217682

Final Report of a National Institute of Justice Grant to Develop an Advanced, Miniaturized Voice Response Translator for Pre-Programmed Law-Enforcement Phrases

Grant Number 2004-IJ-CX-K042

Submitted By:

Integrated Wave Technologies, Inc.
4042 Clipper Court
Fremont, CA 94538
(510) 353-0260
(510) 353-0261 FAX
http://www.i-w-t.com

Executive Summary

This project was to miniaturize and advance to production-ready status the Voice Response Translator developed and demonstrated by Integrated Wave Technologies, Inc. (IWT) with the Oakland Police Department under a previous NIJ Grant. Extensive evaluation by OPD personnel validated the requirement for the device and the usefulness of the technical approach. This evaluation also revealed several deficiencies to be corrected before the device is put into production for wide-scale use by police departments and other agencies.

This grant was extended on a no-cost basis until 2005 as IWT, Inc. worked with NIJ, the Office of Naval Research, the Air Force Research Labs, and the Defense Advanced Research Project Agency to advance the design of the Voice Response Translator. An eyes-free, hands-free, voice-to-voice translator because a national defense priority after the 9/11 attacks. While the Defense Department rushed into production a voice-to-voice translator, this system was not able to achieve eyes-free, hands-free operation. These agencies funded purchases and further development of the Voice Response Translator for national security applications, as well, as law enforcement/corrections work.

IWT has supported field use and evaluation of the VRT during the period of this grant and documented successful use in numerous applications. The technical performance of the system in law enforcement and corrections applications has been very good, with eyes-free/hands-free capability that no other system has, but implementation and use has been complicated by several factors. First, the large

police departments always had some degree of in-house language translation capability, and officers often relied on personnel who spoke the required language.

More recently, field testing has focused on very small departments without language resources. Immigrant influxes have created significant language requirements, and small departments often do not have bilingual officers. Testing with the Shenandoah County (Virginia) Sheriff's Office has produced highly promising results. The office has deployed VRTs in operational use to conduct routine enforcement activities with persons employed primarily at a chicken processing facility. Initial results are the system meets the operational requirements for this law enforcement department.

Military and homeland security use of the VRT has been highly successful, with the VRT being the only system successfully used in direct action combat. "We went on another air assault today and I used the VRT again, and as before when there was no interpreter around it was wonderful. It helped me establish control in an uncontrollable situation without it," according to a January 2006 report from 101st Airborne Division Iraq operations.

The Coast Guard, in law enforcement operations supporting Operation Iraqi Freedom, used the VRT on all cutters operating in Gulf security duty. According to the Executive Officer of the USCG Cutter Adak, "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."

Based on this successful operational use, the Defense Advanced Research Projects Agency has funded advanced development of the VRT, centering on new configurations and limited two-way capability. Law enforcement testing of a limited two-way version is scheduled for late 2006.

Table of Contents

Project Goals and Objectives Summary

This proposal is to miniaturize and advance to production-ready status the Voice Response Translator developed and demonstrated by Integrated Wave Technologies, Inc. (IWT) with the Oakland Police Department under a previous NIJ Grant. Extensive evaluation by OPD personnel validated the requirement for the device and the usefulness of the technical approach. This evaluation also revealed several deficiencies to be corrected before the device is put into production for wide-scale use by police departments and other agencies.

More generally, this extended effort was to produce a voice-to-voice language translation tool to meet a stated law enforcement and corrections requirement. New features unanticipated in the original design have been added in response to field use of the VRT.

The advancement of the development of this translation capability is designed support more effective community-oriented policing.

For the original effort, program goals were to:

- Improve the effectiveness of the VRT by miniaturizing it and making it easier to "train" using the officer's voice; and
- Assess the effectiveness of the improved system using a criminologist to direct the evaluation.

Program objectives are to:

- Deploy up to 12 Voice Response Translators with Department personnel;
- Test the functionality and reliability of the units;

- Conduct user acceptance testing under field conditions;
- Integrate the expertise of a criminologist into the program to develop effectiveness evaluation criteria and collect data; and
- Work with OPD community Advisory Committees on Crime to continue to assess the impact of this device's use in the community.

Proposed Research and Design Methodology

The first-generation VRT prototypes were based on PC104 boards that allow for relatively easy development and testing of design concepts. IWT expanded the work of the previous project to include new hardware design work after it unexpected problems with the PCMCIA sound card interfaces prevented use of palmtop computers as planned.

The use of more advanced hardware allowed for substantial miniaturization of the VRT. This includes a specialized, ultra-low-power-consumption (500 milliamps) 486 PCMCIA form factor motherboard. Use of this board will allow a device configuration that will fit in a shirt pocket.

This and other development work will allow for an overall effectiveness evaluation of the VRT once it is deployed with 12 members of the Oakland Police Department. A criminologist will work with the program to direct this effectiveness evaluation.

Accomplishments Against Goals

This proposed program was designed to meet the need of law enforcement officers to have an effective audio translation capability. The effort built upon a successful National Institute of Justice-funded program to develop a belt-mounted Voice Response Translator (VRT). The VRT uses a unique voice recognition algorithm that is able to recognize an officer's voice with near 100 percent accuracy even in high background noise environments. This proposal was to advance the VRT design and testing to the point where it would be miniaturized to fit in the shirt pocket of an officer. This goal was to provide law enforcement officers with a device capable of hands-free, eyes-free command playing of 500 pre-selected phrases in at least three languages.

The technical achievements with respect to the VRT have significantly exceeded these goals. The current VRT can hold up to 125 languages, with up to 1,000 phrases, record phrases directly on the unit, and understand limited foreign-language responses when programmed to do so.

The VRT, originally required to last at least 8 hours between charges, will go at least 65 hours between charges.

Audience for Product

IWT has found the military market to be an expanding area for sales of its products. The earliest adopters have been special operations forces: Marines, Rangers, SEALs and Green Berets. For the purposes of accurate market segmentation, IWT defines its potential military market narrowly as front-line combat,

medical and force protection personnel. These are a minority of the armed forces, and total 282,487.[1]

The voice-to-voice translator market is driven both by the desire of police department to have enhanced operational language capabilities and by significant public policy considerations. For example, in California, the Bilingual Services Act (32) requires all state public agencies to provide bilingual public contact personnel and written materials in any non-English language that is spoken by at least five percent of those served at any particular office. Local agencies are under a similar requirement where they serve a substantial number of language minorities. Thus, for example, if a Spanish speaker is unable to file a claim or communicate in Spanish at a public office, and at least five percent of the clientele of that office are Spanish-speakers, that agency is required to provide written and oral language assistance.

Non-English speakers are nothing new to America. But as the number of foreign-born residents in the United States has steadily risen in the past decade, so has the number of people who are not fluent in English. Census data from 2000 showed that 1 in 5 U.S. residents speak a foreign language at home. Only a little more than half of these people (55 percent) also reported speaking English "very well."[2]

As of June 2000, there were about 708,000 sworn law enforcement personnel. Of those, 425,000 were uniformed officers whose regular duties included responding

---

[1] Bureau of Labor Statistics, U.S. Department of Labor, *Occupational Outlook Handbook, 2004-05 Edition*, Job Opportunities in the Armed Forces, on the Internet at **http://www.bls.gov/oco/ocos249.htm.**

[2] "Language Use and English-Speaking Ability," *Census 2000 Brief*, U.S. Census Bureau, October 2003. Available online at http://www.census.gov/prod/2003pubs/c2kbr-29.pdf.

to calls for service. This subgroup of potential users is used for the purpose of establishing the VRT market.

In addition, about 11 percent were either corrections or process servers. As of June 2002, Federal agencies employed more than 93,000 full-time personnel authorized to make arrests and carry firearms in the 50 States and the District of Columbia, according to a survey conducted by the Bureau of Justice Statistics (BJS). Compared with June 2000, employment of such personnel increased by about 6%.

There are approximately 3,300 jails in the United States. Correctional officers in the U.S. jail system admit and process more than 11 million people a year, with about half a million offenders in jail at any given time.[3] Correctional officers held about 476,000 jobs in 2002. About 3 of every 5 jobs were in State correctional institutions such as prisons, prison camps, and youth correctional facilities. Most of the remaining jobs were in city and county jails or other institutions run by local governments. About 16,000 jobs for correctional officers were in Federal correctional institutions, and about 16,000 jobs were in privately owned and managed prisons.

There are 118 jail systems in the United States that house over 1,000 inmates, all of which are located in urban areas. A significant number work in jails and other facilities located in law enforcement agencies throughout the country. However, most correctional officers work in institutions located in rural areas with smaller inmate populations than those in urban jails.

---

[3] Bureau of Labor Statistics, U.S. Department of Labor, *Occupational Outlook Handbook, 2004-05 Edition*, Correctional Officers, on the Internet at **http://www.bls.gov/oco/ocos156.htm.**

One particular segment of the consumer market for speech recognition deserves specific mention because of the opportunities for the Company that it creates. The U.S. Government is placing increasing importance on the development of assistive technology for disabled persons, both as a matter of civil rights for these persons and the desire to make them more independent and productive members of society. The passage of the Americans with Disabilities Act (ADA) requires that public and private entities make all reasonable accommodations for disabled persons. Simply put, speech recognition that allows disabled persons to live more independent home lives and more productive work lives would be funded by many government programs or required to be purchased by relevant accommodation laws such as the ADA.

Assistive technology development areas that have lagged are Voice Recognition and Speaker Verification. Useful and effective Voice Recognition (VR) and Speaker Verification (SV) technology would allow many disabled persons to lead more independent and productive lives. But the voices of some disabled persons are unrecognizable by currently available VR and SV systems. These systems rely on the recognition of separate parts of speech — phonemes — to determine the words spoken. Non-standard speech cannot be processed effectively by these systems.

The General Accounting Office, in a 1996 report, "People with Disabilities: Federal Programs Could Work Together More Efficiently to Promote Employment," noted that Congress has worked to find means of reducing the $60 billion spent each year on disability assistance payments. One of the most effective way of making disabled persons more self sufficient is to develop more advanced assistive

technologies, and the GAO specifically cited voice recognition for computers as a promising technology.

Many persons receiving disability payments have impaired speech. The group of persons with diseases or injuries involving impaired speech is large. Approximate numbers for some are:

- 700,000 with cerebral palsy;
- 350,000 with multiple sclerosis;
- 500,000 with Parkinson's Disease;
- 2 million new cases annually of Acquired Brain Injury;
- 4 million with Alzheimer's Disease;
- 4 million with epilepsy; and
- 4 million stroke survivors.

Other important groups for which numbers were not immediately available are:

- Throat and mouth cancer; and
- Other muscular diseases.

For some groups, only some persons suffer from impaired speech. For others, such as cerebral palsy and Parkinson's Disease, virtually all members suffer from impaired speech.

Several laws – notably the Americans with Disabilities Act and relevant sections of the Rehabilitation Act and the Telecommunications Act – require governments and private employers to purchase assistive devices when they are shown to be effective in helping disabled persons have more productive work lives.

IWT anticipates the market for services and high-technology hardware for disabled persons to be a significant business area. Some 52 million non-institutionalized Americans experience some form of disability that impairs their performance of daily activities. While many of these disabilities are relatively minor, approximately 26 million Americans are "severely" disabled.

According to marketing data taken from studies by FIND/SVP, a 25-year-old worldwide research and consulting firm, this market is now estimated to be $796 billion and will top $1 trillion in the next five years. Functional limitations affect 36 million persons, and work disabilities affect 17 million.

The adoption of IWT's voice command capability as a reasonable workplace accommodation under the Americans with Disabilities Act will add to this market size. The majority of people with disabilities are of working age, but only one-third of disabled persons are employed and most of these are in blue-collar occupations. Employed people with disabilities earn less than their peers. Voice command will open new areas of employment in high-tech, computer-based jobs. This will offer new opportunities for many large groups. For example, visual impairments affect more than 10 million, and more than two million have impaired speech.

Impact of Product on User Groups

Capt. Ken Pence of the Nashville Police Department wrote:

> The Metropolitan Nashville Police Department uses the AT&T language line and it costs the city thousands of dollars a year. When we (the

department) reviewed charges that we noticed that most calls only lasted two (2) minutes. Officers in the field can perform better than preliminary questions and get real information in a timely fashion...often better than a language line. These devices are a wonderful public relations tool since governments can show their concern by making them available for the languages in their own community.

* * *

The device responds well in high background noise where other speech recognition systems would not work at all. The device fits well in a closed uniform shirt pocket and the level of cooperation is amazing when your shirt starts talking like a native speaker. 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.

Lt. Michael Blanton, Lexington (KY) Division of Police, Bureau of Training, wrote after testing the VRT that:

The Emergency Response Unit used the Voice Translator with the megaphone system on one barricaded subject incident. The system worked as it was designed to do and projected statements for several hours to the subject that refused to respond to Police. The subject was eventually arrested without incident.

According to a Navy Pakistan Relief After-Action Report:

'The VRT's gave the Navy Corpsman and Medical Doctors a means to communicate with patients from a foreign country that was not previously available.'

The S-2A, of the Marine 22nd MEU wrote after a Kosovo Deployment:

'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.'

A Captain with the 3/75 Rangers wrote after Iraq use of the VRT, in February 2005

Voice Response Translators 'contributed immeasurably to the success of more than twenty direct action raids in Iraq in support of national level objectives. In addition to sparing precious time on the objective, they reduced collateral damage by bridging a tremendous language barrier therein resulting in the detention of more than fifteen members of the local insurgent network.'

'The VRT proved invaluable in multiple roles as not only a tactical questioning tool but also as a force protection multiplier used both on objectives and from blocking positions.'

Other field feedback includes:

'We went on another air assault today and I used the VRT again, and as before when there was no interpreter around it was wonderful. It helped me establish control in an uncontrollable situation without it.'

Landing Zone RTO, 101st Airborne Division, January 2006

"In one case, [Coalition Forces] approached the house and announced via the VRT that we were there and conducting an operation. Several people came out of their houses and provided information on [Anti-Coalition Forces] forces in the area and led us to a weapons cache."

Ranger, Iraq, 20 MAR 05

"It is awesome"

Captain, USA, 3/75TH Ranger BN, 14 JUL 04

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

Marine, Bagram Air Base, Afghanistan, 15 MAY 04

"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

"The IWT 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."

Captain Kenneth Pence, Metro Nashville Police Department

"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

Purchased by Marine Corps Warfighting Lab in 2000 after competitive evaluation rated it "impressive."

"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

"Results indicate that the VRT is the easiest of the three to use. It has the fastest response times and the longest-lived battery, in addition to being the only one with hands-free capability."

Justice Department Report of Comparative Testing of the Voice Response Translator (VRT), the Phraselator and the Ectaco UT103

Reviewers and Testers

Test and evaluation of the VRT was conducted by the Naval Air Warfare Command's Training Systems Division (NAWC-TSD) for the Justice Department.

For military use, test and evaluation of the VRT and other translation devices is conducted by Science Advisors at the Corps and Division level. For special operations units, evaluation is conducted by the Special Operations Language Office in conjunction with individual units' language offices. The G2 (intelligence) officers at the division and brigade level also conduct test and evaluation efforts.

All of these efforts include combat reports received by users involved in current operations.

Dissemination Avenues

IWT plans to continue to work with NIJ, OLETC, the IACP and other organization to provide information on the Voice Response Translator. IWT is self-funding continued test and evaluation of the VRT with an emphasis on helping small departments that do not have any language assets. While larger departments are easier to locate with respect to arranging testing activities, the vast majority of US local

law enforcement departments are very small. IWT's self-funded testing indicates that these departments are in needed of a system such as the VRT.

IWT plans to work over Summer 2006 to document the use of the VRT with small law enforcement and corrections departments in Virginia and provide feedback and contact information to NIJ for review.

Background

No formal study of translation/interpreter problems confronting law enforcement is available. But there is substantial anecdotal evidence that both large and small law enforcement departments are having increasing difficulties dealing with communities of persons who do not speak English. Also, census data indicate that the problems 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 [4]

IWT identified an application of this sound analysis technology that would meet a need identified 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.[5]

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[6] -- and the changing mix of languages

---

[4] 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.
[6] Conversations with police departments and database research.

frustrate attempts to provide officers with the ability to give even simple directions to persons speaking other languages.

Integrated Wave Technologies, Inc., (IWT) worked from June 1, 1996 to March 30, 1998 under a National Institute of Justice Science and Technology grant to develop the VRT. The VRT, on voice command, produces pre-programmed phrases in various languages. This allows officers to identify the language spoken by a person, issue emergency commands to the person and make inquiries to which a person could respond with hand signals. This system was designed for use in both hostile and non-hostile encounters with non-English-speaking persons.

The Oakland Police Department provided extensive support for hardware/software design review, translated phrase selection and translation review, and operational testing. Both the hardware and software have evolved extensively in response to detailed feedback from OPD personnel.

IWT designed and built prototypes of the first belt-mounted, voice command translation unit. IWT worked to develop this design -- based upon the PC104 board format -- after evaluating all available palmtop computers and determining that none was suitable. The highly compact nature of the IWT voice command algorithm allowed for the use of a 386-class processor in the unit, which greatly reduces size and power consumption. Other voice recognition software requires a Pentium-class microprocessor, consuming 10 times the power and requiring a large amount of supporting hardware.

The basis of this unique device is IWT's work with Soviet-conceived sound analysis technology. Scientists and engineers at these laboratories, in seeking to

develop speech identification and other sound analysis programs, took an approach that is fundamentally different from that used in voice recognition systems developed in the United States and other Western countries.

Prototypes for this program were approximately five inches by five inches by two inches, much smaller than any previous voice-command platform capable of this task. Advanced prototypes and production units can be as small as a hand-held calculator when based upon custom-designed computer boards such as has been developed for another IWT application.

The VRT completed initial evaluation by the Oakland Police Department (OPD). OPD worked with the National Institute of Justice and IWT to evaluate the phrases being used, the translation of the phrases and the configuration of the hardware of the VRT. The Department worked closely with its Advisory Committees on Crime to gain pertinent feedback from citizens on the impact and effectiveness of the VRT in operational use.

The Voice Response Translator grant effort broke new ground in several areas. First, it has refined, to the greatest extent ever, scripts needed for communicating with foreign language speakers using a computer-based translator. Second, it has resulted in the design of computer optimized for law enforcement use that is both smaller and lighter than any previous unit able to accept voice commands. Third, it has demonstrated the highly developed capabilities of the Soviet-based voice recognition system by operating in a high noise environment with virtually no externally induced audio-recognition errors.

This work has laid the groundwork for a testing program that will help to provide a better understanding of the effectiveness of this device, in particular, and machine translation in general. A follow-up report based upon the results of planned testing should provide a wealth of data on community acceptance of the device and its ability to communicate with non-English-speaking persons.

The program changes were preliminary driven by the following three factors: 1) the need to design and build a belt-mounted computer rather than using an off-the-shelf unit as planned; 2) the need to conduct extensive community evaluation/relations work prior to field testing; and 3) the need to include approximately 10 times the number of translated phrases as planned.

IWT has been successful in responding to, and meeting, the refined requirements of this program. The design work on the belt-mount computer has been completed and demonstrated to top officials of the Oakland Police Department (OPD). The OPD has devoted considerable resources to making the community evaluation a success, and we believe that this unanticipated part of the program will be of great benefit to this law enforcement technology development effort. IWT has also completed the expanded translator work, in close coordination with the OPD.

This program's goals were to improve the effectiveness of the VRT by miniaturizing it and making it easier to operate.

Even though the VRT developed under the preview NIJ effort was the smallest voice-command computer able to meet this requirement ever developed, further miniaturization is needed for it to become optional equipment for law enforcement officers. This miniaturization is a central aspect of this proposal.

A significant drawback of the VRT as developed was the need for it to be partly disassembled and connected to a video card, monitor and keyboard for officer voice pattern training. A "training" is when a user makes a voice-command imprint computer file used as a template for the recognition algorithm. Because an IWT or EMA technician needed to be present for all training sessions, officers were unable to create the voice sample collections at their convenience. This approach was taken to minimize the initial development needed to field the VRT, but greatly increased the training needed for use of the device.

Effectiveness of the training was also hampered because of the lack of a voice-prompting training system. While the pattern recognition algorithm will correct for a wide range of variances such as the speed with which a word is said, the best voiceprints are those said in the situation in which the device will be used. A good analogy is fingerprinting, as both are wave representations unique to individuals. Law enforcement experts can match fingerprints from even a partial sample. But when taking the sample upon which to base these matches, a clear, complete fingerprint is taken. Clear voice samples, said naturally with no background noise, provide the most flexible and accurate basis for use of the IWT algorithm. Using this sample, the algorithm can match the same word or phrase even if it is said slower, faster, under stress or with background noise.

All persons working initially with voice recognition systems experience a high degree of self consciousness, as talking to a machine is an unusual act. This disappears quickly, but the awkward training structure of the VRT hardware made repeat trainings difficult. An initial design decision to make the units as small as

possible meant that the video card was not included. Though this helped to meet the operational requirements, a software architecture redesign needs to be made to allow for audio prompts that will make the VRT a self-training device. IWT has developed this design architecture for use with another application and can use it in future work with the VRT.

APPENDIX A: Current Law Enforcement Documentation

PRIMARY (1 of 3)

"I'll try" *"I will try"*
"I Thank you" *"Thank you"*
"I say yes" *"Affirmative"*
"I say no" *"No"*
"Family name?" *"Which name is your family name?"*
"Group leader?" *"Who is the leader of your group?"*
"Turn Off Engine" *"Turn off the engine"*
"My name is ... " *"Hello. I am Officer..."*
"My reason" *"The reason I stopped you is..."*
"Go too fast" *"Your vehicle was exceeding the speed limit"*
"Red light" *"Your vehicle went through a red light"*
"Ran stop sign" *"Your vehicle went through a stop sign without stopping"*
"Illegal turn" *"Your vehicle made an illegal turn"*
"Plates Out Of Date" *"Your license plate has expired"*
"Driver's license?" *"May I see your license? Please hand it to me."*
"Registration?" *"May I see the vehicle registration?"*
"Car Owner?" *"Are you the vehicle owner?"*
"Car Insurance?" *"Do you have insurance?"*
"Giving You Ticket" *"I am required to write you a citation."*
"Get out of car" *"Step out of the vehicle"*
"Move over here" *"Move over here"*
"Vehicle search?" *"May I have permission to search your vehicle?"*
"Wait Here" *"Wait here"* (Officer points where)
"No guilt" *"Please sign here."* (point where) *"This isn't an admission of guilt – just that you received a citation"*
"Court Date" *"Court date is here"* (point to date/time) *"or you may mail in the fine"*
"Pull Back" *"Please be careful as you pull back onto the road"*
"Tow Vehicle" *"Your vehicle will be towed"* (officer will write down address)
"Other driver?" *"Will you allow someone else here with a license to drive your vehicle?"*
"Open the door" *"Police. Open the door!"*
"Warrant for Your Arrest" *"We have an arrest warrant for you"*
"Warrant for Search" *"We have a search warrant for you"*
"Face Away" *"Face away from me slowly and put you hands here* (touch where)*" on you? Point to where they are."*
"Take hands" *"Take your hands slowly out so I can see them"*

PRIMARY (2 of 3)

"Any Weapons?" *"Do you have any weapons or drugs on you? Point to where they are."*
"Spread feet" *"Spread your feet. Turn your toes outward. Feet wide apart."*
"Place hands back" *"Hands behind your back. Palms out."*
"See judge" *"I am taking you to see a judge"*
"Lie face down" *"Lie down face down. Don't move."*
"Stop Resisting!" *"Stop resisting! You won't be hurt"*
"Miranda" *"You have the right to remain silent. Anything you say can and will be used against you in a court of law. You have the right to be speak to an attorney, and to have an attorney present during any questioning. If you cannot afford a lawyer, one will be provided for you at government expense."*
"What's your age?" *"How old are you?*
"Write your phone number?" *"Please write your telephone number?"*
"Home address?" *"Do you know your home address? Write it here."*
"Parents work?" *"Where do your parents work?"*
"Work number?" *"Do you know their work telephone number?"*
"Probation?" *"Are you on probation or parole?"*
"Date of birth?" *"Write down your date of birth. Month, day and year."*
"Card number?" *"Write down your social security number if you know it?"*
"Come by car?" *"Did you come here by car?"*
"Did you walk?" *"Did you walk?"*
"Were you drinking?" *"Have you been drinking? Shake your head up and down for yes."*
"How many drinks?" *"How many drinks? Hold up as many fingers as you have had drinks."*
"Do eye test" *"Stand here with your hands to your side. Follow the motion of my pen or finger with your eyes. Do not move your head"*
"Do balance test" *"Stand here with your hands to your side. Do not start until I tell you to start. Raise either foot about six inches (a hand width) off the ground. Point your toe like this and look at your foot. Start counting in your language until I tell you to stop. You may begin"*
"Gang member?" *"Have you ever been a member of a gang? Shake your head up and down for yes. Side to side for no."*
"Any tattoos?" *"Do you have any tattoos? Where? Point to them."*
"Cool tattoo" *"That is a cool tattoo. May I see it?"*

PRIMARY (3 of 3)

"Don't be afraid" *"Don't be afraid. I'm a police officer and I'm here to help you."*
"Are you lost?" *"Are you lost? Can you show me, write down or say the address?"*
"Tow truck?" *"Do you need me to call a wrecker (a tow truck) for you?"*
"Need telephone?" *"Do you need a telephone?"*
"Need a rest room?" *"Do you need a rest room?"*
"Are you thirsty?" *"Are you thirsty?"*
"Are you cold?" *"Are you cold?"*
"Draw you a map" *"I will draw you a map."*
"I'll lead you" *"I will lead you there...follow me."*
"Need a Doctor?" *"Do you need a doctor? Shake your head up and down for yes - side to side for no."*
"I understand" *"I think I understand."*
"I don't know" *"I don't know."*
"Draw it here" *"Draw or write here."*
"Are you in pain?" *"Do you have pain?"*
"Show where hurt?" *"Where does it hurt?"*
"Pain in chest?" *"Are you having chest pain?"*
"Trouble breathing?" *"Are you having difficult breathing?"*
"Diabetic?" *"Are you diabetic?"*
"Pregnant?" *"Are you pregnant?"*
"Caused injury" *"Show me what caused the injury"*
"Please sit" *"Please sit down"*
"You lie down" *"I need for you to lie down now"*
"Doctors numbers?" *"Write the names/phone numbers of your doctors"*
"Any medical?" *"Write down any medical conditions you have"*
"Ambulance" *"The ambulance is coming"*
"Straight line" *"Stand here with your hands to your side. Place your left foot on the line. Take nine steps, heel to toe, down this line. Turn around. Take nine steps back, heel to toe."* **"You lie down"** *"I need for you to lie down now"*

Always Active Commands

"My greetings" *"Hello"*
"Begin directing" *"I'm speaking to you through a device that translates select phrases into your language. Please nod your head for yes, shake your head for no, or write down short answers."*
"Goodbye to you" *"Goodbye"*
"Under arrest" *"You are under arrest"*
"Don't move" *"Don't Move"*
"Show eye-dee" *"Show me your identification"*
"You halt now" *"Stop now"*
"Don't get it" *"I don't get it"*
"What's your name?" *"What's your name?"*
"Speak American?" *"Do you speak English?"*
"Calm down" *"Calm down"*
"Use tear gas" *"Failure to leave immediately will result in our use of tear gas"*
"Illegal" *"What you are doing is against the law"*
"Keep moving" *"Keep moving"*
"Break up now" *"Failure to disperse will result in your arrest"*
"Behind barricades" *"Stay behind the barricades"*
"Write name" *"Please write your name"*

"Phrase Alpha" "_____"

"Phrase Bravo" "_____"

"Phrase Charlie" "_____"

"Phrase Delta" "_____"

"Phrase Echo" "_____"

PEOPLE (1 of 2)

"File Charges?" *"Do you want to file charges?"*
"Who called?" *"Who called the police?"*
"Live here?" *"Do you live here?"*
"Another room" *"Let's talk in another room"*
"Your husband?" *"Is this person your husband?"*
"Go to jail" *"I am taking this person to jail"*
"Youngsters?" *"Do you have children?"*
"Kids in danger?" *"Are the children in danger?"*
"Temporary place?" *"Do you wish to stay temporarily in a shelter for women?"*
"Restraining order? *"Is there a restraining order on that person?"*
"Who has custody?" *"Who has custody of the children?"*
"View order?" *"Can I see a copy of the restraining order or custody order?"*
"Live together?" *"Do you live together?"*
"What's your age?" *"How old are you?*
"Write phone number" *"Please write your telephone number"*
"Home address?" *"Do you know your home address?"*
"Show me?" *"Are you hurt? Show me"*
"Stranger bring? *"Did a stranger bring you here?"*
"Parents work?" *"Where do your parents work?*
"Parents telephone?" *"Do you know their work telephone?"*
"Dad's name?" *"What is your father's name?"*
"Mom's name?" *"What is your mother's name?"*
"The victim?" *"Who is the victim?"*
"When did it happen?" *"When did it happen?"*
"Write all" *"Write your name, address and phone number."*
"Know where suspect lives?" *"Do you know the suspect or where he lives?"*
"Go where?" *"Which way did the suspect flee?"*
"Man or woman?" *"Is the suspect a man or woman?"*
"Show how tall" *"Use your hand to show me how tall the suspect is"*
"Latino suspect?" *"Is the suspect Latino?"*
"Caucasian suspect?" *"Is the suspect White?"*
"Black suspect?" *"Is the suspect Black?"*
"Asiatic suspect?" *"Is the suspect Asian?"*
"Color shirt?" *"What color shirt was the suspect wearing?"*

PEOPLE (2 of 2)

"Color jacket?" *"What color jacket was the suspect wearing?"*
"More than one?" *"Was there more than one suspect?"*
"Go on foot?" *"Did the suspect leave on foot or in a vehicle?"*
"Vehicle's color?" *"What color is the vehicle?"*
"Plate number?" *"Do you know the license number? If so, write it here."*
"Any weapons?" *"Did the suspect have a weapon?"*
"Child's name?" *"What is the child's name?"*
"Boy child?" *"Is the child a boy? Shake your head up and down for yes, side-to-side for no"*
"Child's age?" *"How old is the child?"*
"When last seen?" *"When was the person last seen?"*
"Seen last where?" *"Where was the person last seen?"*
"Child come home?" *"Did the child come home from school?"*
"School on foot?" *"Does the child walk to school?"*
"Child's height?" *"Show me how tall the child is"*
"Got photo?" *"Do you have a photo?"*
"Search house?" *"Can I search your house for the child?"*
"Medical problem?" *"Does the lost person have a medical problem?*
"Checked with relatives?" *"Have you checked with relatives?"*
"You Epileptic?" *"Epilepsy?"*
"Alzheimer's?" *"Alzheimer's Disease?"*
"Child last seen by?" *"Who saw the child last?"*
"Who saw person last?" *"Who saw the person last?"*

Changing from "Primary" to "People"

If these groups are trained with your voice, you can switch from having Always Active + Primary to Always Active + People by using the voice command "Choose Different Group" while the VRT is in active mode. When asked which group, say "People" or "Primary." See Instructions and Default cards for more information on Groups.

LEARNING TO USE THE TRANSLATOR

1. This is a Voice Response Translator Voice-To-Voice English-to-Foreign Language System. The VRT responds to specific, short command phrases you say into the headset microphone. For example, you would say, "Under Arrest" and the VRT would play in chosen foreign language, "*Hello, I am a US Soldier.*"
2. Read all the instructions on this card before starting. Note each User must record his voice so the Translator recognizes his Voice Commands. The first time you use the Translator, it will ask you to do this. Follow the Translator's audio instructions, repeating each phrase as requested. The Translator will only recognize commands you have trained.
3. There are about 350 phrases in the application, separated into nine groups. Remember the names of the groups, because you will use voice commands to switch among them. They are named: Always Active, Primary, Base Duty, Local Aid, Medical, Pharmacy, Training, Devices, and Boarding Commands.
4. There are also voice commands used to operate the unit, called "Control Commands," and the names of the languages. You'll train the Control Commands, the Group names and the Language Names during initial training. You can train any of the groups immediately after that, or later if you like.
5. Read the Voice Commands on the cards out loud at least once before starting, using a strong command voice. Speak clearly And naturally – say the phrases without pauses between words, e.g., "Stand'n'line", not "Stand-In-Line."
6. Put on the headset so the mike is in front of the left side your mouth, about 1/3" in front of your lower lip, with the hoops curved over your ears and the band around the back of your head behind your ears. After practicing the Voice Commands, turn on the Translator. Listen to the commands and watch the lights on top of the Translator for guidance.
7. Turn on the Translator. The Translator will say "Please Input Your ID." Press the button – once for User One, etc., up to User Eight. You may also say a number from One through Eight to select a User Number. YOU MUST ALWAYS USE THE SAME NUMBER, AND NO ONE MAY USE YOUR NUMBER.
8. Wait until the red light goes off before repeating the phrase – if you start too quickly, it will cut off the beginning. You must also speak loudly enough for the light to flicker green.
8. After you train the Initial Commands, you may train the other groups of commands, described on the phrase cards. To train a group of phrases, say "Begin Training" after putting the translator "On Standby," and say the name of the group when prompted by the VRT. Training will then begin.
9. To switch from "Use VRT" (Translate) mode to "On Standby," press the button for five seconds or say "Go On Standby".
10. Retraining the Translator after you've learned to use it will greatly increase effectiveness. If you've trained the Translator in a non-operational environment – a classroom for example – you should retrain the unit in the setting where you'll be using it, using the command voice you will use operationally.
11. If the Translator keeps asking you to repeat a phase – it says, "Training must be repeated ..." – then say the phrase faster, in a deeper voice and loudly enough for the green light to flicker.
12. After you turn on the translator and input your user ID, the "Always Active" commands and the Default Group commands are available to use if you've trained them. Say any of these commands and the selected foreign language output phrase will come out. Other groups are available after you've trained them. Use the command "Choose new group" to switch groups.
13. A green light will blink if the battery needs to be charged.

Phrase Group Names: "Primary," "People"

Control Commands Are:

1. "My Location"
2. "Find Language"
3. "Choose Different Group"
4. "Go on Standby"
5. "Begin Training"
6. "Use VRT"
7. "Turn Volume Down"
8. "Repeat Phrase Loop"
9. "Start Recording"

Police
Version 260

Integrated Wave Technologies, Inc.
Fremont, CA
support@i-w-t.com

Playing Emergency Phrase

The Voice Response Translator is equipped with an Emergency Phrase that can be activated and played continuously by any person without issuing a voice command.

To start the Emergency Phrase, turn the translator on and after it asks for a User Number, but before entering it, hold the Button down for at least five seconds. Let go of the Button and the Emergency Phrase will play continuously until you press the Button again or turn the unit off. You still must press the megaphone button.

The Emergency Phrase, played continuously in available languages is:

"You are under arrest."

Sound Amplifier Attachments

The Voice Response Translator is equipped to play through auxiliary amplification devices.

To use the Megaphone attachment, attach the Translator to the Megaphone using the matching Velcro surfaces. Unclip the short connection cable from the Megaphone ring and insert the small plug into the translator and the larger plug into the jack on top of the Megaphone microphone.

Please note you must press the Megaphone trigger for Translator to play through it. When the Translator is not plugged into the Megaphone, you may use the Megaphone normally.

TURN OFF VRT BEFORE ATTACHING AMPLIFYING DEVICE. TURN ON VRT ONLY AFTER MAKING CABLE CONNECTION.

Changing Default Languages & Phrase Groups

The VRT has a default language and default group of commands that start when you turn it on and enter a user number. While using the VRT, you can temporarily switch to other languages or groups of commands by using the voice commands "Find Language" and "Choose Different Group." Please note that these groups are only available after you've trained them, as described on the card, "First Steps to Using This Translator." Please also note that the VRT will return to the default language and group after you turn it off and back on again.

The VRT lets you know which defaults are selected when you turn it on. For example, it will say, "Iraqi, Primary" when you turn it on (after you've trained both the Initial and Primary commands.)

If the phrases and/or language you plan to use on your next mission are not the default ones, it only takes a few seconds to change the VRT so it will boot up in the language and phrase group needed for your mission.

To change the default language and/or phrase group, do the following. Put on the headset with the microphone next to your mouth. Turn on the VRT and enter your user number. When the unit boots up, it will say the default language and group. Say "Go on standby" or press the button for five seconds to put the VRT on standby. While VRT is in "Standby" mode, say "Choose Different Group." After the VRT asks which group, say the name of the group you wish to select. Once this is recognized, the default group has been changed. To change the default language, follow this procedure and say, "Find Language" and the name of the new default language when asked.

Turn the VRT off and back on again, select your user number, and when it boots up it will say the names of the new language and group.

User Numbers

USER 1: _____
USER 2: _____
USER 3: _____
USER 4: _____
USER 5: _____
USER 6: _____
USER 7: _____
USER 8: _____

Choosing User Numbers

When the VRT is turned on, it will ask you to input a user number. Press and release the Button from one to eight times to correspond to the user ID that you want to enter, or say a number from One to Eight. The unit then will enter either "Initial Training" if this is the first time you've used the unit, or play the name of a language if you've done Initial Training. If you've trained the main command group, it will say the name of that as well when turned on.

Recording over User Numbers

If you are a new user and choose a number, and the translator starts up naming a language and a group of commands, that user number was used already. Either choose an unused number, or reset that number as follows.

Doing a "Hard Reset" of a User Number: To erase all templates in a User Number, hold the button down while turning on the unit. The translator will then ask for a User Number. After you input a User Number, the previous trainings are erased and the unit starts "Initial Training."

Functions Controlled By VRT Button

To erase ALL of a user's trainings: Hold the button down while turning on the unit (a "Hard Reset"). Then selecting a User Number will erase all of that user's trainings.

To stop an output phrase while it is playing: Press and release the button.

To play the Emergency Phrase: After the VRT is powered up and asks for a user number, but before selecting a User Number, hold the button down for five seconds; The Translator then plays the Emergency Phrase continuously until you turn it off or press the button.

To go to "On Standby": After powering up and selecting a User Number, then hold the button down for five seconds; The Translator will then go "on standby".

To retrain JUST the "Initial Training": While in "On Standby" mode, hold the button down for 5 seconds; The Translator will then begin initial training for the User Number selected when it was turned on. If phrases in the Group(s) were trained, these remain trained.

To hear which language/phrase group is active: Press and release the button quickly, and the VRT will play the name of the language and phrase group selected.

Low Battery Warning

When the battery is getting low, a light on top of the translator will blink steadily during use. When charging, the light will be red. When charged, it will turn green and eventually the green light will go off.

Adjusting VRT Speaker Output Volume

The VRT starts at maximum volume. To lower the volume by half, use the voice command "Turn Volume Down." This command reduces the output volume of the device, as many times as the user says it. To reset to maximum volume, press and release the button.

Automatic Phrase Repeat Function

The VRT can be commanded to repeat the same foreign-language output phrase over and over (until the button is pushed) by using the command, "Repeat Phrase Loop." The VRT will then ask "Which phrase," and the user says the command for the phrase to be repeated. The phrase will repeat continuously until the button is pushed or the unit is turned off.

Field phrase recording capability

Users can record up to five foreign-language output phrases in each language directly onto the VRT. The VRT contains five "generic commands" – the commands, "Phrase Alpha" through "Phrase Echo." To record an output phrase attached to one these commands, first turn on the VRT, select your user number and when the unit boots up put it on standby either by voice command, "Go On Standby," or by pressing the button for five seconds.

When unit is "On Standby," saying the command "Start Recording," gets the response, "Which Language." The user says a language (for example "Iraqi") and gets the response, "Which Phrase." The user says one of the generic phrases, "Phrase Alpha" through "Phrase Echo." The VRT says "Press and release the button to begin recording; Press and release the button to end recording." Hand the headset to the interpreter, who puts the headset on and then presses the button once immediately before saying the translated phrase. Immediately after saying the phrase press the button again to end the recording. The new output phrase will be saved automatically.

English Reminders of the Phrases:

Phrase Alpha:_____

Phrase Bravo:_____

Phrase Charlie:_____

Phrase Delta:_____

Phrase Echo:_____

Getting Started

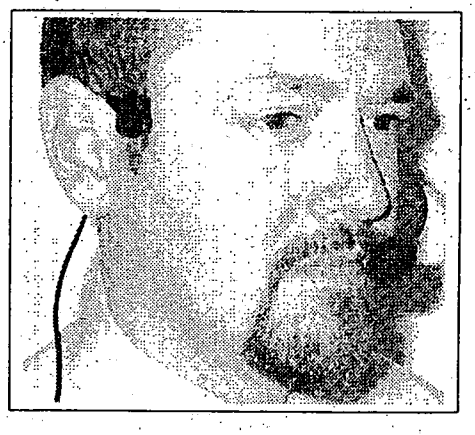
1. Read the card titled, "LEARNING TO USE THE TRANSLATOR." Do not turn on the device until Step 4, below.
2. Put on the headset as shown on the other side of this card.
3. Read out loud the following words in a command voice. When you turn on the translator, it will ask you to repeat these words again. Practicing them allows you to speak them in a natural command voice when training the translator.

"My Location", "Find Language", "Choose Different Group", "Go on Standby", "Begin Training", "Use VRT", Turn Volume Down", "Repeat Phrase Loop", "Start Recording", "Begin Dari Farsi", "Begin Lao", "Begin Viet", "Begin Serbian", "Begin Kurd", "Begin Russian", "Begin Swahili", "Begin Korean", "Begin Spain", "Begin Arabic", "Begin Greek", "Begin Italian", "Begin American", "Begin Mandarin", "Begin Cantonese", "Begin Tagalog", "Begin Creole", "Begin Japanese", "Primary", "People", "My Greetings", "Begin Directing", "Goodbye to you", "Under Arrest", "Don't Move", "Show Eye-Dee", "You Halt Now", "Don't Get It", "What's your name?", "Speak American?", "Calm down", "Use Tear Gas", "Illegal", "Keep Moving", "Break Up Now", "Behind Barricades", "Write Name", "Phrase Alpha", "Phrase Bravo", "Phrase Charlie", "Phrase Delta", "Phrase Echo"

4. With the headset on, turn on the translator and select your user number as described in the LEARNING TO USE THE TRANSLATOR card. Repeat the phrases, taking care not to speak until the red light goes out. Review the instruction card and begin use. The VRT is ready to use those commands just trained, but you must train any others that you want to use in the same manner.

Correct Position for the VRT Headset

Place the headset on your head as shown. Pull the boom around so the end is almost touching your mouth. Place this in the same place each time you use the translator.



DEFENSETECH.ORG

the future of the military, law enforcement, and national security

ENGLISH TO ARABIC, HANDS-FREE

Four-and-a-half years after 9/11, only a teeny-tiny percentage of our troops speak Arabic. And despite advertised plans for increased language training, that's not going to change any time soon. In the meantime, the military is turning to technological fixes -- translator gadgets that let soldiers convey simple commands.

The best known of these is probably the PDA-like Phraselator. Make a couple of stylus taps, or say a few words in English, and out comes an Arabic phrase. "It gets really funny looks from the Iraqis, but they think it's cool," one company commander tells me.

But the Phraselator can be a bit of a pain, too. Because you have to hold the thing in your hands in order for it to work. And that makes it a lot harder to hold an M-16 at the same time.

So Integrated Wave Technologies has come up with a translator that doesn't require a hand to work. Talk English into a headset, and a ammo clip-sized speaker broadcasts out the Arabic equivalent. Check out this video for an example. You'll see, the translators aren't for carrying on conversation; they only interpret a few words at a time. But they seem to work well, when you're yelling at someone to get on the ground while your gun is pointed at his head. About 600 of the things are now in theater, according to the company.

The next step, of course, is to make the translators two-way, so Iraqis can talk back to the soldiers. Integrated Wave Technologies has a Darpa contract to do just that -- one of several translation projects the Pentagon's way-out research arm is funding.

March 10, 2006 10:03 AM

http://www.defensetech.org/archives/002186.html

The Coming of the Portable Universal Translator

By John D. Gresham

There is story from the Old Testament book of Genesis, about a pagan king named Nimrod who wanted to touch the heavens. To achieve this goal, Nimrod put the people of his kingdom, known as Babel, to work and began to build a temple so tall that it would allow Nimrod to reach the sun. Seeing their lack of faith and Nimrod's ambition, God made the workers begin to speak in different tongues, so that they would be unable to continue coordinating their efforts. Thus is explained the beginnings of modern language diversity, and the heart of a problem that has faced military personnel since the first one put on a uniform. How do you communicate in multi-national collation situations with allies, or with civilians in the lands you are operating in?

This question is multiplied many times when one considers just how difficult it is for military forces to recruit or train even a handful of personnel able to speak a limited range of foreign languages. Much of this problem stems from the decline of foreign language instruction being presented in American schools and colleges. In addition, people with foreign linguistic skills have commanded excellent salaries in the booming international economy of the past decade, and are still very much in demand. Another problem is that few of the significant contingencies of the past decade since the end of the Cold War have occurred in places with what might be called "common" languages. The Balkans use a mix of Slavic dialects, while Arabic which is common in the Persian Gulf is one of the most difficult languages to master. Of even more concern is Chinese, which will become the single most spoken and written language on Earth (particularly on the Internet) in the next decade. Finally, as difficult as these languages are for us to learn, imagine the pain speakers of romantic, Slavic, Semitic, and other dialects must have when dealing with the freewheeling nature of American English!

Language skills can make or break the success of an overseas military operation, and translators quickly become attached at the hip to on-the-ground commanders as they deal with local leaders and indigenous personnel. While one might think that the most critical parts of the language problem for soldiers would be in areas like riot control and hostage rescue, in fact the reverse is true. The most critical need for qualified translators has proven to be in day-to-day life and operations. Things as simple as public announcements of scheduling for water and food distribution, arrival of medical personnel, or even the messages of politicians in democratic elections make up the bulk of the language problem for military personnel operating on foreign soil. Unfortunately, there has never been a good solution to the challenge of multi-national communications, despite several thousand years of effort on the problem.

The magnitude of the military language problem might best be gauged by considering the needs of a force which requires every member to have foreign dialect skills: the U.S. Army Special Forces (SF), better known as the "Green Berets." Composed of less than 10,000 highly trained and motivated soldiers, the Special Forces are the only community in the U.S. military, which requires each serving soldier must speak at least one foreign language fluently. Organized into

regionally seven focused groups (five active and two National Guard), the SF soldiers are America's most culturally adept warriors. Each undergoes years of intense selection and rigorous training before they ever are dispatched overseas. This includes training in one or more languages (out of a total of seventeen), and can take up to several years in the case of more difficult dialects like Chinese or Arabic. That is up to 10% of an entire 20-year career for some SF soldiers, something that is unacceptable for normal military personnel. Unfortunately, SF soldiers are hardly the only U.S. military personnel who deploy overseas, making anyone who can speak the local dialect literally worth their weight in gold, given the goodwill and cultural understand that can result from just one good translator in the right place, at the right time.

Technology has tried for centuries to provide some means of universal translation for soldiers and statesmen, with poor to miserable results. The first efforts probably initiated with the Egyptians, whose invention of papyrus scrolls provided a medium for hieroglyphic depictions of commonly understood tasks, goods, and actions. Since that time, the invention of alphabets, wood-based paper and plastics, high-resolution printing technology has provided some improvements on the idea of book-based translation systems. Unfortunately, these are at best slow and cumbersome, hardly what is needed in a fast-breaking military situation. The development of micro-electronics and digital computers following World War II provided some hope of improving on print-based translation systems, though the process of creating an personal "universal translator" has been much slower than one might think. In fact, the problems of creating such devices has thus far proven successful only in *Star Trek* movies and television episodes, much to the frustration of military personnel everywhere. Commercial customers also exist in ever growing numbers, especially in the automobile and medical industries. Rarely has a technology been so desired and anticipated, and yet take so long to gestate. Nevertheless, there has been progress.

The beginnings of the drive to the universal translator really began in the 1950's, with the early work at Bell Laboratories on identifying vowel sounds in single digit numbers. Commercial work on this technology continued at RCA and in Japan in the 1960's, with limited success. Most of the early problems had to do with the limited capacity of early computers to do basic signal processing, and storage technology to provide the library of software and data to accomplish simple speech recognition tasks. The first electronic speed synthesizers also were produced at this time, making use of the technologies developed for the electronic music instrument industry in the late 1960s. Perhaps the most promising work took place in the former Soviet Union during the 1960's, when T.K. Vintsyuk, N.G. Zagoruyko, and other researchers began to actually produce usable algorithms for matching speech input patterns stored in electronic data banks. What made their work so important was that it involved use of the minimal computing and storage resources available in the USSR during that period. It was consideration that would later produce major dividends in the 1990s.

In 1971, the Advanced Research Projects Agency (ARPA - now the Defense Advanced Research Projects Agency (DARPA)) issued contracts to companies and universities to prototype a speech recognition system with a 1,000 word English vocabulary. One of the keys

was that the system only needed to operate in a low-noise environment, something that would prove to be a major impediment to future systems development. Three separate contractors produced six different systems, though only one, the Harpy prototype from Carnegie Mellon University, met all the goals of the ARPA effort. From these systems came most of the commercial systems that would evolve over the next two decades, leaning heavily on the explosive growth of the microprocessor and storage technologies of the 1980s and 1990s. Unfortunately, no commercially viable products emerged from the huge investment in voice recognition. The major problems were things that had not been anticipated when the original concepts had been defined.

These included:

- Background Noise – One of the worst problems of making voice recognition viable in a real-world/field environment is isolating the speech being processed from background noise and conversation. This has proven to be a major impediment, since the solutions have only gradually been overcome, though a combination of hardware (microphone design and masking) and software (active noise cancellation, system "training," etc.). Much of the best work on this area was highly classified, as it was being applied to electronic warfare and anti-submarine technology during the Cold War. Thus, much of what has been successfully applied only became available in the last decade.

- Miniaturization – Despite rapid advances in microprocessor and electronics technology, the actual packaging of such a speech recognition unit into a portable unit has been difficult. A good example of the difficulties can be seen in the development of early Personal Digital Assistant (PDAs), which began with the less-than-successful Apple Newton™, and has evolved into the present-day Palm™ and Windows™ CE operating systems. This technology sells over two million units a year presently into the commercial and military markets, and still must be considered immature. This has taken over a decade in an environment with almost unlimited capitol and personnel resources, which provide some perspective on the issues of a universal translator, which might be carried by a soldier, sailor, airman, or Marine into combat.

- Accuracy/Approach – One major impediment to automated speech recognition has been the overall accuracy of the systems devised thus far. Much of the difficulty has been caused by the basic approach taken in the systems development. Most commercial speech recognition systems are based upon phoneme interpretation, which often has problems with regional accents and dialects, along with requiring ever-greater processing and storage capacity to operate in. This has meant that most such systems have required high-end desktop or laptop computers to run something unlikely to be lugged onto a battlefield or into a semi-hostile peacekeeping environment. At the same time, the problems of grammar and jargon upon the final translation can sometime be maddening, especially when in more complex languages like Chinese, Arabic, and English.

All of these factors have combined to make speech recognition a less-than-profitable technology, in spite of its tantalizing possibilities. However, recent developments seem to be changing that trend.
The passage in the 1990s of the Americans with Disabilities Act (ADA) has provided new funding and priority to the development. Designed to provide disabled citizens with greater access to professional and quality-of-life opportunities, ADA has proven a boon to many different technologies, from robotic wheelchairs to machines to carve ramps into existing concrete curbs. Over fifteen million Americans suffer from disabilities, which might be at least partially overcome by the implementation of a successful speech recognition technology, which could be packaged into a small, PDA-sized unit. That this need overlaps the long-standing military requirement for translation equipment has provided additional funding and interest, along with priority in both commercial and military marketplaces.

What may be the breakout development for both requirements has recently appeared in the form of a new system from Integrated Wave Technologies, Inc. (IWT). John H. Hall, a pioneer in the fields of integrated circuits, microprocessors, and low-power microelectronics, heads IWT. Far from merely being a technology "guru," Hall's real-world inventions range the first practical digital watches to the first computerized/programmable pacemakers. Based in Fremont, CA, IWT's efforts since 1992 have focused on the problems of speech recognition and translation, and bringing a practical package to market. In particular, Hall's work of late has been aimed using actual word/pattern recognition instead of the phonemes previously utilized by existing systems. In pursuit of this goal, IWT has taken advantage of several innovative ideas to create the first really useful translation appliance in history.

As a point of departure, John Hall acquired the existing algorithms of Vintsyuk and Zagoruyko, and funded several several generations of improvements for integration into an extremely small package. As mentioned earlier, one of the really attractive qualities of the Vintsyuk and Zagoruyko-based translation routines is that they require very little computational horsepower to run, meaning that a finished device can run on a less powerful, lower power consumption microprocessor than existing systems. This allowed IWT to build their first device into a PDA-sized board, running on PC-104 processor (an 80286-class microprocessor) clocked at only 5 MHz. This is about the same level of technology as the IBM PC/AT, which first appeared in the mid-1980s! Combined with new front-end noise reduction technology, IWT has produced several PDA-sized devices capable of some impressive tricks.

Working under Marine Corps and Department of Justice (DOJ) contracts, IWT has built a modified bullhorn, with a PDA-sized translation appliance installed in front of the output speaker. The idea was to have a number of pre-arranged key phrases that the appliance would recognize then output in one of a number of desired foreign languages. As might be imagined, such a system would be highly useful to law enforcement officers in the field, or deployed military personnel. The 25 initial DOJ units have been issued to police in California and Tennessee (West Palm Beach will be testing soon also), along with two systems for Marines deployed to the Balkans.

Composed of a single board Motorola HC16 computer running at only 4.7 MHz, the device takes its input from a special microphone, which is designed to minimize the pickup of background noise and conversation. The HC16 then runs the speech recognition algorithms, generating audio output for the bullhorn speaker. It can also produce a digital output, such as that required to send commands to a computer of PDA like a Palm™. The large flash memory connected to the HC16 provides plenty of room for the various words used to form phrases. Because of the need for absolute certainty in the accuracy of the translation, the system is programmed to only accept a select library of key phrases. These are designed to provide the user with an absolutely perfect translation, which is critical in cases where deadly force may be used or lives at stake. These might include things like:

"Do you require medical attention?"

"Please drop the weapon and lay down on the ground."

"Will you please take me to the local police station?"

Amazingly, while the system can store up to 500 phrases in each of 40 separate dialects, the version used by the Marine (which outputs in Serbian, only uses around 195. So far, ten languages have been programmed for the police version, with English as the primary input dialect. While the data stream from the translator could be fed into a voice synthesizer, the current system produces the output audio stream from a digitally recorded library of words, which maximizes clarity to the listeners. In fact, the entire system is designed to be as much like a normal bullhorn as possible, which is familiar to a wide range of military and law enforcement personnel. So far, the IWT system has been an unqualified success, with the only problem being that the users often refuse to return the test units for maintenance and upgrades. High praise indeed for a system that is little more than a prototype, and not even close to commercial release.

The potential of IWT's speech recognition technology is both impressive and groundbreaking. After several decades of relatively slow progress, there now appears to be an open road to effective implementation of speech recognition not only for military applications, but also the civilian marketplace. In particular, the possible applications within the disabled community may not only mean an improved quality of life, but also a new and emerging pool of talent in America's tight employment pool. Clearly, the time is right for this technology.

IWT, with the assistance of Eagan, McAllister Associates, Inc. of Lexington Park, MD, is planning to begin development of a PDA translation appliance, which would have both commercial and military applications. By taking advantage of the greater processing power available with the Palm™ or Windows™ CE operating systems and emerging microprocessors, IWT should be capable of bringing an important new product to market. What that product might control, as part of an emerging new user interface, might well be limited only by the imagination itself. In doing so, perhaps God's curse of Babel will finally be lifted from the earth, and mankind will finally speak with one voice.

The New York Times, July 30, 2000

The New York Times

July 30, 2000, Sunday, Late Edition - Final
SECTION: Section 6; Page 20; Column 1; Magazine Desk

LENGTH: 452 words

HEADLINE: The Way We Live Now: 7-30-00: Salient Facts; Cab Checker

BYLINE: By Cate T. Corcoran

BODY:

* * *

HOW DO YOU SAY 'DILITHIUM CRYSTALS' IN KOREAN?
This is not the final frontier. Since at least the 1950's, researchers in artificial intelligence have been trying to develop a real-time personal interpreter like the Universal Translator in the "Star Trek" shows. In January, President Clinton promised that "soon, researchers will bring us devices that can translate foreign languages as fast as you can talk." The German Research Center for Artificial Intelligence in Saarbrcken, Germany, has demonstrated a prototype system for laptop computers that will instantly interpret what two speakers are saying in Japanese and German, as long as the conversation is limited to travel arrangements. Integrated Wave Technologies of Fremont, Calif., has developed a prototype of a language box for police officers. They will eventually carry the device in their shirt pockets and bark commands into a lapel microphone. The device doesn't really translate, but utters prerecorded phrases like "Stop -- police -- or I'll shoot!" in such languages as Swahili and Vietnamese. --Cate T. Corcoran

GRAPHIC: Photo (Kevin R. Morris/Corbis)

LANGUAGE: ENGLISH

LOAD-DATE: July 30, 2000

Fortune

July 10, 2000

SECTION: FIRST: COPS AND ROBBERS, PART II; Pg. 64

LENGTH: 633 words

HEADLINE: Kapow! Zap! Gizmos Give Superhero Powers

BYLINE: Carol Vinzant

BODY:

Imagine the scene: A tight, fist-sized ball of yellow string gets shot 30 feet at a suspect fleeing a crime. As it closes in on him, the ball opens into a 16-foot net that ensnares the bad guy. This is no Spiderman cartoon; police are now using Capture Net--just one of the devices on display at a Denver convention touting technology for cops, paramedics, and firefighters. Like the Capture Net, a lot of the products were backed by the National Institute of Justice, an obscure federal agency charged with boosting research into public safety.

NIJ exists because crime fighting doesn't pay. Local police departments--90% of which have fewer than 30 officers--don't have the budget to develop their own new products. So NIJ pays for research that turns private or military technology into products cops can use. The agency gives out research grants, then serves as a clearinghouse to make sure the findings are available to other companies. (Research that the government funds isn't proprietary.) Recent grant recipients include large public companies like Raytheon, which is developing thermal-imaging equipment able to locate people in the dark, **as well as smaller private firms like Integrated Wave Technologies, which has built a mini-translator to spit out phrases like "You have the right to remain silent" in three languages.**

Some of the gadgetry rises to James Bond levels. At the Denver convention, NIJ demonstrated a prototype radar flashlight designed to detect subtle movement--even quiet breathing--on the other side of a wall. The agency is also sponsoring research to develop a handheld device that can detect, from seven feet away, whether a suspect is carrying a weapon. It uses audible sound waves that bounce off metal or plastic.

In addition, NIJ has become a kind of Consumer Reports for cops, testing products that are already on the market. One goal is to cut down on sham companies selling bogus devices to the police. For example, it rates bullet-resistant vests and tested an $ 8,000 device called the Quadro Tracker, which, when powered by the static electricity of a user's breath, purportedly could detect drugs with an oscillator. That sounded too good to be true, and it was. The Quadro

Tracker didn't work, leading the agency to issue a warning to police departments across the country.

Also nixed by NIJ--Sticky Foam, the much hyped goo that can be hosed onto fleeing suspects to immobilize them. "It turned out to work a little too well," says David Boyd, director of NIJ's Office of Science and Technology. Boyd's researchers found that it took about ten minutes per square inch to remove the foam from a person's skin, making it too much of a hassle for the police. (The military didn't think so--it has authorized the use of Sticky Foam.)

Some NIJ-funded research goes toward computer applications. The agency gave $ 3 million to Anser, a not-for-profit research group, to develop software that recognizes people by 35 or so key points on their face. Anser found the facial points work as well as fingerprints in identifying people and originally planned to use it searching the Internet for photos of missing kids. But now Miami drug cops are testing Anser's system to match suspects to a database of mug shots. It's also testing a camera that "learns" the people it watches and alerts a guard if, say, the wrong person takes a bike under its surveillance.

One of the most powerful cop gadgets now in testing is an electromagnetic Auto Arrester that the police can shoot under cars during high-speed chases. Get it close enough, and the Auto Arrestor's magnets shut down the getaway car's engine. That's bad news for would-be escapees, but good news for everyone else on the road. Now if NIJ could only find something to clear up traffic jams.

GRAPHIC: TWO COLOR PHOTOS: JEFFREY LOWE, These gadgets resemble hair dryers, but they detect motion and weapons.; COLOR PHOTO: JEFFREY LOWE, This virtual-reality simulator shoots back. (Don't worry: It fires only pellets.); COLOR ILLUSTRATION: MARTIN KOZLOWSKI, THE ECONOCLAST

LANGUAGE: ENGLISH

LOAD-DATE: June 28, 2000

AP Online

December 4, 1999; Saturday 13:32 Eastern Time

SECTION: Domestic, non-Washington, general news item

LENGTH: 546 words

HEADLINE: Police Test Voice Translator

BYLINE: JORDAN LITE

DATELINE: OAKLAND, Calif.

BODY:

Hoping to get a Spanish-speaker behind the door to open up, the English-speaking police officer makes his request to a little hand-held box.

In a flat tone, reminiscent of the spaceship computer Hal in Stanley Kubrick's "2001," the box repeats back the message.

Then: "Policia! Abra la puerta en esta momento!" the machine says, in the voice of a Spanish-speaking cop with considerably more emotion.

The scenario may one day become a reality. Big city police officers routinely encounter speakers of many different languages, and often have trouble communicating effectively.

So police in Oakland have begun testing a battery-powered language interpreter, the Voice Response Translator, which responds to as many as 125 vocal commands, spitting out statements and questions in Spanish, Cantonese or Vietnamese that demand yes-or-no answers.

During a demonstration last week, Everett James, a police department community liaison officer, showed off the machine. Saying "Miranda" into the translator's microphone, for example, produces the message, "You have the right to remain silent ...," in a foreign language.

So far, James has only been able to try it out on police station walk-ins, who have had few complaints. "They're obviously pretty perplexed when I pull this out," he said. "No one knows what it is."

The National Institute of Justice began developing the device after a 1994 task force noted that officers in places like Oakland's Chinatown, where 70 percent of people are older than 65 and not proficient in English, are increasingly unable to communicate effectively.

James is helping scientists enhance the tool, which at a boxy 4 inches by 6 inches can't easily be carried by foot patrols. An upgraded version smaller than a pack of cigarettes is in the works, and Fremont-based Integrated Wave Technologies hopes to begin marketing it next year at a retail price of $900.

"This is a big focus of community policing," said Tim McCune, an analyst with Eagan McAllister Associates, a Washington area defense consulting firm working on the device. "Being able to say hello, thank you, good bye officers say it really helps because they want to be viewed as accessible to the community."

An infinite number of phrases and languages can be programmed into the device. James, as well as officers in Nashville, Tenn., are working out the machine's kinks and deciding what phrases they need.

"It would be a good tool for any officer," said Oakland police Officer Barry Ko, who often speaks Cantonese during his patrols. "It's more personal. 'I'm out there, I want to help you even though I can't understand you.'"

But Nelleke Van Deusen, a Berkeley sociolinguist who works with medical providers to improve their communication with members of Laotian mountain tribes living in the United States, said the device has "serious shortcomings" among them the machine's inability to understand anything other than its lexicon of English commands spoken by an officer.

"In a tense situation with a police officer, a nonnative speaker might get suspicious about questions that are asked and might not be given the chance to respond appropriately given they only have a yes-or-no question to answer," she said.

LANGUAGE: ENGLISH

LOAD-DATE: December 4, 1999

National Public Radio (NPR)

SHOW: ALL THINGS CONSIDERED (8:00 PM ET)

December 2, 1999, Thursday

LENGTH: 692 words

HEADLINE: CAPTAIN KEN PENCE OF THE NASHVILLE POLICE DEPARTMENT DISCUSSES A TRANSLATION DEVICE BEING TESTED IN HIS CITY

ANCHORS: NOAH ADAMS

BODY:

NOAH ADAMS, host:

In Nashville, Tennessee, police are testing an electronic language device that could help in dealing with people who don't speak English. It's called a Voice Response Translator. The officer speaks a key phrase, and the machine responds. The first models have been in use in Oakland, California, and now the testing will include San Diego and Nashville. Captain Ken Pence is with the Nashville Police Department. He says the translator is necessary because so many different groups of people have come to Nashville to find jobs.

Captain KEN PENCE (Nashville Police Department): We're going to try it with Hispanic and Vietnamese, generally. We have a Laotian population, so when we get the new units, which are smaller, we'll be programming them with Laotian, too.

ADAMS: How many different languages are your officers dealing with on a regular day in Nashville?

Capt. PENCE: Oh, probably 20 languages. And to change languages, you can have--there's a little disk inside that--you just simply replace this disk and you have three more languages.

ADAMS: Tell us what the instrument looks like, actually.

Capt. PENCE: Right now, the prototype is about the size of a small AM-FM radio. Fits in a shirt pocket, or clips to the belt, and the microphone is a little tube. Looks like a Bic pen, about two inches long. This one does Vietnamese, Spanish and Cantonese.

ADAMS: All in the same unit?

Capt. PENCE: Yes.

ADAMS: Right. So you have to switch between those languages?

Capt. PENCE: No, you just say, 'Change language.'

Voice from Translator Machine: Which language?

Capt. PENCE: Start Spanish.

Voice from Translator Machine: Start Spanish.

Capt. PENCE: See, that's what you do, and you're good to go.

ADAMS: What if you wanted to say to a person that an officer has stopped--what if that officer wanted to say you were driving too fast?

Capt. PENCE: Too fast.

Voice from Translator Machine: Too fast. (Spanish spoken)

Capt. PENCE: And it's made for police. This isn't your generic conversational. This is made for police, so the phrases are very, very useful and very specific to domestic disputes, lost children, medical problems; do they need help, or just conversational greetings.

ADAMS: Can you switch to Vietnamese for us, please?

Capt. PENCE: Sure. Change language.

Voice from Translator Machine: Which language?

Capt. PENCE: Start Viet.

Voice from Translator Machine: Start Viet.

Capt. PENCE: OK. Now I'll give you the same thing. How about I ask permission to search their car?

ADAMS: OK.

Capt. PENCE: Vehicle search.

Voice from Translator Machine: Vehicle search. (Vietnamese spoken)

ADAMS: So it is not recognizing--it's not really speech recognition; it's recognizing certain code words?

Capt. PENCE: What it does, it does Russian phrase recognition. So it's taking a phrase, instead of phonemes in the speech, and it recognizes a pattern. So it only recognizes the officer that trains with it. But you can go out in pretty much a loud-noise environment and it'll pick up your speech pretty well. And it's loud enough that the other person can hear it.

ADAMS: So you would have one, and I would have one, and we couldn't use each other's?

Capt. PENCE: Right. Well, I mean, you could spend 30 minutes training it, and then you could.

ADAMS: What about in a really tense situation? If somebody were running away from an officer--this has happened in many cities--and they would say, 'Stop, or I'll shoot,' and the person running wouldn't know that language, could it be useful at all in that situation?

Capt. PENCE: I'm not sure it could. Some people have said, 'Halt, police,' and, of course, you might know stop in Spanish, but you might not know halt, or--especially in Vietnamese. You know, you wouldn't--or Hmong.

ADAMS: Captain Pence, what do you have--if you don't have this, what do you have? What is your alternative?

Capt. PENCE: A lot of grinning, waving hands, shuffling of feet and drawing pictures on little scraps of paper.

ADAMS: You could, I suppose, bring somebody back to the station house and get a translator, or do interpretation by telephone?

Capt. PENCE: Right. We use one of the language lines, but that costs from $2 to $ 5 a minute. And then in the middle of the night if you need somebody that speaks Hmong, it's very--you know, it might be an hour. And if you're on the Interstate, that's--when somebody runs up to you, this gives you a first line of defense to serve the public, and our chief's very technologically minded. And we want to provide some service to the public, and the way to do that is you need to provide that service not just in English.

ADAMS: Captain Pence, thank you for your time.

Capt. PENCE: Thank you.

ADAMS: Ken Pence of the Nashville Police Department.

You're listening to NPR's ALL THINGS CONSIDERED.

(Soundbite of music)

LANGUAGE: English

LOAD-DATE: December 3, 1999

APPENDIX C: IWT Testing Paper

Cost/Benefit Analysis Of Language Translation Devices Through Improved Testing

Integrated Wave Technologies, Inc.
4042 Clipper Court
Fremont, CA 94538
(510) 490-9160
(510) 353-0261 FAX
http://www.i-w-t.com

Overview and Introduction

Development of speech-to-speech language translation devices was spawned to a large extent by the 1990-1991 Gulf War and accelerated significantly after the Sept. 11, 2001 attacks. Systems were deployed in small numbers after 1997 and in the thousands since 2001, but development and procurement have outpaced developmental and operational testing.

Testing to date with speech-to-speech language translation devices has consisted mostly of characterization, experimentation and demonstration rather than true testing and evaluation, something one US Army observer called "drive-by fieldings." This has hurt the voice-to-voice translation field by allowing unproductive programs to continue while missing the opportunity to justify expenditures by documenting actual and potential operational contributions from systems that have performed well.

Testing allows evaluators to determine whether speech-to-speech language translation devices have or can make contributions to operational activities by focusing on tasks where users have identified language communication issues.

To conduct a useful testing program, personnel with operational experience should diagram tasks where language communications difficulties exist and describe the problems that cause either friction or fail points. Using a task documentation/segmentation/analysis approach, evaluators can measure improvements in users' ability to accomplish tasks when they have voice-to-voice devices to assist them.

Potential benefits – e.g., reduced search time per vehicle because of increased occupant understanding/cooperation and more effective handling of persons during house searches – can be documented effectively based on test and evaluation. Alternatives for reducing or eliminating fail/friction points such as live translators, graphical aides and differing kinds of voice-to-voice translators should be identified and the relative costs/benefits of each examined.

Discussion

Past assessment efforts have fallen short of being realistic operational testing and evaluation reviews. Users were trained in a classroom setting to use voice-to-voice devices and then encouraged to demonstrate and use the devices in their missions.

The data that emerged from these assessments consisted mainly of descriptions of simple demonstrations for senior personnel. Also included are operator impressions of how the systems might contribute in future missions. Actual integration into the repetitive tasks such as house searches, vehicle searches, entry control points and force protection was not done. Reasons given by users for failure to use the voice-to-voice devices (if feedback is received at all) are that using them would be disruptive in

the field exercise, or that there was some difficulty in using the device that they hoped would be solved by more training time.

The basic problem is that voice-to-voice translators are a new class of equipment and gaining user acceptance and device integration into operations has proven to be more difficult than anyone anticipated. Users cannot envision how the devices would work operationally and don't like to be the high-visibility guinea pigs for initial use at the unit level.

User acceptance and device integration is the essential component to testing and evaluation. One successful tool for gaining initial user confidence is to show videotapes of field exercises where the devices are employed. Another is to focus user training on working with the device to accomplish specific tasks in role-playing situations. Testing can then proceed with early adopters willing to employ the systems in operational or realistic settings.

Developing Testing Criteria

Operational and Developmental[7] Testing criteria fall into two areas:

1) Functionality related to the basic task the device must do: and
2) Environmental Suitability/Survivability Criteria related to the devices suitability/survivability in the user's operational environment.

Functionality criteria need to be developed by first selecting critical user tasks being frustrated by language-based communication problems. These range from simple tasks such as warning persons away from secure locations to complex ones such as interrogations of hostile individuals.

User tasks then need to be broken down into task elements so failure points without the system can be identified. Though there are different correct ways of conducting tasks such as house and vehicle searches, the failure points will often be the same even if a user does not conduct the task in the same way. Task criteria must focus on whether the system enables the user to overcome a validated task fail point.

Task example:

1) House Search

[7] **Developmental testing** involves determining the extent to which a system/device meets performance parameters set during a formal requirements process. System performance levels are characterized either as "threshold" or "objective", with the latter being scheduled for later in a system's development. These are often linear measurements – e.g., a system will have to demonstrate the ability to work in XX dB of noise early in its development, but (XX+35) dB before entering advanced testing and/or production. **Operational testing** involves putting a system into a realistic (or real) environment and made to perform the tasks for which it was designed: "A continuing process of evaluation that may be applied to either operational personnel or situations to determine their validity or reliability."

a. Occupants must be informed that persons coming to search the house are US military personnel
    i. Fail Point without device: Occupants might not know initially that persons are authorities and resist entry.
b. Occupants must be divided into two groups – women/children and men – and directed to go to separate areas.
    i. Fail Point without device: Operator is unable to direct persons effectively by voice command and must physically move them.
c. Occupants are told they are going to be blindfolded. After being blindfolded, they are told that if they can provide information about weapons caches or anti-coalition individuals, they should indicate this by putting their chin on their chest.
    i. Fail Point without device: Operators cannot ask this question and must bring forward a human interpreter or bring house occupants back to a secure area.
d. Individuals are separated so the identities of those providing information are not known to the others in the group.
    i. Not applicable if questions cannot be asked.

Survivability Testing

Environmental suitability/survivability criteria need to be generated from actual data. Testing for ambient temperature survivability is simple, but often not done. Testing for problems resulting from other factors such as high levels of sand/dust clogging equipment cooling systems is more difficult and needs to be done in the operational environment.

Alternatives to electronic language translation systems include on-the-spot human translators, translators connected telephonically, and other translation devices such as simple graphical cards.

Conducting Tests

Environmental Suitability/survivability criteria testing should be addressed first by independent laboratories, i.e., "shake and bake" testing. Much of the relevant work will be relatively low cost. Simple oven tests, where devices are operated at temperatures representative of some current operational environments (140 degrees F) and drop tests can reveal the limitations of equipment before significant numbers are procured and deployed. Other tests, such as battery life/power consumption, can validate or refute vendors' statements about critical performance areas such as time between recharging.

Please note there is a tendency to believe that performance in areas such as power consumption and ruggedness can be added onto a system after task functionality problems are solved. Such improvement efforts are usually only marginally successful,

and involve tradeoffs such as additional weight and/or size. Therefore performance areas critical to system deployment should be tested as threshold requirements prior to significant investment in development and/or acquisition of systems.

Conclusion

Voice-to-voice language translation devices are tools developed in response to specific operational shortfalls cited by users. System development/deployment/testing must be tied back to specific operational problems. Systems/devices must address specific tasks and provide solutions to problems in completing those tasks so that the benefits of system use are documented.

Detailed testing plans must be prepared before expending resources on field activities. These plans should include:

1) A description of tasks for which users have provided language-communication-related fail point feedback;
2) A description of translated phrases that would address these language-related fail points;
3) A description of system/device operation in the task setting discussing how the device would be integrated by the user;
4) A description of the expected successful outcome of system/device insertion into the task; and
5) The method of documentation/quantification of benefits from use of the system/device.

APPENDIX D: TECHNICAL APPROACH

Introduction:

Integrated Wave Technologies, Inc., has made significant hardware and software advances related to speech recognition.

IWT voice recognition technology performs in a robust manner using novel signal processing methods. The accuracy of the system exceeds 99% in adverse conditions using different communication channels and in the presence of background noise. The core technology is very efficient and inexpensive to implement: A standard 8-bit audio/digital converter and a 5 MHz controller chip is sufficient to run the program.

IWT began ten years ago on an R&D track radically different from that of other speech recognition companies. Companies such as Dragon Systems, Inc., were founded by highly talented linguists who attempted to mechanize their knowledge of how humans process speech into computer systems. Their systems try to recognize phonemes – parcels of speech such as consonants and vowels peculiar to each language – and then assemble them into words and words into sentences using contextual analysis, much like humans do. Phonemes are extremely subtle, often only 100 milliseconds long. The software developed to do this was supposed reside on top-line personal computers and allow persons to convert conversational speech into text. Their hope was that documents such as letters, memos and reports might be done with little or no use of a keyboard.

This phoneme-based R&D – now dating back over 30 years – has not been technically or commercially successful. These speech recognition companies have been able to eke out incremental improvements in the products, increasing vocabularies and accuracy in some situations, but noise immunity and overall accuracy have not increased to the point where these systems are useful for real-world operation. These companies have placed their technical hopes in the development of more effective noise-canceling microphones, noise cancellation from digital signal processors (DSPs), and advanced software dependent on more powerful PC processors.

IWT is a company with a core expertise in miniaturized electronics and pursued a different path than that of these linguists. Rather than emulating human speech recognition, the Company approached this problem as its founder approached the challenges of producing the first electronic watch, the first computerized heart pacemaker and many other technological firsts. This approach was to analyze precisely the delicate audio signals produced by human speech and develop innovative ways of extracting this sound from background noise and recognizing it with high accuracy. While perhaps insurmountable roadblocks were encountered in

pursuing the linguistic approach, the Company was able to reach its technological goals.

Key to this approach and its success are the talents of a team of scientists and engineers who have been employed by IWT for the past ten years. Algorithms that, as pure calculus equations, would take literally a supercomputer to execute have been refined so that they are performed on a 5 MHz controller chip in milliseconds. These individuals, working as a closely knit team, have been able to produce complementary advancements in pattern-matching mathematics, signal analysis and software/hardware implementation.

A key to the Company's current product development position was its strategic decision in 1991 to develop recognition systems for miniaturized applications such as cellular phones and personal digital assistants rather than for personal computers. This decision resulted from IWT's analysis that the greatest value-added for speech recognition is for devices that, unlike desktop units, do not have useful keyboards. This decision has allowed the Company to create a capability that converges with the emergence of web-capable personal digital assistants and mobile phones and provides IWT with its overwhelming competitive advantage.

IWT decided also to work closely with high-end law enforcement and other government users during its development process rather than market its technology in its incrementally capable stages. The Company believes that speech recognition technology must cross performance thresholds to be useful and commercially viable, and that other companies have damaged the market for speech recognition by trying to sell products that do not live up to promised minimum performance levels.

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 in inventing new solutions for electronics problems that had eluded other developers. 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. A full list of Hall's technical innovations is included below.

This 40-year technical resume indicates why the Company has been able to make groundbreaking advances in speech recognition where other firms – including large ones such as Intel and Microsoft – have failed. His unique perceptions and

insights related to analog/digital signal processing – technology at the center of the speech recognition challenge – have led to the significant advancements described in this plan.

The Company's founder has 40 years experience in electronics manufacturing start up and continued growth operations, having designed and built factories in Ireland, Japan, Finland and the US. For the products described in this plan, the Company will outsource the manufacturing to entities with which Hall has long-standing relationships. IWT does not foresee needing to build any manufacturing facilities.

Technology's Background

Speech recognition development began in the 1950s. In 1952, Bell Laboratories built a system based on measuring spectral resonance to identify the vowel sounds of single digits. A program at RCA in 1956 also used this vowel sound approach, and other work was begun in Japan in the 1960s.

An approach important to this proposal was begun in the 1960s in the former Soviet Union. During that time, research by T.K. Vintsyuk and others produced work related to matching speech inputs to patterns stored in a databank. The Company's recognition algorithm is a unique and highly refined example of this basic approach.

In 1971, the Advanced Research Projects Agency (ARPA), now called the Defense Advanced Research Projects Agency, challenged American companies and universities to develop a speech-understanding system with a vocabulary of at least 1,000 words capable of processing connected speech with an error rate of under ten percent in a low-noise environment for use by many cooperative speakers. Most speech recognition systems – other than the Company's – descend from these ARPA efforts, backed by private investment in addition to some continued public funding. Companies ranging from IBM and Philips as large integrated high-technology firms to Dragon Systems, Inc., Kurzweil and Lernout & Hauspie (L&H) as entities dedicated to speech recognition technology have labored to produce saleable consumer products. A host of smaller firms have also devoted significant R&D efforts in the field.

High-technology leaders have made investments in these firms to promote product development and to secure access to this work. For example, Lernout & Hauspie Speech Products announced last year that Intel Corporation has signed a binding letter of intent to invest $30 million in L&H. L&H also announced last year that Microsoft elected to exercise its warrants to purchase an additional 857,142 shares of L&H Common Stock at an exercise price of $17.50 per share, for a total exercise price of approximately $15 million. This purchase increased Microsoft's investment in L&H to approximately 3.75 million shares, or approximately 7% of L&H outstanding Common Stock. These warrants were issued to Microsoft in connection with Microsoft's $45 million investment in L&H in September 1997.

Despite this impressive dedication of resources, no company has been able to market a commercially successful speech recognition-based desktop product or been able to produce significant revenue. L&H – the parent of Dragon, Kurzweil and Dictaphone – is currently attempting to reorganize in bankruptcy courts in the U.S. and Belgium.

Development of successful speech recognition devices has been hampered in general by problems of background noise, miniaturization and accuracy. Phoneme-based systems such as those produced by L&H, Dragon, Philips and IBM have the further problem of being language and accent-specific, so that new versions have to be developed at great cost to serve desired markets.

Until recently, speech recognition development focused on software dedicated to the task of large vocabulary speech-to-text dictation on desktop computers. The technically poor performance of this software, combined with the advent of widespread Internet and wireless phone service, have led speech recognition companies to look toward the development of large-scale recognition centers on one hand and to miniaturized speech recognition for cellular phones and personal digital assistants on the other.

Some success, under carefully controlled conditions, has been achieved by Nuance and others in the field of large-scale recognition centers, generally described as having Interactive Voice Response (IVR) capability. Instead of using touch-tone responses, the IVR uses key words and phrases to direct the flow of the call. If callers clearly understands what phrases and words to use, they will find the applications easier to use than touch-tone systems. The performance of these systems has been limited and this market has not proven to be profitable.

Voice Information Associates, Inc. has recently released a study, *Automatic Speech Recognition for Telephony Applications: The World-Wide Market: 1995 – 2003,* which predicts the growth in end-user revenue obtained from automatic speech recognition products in telephone applications. The unified messaging subsegment will show the strongest growth with an cumulative annual growth rate (CAGR) in excess of 80%, while operator services is projected to have a more modest CAGR of 27.5%. The service revenue obtained by the carriers for ASR-based services is anticipated to be in excess of $9.5 billion in 2003.

While this segment of the speech recognition market might eventually prove to be technically and commercially viable, its limitations are apparent. These devices are susceptible to background noise interference, and users overcome this by calling from relatively quiet areas. As important, these systems require large-scale hardware resources for implementation, limiting them to well-funded applications.

Background:

The Company's original core software technology was developed in the former Soviet Union, in an atmosphere where expensive and complicated resources were limited. Russian scientists were forced to use inferior (by Western standards) computing machinery. To get results, they had to rely on elegant, yet parsimonious, algorithms to achieve comparable results being accomplished in the West with more powerful computers. IWT's current generation of software is an entirely new creation, developed by the Company's employees and a generation ahead of the already-impressive work acquired eight years ago.

In the 1960s, Vintsyuk first proposed the use of dynamic programming methods for time-aligning a pair of speech utterances.[8] Although the essence of the concepts of dynamic time warping, as well as rudimentary versions of the algorithms for connect-word recognition, were embodied in Vintsyuk's work, it was largely unknown in the West and did not come to light until the early 1980s -- long after more formal methods were proposed and implemented by others.

A significant milestone in voice recognition work was achieved in the 1970s by Velichko and Zagoruyko.[9] They created perhaps the first viable and useful voice recognition system. These Russian studies helped advance the use of pattern-recognition ideas in speech recognition. It should be noted that these studies predated those by Sakoe and Chiba in Japan[10] and Itakura in the U.S.[11]

The work in the Soviet Union continued on with an emphasis in robust voice recognition and voice identification for use in military and covert operations. A wealth of commercially available potential research soon became available after the fall of the Soviet system. IWT secured the commercial rights to the most significant and applicable research. The technical details have not been published so as to protect these rights.

The Company has achieved its significant technical breakthroughs partly due to its aggressive development-deployment-evaluation-redesign-redeployment-reevaluation program. IWT has created new generations of its devices in as little as six months, incorporating new scientific insights by Dr. Hall and other staff members. IWT's practice is to field new technology quickly in demanding situations and draw lessons from it rather than attempting to provide it to consumers prematurely. IWT's products enjoy a vast technological lead over those of its competitors and an unmatched reputation among sophisticated government users.

---

[8] T.K Vintsyuk, "Speech Discrimination by Dynamic Programming," *Kibernetika*, 4(2): 81-88, Jan./Feb. 1968.

[9] V.M Velichko and N.G. Zagoruyko, "Automatic Recognition of 200 Words," *International Journal of Man-Machine Studies*, 2:223, June 1970.

[10] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," IEEE Tans. Acoustics, Speech, Signal Proc., ASSP-26 (1): 43-49, February 1978.

[11] F. Itakura, "Minimum Prediction Residual Applied to Speech Recognition," IEEE Tans. Acoustics, Speech, Signal Proc., ASSP-23(1): 67-72, February 1975.

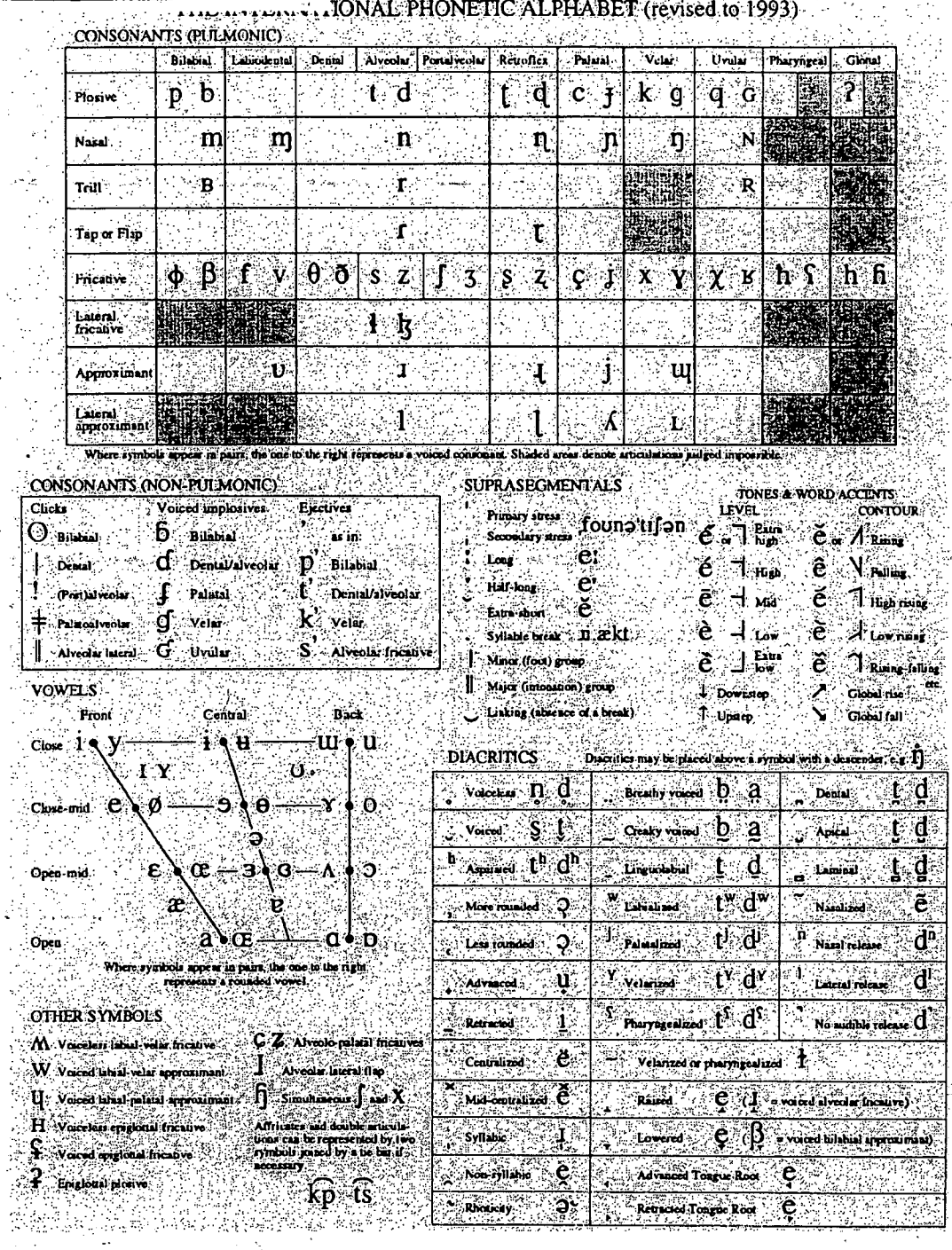
The Company's technological breakthroughs also have come because it has taken an approach fundamentally different from developers such as Lernout & Hauspie (L&H), Dragon Systems (now part of L&H) and IBM. These companies, or their speech recognition divisions, were founded by highly talented linguists who attempted to mechanize their knowledge of how humans process speech into computer systems. Their systems try to recognize phonemes – parcels of speech such as consonants and vowels peculiar to each language – and then assemble them into words and words into sentences using contextual analysis, much like humans do.

Phonomes are subtle variations in speech peculiar not only to each language, but each accent and/or dialect within that language. Each phoneme is perhaps only a hundred milliseconds long, and recognition software based on them must separate them from background noise and each other to identify them continuously. These recognized phonemes are then assembled into a word, and then words into sentences.

This recognition approach is dependent on a continuous string of tasks being done correctly. If a "v" sound is misrecognized as an "f", then the entire word will be wrong even if the phonemes that follow are recognized correctly.

This problem has driven the complexity of phoneme-based speech recognition software. To compensate for phoneme misrecognition, this software uses a probability analysis to attempt to identify words from phonemes, and contextual analysis to assist in selecting words. For example, if the system recognizes "the dog" as the beginning of a sentence, it will conclude the next word is "barked" rather than "borrowed", though the recognition part of the software might not be able to discriminate between those words. This method of improving accuracy is very limited – many possible choices exist for each word positioned in a sentence – and it is of no use for command/control recognition as there is no context for words such as numbers.

Figure 1: Phoneme Chart



IWT's analysis, reached initially eight years ago, has been that this approach is fundamentally flawed. Other research supports this analysis. Dr. Steven Pinker, a noted language expert and director of the Center for Cognitive Neuroscience at the Massachusetts Institute of Technology, described the fundamental problems with the phoneme/linguistic approach in his book *The Language Instinct*. He wrote that, "no human-made system can match a human in decoding speech." He elaborated, adding:

> Sentences and phrases are built out of words, words are built out of morphemes, and morphemes, in turn are built out of phonemes. Unlike words and morphemes, though, phonemes do not contribute bits of meaning to the whole. The meaning of *dog* is not predictable from the meaning of *d*, the meaning of *o*, the meaning of *g*, and their order. Phonemes are a different kind linguistic object. They connect outward to speech, not inward to mentalese: a phoneme corresponds to an act of making a sound.[12]

IWT identified in its early analysis of speech recognition technology that phoneme-based systems are also highly susceptible to background noise and require large computer processing resources to operate. A key flaw in the phoneme approach is that the processors needed to implement it create system noise that interferes with speech recognition. Each generation of phoneme-based software has required increasingly powerful processors, which in turn interfere with the system's ability to recognize speech, evolution that is partly self defeating.

Similarly, Pinker describes the futility of trying to guess words from the sentence context because of the "sheer vastness" of language. He wrote:

> Go into the Library of Congress and pick a sentence at random from any volume, and chances are you would fail to find an exact repetition no matter how long you continue to search. Estimates of the number of sentences that an ordinary person is capable of producing are breathtaking. If a speaker is interrupted at a random point in a sentence, there are on average about ten different words that could be inserted at that point to continue the sentence in a grammatical and meaningful way. (At some points in a sentence, only one word can be inserted, and at others, there is a choice from among thousands; ten is the average). Let's assume that a person is capable of producing sentences up to twenty words long. Therefore the number of sentences that a speaker can deal with in principle is at least 10^{20} (a one with twenty zeros after it, or a hundred million trillion.) At a rate of five seconds a sentence, a person would need a childhood of about a hundred trillion years (with no time for eating or sleeping) to memorize them all.[13]

Rather than emulating human speech recognition, the Company approached this problem as its founder approached the challenges of producing the first electronic watch and the first computerized heart pacemaker. This approach was to analyze precisely the delicate audio signals produced by human speech and develop innovative ways of extracting this sound from background noise and recognizing it with high accuracy. While perhaps insurmountable roadblocks were encountered in

---

[12] Pinker, Steven, "The Language Instinct: How the Mind Creates Language," William Morrow and Company, New York, 1994, pp. 162-163.

[13] Pinker, op.cit., p. 87.

pursuing the linguistic approach, the Company was able to achieve the specific results described below.

IWT's technologies are not merely superior to those of other companies. They cross performance thresholds that will allow them to be the basis of new products and new markets. In addition to securing ownership of the algorithms, IWT has pursued an aggressive strategy of developing essential implementation technologies. The Company is in the process of completing patent application documentation for these technologies and believes that the resulting patents will prevent competitors from developing similarly capable products.

These technologies also can be combined with existing innovations such as the Universal Serial Bus standard and emerging ones such as the Bluetooth radio frequency interface standard to become important integrated parts of the next generation of computing systems.

Background noise is simply the everyday noise that surrounds us. The noise level exceeds 40 decibels often even within the perceived quiet of an office because of ventilation systems, equipment and other people. City street noise is generally around 80 decibels, while the noise within moving vehicles rises to about 100 decibels at highway speeds. The military and police applications in which the Company's products are being demonstrated routinely experience background noise over 100 decibels.

The Company has developed systems based on its unique intellectual property that have unprecedented capabilities in background noise situations over 100 decibels. This capability, described below in detail, is key to IWT's competitive advantage. Other speech recognition systems being marketed cease to recognize generally at about 30 to 40 decibels of extraneous noise. These systems "lock up" under the pressure of noise above that level, making them useless.

The Company has made similar strides ahead in miniaturization of speech recognition. Systems such as those made by Lernout & Hauspie, the largest speech recognition software manufacturer, require a Pentium III computer with 128 megabytes of memory. IWT has reduced the size of its complete speech recognition system to a two-inch by three-inch board that weighs less than an ounce. In addition to being a highly capable device capable of integrating the Company's advanced speech recognition into many products, this board also demonstrates that IWT could integrate its technology into PDA running the Palm OS or Windows CE without increasing processor power or overall size.

IWT has also produced systems that approach 100 percent in recognition accuracy, even when using short commands such as numbers.

The problems of background noise susceptibility, large system size and poor accuracy have prevented the emergence of successful desktop, laptop and handheld speech recognition products, despite intensive marketing efforts by L&H, Dragon

Systems, Inc. (purchased by L&H) and others. Reviews of these products are critical of poor accuracy rates, long times needed for setup/training, and the high level of computing resources needed for basic operation.

A hardware-based competitor of IWT is Sensory, Inc., founded by well-regarded innovators who produced successful speech synthesizer integrated circuits. The work done by Sensory has been impressive, but suffers from the limitations of systems such as L&H and Dragon. The Company can demonstrate also that its products have significantly higher performance that is at least a generation ahead of systems using Sensory's chips.

Sensory and others attempting to produce useful handheld speech recognition are trying to augment the performance of their technology by using digital signal processors (DSPs) to reduce background noise. These companies hope that using the DSPs to remove noise will allow them to achieve breakthroughs in performance. IWT's analysis is that this approach will allow for noise rejection of only 30 dB, far less that the Company can do with its technology. Further, DSPs work best reducing repetitive noise, while the noise encountered in using speech recognition is non repetitive. IWT therefore believes that it will maintain its lead in speech recognition technology for a significant amount of time.

A speech recognition application area involves the area of telephonic speech recognition. This application area, developed by Nuance and other companies, involves recognizing words and phrases spoken over telephones to computer-based systems. These systems base their success on being able to devote high-level computing resources to the recognition of a limited number of words and phrases, generally spoken in relatively low noise environments. Even using these constraints to increase performance, call center systems have not met technical or commercial expectations.

Call centers are attractive to many developers because they appear to allow mobile devices to be used with phoneme-based speech recognition systems that require large-scale processors. Developers state correctly that they can use the most advanced phoneme systems and software-based noise reduction. Deeper consideration reveals significant problems with this approach. First, the commercial requirements are for these systems to handle many thousands of recognition tasks simultaneously. Thus, what might be made to work for a single call or for very small numbers would require perhaps thousands of Pentium equivalents to succeed. Second, such systems are victims of input variations. Telephones are designed to operate within a specific band of frequencies, but otherwise produce vastly different characterizations of the same voice. Further, noise is introduced into the system in many ways: from the background; from the telephone set; from the landline or mobile connection; and from long-distance switching systems.

For these reasons, IWT has no plans to invest resources in the call center market area.

The Company believes the best marketing opportunities for speech recognition exist in the area of handheld devices of various types. The miniaturization of computing and telephonic resources have created a greatly increased need for speech recognition command/control and information entry. Unlike desktop and laptop computers, PDAs and mobile phones have tiny, awkward keyboards and no complementary mouse.

Several new market opportunities are based upon IWT's ability to build very compact, low-cost devices that recognize all languages, dialects and impairments in speech in environments with loud and unpredictable background noise. The core technology was developed in the former Soviet Union in a atmosphere where expensive and complicated resources were limited. Russian scientists were forced to use inferior (by Western standards) computing machinery. To get results, they had to rely on elegant, yet parsimonious, algorithms to achieve comparable results being accomplished in the West with more powerful computers. IWT acquired the applicable rights to this technology in 1991 and continues to fund related research. Its current generation of software is not related as intellectual property to this earlier work and is being patented separately.

IWT's system analyzes the frequency and energy characteristics of the sounds rather than the phonemes. This allows IWT to match sounds directly and precisely to templates of voice commands and also to distinguish the voice command sounds from all types of background noise, even other human speech.

The second characteristic is the highly efficient architecture of the Soviet algorithm. Software based on this algorithm can run on relatively modest hardware – a 286 processor equivalent compared with Pentiums require for Western voice recognition systems – which means that IWT has been able to create systems that are vastly smaller, more power efficient and cheaper than systems relying on Western systems. The algorithm can be embedded in a single chip, which can then be integrated directly into devices where voice command is desired. Western systems were forces to rely on large, costly laptop Pentium computers that use a significant amount of power to integrate voice command into other systems.

Based upon these two characteristics, IWT has been able to create new market opportunities by meeting the stringent requirements of carefully selected applications. Currently available voice recognition systems are unable to meet the demands of these applications, and IWT's technical success can create a monopoly in these specific areas. IWT has worked closely with federal government technology managers, federal laboratory engineers, university technology application specialists, industry experts and non-profit organization experts to ensure ability of its voice recognition to meet the requirements of these market opportunities.

IWT adopted this strategy of pursuing high-end, demanding requirements unmet by current voice recognition systems after determining that the poor performance of devices put on the market by other companies had created a negative image of all

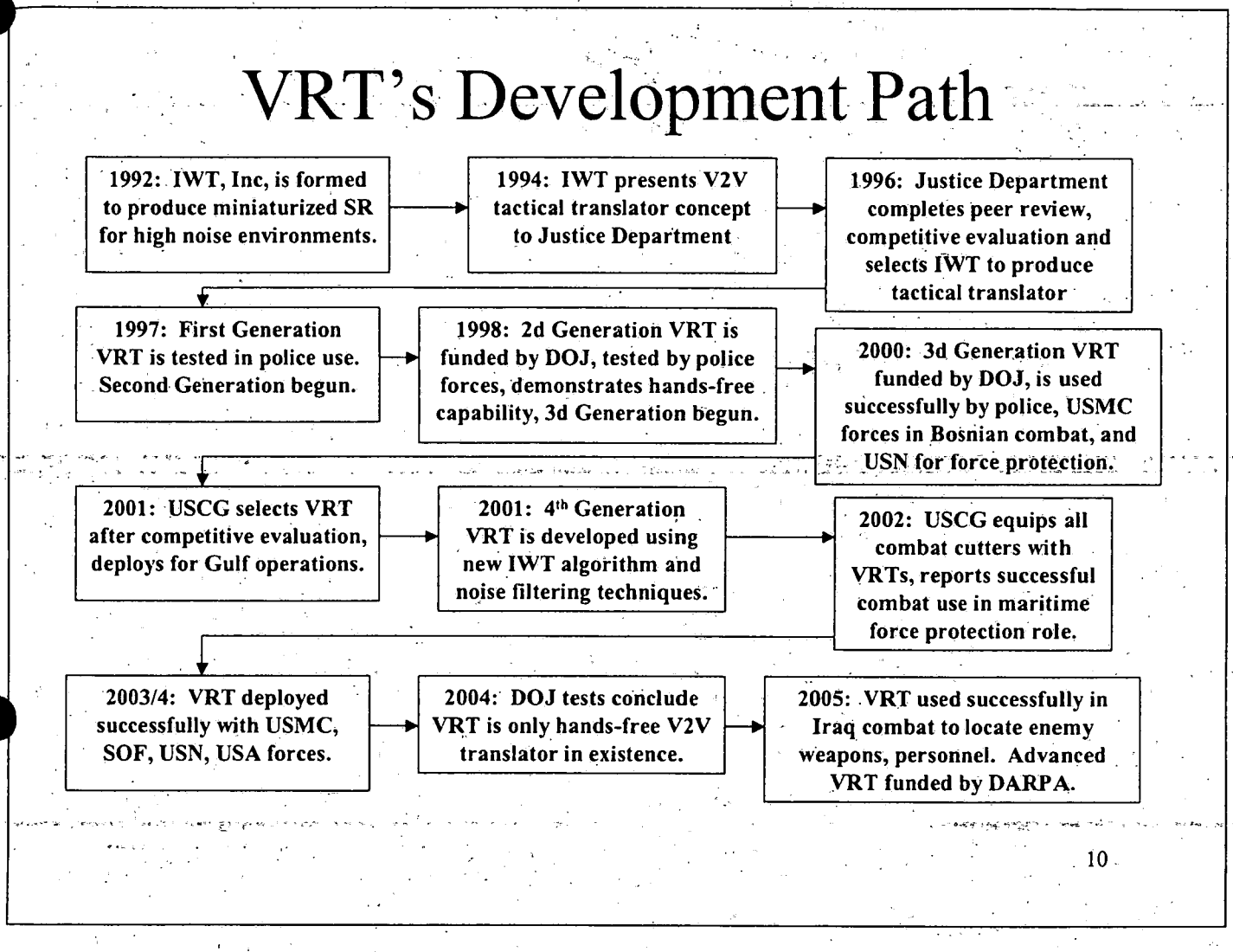
voice recognition systems in the minds of many computer users. These projects both create important products and benchmark this technology as being clearly superior to all other voice recognition. IWT believes that after it demonstrates its capabilities through demanding, high-profile applications, it will be able to sell applications for general use, either through computer manufacturers or as discrete items through software and hardware retailers.

The Company has developed increasingly cost and performance effective mechanizations of its speech recognition technology. Stand-alone sound/data processing boards implementing the technology at a size of two-inch by three-inch have a base production cost of $30 each. This technology is being transferred to a chip-on-board package that reduces size to one-inch square and cost to $15 while increasing performance. These boards and chips can be used as add-on speech recognition modules or integrated into various hand-held devices by Original Equipment Manufacturers purchasing IWT's technology and/or components.

The Company's technology is capable of recognizing impaired speech with an effectiveness that is unmatched by any other speech recognition technology. IWT is testing new prototypes with the Cerebral Palsy Association of Greater St. Louis. The national United Cerebral Palsy Association has evaluated this technology also and has promised to market devices produced through the 160 member associations and also through its Internet catalog of assistive devices. IWT is confident that once produced, this technology will become a standard for "reasonable accommodation" under the Americans with Disabilities Act and the Individuals with Disabilities Education Act and will be required for use in workplaces and schools for disabled persons with and without impaired speech.

The Company's development timeline below underscores the advancements it has made in speech recognition.

Figure 2: IWT R&D Timeline



IWT Approach to Voice Recognition

Broadly speaking, there are three approaches to speech recognition:

- The acoustic-phonetic approach.
- The artificial intelligence approach.
- The pattern recognition approach, which is used by IWT.

The acoustic-phonetic approach is straightforward. The machine attempts to decode the speech signal in a sequential manner based on the observed acoustic features of the signal and the known relations between acoustic features of the signal and the known relations between acoustic features and phonetic symbols. It is a viable approach and has been studied in great depth for more than 40 years.

In 1971, the Advanced Research Projects Agency (ARPA), now called the Defense Advanced Research Projects Agency, challenged American companies and universities to develop a speech-understanding system with a vocabulary of at least 1,000 words capable of processing connected speech with an error rate of under ten percent in a low-noise environment for use by many cooperative speakers. The systems were allowed to have an artificial syntax and a highly constrained context and were not required to operate in real time. ARPA deliberately used the word understanding, as opposed to recognition. Understanding, when used in this way, came to mean that once input was recognized, or partially recognized, it would be further processed. If a question were posed, the system would be required to answer it; if a request were made, the system would have to fulfill it.

At the end of the project in late 1976, three contractors, Carnegie Mellon University (CMU), Bolt Beranek and Newman (BBN), and System Development Corporation (SDC) - Stanford Research Institute (SRI), had produced six systems. The three most viable were the Harpy and Hearsay II systems of CMU and the HWIM ("Hear what I mean") system of BBN. Of these only Harpy fully met the five-year goals of ARPA. The ARPA project pioneered the use of linguistic knowledge. Hearsay II borrowed the "blackboard" notion from the artificial intelligence field. Blackboard is jargon for a database of information made available to the diverse processes of a software system. Hearsay II had various subparts that checked on whether a potential sound sequence was consistent with syllable structure, whether a potential syllable combination was a legitimate word, whether a potential word combination was a legitimate phrase, and so on.

Through the blackboard, information from these various levels of knowledge sources could be exchanged. Thus, if a potential word was found in Hearsay II's dictionary of allowable words, the system could back up and substitute a different, sound or syllable, forming a different word, which it could then try out. HWIM employed a syntactic analyzer called an "augmented transition network" that eliminated phonetic

choices that led to ungrammatical sentences. Harpy achieved a similar end by means of a "finite state grammar." In both systems, if the recognizer's best guess was ill-formed, say, John green its dog, the syntactic component would ask the recognizer for its next best guess and continue to do so until a grammatically acceptable sequence occurred. If no well-formed sentence could be found, the system rejected the input as unrecognizable. All large speech recognition systems developed after ARPA had ways to restrict recognition choices based on the syntactic constraints of the language.

The ARPA projects were concerned chiefly with the kinds of fundamental problems of recognition and understanding, but none worried about noise. Experiments took place in quiet environments using high-quality electronics. The quest for practical, usable systems led to an investigation of the effects of noise, which can be devastating. Systems with five percent error rates in quiet environments found themselves with 35 percent error rates when background was introduced. Channel noise plays havoc with the recognition process as does noise introduced by the speaker such as coughing, throat clearing, snuffling, snorting, sputtering, spluttering, stuttering, stammering, slurring, lisping, lip smacking, and nonlinguistic vocalizations such as hemming, hawing, uh-ing, and er-ing. These difficulties were addressed throughout the 1980s through the use of noise canceling microphones and internal noise reduction systems based for the most part on digital signal processors (DSPs).

However, for a variety of reasons, the acoustic-phonetic approach has not achieved the same success in practical systems. The central problem is the extreme difficulty in getting a reliable definitions of phonemes, i.e., segmenting the speech into discrete regions where the acoustic properties of the signal are representative of one (or possibly several) phonetic units (or classes) and then attaching one or more phonetic labels to each segmented region according to acoustic properties.

A second problem is that, once the labels have been defined, a valid word must be determined from the sequence of phonetic labels (usually in the form of a phoneme lattice) that can have many permutations for a given word or phase.[14]

The artificial intelligence (AI) approach attempts to combine the above phonetic approach with the power of an expert system that integrates phonemic, lexical, syntactic, semantic and pragmatic knowledge. Although some of the limitations of the acoustic-phoneme approach can be overcome using AI, the complexity of the task makes it unsuitable for small, portable applications, or in applications where costs must be kept low.

The pattern-recognition approach is the basis for the IWT speech recognizer. It has three qualities that lead to superior performance in applications:

1. Simplicity of use. The method is easy to understand, rich in mathematical and communication theory, and is widely used and understood.

[14] L. Rabiner and B. Juang, "Fundamentals of Speech Recognition," Prentice Hall Signal Processing Series, 1993.

2. It is robust and invariant to different speech vocabularies, users, languages, word vocabularies, talker populations, background environments, and transmission conditions.

3. Proven high performance. The pattern-recognition approach to speech recognition consistently provides high performance on any task that is within its technological parameters and provides a clear path for extending the technology in a wide range of directions.

The pattern-recognition approach is better suited for the conditions to which hand-held devices will be subjected for the following reasons:

1. The signal processing front end provides a set of unique filter bank parameters that are consistent over a wide range of speakers and communication channels.

2. The Filter Bank parameters are transformed into a set of Principal Features (PF) that is statistically determined to remove redundant data across the vocabulary.

3. The PF is transformed into frame pairs that model the statistical correlation between nearby speech frames.

4. The system employs a modified dynamic-time-warping (DTW) process in which all templates are scanned continuously. The system then relaxes end-point constraints of the input utterance and updates allowable paths of the utterance.

5. The algorithm works for speaker-dependent and speaker-independent recognition.

6. The system works in a fast and efficient manner.

Acoustic waves are converted with an 8-bit analog-digital converter (ADC) at a sample rate of 12.8 K/sec. The PCM data is placed into a circular buffer that is continuously updated. The input data is converted into a stream of parameters in the preprocessor. This secondary stream of data is converted into 8-dimension (8-D) feature vectors every 20 ms.

A word to be recognized is recognized against a "template" that is initially recorded during the "training" process. These are stored in external memory. A resident set of templates in memory defines the vocabulary.

Consider the input "utterance" as a set of feature parameters that stream in continuously. To consider this utterance as a candidate for recognition, a front-end processor is needed to "grab" the utterance.

Once the utterance is captured, it is compared against the templates in memory using a comparison technique known as a dynamic time warping (DTW) algorithm. The DTW provides the best time alignment of two utterances (unknown and template).[15]

However, instead of the common DTW algorithm, the comparison is performed continuously. This means that the input is estimated every time a feature vector comes from the preprocessor, i.e., every 20 ms.

Accurate end-point detection is crucial for accurate voice recognition. Tests have shown that small variations in end-point detection, such as +/- 40ms, can reduce accuracy by 3%.[16] The method used in this algorithm reduces these end-point errors, quickly sorts out unlikely templates, and shows promise for continuous speech recognition.

Inside the Pre-Processor

The pre-processor part of the speech recognition algorithm converts the input signal waveform into a stream of feature. there are two stages to this: the primary transformation from the time to the spectral domain; and the statistically based method to obtain a more compressed and reliable feature vector. The first is realized by means of a quasi-synchronized (with FO, the fundamental or glottal frequency) 17-band filter bank. The second is a frame-pair conversion using a Karhunen-Loeve transformation (KLT or principal feature method).[17]

Conclusion

The speech recognition algorithm used by IWT is very accurate and fast. It encompasses many of the "proven" techniques used in commercial speech recognizers, along with many novel techniques that have been added to improve system performance. It uses low cost hardware (8-bit analog-digital converter) and low computational overhead, typically well under 5% total on a 486-33 PC.

It should be noted that other methods, such as using Linear Predictive Coding for the preprocessor, have been investigated thoroughly, but shown to have lower performance due to added complexity. In addition, the use of "hidden Markov models" (HMM) has also been investigated. HMMs are widely used for large vocabulary systems and for some speaker-independent systems. However, the reliability of using

---

[15] L. Rabiner and B. Juang, "Fundamentals of Speech Recognition," Prentice Hall Signal Processing Series, 1993.

[16] J.G. Wilpon, L.R. Rabiner, and T.B. Martin, "An improved word-detection algorithm for telephone-quality speech incorporating both syntactic and semantic constraints," AT&T Tech. J., 63(3): 479-498, March 1984.

[17] E.L. Bocchieri and G.R. Doddington, "Frame-Specific statistical features for speaker independent speech recognition," IEEE Trans. on Acoustics, Speech & Signal Processing, 34(4), August 1986.

HMMs for reliable and robust command-and-control voice recognition does not perform as well as template-based approaches.

Appendix E: System Under Development

The capability being developed here would provide effective combat voice-to-voice interactions for combat operators who must maintain weapon readiness and visual situational awareness. Users would be able to issue a variety of instructions, statements and questions. Non-English-speaking interview subjects would be able to provide answers in their language in a limited-but-useful domain of words and phrases. The effort builds on IWT's success in producing combat-proven one-way tactical translators.

This project is to develop a voice-to-voice translation system enclosed in modified Modular Integrated Communications Headset (MICH)-type combat hear-through[18] headsets. The system will recognize a user's voice commands to issue any necessary number of foreign-language output phrases. The system would also receive a limited number of foreign-language responses (about 50 in the initial version) to questions in one or more foreign languages. The headsets will include speakers to play the foreign-language output phrases and microphones to provide sound pickup of foreign language speakers.

The translation system will be integrated seamlessly into the MICH and not interfere with the headset's communications or hear-through functions.

The DARPA payoff from a successful program development effort would be significantly advanced combat voice-to-voice translation capability with eyes-free, hands-free operation. Combat operators would be able to issue instructions more effectively, potentially lessening danger for both US military personnel and foreign national civilians. Combat personnel would also be able to ask questions such as those regarding wanted personnel and weapons caches in tactical situations.

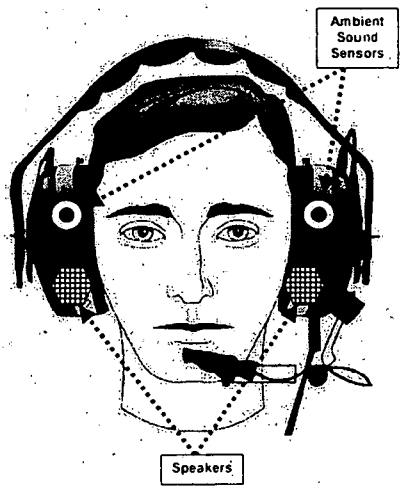
Major program elements are: 1) the design of a new circuit board miniaturized further to fit in the headset; 2) incorporation of further sound analysis features such as phased array processing; 3) further development of the recognition and application software; and 4) design and production of the headset form factor. This form factor will be based on the chassis of an existing MICH headset, with the outer body and electronics modified to include the translation capability.

Primary program risk centers on the ability to recognize foreign-language utterances in operational environments. Secondary program risk centers on miniaturization and headset integration of the recognition hardware. Neither capability has been developed previously by any other effort.

[18] The hear-through function means that headset microphones pick up sound around the user and process it. Loud sounds are reduced in volume, while quiet ones are amplified.

This program will develop 15 test articles – to be based on one or more types of MICH headset frames currently in use – during the first 270 days of this effort. Refinement of the systems based on test and evaluation will be done during the second 180 days of the program.

Headset Integrated Translator



Current Mission:

• Tactical (One-Way+Limited Response), eyes-free, hands-free voice-to-voice translation. Applications include providing instructions and receiving limited responses during: house/vehicle searches, patrol, civil aid missions, entry control duty

• Target User: Soldiers who have repetitive interactions with local population and need to maintain eye contact and weapon readiness.

System Features:

- Minimal increase in ACM/MICH headset size in final system package and ultra-low power usage.

- Only hands free & eyes free technology in existence

- Highly accurate speech recognition that works in operational environments

Background

- Base technology has received strong positive feedback from Rangers, Green Berets, Marines, USCG personnel, other users after deployment.
- Development team has worked together for 12 years, team leader is Silicon Valley pioneer with 45 years experience related to current project.

Program

• Development of Integrated Headset Hardware and Software (9 months)

• Development/Field Testing of initial One Way + Limited Response foreign language system (3 months)

• ROM: $550k