Mr. Charles Wayne
Information Awareness Office (IAO)
Human Language Technology: TIDES, EARS, Babylon

Buenos dias! Ni men hao! Sabah el kheir! Or in English, Good morning!

I look forward to telling you about three key programs in the area of Human Language Technology- d EARS are aimed at the problem of exploiting human language to find out what is going on in the world; Babylon, to communicating with people who do not speak English.

All three programs deal with multiple languages. And all support IAO's overarching goal of Total Information Awareness. I am the program manager for TIDES and EARS. LTC Jim Bass is the program manager for Babylon. I will tell you about all three programs.

HUMAN LANGUAGE TECHNOLOGY

Let me first say something about the field of Human Language Technology (HLT). As this slide shows, human language can be partitioned into four major areas—depending upon the medium (speech or text) and the communicants (whether people are talking to one another directly or communicating with or via a machine). TIDES and EARS fall under human-human communication; Babylon, under human-machine communication.

Although the problems—and the solutions—
are very different, there are natural synergies among the programs. They build off one another and the accomplishments of previous programs.

On the human-human side, to achieve Total Information Awareness, the U.S. must exploit vast volumes of naturally occurring speech and text in many languages. This includes newswires, news broadcasts, telephone conversations, et cetera. More and more of this material is becoming available electronically, but the staggering volumes make it increasingly difficult to find and interpret the relevant portions.

Exploiting human language is currently a very labor intensive process. And much of it requires foreign language skills that are in very short supply in the defense, intelligence, and law enforcement communities.

It is clear that the U.S. cannot succeed simply by adding more people. To obtain timely, actionable, mission-critical information, we absolutely must have effective language exploitation technology to magnify greatly the capabilities of a necessarily limited set of analysts. This is what TIDES and EARS are about.

In addition, when our forces are on the ground overseas, they must be able to interact effectively with local military personnel, civilians, and prisoners who do not speak English. They must do this to protect themselves, to gather vital information to defend the U.S., and to help others. This is what Babylon is about.

It is clear that human language technology able to exploit natural speech and text will be a key enabler for a HUGE number of defense and national security applications. It will revolutionize the way terrorists are detected, conflicts avoided, wars fought and won.

HLT becomes increasingly important—as volumes increase, as threats emerge from new directions, as U.S. forces are required to operate in new places, and as pressures mount to accomplish missions with fewer and fewer personnel. DARPA has responded to these challenges with three ambitious programs. Let me now tell you about those programs.

TIDES

TIDES stands for Translingual Information Detection, Extraction, and Summarization. The goal is both simple and daring: to enable English speakers to find and interpret mission-critical information quickly and

effectively regardless of language or medium. We want to make a night and day difference—to change the world from the way it is today—where huge volumes of data are available electronically but English speaking operators and analysts can exploit only a small fraction of it—to where the same people can exploit a much larger fraction of the English data plus many other languages.

As indicated here, the input to TIDES could be speech or text, stationary or streaming. Key information may span one or more documents, one or more sources, and one or more languages. If the source is speech, it must first be converted to text automatically.

To support commanders and policy makers, English-speaking operators and analysts must be able to interact with the TIDES technology and understand its output. Creating this technology is an enormous challenge, with enormous payoff. To address the challenge, TIDES is conducting research on powerful, broadly useful component technologies that could be employed alone or in combination.

The component technologies are:

> Detection—to find or discover needed information. The desired information could be specified by a user, or discovered automatically by the system. (New events fall into this category.)
>
> Extraction—to pull out key facts about entities, relations, and events. These facts could be passed on to other TIDES components or placed in a database.
>
> Summarization—to substantially reduce the amount that a person must read to ascertain the essence of what was spoken or written. A summary could describe a single document or a set of documents.
>
> Translation—to convert text, transcripts, or summaries from some other language into readable English.

I am happy to report that we have made solid progress on all of these challenges. For example, and this is a very incomplete list:

- In the Detection area, we improved the precision of English information retrieval and demonstrated the ability to find Chinese and Arabic documents using English queries. We showed that we could find news stories matching a set of sample stories, both within and across languages. And we have begun to discover the emergence of new events.

- In the Extraction area, we teamed up with the Intelligence Community's ACE (Automatic Content Extraction) program and DARPA's EELD program, demonstrating the ability to identify entities (both named and unnamed) and starting work on methods for identifying relationships among entities (who works for whom, where people are, et cetera).

- In the Summarization area, we figured out how to shorten individual documents and have made a good start on the more difficult problem of describing succinctly the content of collections of documents. (You may have seen references in the press to the Columbia NewsBlaster system.)

- Translation is the most difficult and, probably, the most important research area. As a result, we are working especially hard on it. Last month we ran our first test of Chinese-to-English and Arabic-to-English translation using a novel automated evaluation paradigm with very encouraging results.

TIDES is emphasizing work on English, Chinese, and Arabic because these languages are clearly important and distinctly different from one another. (On 9/11, we accelerated our work on Arabic.)

As we progress with these languages, we will find ways to port the technology rapidly and inexpensively to other languages, including languages with limited linguistic resources. (These are sometimes called low density languages.)

This will be feasible, because we are using new approaches that are reasonably language-independent, approaches that can learn from data and take advantage of the rapidly growing volumes of speech and text that are accessible electronically.

In addition to developing robust multilingual components, TIDES is:

- Integrating them—with one another and with other technologies—to produce synergistic, end-to-end technology demonstration systems;

- Conducting experiments on real-world problems with real data and real users; and

- Transitioning successful elements to various government customers, including the Information Awareness Center at INSCOM that Dr. Poindexter mentioned and the Evidence Extraction and Link Discovery program that Ted Senator described.

These investments have already produced successes. In terms of systems:

- Open Audio Source Information System units have been deployed overseas. The units work on accented English and Arabic. They speed up business by a factor of 8-10. OASIS was the first TIDES Deployment Unit.

- The On-Line Text and Audio Processing (OnTap) system combines leading edge detection and extraction technologies for English and Arabic. Later this year, we will add Chinese to the mix. OnTAP works on both speech and text and includes component technologies from several sites.

- The MITRE Text and Audio Processing system gathers information from a wide variety of text sources in a number of languages. You can see a nice demonstration of it at the IAO booth. It is also used to produce the TIDES World Press Update that is distributed to a number of government organizations.

TIDES runs Integrated Feasibility Experiments (IFEs) to assess the value of the evolving systems, to determine where improvements are needed, to develop effective concepts of operation, and to jumpstart the transfer of the most effective technology into operational use:

- The IFE-Bio series of experiments explored the power of the MiTAP system to find information related to bio-security and the spread of infectious disease. MiTAP provided useful information that would not otherwise have been found.

- The IFE-Arabic experiment currently underway is focusing on terrorist-related information described in Arabic news sources (both speech and text). It uses the OnTAP system.

- At the end of this year, we will start an integrated feasibility experiment known as IFE-Translingual. It will utilize components from various sites in an improved OnTAP system operating on English, Chinese, and Arabic (speech and text).

The TIDES program has another three years to go. It will have a major impact on national security.

EARS

EARS stands for Effective, Affordable, Reusable Speech-to-Text. Its goal is to produce rich, accurate transcripts of natural human-human speech that will be useful to both people and machines. Like TIDES, EARS is working hard on broadly useful component technology—in this case technology that will revolutionize the way that speech is processed by the military, intelligence, and law enforcement communities.

At present, skilled linguists struggle to listen to lots of audio recordings. Automatic filtering helps somewhat in focusing their attention, but it is of limited value because of the inaccuracy of current speech-to-text technology. More precise detection—plus extraction, summarization, and translation—are currently impossible due to the error rates for converting speech to text.

In addition, linguists must listen laboriously to the selected recordings. On unclassified test data, speech-to-text error rates now run from about 15-30% for broadcasts (depending on the source) and 25-50% for conversations.

The EARS BAA was published last fall, contracts were awarded this spring, and the program had its kickoff meeting in May. During the next five years, we expect to reduce word error rates dramatically, both for broadcast speech and for conversational telephone channel speech. We aim to get the error rates down to the 5-10% range in English, Chinese, and Arabic and to devise methods for porting the technology quickly to other languages that suddenly become operationally important.

In addition to reducing word error rates, EARS will extract useful metadata from audio signals—information about speaker identity, phrase and sentence boundaries, emphasis, verbal corrections, and so forth. The resulting output (words + metadata) will be what we call a "rich transcript." It will be packaged in a standard markup format and fed into various downstream processes (such as those being developed by TIDES) and presented to users in a neatly formatted, easily readable form.

The payoffs will be enormous:

- Machines will be able to do a far better job of detecting the material that analysts are interested in--particular stories in a broadcast, particular conversations (or portions of conversations) in streams of conversations.
- Machines will also be able to:
  - Extract useful information about entities, relationships, and events.
  - Produce compact summaries of stories and conversations.
  - Translate the transcripts and summaries into passable English.
- People—our scarcest resource—will be able to use their eyes, not their ears, to examine material that machines identify as potentially interesting. Because listening is a very slow process, reading will greatly speed up analysis and reporting. It will magnify the effectiveness of the operators and analysts and speed the delivery of time critical information.

Combined with other technologies, EARS will enable enormous productivity gains—on the order of 10 to 100 fold. This will have a huge impact on the War on Terrorism and on law enforcement and national security in general.

BABYLON

Babylon gets its name from the Tower of Babel. This program aims to make it possible for U.S. forces to communicate effectively with people who do not speak English. It is speech-to-speech machine translation for constrained domains. Babylon builds off the success of the "DARPA One-Way translation technology that has been used to good effect in Kosovo, Afghanistan, and other places.

One-way translation combines spoken phrase recognition with audio play out. An American soldier speaks one of a few thousand possible phrases in English, the machine recognizes what he has said, then plays out the pre-recorded equivalent phrase in the other language.

In Afghanistan, DARPA One-Way technology (called "Phraselator") has been used for force protection, medical services, and civil affairs. Babylon aims to go a lot further producing "One-plus-One-Way Translation" and then full "Two-Way Translation" capabilities. One-plus-One will enable an English speaker to understand what another person is saying.

It will work as follows: The English speaker asks a question or makes a statement using One-Way technology; the other speaker gives a response; the system, knowing what the English speaker has said, knows the set of things that the other person is likely to say; if he or she says one of those things, the system plays out the corresponding English phrase.

Thus, instead of forcing non-English speakers to communicate only via gestures (as in the One-Way Phraselator), One-plus-One Way technology will let them use their voices, and thereby convey more information more quickly.

To support Operation Enduring Freedom, we intend to have prototype One-plus-One devices in the field within the next six months. Full Two-Way translation is more difficult and will take longer. It will be somewhat domain constrained, but will allow much greater freedom of expression, from both the English speaker and the non-English speaker. The idea here is to use speech-to-text technology trained on domain-relevant vocabulary, parse the recognizer output, use an interlingual meaning representation, then generate the appropriate text-to-speech output. We hope to have functioning Two-Way prototypes in approximately 18-24 months. These will operate in Arabic, Chinese, Dari, Farsi, and Pashto.

Babylon is a wartime program geared for rapid prototyping and quick insertion tied to defense mission requirements. One-plus-One and Two-Way translation will provide tactical warfighters with real-time, face-to-face speech translation capabilities. As time goes on, these will accept increasingly flexible and fluid speech to facilitate normal human speech communication and, therefore, greater information awareness.

EVALUATION

Before closing, I would like to say a few words about evaluation. For more than a decade, DARPA Human Language Technology programs have sponsored rigorous, objective, metrics-based performance evaluations at regular intervals. These evaluations have been an enormous help to everyone in moving the technology forward.

The National Institute of Standards and Technology develops and runs the evaluations with advice from the research community. Many groups not funded by DARPA participate. During the next year, we will be sponsoring evaluations in rich transcription, information retrieval, topic detection and tracking, automatic content extraction, summarization, and machine translation. We would very much welcome the participation of other organizations in these evaluations. Although we cannot pay directly for the participation of outside groups, we do underwrite the costs of the evaluation infrastructure, and researchers find that participation helps them as well as the government.

To summarize, I have told you about the need for strong Human Language Technology. I have described three important programs. And I have explained the important role that evaluation plays.

TIDES, EARS, and Babylon are all quite valuable. They address the needs of many customers, including other DARPA programs. These programs are also important steps towards the "Grand Challenge" goal of teaching computers to hear, read, and understand human language in all its forms.

Thank you very much for your attention.