-Arabic males
- Three Moroccan-Arabic males
- Two Palestinian-Arabic males
- Five Saudi-Arabic males
- Two Moroccan-Arabic females
- Two Egyptian-Arabic females
- Two Saudi-Arabic females

Each speaker was instructed to say the word list in Modern Standard Arabic as spoken in his country and to repeat each of the words eight times. Each of these recordings was then coded and converted into a template form used by IWT's recognition algorithm.

Template creation is a central issue for the performance of this system. This effort addressed the following questions related to that issue:

- What is the minimum number, variety of samples necessary for accurate speaker-independent recognition?
- What types of speakers are best for sample pool, i.e., educated vs. uneducated, male vs. female, etc.
- At what user level is sample gathering appropriate?
- What hardware produces the best templates?

None of these questions was answered conclusively by this effort, but this experience provided a solid basis for proceeding. Twenty-five speakers appear to be enough for creating speaker-independent templates, but for the purposes of IWT's system male and female phrases need to be treated separately. This means that there should be 25 male and 25 female sets collected. The quality of the samples was not uniform, which affected results.

IWT also hopes that advancement of its algorithm will reduce the number of phrases needed.

The types of speakers should be varied somewhat so that a wider range of utterances will be recognized. Interviewees will mimic somewhat the list of prompts read to them, so the same prompts should be used for template collection.

IWT's analysis indicated that the prompts collected by tape recorder and then converted to templates were different in sound composition from templates collected directly using the translator. For more effective and probably easier template collection, we developed a software tool that allows the VRT to be used to collect the templates.

Application Development

The application was created so users would be taken through response dialogues as simply as possible. One goal was to provide a device that would require little training time. Another was to communicate with clarity and precision to interviewees which words/phrases they could speak in response to questions.

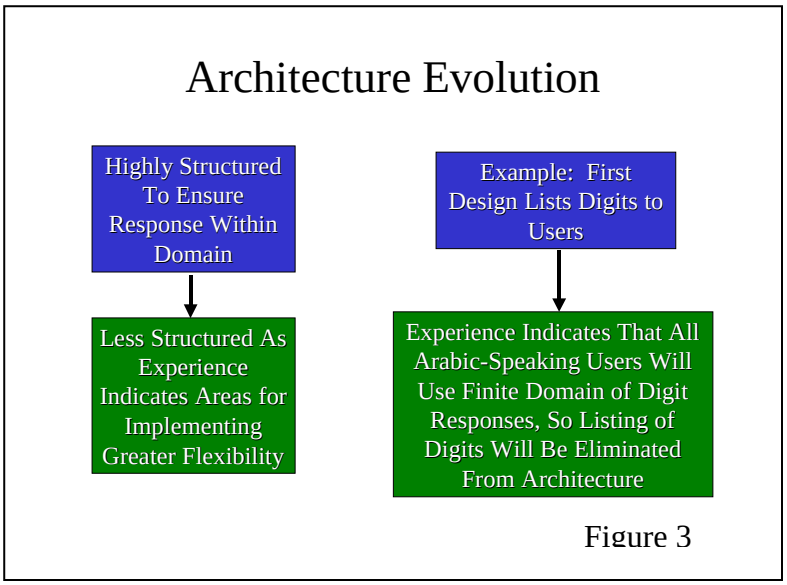

Figure 3

The initial system used voice prompts that for the sake of clarity were long and described to both the interviewer and the interviewee what to do. For example, as described in Figure 3, the unit lists the numerals in Modern Standard Arabic prior to engaging the Arabic recognizer to hear the response. The listing of available responses begins automatically after the interviewer commands a question, but the response list can be abbreviated at any time by pressing the reset button. The translator then stops playing the list and engages the Arabic recognizer.

This structure is straightforward and effective for simple questions and answers, but limits the overall usefulness of the two-way translator. As experience is gained in developing the system and applications, we expect there to be an evolution to more flexibility. The greater the response domain, the more flexibility the two-way translator will have in asking questions and receiving

useful answers. As described in Figure 4, this could involve spotting a word within a non-prescribed response and then making a statement to the interviewee to confirm his response.

We expect that a word-spotting system will need a system as described above as a back up, at least during initial development and deployment.

Broad Expectation-Based System

- IWT technology will allow for future development of a wide range of word spotting in different dialects – much broader range of responses for detection.

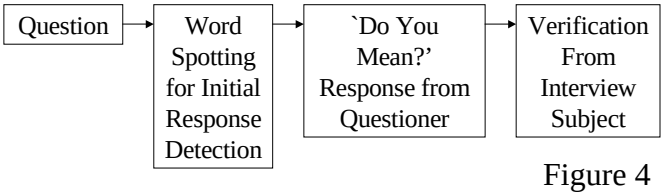

Figure 4

Application Testing

Testing of the prototype device involved both native Arabic speakers who were not sample providers and native English speakers who learned the target words in Modern Standard Arabic. Some Arabic words, such as "HANsa" (five), were easy for native English speakers to mimic. Others were more difficult. Use of native English speakers was only for demonstration purposes, and data were not collected for evaluation of the device. Data collected from native Arabic speakers are presented in Figure 5.

Figure 5: Test Results for MSA Numerals

Numeral	Background	1	2	3	4	5	6	7	8	9	0	Average
Tester 1	Saudi M	90	100	90	100	100	100	100	100	100	100	98
Tester 2	Saudi F	60	70	70	60	50	40	80	50	60	60	60
Tester 3	Moroccan F	30	40	40	20	50	40	30	40	50	40	38
Tester 4	Moroccan F	20	40	40	30	40	50	20	50	40	30	36
Tester 5	Moroccan M	100	100	100	90	80	90	100	100	100	90	95
Tester 6	Egyptian M	90	80	80	70	90	80	80	90	100	90	85
Tester 7	Egyptian M	100	100	100	90	90	80	90	100	90	90	93
Tester 8	Jordanian M	90	100	90	90	100	100	100	90	90	80	93
Tester 9	Jordanian F	50	60	60	70	40	70	70	80	60	70	63
Tester 10	Jordanian M	90	100	100	100	80	90	90	100	100	100	95
Tester 11	Saudi M	100	100	100	90	80	100	80	100	100	100	95
	Average	80	86.67	86.67	86.67	66.67	86.67	80	93.33	86.67	90	84.33
										F Average		49.25
										M Average		93.43

One of IWT's challenges in developing its one-way translator has been to adapt its recognition to work as well with female voices as male voices.[13] This proved to be particularly true with this effort – recognition of female voices was about half as effective as with male speakers. This was due partly to the database of female utterances being smaller than that for male speakers, and partly due to problems inherent in recognizing female voices, which typically have less sound energy. Native female Arabic speakers also appeared to be more hesitant in using the device, which could account for the lower accuracy/effectiveness.

The basic VRT form factor, which potential users said they preferred to the dual handset version originally proposed for this effort.

Since this test effort, IWT has produced improvements in its basic recognition algorithm that have improved training/recognition with all users, particularly women. There has been no formal report on this yet[14], but extensive training/testing/deployments with military and police users have generated anecdotal evidence supporting the effectiveness of these improvements. In one instance, an employee of the Special Operations Forces Language Office (SOFLO) had some difficulties working with the VRT during a training exercise in Hawaii in February 04. After IWT improved its recognition/training program, he tried the device in June 04 and said that it worked perfectly for him.

Similarly, training with a group of 10 soldiers, including two females, at Ft. Huachuca resulted in effective use of the VRT by all participants. A training session with the Lexington, KY with 25 participants, including 2 females, also had no problems. Training with 51 1st Marine Division participants at Camp Pendleton, including three females, resulting in complete success.

TP[13]PT Since the end of this effort, IWT has developed much better techniques for training and recognizing female voices. This has involved making the training/recognition algorithms less sensitive to minute gaps in utterances. Tests with USMC, USA and local law enforcement personnel have since have a perfect success rate in working with female subjects.

[14] A report including some of these results is due from Special Operations Command in August 2004.

IWT believes that the changes in its algorithm, along with a wider collection of female voice samples, would result in much better female recognition for this effort. Further, such improvements could bring recognition for male speakers up to 100 percent.

Testing for noise immunity was done with native English speakers mimicking the Modern Standard Arabic words. Noise immunity was at or above 95dB, slightly less than in one-way VRT operation.

Conclusions: Program Accomplishments and Deficiencies

The project was to demonstrate hardware and software technology to produce a miniature two-way voice-to-voice translator capable of operating in high noise environments.

The initial milestone was to produce a one-way, speaker-independent unit that retains IWT's noise immunity and accuracy. The second milestone was operating and demonstrating a dual-handset system with acceptable overall system accuracy and noise immunity. The third milestone was incorporation of an initial Q&A software architecture that allows for speaker-independent, foreign language responses in realistic background noise settings. The fourth milestone was the inclusion of a mission-driven, expanded set of phrases to allow a useful Q&A. The fifth milestone was operational testing of the developmentally tested device.

Development to reach the first milestone was successful. IWT's integrated recognition board provides an effective basis for this limited two-way voice-to-voice translation. Further, the basic recognition algorithm, as implemented in the new two-way application, worked for this purpose.

Discussions with military and law enforcement users determined the dual-handset approach would not be useful in field applications. Users stated this form factor could be used as a weapon against the person conducting the interview. User groups said IWT's current Voice Response Translator form factor, modified with an additional microphone for two-way translation, was preferable. For that reason, the dual handset form factor was abandoned and a modification of the existing form factor used. This had the additional benefit of allowing for hands-free, one-way voice-to-voice communications by the user.

The third milestone was reached successfully. The application as developed was able to work in high (greater than 95dB) background noise for the short two-way responses.

Reaching the fourth milestone was frustrated by difficulties in obtaining sufficient quantities of Arabic foreign language utterances, particularly in the instance of female speakers. Official Arabic sources were in short supply for this research. Language samples from private sources were obtained, but availability and quality was limited. Part of the quality problem was IWT's inexperience in gathering samples for this purpose.

Our inability to reach Milestone Four prevented our deploying the device operationally and reaching Milestone Five.

This effort was based on IWT's development of the one-way Voice Response Translator, and much of the continuing work to improve this device will be applicable to an improved two-way device.

Also, approximately 300 one-way VRTs have been deployed with military and law enforcement users. Some anecdotal reports on the performance of these devices are included as APPENDIX B as they provide one indication of the potential of a two-way VRT. A formal report on the most recent VRT deployments is scheduled to be released in August 2004 by the Naval Air Warfare Center, Training Systems Division, POC Ms. Dee Sheppe.

Together with the lessons learned from this effort, continued VRT development and improvements based on user feedback could provide a reasonable basis for continuing this work in some form. IWT will continue this research independently in the near term.

APPENDIX A: Project Team

IWT's technical development for this project was led by John H. Hall, Ph.D., h.c. The fundamental aspects of IWT's technological breakthroughs are based on the specific analog-digital insights and miniaturization expertise used by Hall to create semiconductor development breakthroughs over the past 40 years.

Hall's start-up career began in 1962 when he was hired to help found Union Carbide's semiconductor operation. Since that time, he invented the semiconductor design and process technology for a series of groundbreaking, successful commercial products, including: the first electronic watch; the first LCD digital watch; the first CMOS liquid crystal display hand-held calculator; the first electronic camera shutter, voice synthesizers; color autofocus cameras; low-power programmable heart pacemaker; and the first computerized heart pacemaker.

Hall also provided services to the U.S. Government for important new military technologies, including: a combination linear/digital low-cost sonobuoy IC; the phased array radar module for the B-1B bomber; the first radiation-hardened computer for a classified program; and a high-speed data acquisition system for a long-range infrared missile detection system.

Each of these commercial and military programs involved Hall personally inventing new solutions for electronics problems. Many of these solutions included making fundamental advances in semiconductor technology. For example, Hall invented the low-power CMOS technology that now forms the basis for virtually all of the consumer electronics products being produced today. A company he founded and led, Micro Power Systems, Inc., produced devices based on this technology for 10 years before it was adopted by Intel for use in its microprocessors and other products.

Hall was co-founder of Intersil with Fairchild Eight member Jean Hoerni in 1968, heading its research and development, with work that included a breakthrough in coating silicon oxide gates with silicon nitride and creating the first practical metal oxide semiconductor (MOS) processes. Intersil also developed the first N Channel memory chip, which was later adopted as an industry standard.

Hall founded Micro Power Systems in 1971 with work that included low-power CMOS integrated circuit designs that he used in the first computerized programmable heart pacemaker and the first electronic camera shutter, the first low-cost ICs highly resistant to nuclear radiation, stationary phased array radar systems, frequency synthesizers, handheld digital voltmeters, hand-held LCD

calculators, molybdenum gate MOS process used for cellular phone construction, and the first one-chip analog-to-digital converters.

Hall continues to produce specialized analog and digital devices at his semiconductor company, Linear Integrated Systems, Inc. He has turned over day-to-day operation of Linear Integrated Systems to his General Manager so that he can focus on the operations of IWT.

The Project Manager was Tim McCune. He managed technology development efforts for Eagan, McAllister Associates, Inc., a Washington Technology Top 50 Federal Information Services contractor, prior to becoming president of IWT. He managed requirements analysis, product development and developmental testing for the Justice Department-funded Voice Response Translator. He has an MBA with an emphasis on Technology Development.

APPENDIX B: IWT One-Way Translator User Feedback

The two-way VRT is based substantially on the technology, including form factor, of the one-way device. For this reason, user reports concerning the one-way device are included here to provide one indication of the potential of a two-way VRT.

The U.S. Coast Guard purchased a total of 75 VRT systems equipped with megaphones after testing and comparison with other devices. No formal report has been compiled as this was considered an off-the-shelf procurement rather than a testing effort. The following feedback was obtained from one user on station in the Gulf as part of Operation Iraqi Freedom.

> *`It has proved to be the best interpreting tool that we have used to date. Others have been purchased for us, but yours is used all of the time. It is simple to program, easy to use and the voice that results from the unit is clear and understandable to the end user-the Arabic vessels that we encounter each day.'*
>
> XO, CGC Adak, 11 SEP 03

Marine Forces Pacific (MARFORPAC) received 35 VRTs as part of Special Operations Command procurement of 100 units. The MARFORPAC Experimentation Center (MEC) conducted training and evaluation for 1st Marine Division personnel at Camp Pendleton in January 2004. The 1st Marine Division then purchased 50 more VRTs for OIF deployment.

The evaluation effort included the VRT, the Phraselator and a non-speech translation device that used a stylus. Some feedback is below.

> *"I like it"*
>
> Marine, Fox 2/1, 14 JAN 04

> *"Good piece of equipment."*
>
> Marine, H&S COMM PLT, 14 JAN 04

> *"In my opinion, this was my favorite device. It was the easiest to use (2 button) there wasn't a fragile LED screen or one of those pen touch screen things to lose. It's hands free and the commands are short and it understands your normal talking. Plus it easily adapts to a loudspeaker."*

Marine, 2nd Bn 1st MAR WPNS Co., 14 JAN 04

"It is super easy to use, small and hands free. This is the best of the three [evaluated] for squad-level missions."

Marine, Lima 3/1, 14 JAN 04

"The [IWT] device works great … this is a very nice and unexpected addition to [the ship's] force protection capability."

Weapons Officer, a U.S. Navy Destroyer

"It is awesome"

Captain, USA, 3/75TH Ranger BN

"The translators have arrived, and let me tell you, they work superbly!

Marine SGT, Bagram Air Base, 15 MAY 04

"The device responds well in high background noise where other speech recognition systems would not work at all. The device is highly adaptable for paramedics, hospital triage, retail stocking or other situations and trains well with speakers who may have serious accents or speech impediments where other speech recognition systems would not work at all."

Kenneth Pence
Captain - Metro Nashville Police Department

Two VRTs were purchased by Marine Corps Warfighting Lab after competitive evaluation rated it "impressive". These units were delivered to the 22nd MEU for testing, equipped with detachable bullhorns for outdoor environments and Serbian phrases to support S-2. The following feedback was provided:

"The Marines who employed the VRT give it credit for being a very rugged unit that can stand the rigors of being a permanent part of battle gear, getting bumped and dropped, and still function properly.

"It was exposed to extreme heat in excess of 95 degrees F. with greater than 80 percent humidity. It weathered rain and thunderstorms for up to one hour in the open. It showed no signs of problems.

"It empowers junior leaders to control a situation on their own without having to have an interpreter. This is an important point because these men never get interpreters. The system allows the small unit leaders and commanders the ability to 'speak' to the populace/enemy in his own tongue.

"Soldiers in Kosovo that I visited were often patrolling out of platoon fire bases, controlling small towns and villages alone, far removed from even their own company commanders, with no interpreter support.

"Those men would love to have a system such as the VRT which would allow them a much greater freedom of action and ability to control and diminish an escalating bad situation.

"Every rifle platoon in the MEU should have one in our opinion at a minimum, and one at each squad would be ideal."

Captain, 22nd MEU, S-2A

APPENDIX C: Discussion of Language Translation Requirements

Dealing effectively with non-English-speaking persons has become a top priority for law enforcement departments. Census data indicate that the problem is widespread: In 112 American cities, one of every four residents is foreign-born, nationwide, 31.8 million people over the age of 5 speak a language other than English at home. [15]

IWT has identified an application of this sound analysis technology that would meet a need voiced by law enforcement officials. The National Institute of Justice's Technology Assessment Program Advisory Council (TAPAC), at its December 3, 1993, meeting heard from its Weapons and Protective Systems Committee, which identified instant language translation as one of six "immediate" law enforcement technology priorities.[16]

There is a strong public policy-driven requirement for voice-to-voice translation. For example, Assistant Attorney General Viet Dinh, of the Office of Legal Policy, specifically cited the Voice Response Translator as a key new technology to prevent racial profiling. At a July 19, 2001 hearing before the House Government Reform Committee, Dinh said:

> The National Institute of Justice is also engaged in research on a variety of technologies to enhance police capabilities and improve efficiency. Some of the technologies may also help to make police stops less personally intrusive and allow for a more objective determination of the need for a stop. Among these research subjects [is] the [Integrated Wave Technologies, Inc.'s one-way] **Voice Response Translator,** a small device that allows officers to communicate one-way in the same language as the subject being questioned[.]

Law enforcement officers often encounter situations in which suspects and other persons do not speak English. Departments spend considerable resources on developing multilingual resources. The large number of languages involved -- often more than 10 and sometimes more than 20[17] -- and the changing mix of languages frustrate attempts to provide officers with the ability to give even simple directions to persons speaking other languages.

TP[15]PT Miami Herald, January 2, 1994, "Immigration Overload; United States has lost control of its borders."

[1] Minutes of the December 3-4, 1993 TAPAC meeting, p.6.

[17] Conversations with police departments and database research.

Further research with military users confirms this requirement as well. The Military Sealift Command in conducting its mission routinely encounters approximately 50 languages.[18]

[18] Conversation 9 AUG 01 with Mr. Andrew Troy, MSC HQ.