REPORT DOCUMENTATION PAGE

Form Approved OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE 15 Jan 1999

3. REPORT TYPE AND DATES COVERED Quarterly Progress Report December 22 - January 15, 1999

4. TITLE AND SUBTITLE
Wearable Language Translation System
Topic Number N98-144 Dated August 18, 1998

5. FUNDING NUMBERS
Contract/Purchase Order No.
N00014-99-M-0009

6. AUTHOR(S)
Wendell Carroll
Robert Palmquist

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

ViA, Inc.
11 Bridge Square
Northfield, MN 55057

8. PERFORMING ORGANIZATION REPORT NUMBER
0001AA Progress Report

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Program Officer
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5660
Attn: Joel Davis, Code 342CN

DCMC Twin Cities
3001 Metro Drive
Bloomington, MN 55425-1573

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
0001AA Progress Report

19990120 102

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)* ViA Team Mission Statement: To develop a near real-time, two-way, mobile, lightweight, robust and low-cost multi-lingual language translation device that can be operated with minimal training in a hands-free manner.

This report describes the current status of ViA's Phase I SBIR to develop a mobile language translator. The objectives of this Phase I research effort are as follows: Investigate the scientific, technical and commercial merit and feasibility of the system described in the preceding mission statement; use the results of this investigation to select the best approach for the language translator system; Develop a working prototype that will be delivered to the government representative (Dr. Joel Davis).

There are three technical areas that are being investigated: the mobile computer platform; the operator interface; and the language translation software. The commercial feasibility study includes identifying potential applications, languages to be supported, cost, and user requirements such as performance specifications, system weight and acceptable battery life. By combining both the commercial and technical elements, the ViA team is developing a complete definition and functional prototype of a mobile, near real-time language translation device.

14. SUBJECT TERMS
Wearable Computers;
Language Translation; Voice Recognition; Hands Free Interface

15. NUMBER OF PAGES 33

16. PRICE CODE

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

Table of Contents

A. Language Translator Overview

ViA Team Mission Statement
To develop a near real-time, two-way, mobile, lightweight, robust and low-cost multi-lingual language translation device that can be operated with minimal training in a hands-free manner.

The objectives of this Phase I research effort are as follows:

- Investigate the scientific, technical and commercial merit and feasibility of the system described in the preceding mission statement.
- Use the results of this investigation to select the best approach for the language translator system.
- Develop a working prototype that will be delivered to the government representative (Dr. Joel Davis).

There are three technical areas that are being investigated: the mobile computer platform; the operator interface; and the language translation software. The commercial feasibility study includes identifying potential applications, languages to be supported, cost, and user requirements such as performance specifications, system weight and acceptable battery life. By combining both the commercial and technical elements, the ViA team is developing a complete definition and functional prototype of a mobile, near real-time language translation device.

B. Language Translator Status

As per the contract, this Phase I project has a duration of six-months, with an option for an additional three-month effort. However, ViA is optimistic that the project will be completed in a shorter time frame. A summary of the Project Plan is given in Figure 1. Details of each activity and the current status are described in the document section referenced in this figure and in the following text.

B.1. Work Plan Overview

As per the Project Schedule, the research efforts commenced with a Design Requirements Meeting between ViA and the government's contract representative, Dr. Joel Davis. Representing ViA in this meeting was Bob Keene, and by telephone, Robert Palmquist. In this meeting, the design approach (which is outlined in the following paragraphs) was reviewed and performance expectations discussed. In particular, ViA's approach was compared to alternative approaches that have been developed. During a later conversation, the language pair to be used for the prototype system was selected to be English and German. These decisions were combined into the *Design Requirements* document, which has been submitted for approval.

Research is continuing on the three technical aspects of the mobile language translator: the mobile computer platform; the operator interface; and the language translation software. The results of a survey of available technologies for each of these three areas is included in this report (Sections B.2, B.3, B.4). In a parallel process, ViA is interviewing potential users of the language translator to determine their needs and desires for such a system. Some of the preliminary results of this survey are included in this document (Section B.6). A complete write-up will be included in the next report. By combining both the commercial and technical elements, a complete definition of a successful mobile, near real-time language translation device is being achieved. ViA is using all of these results and their extensive in-house knowledge of mobile PC systems to develop a complete Prototype System Design that meets or exceeds the specifications outlined in the *Design Requirement* document. This document will be submitted for approval prior to February 15, 1999. Once the design is approved, a prototype system will be fabricated, demonstrated and delivered to the government representative (Dr. Joel Davis). This delivery is scheduled to occur on April 15, 1999. In addition to this working prototype, a Phase I Final Report will be delivered detailing the results of the project and the recommendations for Phase II activities.

Two options are included in Phase I: one for the integration of an additional language pair and the other for adding application specific vocabulary to the dictionaries. These options, which are described in detail in section E.7 of the proposal, are not being pursued at the current time pending notification of the government representative.

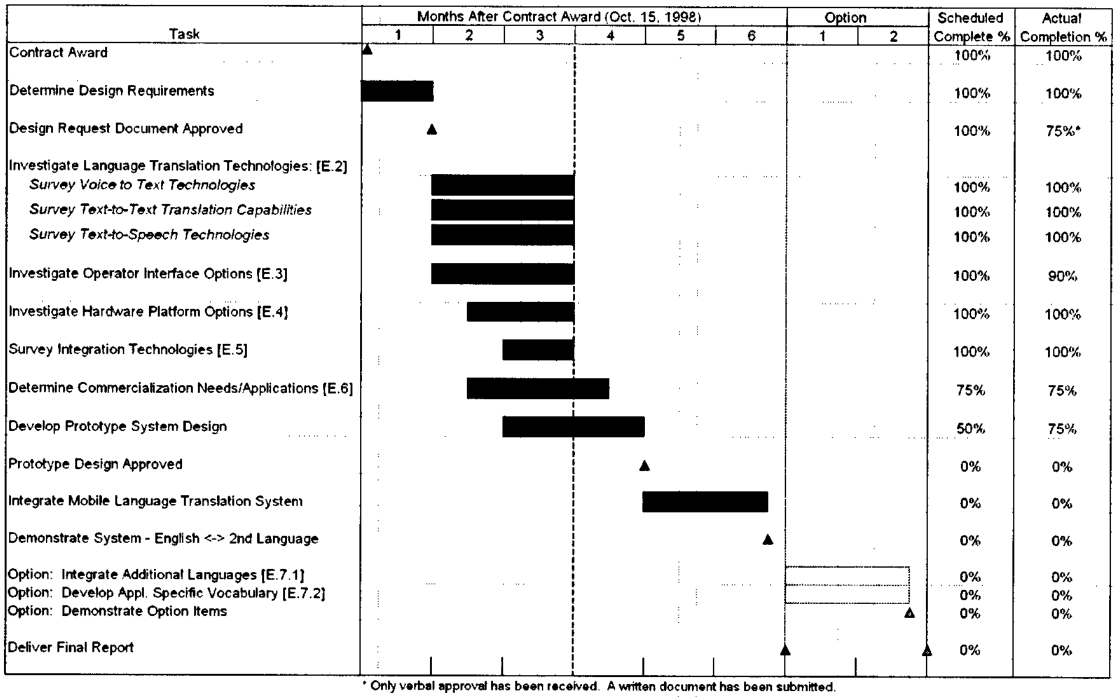


Figure 1 – Project Work Plan and Current Status

B.2 Language Translation Technologies

As per the previously stated mission statement, one of the goals of this project is to develop a near real-time mobile translation capability. The metric for this element is to have the computing system start speaking the translated words within two seconds of the initial speaker completing the sentence. In order to achieve this goal, a direct voice-to-voice capability must be developed. This ambitious goal will be a focus of the Phase II effort. In Phase I, insight is being gained towards this "voice-to-voice within two seconds" goal by implementing an easier, albeit much slower, alternative approach. In Phase I, a three stage translation process is being implemented using voice-to-text software, then translating the text into text of the foreign language and finally having the computer speak the resulting foreign text. Each of these three steps is described in more detail in the following paragraphs. It is anticipated that this process will take approximately ten seconds and that it may involve operator assistance to complete the translation process. By implementing this shorter-term solution, a solid foundation will be gained towards the ultimate goal of having a direct voice-to-voice system.

B.2.1 Voice-to-Text:

In the past two years there has been a significant improvement in the performance of voice engines. This is a result of technological advancements along many fronts such as the voice recognition algorithms, the processing power of PC platforms and anti-noise canceling microphones. Each of these three areas are being investigated to determine the best voice system to use for the language translator. For the language translator, ViA will be coupling their own voice engine enhancement software (called *SonicBoom)* with commercially available voice recognition engines. These software packages are described in the following two sections.

B.2.1.1 SonicBoom Software:

ViA has developed and deployed several wearable computer applications that use voice-based interfaces. To assist in the development and support of these systems, ViA has developed their own voice engine enhancement software package, which is called *Sonic Boom.* This package improves the performance of any Speech Application Programming Interface (SAPI) compliant voice engine by providing the following capabilities:

- *Concurrent Multiple Dictionary Referencing*: In Sonic Boom, each context has a set of associated dictionaries. Therefore, when a given context is enabled, all of the necessary dictionaries are loaded, enabled and complied. By using this approach, all of the required dictionaries are pre-processed, thus improving the overall speed of the process. This multiple dictionary capability is required for the direct voice-to-voice language translation system that will be developed in Phase II.
- *Automatic Gain Control*: SonicBoom's volume control routine provides instantaneous compensation for ambient background noise. This allows the language translator to be used in noisy environments (e.g., outdoors, in airports, etc.).
- *Echo Canceling*: Automatic switching between full- and half-duplexing modes of operation provides an improved echo canceling capability over other commercially available products. This further enhances the robustness of the speech recognition software.
- *Remote Program Support Tools*: SonicBoom's web-based format allows remote loading of data and new program files. Thus, if additional words need to be added to a dictionary (e.g., words specific to a particular application), this can be accomplished wirelessly in a mode that is transparent to the operator.

B.2.1.2 Evaluation of Commercially Available Voice Engines:

As part of the Phase I activities, five commercially available SAPI engines were evaluated; Lernout & Hauspie's (L&H) *VoiceXpress*; Conversa's *Lingo*; IBM's *ViaVoice*; Dragon's *Naturally Speaking* and Microsoft's *Whisper*. Each of these systems was tested during the month of November and rated as to their suitability for the language translator system. The performance parameters included robustness in noisy environments, speed, accuracy, product cost, hardware requirements and, of special importance, the product's ability to support foreign languages.

Table 1 – Comparison of Voice Engines

Dictation Software			
Product Name	Dragon Naturally Speaking 3.0	Lernout and Hauspie Voice XPress 2.02	IBM ViaVoice
Active Words	35,000/64,000	30,000/ 50,000	32,000/64,000[3]
Training time*	35.5 Minutes	47.3 Minutes (Low tolerance)	
Performance	133Mhz (200 Recommended)	166Mhz (200 Recommended)	166Mhz (200 Recommended)
Memory	32MB (64MB Recommended)	32MB (48 MB Recommended)	32MB (64MB Recommended)
Accuracy (out of the box)[1,5]	8.4	10.5	7.1
Accuracy (trained)[2,5]	3.25	3.45	3.75
Speed of input[4]	15.3 Seconds	13.5 Seconds	14.6
Number of Languages Supported	6 = English (British and American), French, Dutch, Italian, Spanish and Swedish	4 now, 9 by end of 1999. Available now = English, German, Mandarin Chinese and Cantonese Chinese; Available later this year = Dutch, French, Spanish, Protuguese and Japanese	1 = English

[1] – Installation of software, but the training was skipped.
[2] – Suggested training completed. Nothing additional was trained. All words were verified in the engines active vocabulary before dictation began.
[3] – ViaVoice's statistics are unpublished. The dictionary size can be confirmed, but the active vocabulary is a guess formulated by industry experts.
[4] – Determined by voicing in the sentence, "Voice dictation has progressed to a level where it is feasible to dictate flawlessly into Microsoft Word, or any other productivity application." – Ignoring any errors. Tested on a Pentium 200 running 64MB of memory.

[5]– Accuracy is reported as mistakes per paragraph. The sample text used was "Startup" by Jerry Kaplan (Penguin books), P. 170-172

B.2.1.3 Selected Approach for Voice-to-Text:

Test results show that IBM's *ViaVoice* is not a viable product for this project. The recognition rate is slow, the accuracy (even when trained) is poor compared to the L&H and Dragon products and the programming interface requires development of an intermediate application. L&H *VoiceXPress* and Dragon *Dictate* are both suitable products for the language translator, however, L&H was determined to be the superior solution. There are three reasons for this selection: better run-time performance; number of supported foreign languages; and commitment to support language translation issues. As for run-time performance, L&H software ran very well on the recommended 200Mhz computer whereas with Dragon, 200Mhz was the minimum speed to produce decent results. L&H currently supports four languages, with plans to add five additional languages by the end of 1999. This compares with Dragon's support of six different languages. The release dates for *VoiceXpress* languages are provided in Table 2. Finally, Lernout and Hauspie's support of langauge translation capabilities is better than Dragon's. For example, as a company L&H's Mendez Division has over 500 employees supporting 20 different languages. The knowledge and results of these individuals (e.g., over 100,000 documents have been translated), are being coupled into development of Machine Translation algorithms. Also, L&H's Software Development Kits (SDKs) for their voice engines, have a superior interface to translators than Dragon's (see Section B.5 for further information regarding SDKs).

Table 2: L&H VoiceXpress Languages

Language	Availability
English	available today
German	available today
Mandarin Chinese	available today
Cantonese Chinese	available today
Dutch	2Q '99
French	2Q '99
Spanish	3Q '99
Portuguese	4Q '99
Japanese	4Q '99

As a result of the decision to use *VoiceXpress*, ViA personnel traveled to L&H Headquarters during the week of December 14^{th} (the expenses for this trip were not charged to the SBIR project) to discuss teaming arrangements and forthcoming product from L&H. During this visit, ViA obtained L&H beta software of their forthcoming *VoiceXpress Software Developers Kit (SDK).* This SDK includes an API that allows *VoiceXpress* to use native Windows programming calls instead of using a pass-through such as Microsoft Word. This will ultimately increase the speed of the translation, thus increasing the performance of the language translator.

B.2.2 Text-to-Text Translation:

One of the challenges of Text-to-text translations is interpreting the context of the phase. In order to understand this context, the translator must be familiar not only with both languages, but also the culture and idioms of the each language's country and the vocabulary that is specific to the topic area being discussed. Simply translating on a word-for-word basis often results in a translated sentence that incorrectly states the original meaning. Here are two examples that are commonly referenced:

> *English:* "I am full." (as in, after a good meal)
> *French literal translation:* "Je suis plein."
> *Meaning of literal translation:* "I am pregnant."
>
> *English:* "I am a Berliner."
> *German translation without cultural context*[1]*:* "Ich bin ein Berliner."
> *Meaning of translation:* "I am a jelly donut."

Developing software that understands context subtleties is an extremely difficult task. However, there are numerous commercially available software packages that are coming close to making this capability a reality. These software packages can be classified into three areas: Terminology Managers; Machine Translation packages; and Translation Memory software[2]. Each category has inherent strengths, shortcomings and price points that make it necessary to do a careful assessment of which technology, or which combination of technologies, is the best solution for the mobile translator. Each of these approaches, plus the opportunity to combine them to form a hybrid system, is discussed in the following paragraphs.

B.2.2.1 Terminology Managers:

[1] This of course is the phrase spoken by President Kennedy during the Berlin Crisis. The correct phrase that should have been stated is simply "Ich bin Berliner."

[2] Several references repeat this breakdown of translation software technologies. One such source is Language Partners International of Evanston, Illinois.

One of the difficulties in translation is appropriate handling of industry-specific terminology. For example, the military, legal and medical domains, each have significant amounts of terminology that are specific to their applications. Translating these terms to a different target language is often a tedious task of researching the word to determine its meaning. Terminology managers assist with this translation process by providing four elements: terminology repository; rapid term lookup; automated terminology insertion and terminology extraction.

- Terminology repository: Terminology managers serve as a collection point for gathering and storing domain specific words and their translations.
- Rapid term lookup: Basic terminology managers translate domain specific words in a unidirectional, one-to-one correspondence. More sophisticated term managers store objects in a "concept" orientation with multilingual mapping in multiple directions. Some allow narrative term definition/description and even the storage of graphics to represent the concept. Searching mechanisms can range from matching on simple word look-up to more advanced approaches that employ "fuzzy" searching techniques looking for matches at a conceptual level.
- Automated terminology insertion: Some terminology managers will insert the translated term into the target document without the need to re-type or cut-and-paste.
- Terminology extraction: Tools with this feature will linguistically analyze source and target documents of previous translations to more easily identify and extract terminology for import into the terminology manager.

Current commercial Terminology Manager products include L&H's VoiceXpress for Medicine, VoiceXpress for Clinical Reporting, VoiceXpress for Legal and VoiceXpress for Safety, MTX's Termex, Trados's MultiTerm and TTT.

B.2.2.2 Machine Translation Software:

Machine Translation (MT) tools linguistically process source documents to create a translation "from scratch." Up until several years ago, these tools required large mainframe computer platforms for timely execution. However, with recent advances in PC and UNIX based systems, many of these high-end solutions are available in affordable versions with quality and accuracy that compares favorably with their mainframe parents.

Because the linguistic rules for parsing and analyzing source text vary by language, the number of languages supported by MT systems is more limited than other approaches. Additionally, there is a need for a sufficiently large core dictionary for the target language to obtain a minimum level of accuracy/quality. MT solutions are best applied in the following areas:

- "Gisting," where the user would like to understand the general meaning of the text.

- Screening large amounts of documentation in order to identify documents that warrant more accurate human translation.
- Conveying simple instructions or non-complex information.

There are numerous groups, both from industry and academia, performing research and development activities on MT. For example, the University of Maryland's *Computational Linguistics and Information Processing Laboratory* (CLIP) is developing MT systems targeted towards syntactic realizations and underlying semantics words across different languages. In particular, they have developed extensive capabilities in ChineseEnglish language pairs. This work will improve the robustness of MT systems across multiple language domains. Another effort of note is New Mexico State's *Artwork* Program. *Artwork* is investigating the machine translation of spoken dialogue. The focus is developing approaches to providing robustness by exploiting models of the task domain and of conversational interaction to generate relevant expectations against which the input can be interpreted. This effort may provide a solution for a direct speech-to-speech system in the not too distant future. Representative commercial products in this category include Langenscheidt's T1, Globalink GTS Power Translator, Intergraph Transcend, LOGOS Intelligent Translation System, PC Translator and SYSTRAN PROfessional for Windows.

B.2.2.3 Translation Memory:

Translation Memory (TM) tools are based on the automated re-use of previously translated terms and sentences. These tools assist, rather than replace, the translator. For example, when using a TM-based tool, typically 20-50% or more of a document will require manual translation. With TM tools, the level of benefit is directly proportional to the amount of repetition in the document. Therefore long, technical manuals tend to be good candidates for TM whereas the use of TM for a mobile language translator is very limited. Thus, TM tools will not be used for the language translator. They will be included in ViA's survey for the sake of completeness, but will not be tested to the same extent as the other software packages. TM tools are especially helpful in translating updated versions of previously translated documents. Other benefits include:

- Better translation consistency across an entire document, especially valuable when multiple translators are involved.
- Ability to begin translation projects before source documents have been frozen.

TM-based systems are less sensitive to language directions than the other approaches and thus a wide range of languages are supported.

The development of efficient TM systems is being conducted both in industry and academia. One such effort is the *Deductive and Object-Oriented Databases* being developed by University of Toronto's *Computer Science and the Computer Systems Research Institute*. Representative commercial products in this category include EUROLANG Optimizer, Trados Workbench and IBM Translation Manager.

B.2.3.4 Hybrid Systems:

Many vendors are coupling aspects of these three approaches into a single package, such as Langenscheidt's T1 Professional and Transcend Natural Language Translator. Additional new approaches to language translation are being developed using artificial intelligence. For example, L&H is developing neural networks that will perform post-processing of Machine Translations. This capability, if successful, will make significant strides in completely automating the translation process. The neural nets are constructed by comparing the final version of a document that is manually translated by L&H's Mendez division, with that of the same document processed by the Machine Translator. By forming this comparison, translation errors are detected and algorithms developed (i.e., a neural net) to automatically perform the post-editing process. Another effort of note is Pangloss, which is being jointly developed by Carnegie-Mellon University, New Mexico State University and the University of Southern California. This system combines three different translation engines to formulate a "best-output" translation. The goal of this effort is to develop software for direct speech-to-speech translation.

B.2.3.5 Evaluation of Commercially Available Text-to-Text Translators

In order to evaluate available Text-to-Text translation software, ViA created a matrix of phrases that tested all aspects of language translation. To determine which engine could produce the more accurate translation, ViA used a third-party consultant and generated a list of thirty-four sentences that tested different verb-forms, tenses, idioms, conditionals, passive vs. active voice, reflexive verbs, dative clauses, accusative clauses, and negative response.

The primary considerations for selection of the software package were language options, programmability, and speed. Many translator packages are not suitable for the mobile system. For example, most translators require a server class machine, responsible for translating for up to 50 clients. This requirement makes these engines impractical. Thus, in order to maximize test times, ViA performed preliminary tests and evaluations to determine the two best engines, then put these packages in a head-to-head competition with our grammar matrix. Globalink's *Power Translator Professional* (Version 6.4) and Systran's *Systran Personal for Windows* were selected as the two finalists. To evaluate their performance, the selected phrases were evaluated bi-directionally, and in some cases using multiple phrases to say the same sentence (different grammatical structures). After the translations were complete, they were examined and scored. In the attached chart, red indicates a severe shift in translated meaning, whereas a purple indicates a minor a word transposition, or omission that reduces readibility. Yellow indicates a minor word choice error. Finally, blue indicates a failure to properly conjugate a tense.

The anticipated translations (the ideal meaning of the text) are listed in the left column, The test results are provided in the following two columns.

Table 3: Comparison of Language Translation Software

TRANSLATIONS FROM GERMAN TO ENGLISH	GLOBALINK	SYSTRAN TRANSLATION
Present Tense		
I would like to buy a ticket	I would like to buy a ticket	I would like to buy a ticket
This bread is fresh.	The bread is fresh.	[redacted] is fresh.
Do you have fresh bread?	Do you have fresh bread?	Do you have fresh bread?
Present Time Subjunctive		
You would like to see this movie. If only I had more time! Could you please help me? He said he was coming tomorrow.	You/they would like to see this film. If I had only more time! Could you please help me? He/it said, he/it would come tomorrow.	They would see this film gladly. If I [redacted] more time! Could you help me please? He said, he would come tomorrow.
Past Time Subjunctive		
I would have helped you. Would you have come along? We would have visited you. She would like to have traveled to Germany.	I would have helped you. Would you have come along? We would have visited you. She/it would have liked **to travel** to Germany.	I would have helped you. Would you have come along? We would have visited you. It would have traveled gladly to Germany.
Present Subjunctive		
The politician said he wasn't satisfied with the law, but he had voted for it anyway. Long live the king!	The politician said, he/it is not content with the law, but he/it has voted for it however. Long, the king lives!	The politician said, it was not content with the law, but it was [redacted] for it. Long live the king!
If/Then conditionals using wurde		
Would you please explain to me, what that means? If I were you, I would buy the red dress.	Would you please explai[redacted] me, what the means? If I was you, I would buy the red dress.	Would you [redacted] to me [redacted], what that means? If I were you, I would buy the red dress.
Impersonal Passive		
No parking here. No smoking in the corridor. There was lots of dancing and singing.	One doesn't park here. One doesn't smoke in the walk. Much was danced and was sung.	[redacted]
Reflexive Conjunctions		
This matter will be cleared up soon. The gates were being opened. That will be hard to understand	This matter is explained soon. This matter states soon. The [redacted] were opened. The [redacted] opened. That will be understood [redacted]. That will understand itself [redacted].	This thing is soon explained. This thing explains itself soon. [redacted] were opened. [redacted] opened. That will be [redacted] understood. That will understand itself with difficulty.

Idioms		
That can be done His car could not be repaired. Everything can be easily explained	That can become done. That can be done. His/its car could not be repaired. His/its car could not be repaired. Everything can be explained easily. Everything can be explained easily.	That can be made. That can be made. Its auto could not be repaired. Its auto could not be repaired. Everything can be easily explained. Everything can be explained easily.
Accusative		
I see the man. We go through the house. She goes into a store. He does that every evening. We finally are rid of him	I see the man. We go through the house. She/it goes into a store. He/it does that each evening. We are finally free him/it!	I see the man. We go through the house. It goes into a [redacted]. It [redacted] evening. We are finally the [redacted] it!
Dativ		
His pencil is lying on the table. We are going into the store. You are standing behind them. My mother lives next to us.	His/its pencil lies on the table. We go in the store. You **stand** behind them. My mother lives beside us.	Its pencil is situated on the desk. We go into the [redacted]. You [redacted]re behind them. My [redacted]/mother lives beside us.
Negative Responses		
No, I am not buying a battery. Does the store have aspirin? No, it has no aspirin.	No, I **buy** no battery. Does the business have aspirin? No, it has no aspirin.	No, I do not **buy** a battery. Does the business have aspirin? No, it does not have aspirin.

	Key
	Word Transposition or missing word
	Minor word choice error
	Major recognition loss of improper translation
	Improper grammar structure (singular/plurals)

TRANSLATIONS FROM GERMAN TO ENGLISH	GLOBALINK	SYSTRAN TRANSLATION
Present Tense		
Ich möchte eine Fahrkarte kaufen	Ich würde gern eine Karte kaufen.	Ich möchte eine Karte kaufen
Das brot ist frisch.	Dieses Brot ist frisch.	Dieses Brot ist frisch.
Haben Sie frisches Brot?	Haben Sie frisches Brot?	Haben Sie frisches Brot?
Present Time Subjunctive		
Sie sähen diesen Film gern. Wenn ich nur mehr Zeit hätte! Könnten Sie mir bitte helfen? Er sagte, er käme morgen.	Sie sähen [redacted] diesen [redacted]. Wenn [redacted] mehr Zeit hätte! Könnten Sie mir bitte helfen? Er sagte, er **kam** morgen.	Sie möchten diesen Film sehen. Wenn [redacted] mehr Zeit hatte! Konnten Sie mir bitte helfen? Er sagte, daß er morgen kam.
Past Time Subjunctive		
I Ich hätte Ihnen geholfen. Wären Sie mitgekommen? Wir hätten euch besucht. Sie wäre gern nach Deutschland gereist.	Ich hätte Ihnen geholfen. Wären Sie mitgekommen? Wir hätten Sie besucht. Sie hätte gern [redacted] nach Deutschland.	Ich **würde Ihnen** geholfen haben. Würden Sie entlang gekommen sein? Wir **würden** Sie besucht haben. Sie möchte nach Deutschland

		gereist sein.
Present Subjunctive		
Der Politiker sagte, er sei nicht mit dem Gesetz zufrieden, aber er habe doch dafür gestimmt. Lang lebe der König!	Der Politiker sagte, daß er nicht mit dem Gesetz zufriedengestellt wurde, aber er hatte jedenfalls dafür abgestimmt. Lang leben Sie den König	Der Politiker sagte, daß er nicht mit dem Gesetz zufrieden, aber er für es irgendwie gewählt hatte. Leben lang der König!
If/Then conditionals using wurde		
Würdest du mir bitte erklären, was das bedeutet? Wenn ich Sie wäre, würde ich das rote Kleid kaufen.	Würden Sie bitte zu mir erklären, das was diese Mittel? Wenn ich Sie wäre, würde ich das rote Kleid kaufen.	Würden Sie bitte mir erklären, was dieses Mittel? Wenn ich Sie war, würde ich das rote Kleid kaufen.
Impersonal Passive		
Hier wird nicht geparkt. Im Gang wird nicht geraucht. Es wurde viel getanzt und gesungen.	Kein Parken hier. Kein Rauchen im Korridor. Es gab viel Tanzen und das Singen.	Kein Parken hier. Kein Rauchen im Flur. Es gab Lots des Tanzens und des Singens.
Reflexive Conjunctions		
Diese Sache wird bald erklärt. Diese Sache erklärt sich bald. Die Toren wurden geöffnet. Die Toren öffneten sich. Das wird schwer verstanden werden. Das wird sich schwer verstehen.	Diese Sache wird bald geklärt werden. Die Tore wurden geöffnet. Das wird schwer zu verstehen sein.	Dieser Stoff wird bald aufgeräumt. Die Gatter wurden geöffnet. Das ist hart zu verstehen
Idioms		
Das kann gemacht werden. Das läßt sich machen. Sein Auto konnte nicht repariert werden. Sein Auto ließ sich nicht reparieren. Alles kann leicht erklärt werden. Alles läßt sich leicht erklären.	Das kann gemacht werden Sein Auto könnte nicht repariert werden. Alles kann leicht erklärt werden	Dem kann getan werden sein [redacted]uto könnte nicht repariert werden. Alles kann leicht erklärt werden
Accusative		
Ich sehe den Mann. Wir gehen durch das Haus. Sie geht in einen Laden. Er macht das jeden Abend. Wir sind ihn endlich los!	Ich sehe den Mann. Wir gehen durch das Haus. Sie geht in einen Laden. Er macht, daß jeder Abend. Wir sind schließlich, befreien Sie von ihm	Ich sehe den Mann. Wir laufen das Haus durch. Sie steigt in einen Speicher ein. Er tut daß jeder Abend. Wir schließlich werden von
Dativ		
Sein Bleistift liegt auf dem Tisch. Wir gehen in den Laden. Du stehst hinter ihnen. Meine Mutter wohnt neben uns.	Sein Bleistift liegt auf dem Tisch. Wir gehen in den Laden. Sie stehen hinter ihnen. Meine Mutter lebt neben uns.	Ihm gereinigt Sein Bleistift liegt auf der Tabelle. Wir steigen in den Speicher ein. Sie stehen hinter ihnen. Meine Mutter lebt nahe bei uns.
Negative Responses		
Nein, ich kaufe keine Batterie.	Nein, ich kaufe keine Batterie.	Nein, kaufe ich nicht eine

Hat das Geschäft Aspirin? Nein, es hat kein Aspirin.	Hat der Laden Aspirin? Nein, es hat kein Aspirin.	Batterie. Hat der Speicher Aspirin? Nein, hat er kein Aspirin.

Key	
	Word Transposition or missing word
	Minor word choice error
	Major recognition loss of improper translation
	Improper grammar structure (singular/plurals)

B.2.3.6 Selected Approach for Text-to-Text Translation:

Globalink's Power Translator Professional (Version 6.4) is the best selection for the language translator. The product was one of only a handful that went beyond simply using a word-for-word dictionary lookup, but also included the context, grammar and construction of the phrases to form the translation. This software's performance was a little slower than competitive products, but translation accuracy more than justifies the performance cost. Globalink also includes an API that allows the translator to use native Windows programming calls instead of using a pass-through such as Microsoft Word. Once again, this increases the speed of the translation thus increasing the performance of the language translator.

Systran Professional was a close second to Globalink. Advantages to Systran include support for eight different language pairs (compared to Globalink's five), speed of translation and industry specific languages. The significant disadvantages are Systran's cost ($3,350 versus $149) and lack of integration support with voice engine software. Currently, the Globalink software will support translations bi-directionally in the following languages:

- English
- French
- German
- Spanish
- Italian
- Portuguese

Globalink plans to add additional languages in 2Q '99, although they have not yet released a list of which languages will be included.

B.2.4 Text-to-Speech:

B.2.4.1 Overview:

Text-to-speech, also referred to as speech synthesis, is the technology the computer uses to produce the sounds an individual would make if he/she were reading the text aloud. Of all the technologies required for the mobile language translation system, speech synthesis requires the least computing power. There are two basic approaches that are used in speech synthesis: pulling voice wavefiles from a database and processing text-based command strings. For the former, large wavefile databases are assembled with an entry for each word. If different pronunciations of the word are desired (e.g., a male and a female voice), then multiple entries for each word are required. Examples of this type of speech synthesis approach include IBM's *ViaVoice Outloud* software and Talx's *TalxWare*. The alternate approach, called "formant synthesis," uses a mathematical model of the human vocal tract to reproduce the correct sounds. The technology is based on parameterized segment concatenation algorithms, where human voice samples such as diphones, triphones and tetraphones are stored and used to convert the text into speech. In-depth linguistic processing is used to intelligently convert spoken text to its correct pronunciation, combined with advanced prosody rules that provide natural sounding intonation. An example of this technology is L&H's *TruVoice TTS3000/M* software. This saves disk space, at the expense of increasing the computational requirements. Both of these approaches were investigated to determine which one is best suited for use in the mobile translation system.

B.2.4.2 Evaulation of Text-to-Speech Software

Formulating acceptable test criterion proved to be most difficult with the text-to-speech software. Accuracy, speed and flexibility are all-important parameters in selecting the best package. In ViA's design of the system, each dictation engine keeps an accurate profile of the user's age-bracket and gender, which would ideally reflect the sound of the synthesized voice. Thus, the text-to-speech engine should at a minimum support both male and female sounding synthesis. This allows some personalization when using the system. After extensive testing, ViA identified five engines as being suitable for our project. Each of these engines had similar performance levels.

The determining choice factor was a combination of performance and foreign language support. The best two packages are L&H and Eloquent. Each offers numerous languages and has robust interfaces. Unfortunately, Eloquent does not support a SAPI interface. This means if there was a need to switch away from Eloquent, extensive programming would be required. With a SAPI engine, or even one that partially supported SAPI, the language translator would be free to choose any SAPI compliant voice engine. For that reason, ViA has chosen the Lernout and Hauspie TTS3000 to be the voice synthesis engine for this project, and recognizes that a trade has been made for flexibility versus immediate performance. The long-term goal will be to integrate Lernout and Haupsie's RealSpeak product (as availability will dictate).

ViA invites you to test the engines performances for yourself. Each of the following engines (with the exception of the L&H engine) has a web-site that allows anyone to sample the synthesis. The web-addresses can be found on the comparison table below.

Table 4: Comparison of Text-To-Speech Software

TEXT-TO-SPEECH ENGINES	Lucent Technologies Bell Labs Innovations	L&H	Eloquent Technology	AcuVoice	BaBel Technologies SPEECH & IMAGE SOLUTIONS
Company	Lucent/Bell Labs	Lernout and Hauspie	Eloquent Technology	AcuVoice	Babel Technologies
Web Address	http://www.bell-labs.com/project/tts/voices.html	http://www.lhs.com/speechtech/pcmmdevtools/tts.asp	http://www.eloq.com/	http://www.acuvoice.com/samples.html	http://www.babeltech.com/
Sound Quality	16-bit Stereo	16-bit Stereo	16-bit Stereo	16-bit Stereo	16-bit Stereo
Supported Languages	English German Mandarin French Italian Spanish	English UK English German French Italian Spanish Portugese Dutch Mexican Spanish	English UK English German French Italian Spanish Mexican Spanish	English	English German
API Available	Yes	Yes	Yes	Yes	Yes
Quality of Speech	6	7	8	3	7
SAPI Compliant	*	Yes	No	*	*

* - SAPI compliance is not wholly supported, and therefore the programming interface may not be suited for integration into the language translator.

B.2.4.3 Selected Approach for Text-to-Speech

Five different Text-to-Speech packages were evaluated with respect to overall performance, cost, languages supported and their ability to be integrated seamlessly into the language translator software. Based on these tests, the L&H TTS3000 engine has been selected. This package includes an exceptional API that improves the speed of the system and the ability to expand to more languages than just the original selected pair (German-English). Currently, the TTS300 text-to-speech engine supports eight languages:

- US English
- UK English
- Dutch

- German
- French
- Spanish (Mexican and Native)
- Arabic
- Italian

L&H plans to add six languages to this list by 2Q '99.

There is a high probability that this package will be replaced by L&H's RealSpeak software in December, 1999. RealSpeak will place additional demands on the computational hardware (especially memory), but the natural sound of the language is far superior to any other current package.

B.3 Operator Interface Options

There are several Operator Interface options that are being investigated for use with the mobile translator system. The goal is to make these interfaces unobtrusive (e.g., lightweight, comfortable, easy to access, etc.). In early prototypes, it is expected that some translations (e.g., those with domain specific terminology) will at times require manual assistance. Thus, in addition to the required microphone and speaker system, some form of display may be needed. Interfaces that are being investigated include pocketable systems, wireless wrist-mounted designs, collar worn microphones and headsets. Each of these interfaces is described in the following paragraphs.

Note that having a display allows the mobile translator system to be also used as a common PC. Documents can be viewed and edited, databases such as phone numbers accessed, email exchanged and the web accessed using wireless modems, all providing a multi-dimensional benefit to using the mobile translator.

B.3.1 Pocket-Sized Touchscreen Displays:

ViA's current touchscreen displays, which work very well for detailed images such as diagrams and maps, are approximately 8.5" x 5" x 0.75". One such display, with a 6.5" screen, is shown in Figure 2. An 8.4" unit with a highly reflective color display that will be readable in bright sunlight is currently being developed and will be commercially available in 1Q '99.[3] A linear array microphone could be embedded into such a display to provide mounting for the microphone/speaker. ViA also has developed a prototype pocketable display called the Optical Viewer. This display is shown in Figure 3. When positioned approximately one-inch from the eye, this system provides the equivalent viewing capabilities of a 17" diagonal desktop display.

---

[3] Funding for the development of new displays is being covered by ViA and its partners. Funding from this SBIR is not being used to develop new display systems and technologies.

B.3.2 Wrist-Mounted Displays:

Under a DARPA-funded research effort, ViA is developing a wireless wrist-mounted interface. The system, shown in Figure 4, uses a low power RF interface to communicate from the wrist to the "belt" (a wearable computer). The screen itself will be readable in bright sunlight. The microphone/speaker will be embedded directly into the device. This system is expected to be available in 4Q '99.

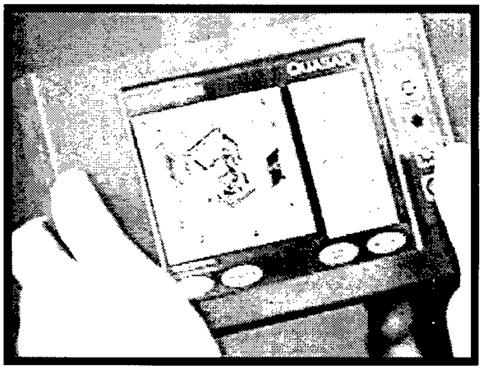


Figure 2 – Touchscreen Display

Figure 3 – Optical Viewer

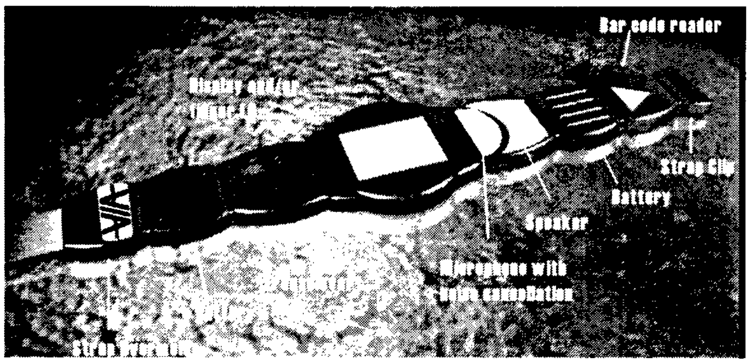


Figure 4 – Wrist Interactive Device

<u>B.3.3 Microphones</u>

To provide a robust voice-to-text capability that will work in all types of environments, an Anti-Noise Canceling (ANC) microphone will be used. These microphones have the ability to separate spoken words from background noise thus dramatically improving the recognition rate of the voice-to-text software.

<u>B.3.3.1 Evaluation of Microphone Systems:</u>

There are several vendors and research groups that provide ANC systems. Some of these systems must be worn close to the mouth, other involve pointing a microphone towards the speaker, while others attempt to automatically "lock-on" to a particular speaker's voice. Several different configurations, such as headsets (e.g., Andrea's ANC-1000), collar mounts (e.g., Labtec's LVA-7370), wrist mounts (e.g., ViA's Wrist Interactive Device) handheld directional units (e.g., Logicon's ABF-4) and intelligent remote microphones (e.g., Ted Berger's work at the University of Southern California) were investigated for their suitability in the mobile translator system. Some examples of these systems are provided in Figure 5.

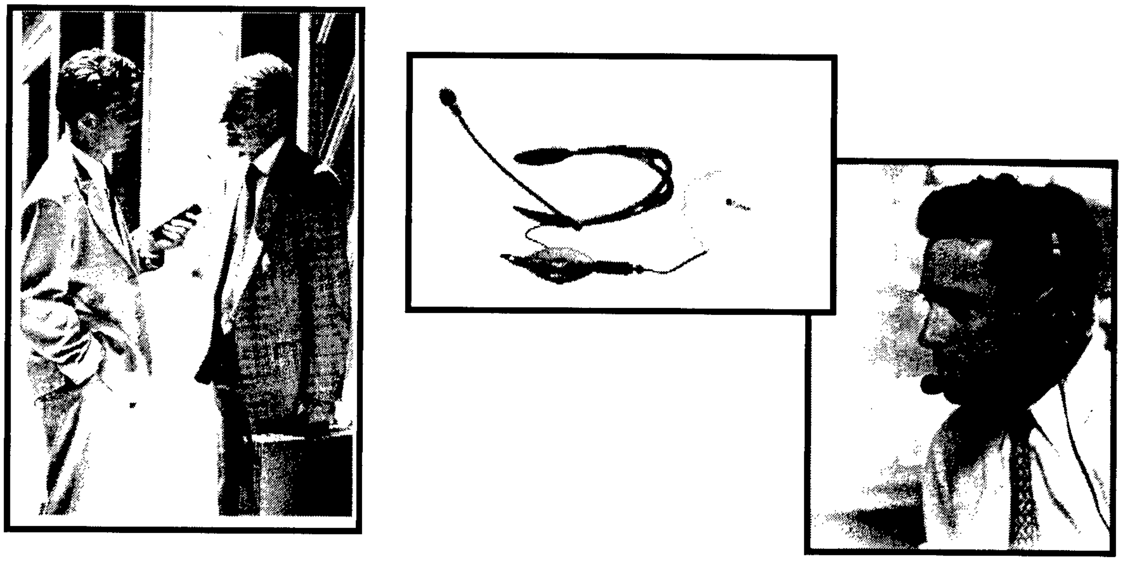


Figure 5 – Sample Microphone/Speaker Systems

It was determined that headset designs provide the best performance in noise canceling. However, this is an unacceptable form factor for the language translator since either two headsets would be needed, or the participants would have to share a single unit. Both situations would make the language translator difficult to use. The alternative approach is to use directional microphones. There are two approaches that are used for these systems: hardware implementations where multiple microphones are used to determine the direction of the sound and filter-out unwanted noise; and pure software approaches where neural networks are trained to simulate a human's ability to filter-out unwanted noise.

B.3.3.1.1 Hardware Approaches:

The best directional microphone was determined to be Logicon's ABF-4. This is shown in Figure 6. The ABF-4 is designed to assist hearing impaired individuals. In its current

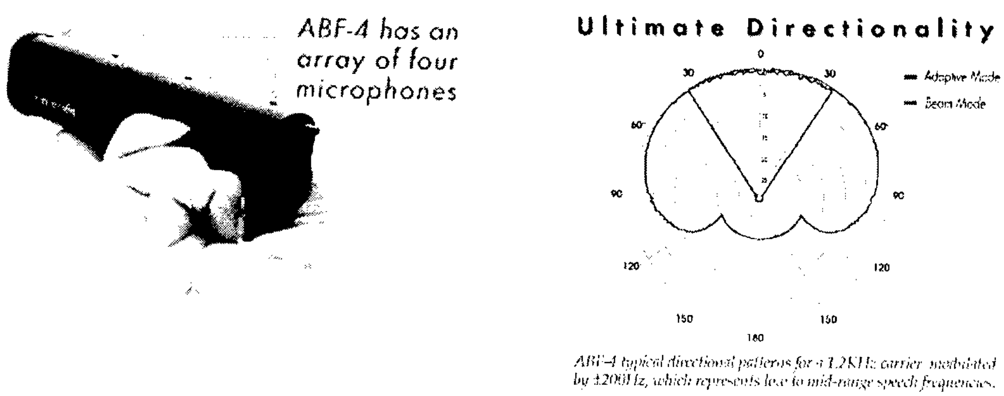


Figure 6: Logicon's ABF-4 Directional Microphone

format, it uses an FM wireless link to connect to a users hearing aid. This same microphone is being repackaged by Andrea Electronics to be used with PC platforms. Beta units of this new configuration are just now becoming available, with commercial release scheduled for 3Q '99.

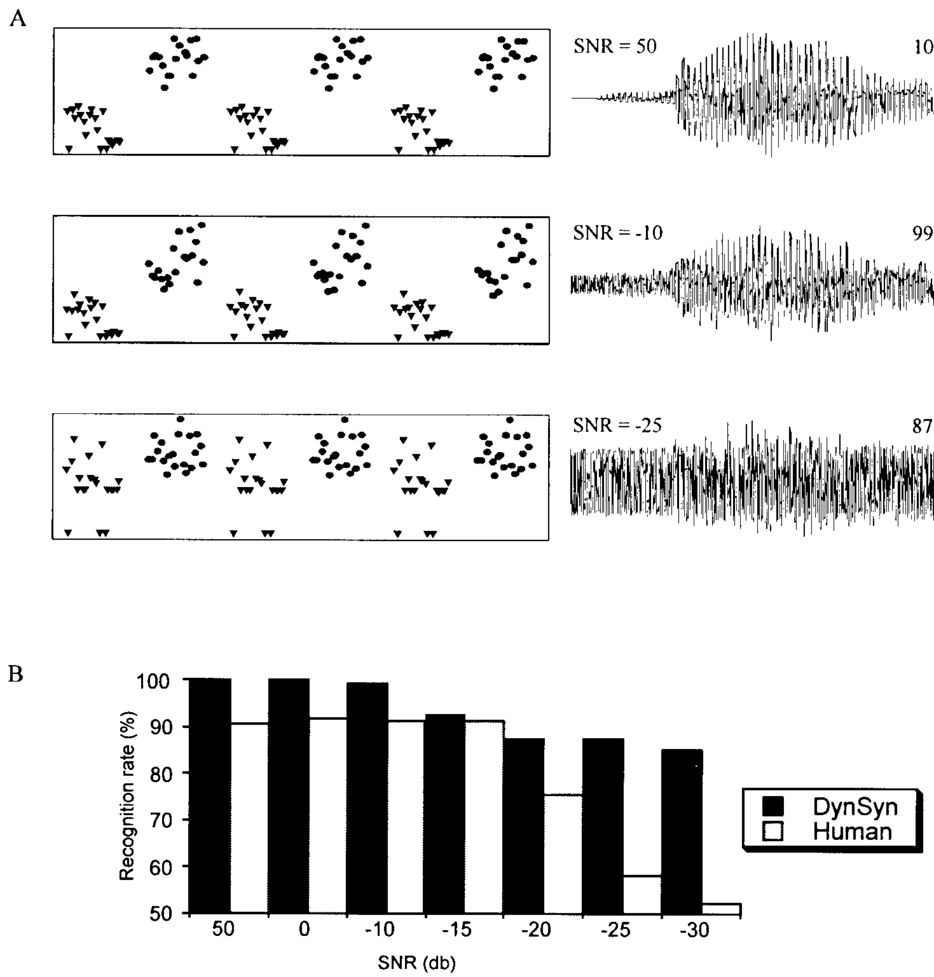
Andrea provided ViA a test unit, which ViA was able to use for a three week period. This unit is shown in Figure 7. Testing was performed in a variety of environments, including office, city and in automobiles/trucks. Noise was deliberatly added to the test environments (e.g., music played over speakers in the office and vehicle environments, standing next to moving trains for the city environments, etc.). Overall, the system performed extremely well with the recognition rates around 90%.

Figure 7: Andrea Directional Microphone Used for Testing at ViA.

B.3.3.1.2 Software Approaches:

For the neural network approach, ViA is pursing collaboration with Ted Berger of USC. Dr. Berger is developing a concept called *Dynamic Synapse*. The intent of this approach is to mirror the principles of how the human brain processes information. A neural network derives its computing capability from the interaction of the neurons in the network. This interaction is regulated by the connections (synapses) between neurons. A neural network can be trained to perform a desired task by changing the synapses according to some learning rules. Neurons communicate with each other by transmitting sequences of electrical impulses and a number of dynamic processes have been known to exist in the synapse. Dr. Berger's concept of a dynamic synapse asserts that with these dynamic processes, a synapse transforms the sequences of electrical impulses into another sequence of impulses. Furthermore, variations across the many synapses of a neuron give rise to different transformation functions. As a result, dynamic synapses allow a neuron to transmit multiple output signals, giving rise to an enormous gain in coding capacity (in conventional neural networks, each neuron generates only a single output). He has used these results to developed a dynamic learning algorithm that trains each dynamic synapse to perform a proper transformation function such that the neural network can achieve highly complex tasks, in this case extracting invariant features embedded in the input signal of each dynamic synapse. This result can be used to filter-out unwanted noise.

Dr. Berger has demonstrated these results by performing speaker-independent word recognition from raw speech waveforms using a small network of neurons connected by dynamic synapses. When tested with speech signals corrupted by noise, the system performed better than human listeners under some conditions; marking the first time ever that a physical device outperforms human listeners in speech recognition task. The results of these tests are shown in Figure 8.



A. Robust speech recognition. Recognition of the utterance "yes" or "no" at three levels of SNR is shown. In the left pannels, the horizontal axis represents the number of trials and the vertical axis represents the total number of spikes generated by the output neuron. The circles indicate the response to "no" and the triangles to "yes."

B Comparison with human performance. The solid bar represents recognition rate by the neural network with dynamic synapses (DynSyn). The model performs significantly better with high noise levels.

Figure 8: Test Results for Dr. Ted Berger's Dynamic Synapse Noise Filtering Algorithms

<u>B.3.3.2 Selected Microphone Approach:</u>

Andrea's version of Logicon's ABF-4 directional microphone has been selected as the microphone to use in Phase 1 of the Language Translator program. This is because of its overall suitability for use in the system and its near-term commercial availability. However, the *Dynamic Synapse* work being performed by Dr. Berger does show tremendous promise for future application, and thus will be pursued further as part of the Phase II activities.

B.3.4 Speaker System:

Most vendors include a speaker system with their ANC microphone. However, such a configuration may not be the best solution for an unobtrusive interface. Thus, alternative speaker systems were investigated. The speaker system must provide clear audio of spoken words, be small and lightweight, robust enough to survive outdoor use (e.g., water and dust resistant), be low in cost and not require high power.

B.3.4.1 Evaluation of Speaker Systems:

Most of the portable speaker systems designed for computers do not have an acceptable form factor for the Language Translator system. They are either designed to be placed on a flat table-top or attached to the edge of a laptop. Some systems with an acceptable form factor, such as HyperSpectral's pizeo-electric speaker system, were determined to be unsuitable because of their high power requirements. Three systems were selected for potential use in the Language Translator: Pryme's *SMP-100*; Kodel's *FlatOut Traveler*; and Mouser's *Mylar 253-5008.* The Pryme design, which is ViA's selection for the Phase 1 system, is described in Section B.3.4.2). The Mouser speaker is the best suited design for the Language Translator (see www.mouser.com for further information). It provides sufficient frequency response for a normal speaking voice (550Hz to 7KHz), requires low-power (100mWatts) and is very small in size (0.8" diameter and 0.1" depth). The significant downside of this speaker is that a separate electronics amplifier will need to be developed. It is possible that this development will be accomplished under a separate contract that ViA has with DARPA. If so, then this speaker will be the best selection for the Language Translator. The Kodel system also has significant potential for use in the Language Translator (see www.kodel.com for further information). The advantages to this system over the Mouser design is its broader frequency response. The disadvantages are its larger size and that it is not yet commercially available in a suitable form factor. ViA is currently discussing potential collaboration with Kodel regarding the use of its speaker technology in future products. If this technolgy is designed into a suitable form factor, then this will provide a high-performance speaker for the Language Translator.

B.3.4.2 Selected Speaker System:

As per the preceeding section, both the Kodel and Mouser systems have significant potential for use in the Language Translator. However, because suitable versions are not currently available, they will not be used in Phase 1 of this project. Both technologies will be followed and may be

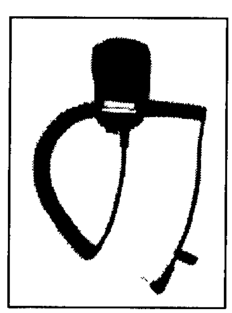


Figure 9: Pryme SMP-100 Speaker

used in Phase 2. The speaker that has been selected is the Pryme SPM-100. This system, which is shown in Figure 9, is designed to amplify voice in outdoor environments. It has an acceptable unit cost of $34.00, and does not require significant power.

B.3.5 Summary of System Components:

Table 5: System Components

Voice-To-Text Software	Sonic Boom with Lernout and Hauspie's Voice XPress
Text-To-Text Language Translation Software	Globalink's Power Translator Pro
Text-To-Speech Software	Lernout and Hauspie's TTS3000 (moving to RealSpeak when it becomes available)
Microphone	Logicon ABF-4 (with continued investigation of Ted Berger's *Dynamic Synapse* software).
Speaker System	Pryme *SMP-100* (with continued investigation of Kodel and Mouser systems).
Computer Platform	ViA II 266Mhz wearable computer with 128 MB RAM
System Integration Software	Lernout & Hauspie Voice XPress SDK

In addition to the above baseline items, a handheld display will be included in the prototype system. This display will not be used during normal operations. However, it will be included in the prototype design to add additional flexibility to the system and serve as an additional communications tool. For example, the operator could show an individual a picture or a video and then ask questions about that particular item. Because the system will often be used in an outdoor environment, ViA's new 8.4" highly reflective display has been chosen. In addition being outdoor readable, this display uses only 1/5th the power of backlit screens. This will substantially increase the battery life of the resulting system.

B.4 Computer Platform Options:

As per the original proposal, the mobile computer that will be used to demonstrate the mobile translator will be ViA's latest generation of wearable PCs. Currently this is a 180 MHz platform. However, in March of 1999, this will be upgraded to a 266 MHz GXm chipset. A picture of this system is shown in Figure 10. The *ViA II GXm*™, which consists of two modules connected with flexible circuitry, is approximately 9¾ inches in length, 3⅛ inches in height, and one inch thick. Its total weight, including batteries for four hours of continuous operations, will be 3 pounds. By using the *ViA II*™, the mobile translator will have at a minimum the following capabilities:

- A Pentium class 266 MHz processor running Microsoft's Windows 95 Operating System. This will provide ample processing for processing the language translation software.

- 128 MB RAM and a minimum of a 3.2 GB Hard Disk (this disk size will most likely be increased to 6 GB to support the system requirements for the RealSpeak Text-to-Speech software package).
- PS/2 keyboard and mouse, standard audio I/O, PC-card (CardBus), RS-232 and USB serial Ports.
- Docking capability, including connection to a port replicator which allows connection to standard PC desktop interfaces (e.g., monitor, speakers, microphone, keyboard, mouse, serial and USB devices), and connection through the CardBus interface to a docking station with standard drive bays for CD-ROM, floppy-disk and additional hard-drives.
- A dual smart battery system providing at least 8 hours of continuous operation or a single battery covering 4 hours of operation.
- A digital wireless RF interface providing remote communication capabilities.

As the leading commercial supplier of wearable computers, ViA is committed to continually updating its product line to include the latest in PC technologies. Thus, by using the ViA wearable platform, a clear path is provided for upgrading the mobile translator to include new technologies and commercial PC components (e.g., processors, memory storage, peripheral interfaces, etc.) as they become available.

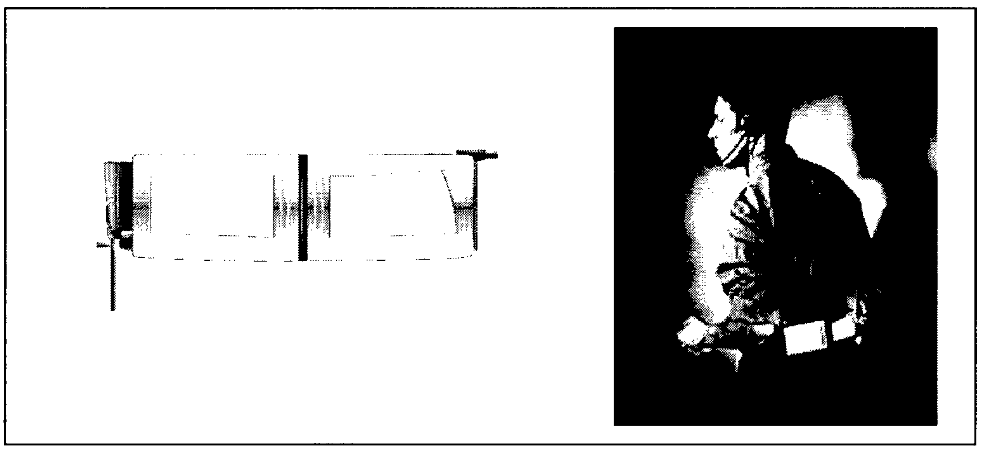


Figure 10 – *ViA II* ™ Wearable Computer

B.5 System Integration Technologies

One of the significant challenges of this research effort is enabling a seamless transition between each of the three phases of translation: voice-to-text dictation, text-to-text language translation and text-to-speech output. In current applications, manual intervention is often required to assist with the translation process. ViA investigated potential solutions to this challenge.

B.5.1 Evaluation of System Integration Technologies:

ViA's investigation of system integration techniques led to two different methodologies. The first approach is to use an intermediate application to transfer results from one step to the next. Most voice engines and translation engines offer integration support for Microsoft Word, and Corel WordPerfect. With this approach, ViA needs to develop an intermediate application that watches any active Microsoft Word documents for incoming text, and then pass the text to the translation process. This approach has two extreme drawbacks. The first arises due to the intermediate application. If the translator utilizes Microsoft Word, the speed of the application is severely limited. Secondly, no text-to-speech engines offer support outside of specialized programmer interfaces. As a result, one third of the application will not be well-served by this approach.

The second method requires ViA to obtain software development kits to simplify the integration process. In this scenario, ViA will take the text directly from one software development kit, and pass it to the next software development kit, and continue the process until all steps have been completed. Each of the selected products offer software development kits (most in beta format) that facilitate this approach. At this phase, ViA anticipates the translation process will not require any user interaction to signal the process to begin.

B.5.2 Selected Approach for System Integration:

The selected approach will create an application that integrates all three software development kits (voice recognition, translation and text-to-speech). To maximize the usability of this system, the software will support both single and multiple platforms. Under this approach, each system will be loaded with all three software development kits. In stand-alone mode, the application will allow users to dictate any amount of text. After a 3 second pause between sentences, the translation will begin. When the translation is in progress, the voice recognition engine will switch to the second language being used with the translator. Therefore, as soon as the text is spoken in the translated language, the other user can formulate a response, and begin dictation.

The distributed approach will maximize the response time of the system. As a user speaks, all systems within network range of the primary machine will receive the text, un-translated. In this approach, the receiver is responsible for translation, and playback. This will allow the real-time use of the system. Multiple users can be speaking at once. Also, this approach will allow for faster operation.

B.6 Commercialization Needs/Applications

ViA has a proven track record in successfully marketing product to the commercial market sector that is being used for the mobile translator. This approach is outlined in the following paragraph.

B.6.1 Background

There are several aspects to ensuring that a technologically successful system will also become a commercial success. For example, ensuring that the product's value to the customer is greater than its cost, and that the customer can afford this cost.

- *Determining Value*: Value to the customer means that the system provides the customer with a solution or service that is an improvement over his/her current means of filling this need. The solution and service that this system will provide is near real-time language translation. The value of having this service, such as enabling new operational capabilities for military missions, needs to be determined.
- *Determining Cost*: In parallel with determining value, market research needs to be performed to determine current approaches (e.g., the use "flip-cards" with pre-defined phrases, phone calls to an on-duty translator and/or miniature text-only translation devices) and their associated costs. Cost in this case is not only the capital expenditure, but also the costs associated with using the product. For example, the costs/savings associated with training, maintenance and time savings.
- *Optimizing Value-to-Cost*: The language translation approaches that are currently used are then compared with the new mobile translator to ensure that its value exceeds its cost. This cost needs to be minimized by ensuring that each feature designed into the system adds greater value then its expense. For example, if a graphic display costs more than the customer value it provides, then it should not be included as part of the baseline design.
- *Affordability*: The resulting system must be one that the customer can afford. Even though a proposed design may be a better solution that in the long run will provide a cost savings, if the customer cannot afford the initial purchase price, then the system will not be a commercial success.

In summary, understanding what the customer needs, would like, and can afford, and designing a system that meets these aspects, is crucial to making the system a commercial success. All of these elements, plus aspects such as marketing, distribution and product support are part of this Phase I activity.

B.6.2 Current Status:

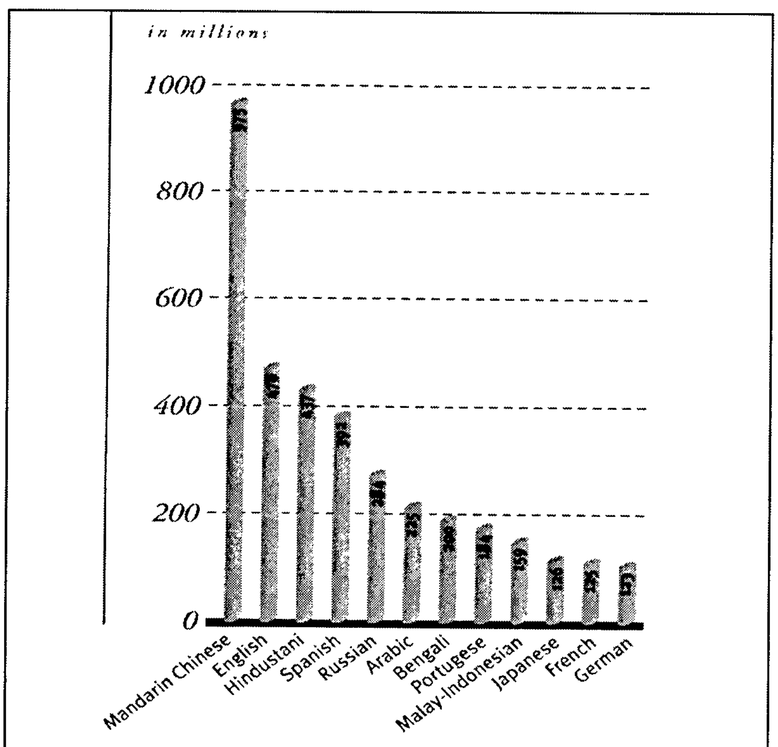
A list of potential customers for the language translator has been generated and ViA is in the process of discussing such a product with these groups. The intent of this effort is not to contact representatives from every group that may use the language translator, but rather to contact enough representatives to generate a sufficient description of the system's design. This list is given in Table 6. Initial discussion with each of these groups has been made.

Table 6 – Potential Customers for Language Translator

Industry	Representative Customer
Military	General Dynamics, Navy, SOCOM
Retail	BestBuy
Airlines	Northwest
Package Delivery	Federal Express and Schwan's
Food Processing	IBP
Manufacturing	Ford and 3Com
Restaurants	Starbucks Coffee
Hotels	Radisson, Hogadata
Cruise Ships	Diamond Lines
Law Enforcement	SantaCruz Police Department
Insurance Agents	Allstate

Special emphasis is being given to the military community to ensure that the system will meet their demanding requirements. Towards this end, ViA has contracted with General Dynamics Information Systems (GDIS) to market their wearable computers to the military. GDIS has knowledge of military needs, performance requirements, deployment, training and support issues and is ideally suited to ensure that this marketing effort is successful. ViA will work with GDIS in designing the mobile translator to ensure that it meets military requirements. Additional contacts have been made with USSOCOM (Orlando) and CINCPAC (SanDiego).

One interesting result of ViA's commercialization efforts has been an identification of the most widely spoken languages in the world. These results are shown in Figure 11.



Source: Univ. of Washington estimates, 1995

Figure 11 – Most Widely Spoken Languages in the World

A complete summary of ViA's market research activities will be included in the next report.

C. System Requirements Document

C.1. Summary

This document defines the requirements for the mobile language translator software under development by ViA Inc. The scope of this document is limited to the Phase I research and implementation effort.

C.2. Requirements

C.2.1 Design Requirements

The goal of the language translator Phase I project is to develop a near real-time, two-way, mobile, lightweight, robust and low-cost multilingual translation device that can be operated in a hands-free manner.

C.2.2 Specific Design Requirements

C.2.2.1 Usage

a. Upon installation, a brief voice-profile training process must be undergone in order to guarantee accurate recognition.
b. A user profile will also be configured that will include an approximate age group for the user, as well as gender. This will increase the recognition capabilities.
c. The user that is speaking the native language (herein referred to as the primary user), will speak either English or German.
d. The computer will then receive the spoken data, and with no interaction from either the primary user or the user that desires the translated text (herein referred to as the secondary user), translate the recognized data to the opposite language pair. (English ←→ German)
e. Upon successful translation, the language translator will then speak the translated data using a voice synthesis product to the secondary user in the translated language.
f. The System will be full duplex, therefore either user could speak as they receive a translated voice response.

C.2.2.2. Time Specifications

a. After the primary user speaks a phrase or sentence, the translation will begin either after the user voices in an end-of-sentence sentinel ("period", "Question mark," "Exclamation point") or after a two-second pause.

b. When the translation begins, all external processing will cease in order to facilitate a quick translation. Since a machine translation approach is being used, a single sentence could take between 5-10 seconds to translate.
c. Upon translation, the text will immediately begin the synthesis and playback process.
d. All of these described steps will take place with no user interaction.

C.2.2.3 Audio Head-set

a. Since the system is designed to be mobile, external, battery-powered speakers will be used to broadcast the translated speech.
b. A mobile array microphone will be used to facilitate a more natural mobile environment.

C.2.2.4 Hardware Platform

a. The system will be robust enough and optimized to run in combination of multiple ViA II computers, but ideally will run locally on a single machine.

C.2.2.5 Commercialization Plan

a. ViA has undergone specific research to ensure the mobile translator will deliver inherent value to our customers at an affordable cost.

MATERIAL INSPECTION AND RECEIVING

Form Approved
OMB No. 0704-0248

1. Proc. Instrument Iden (contract)	(Order) No.	6. Invoice No./Date	7. Page	Of	8. Acceptance Point
N00014-99-M-0009		10141 1-15-99	1	1	Other

2. Shipment No.	3. Date Shipped	4. B/L	5. Discount Terms
FPC0002	99 Jan 15	B/L N/A IDT N/A	N/A

9. Prime Contractor — Code 1DKH7 ViA, Inc. 11 Bridge Square Northfield, MN 55057	10. Administered By — Code S2401A DCMC Twin Cities 3001 Metro Drive Bloomington, MN 55425-1573
11. Shipped From (If other than 9) — Code SAME	12. Payment Will Be Made By — Code HQ0339 DFAS COLUMBUS CENTER DFAS-CO/West Entitlement Operations PO Box 182381 Columbus, OH 43218-2381
13. Shipped To — Code N00014 Office of Naval Research 800 North Quincy Street Arlington, VA 22217-5660	14. Marked For — Code Attn: Joel Davis, ONR 342 Ballston Tower One

15. Item No.	16. Stock/Part No. Description (Indicate number of shipping containers-type of container-container number)	17. * Quantity Ship/Rec'd	18. Unit	19. Unit Price	20. Amount
0001AA	Second Contractor's Progress Report		1	17,498.00	17,498.00

21. CONTRACT QUALITY ASSURANCE

A. ORIGIN	B. DESTINATION
☐ CQA ☐ ACCEPTANCE of listed items has been made by me or under my supervision and they conform to contract, except as noted herein or on supporting documents.	☐ CQA ☐ ACCEPTANCE of listed items has been made by me or under my supervision and they conform to contract, except as noted herein or on supporting documents
Date — Sig. of Auth Govt Rep Typed Name And Office	Date — Sig. of Auth Govt Rep Typed Name And Title — Joel Davis, ONR 342

22. RECEIVER'S USE

Quantities shown in column 17 were received in apparent good condition except as noted.

Date Received — Sig. of Auth Govt Rep

Typed Name And Office

* If quantity received by the Government is the same as quantity shipped, indicate by (√) mark, if different, enter actual quantity received below quantity shipped and encircle.

23. Contractor Use Only