A Natural-Language Speech Interface Constructed Entirely as a Set of Executable Specifications

R. A. Frost

Department of Computer Science
University of Windsor, Windsor
Ontario, Canada N9B 3P4
richard@uwindsor.ca

Abstract

SpeechNet is a collection of speech-accessible hyperlinked objects called sihlos. Sihlos are deployed over the Internet and are accessed by remote speech browsers. When a speech browser accesses a sihlo, it begins by downloading a grammar file which is used to configure the browser in order to achieve high recognition accuracy. Sihlos are hyperlinked in a manner that is similar to the linking of html pages. SpeechNet provides non-visual access to knowledge which is analogous to visual access provided by the web. One of the sihlos can answer thousands of spoken pseudo-natural-language questions about the solar system. This sihlo has been constructed entirely as a set of executable specifications of the language that it can process.

SpeechNet

The world-wide web is primarily based on visual browsing of hyperlinked text and graphical objects. The huge knowledge base available on the web is largely inaccessible to visually-challenged users. The SpeechNet project at the University of Windsor (Frost 1999b) overcomes this problem by augmenting the web with a network of speech-accessible hyperlinked objects called sihlos.

Each sihlo consists of a language processor and a grammar defining the syntax of the input language. When a remote speech browser contacts a sihlo, it downloads the grammar in order to configure the speech recognizer. This alleviates a major difficulty in building accurate speech interfaces to large knowledge bases. Sihlos enable the knowledge to be divided into small modules each having a "small" language associated with it. This allows highly accurate user-independent continuous-speech recognition. Hyperlinking sihlos allows the user to navigate through a complex interrelated body of knowledge using spoken prompts in a way that is analogous to the way in which the web is navigated using visual prompts.

Although sihlos can alleviate the problem of building sophisticated speech interfaces, their design and construction can be quite difficult. The task of building the language processors is compounded by the concurrent task of designing a grammar which has high recognition accuracy. It has been shown (Frost 1995) that this process can be facilitated if the interpreter and speech recognizer are constructed as executable specifications. In the following, we describe an example of applying this approach.

The Natural-Language Query Interpreter

The `solar man` sihlo is a natural-language interface to a database containing information about the planets, moons and people who discovered them. This sihlo is constructed as an 800–line executable specification of an attribute grammar defining the syntax and semantics of the input language. The specification is written in a language called W/AGE developed at the University of Windsor. (Frost 1994). The specification includes:

- A dictionary, where meanings of words are defined directly in terms of database constructs, or indirectly as being equivalent to the meaning of other phrases.
- A set of production rules (syntax rules) augmented with semantic actions which define how attributes of a compound expression are computed from the attributes of its component expressions.
- A set of Chomsky-like syntactic re-write rules.
- A set of equations defining the semantic model on which the processor is based.

The semantic theory underlying `solar man` is loosely based on Montague's approach to semantics (Dowty, Wall and Peters 1981). The processor can answer several thousand well-formed queries expressed in a sub-set of English. For example:

```
who discovered a moon that orbits mars
did Hall discover every moon
how many moons were discovered by
                        Hall or Kuiper
was every moon discovered by a man
```

```
which planets are orbited by a moon that
                was discovered by Galileo
which moons were discovered by nobody
is every planet orbited by a moon
which planets are orbited by two moons
who was the discoverer of phobos
Hall discovered a moon that orbits mars
does every moon orbit every planet
does phobos orbit mars
```

In addition to access trhough a speech browser, `solar man` can also be accessed through an html web page, which contains a full listing of the executable specification and a paper describing the W/AGE executable specification language.

```
www.cs.uwindsor.ca/users/r/
   richard/miranda/wage_demo.html
```

W/AGE and Executable Grammar Objects

W/AGE allows language processors to be constructed by defining the syntax and semantics of the language to be processed. The notation is a variant of standard textbook notation for attribute grammars. Advantages of the "attribute-grammar programming paradigm" are discussed in (Paaki 1995). W/AGE has been used in the construction of various types of language processor including VLSI design transformers, theorem provers, as well as the natural-language speech interface discussed here.

Current work includes the implementation of W/AGE in Java. This involves the development of *executable grammar objects* which allow object-oriented language processors to be built with structures that are closely related to the grammars defining the language to be processed (Frost 1999a).

The Speech Browser

The prototype speech browser is written in Java. It uses the IBM Via Voice speech-recognition engine and IBM's implementation of the Java speech APIs. The browser runs on regular PCs, needs no training (is user-independent), and accepts continuous-speech input. Speech-recognition accuracy is very high for the `solar man` example. Response from the browser is in real-time and the downloading of the grammar and reconfiguration of the speech recognizer (when sihlos are changed) is not noticeable.

Concluding Comments

The real potential of SpeechNet will only be attained if tools are developed to assist people with the task of designing grammars, constructing language processors, and deploying sihlos in an appropriate manner over the Internet. The approach described in this paper goes some way towards achieving the second of these goals.

Current work is investigating the use of the Common Object Request Broker Architecture in the deployment of sihlos. This will facilitate resource sharing, openness, and scalability. It will also enable sihlos to have "sessions" with end users thereby supporting human/computer speech dialogue.

If tools are available to facilitate the construction and deployment of sihlos, it may be possible to create a huge hyperlinked body of knowledge that is accessible through remote speech interfaces anywhere that has access to the Internet.

Acknowledgments

Many people at the University of Windsor have contribute to SpeechNet and to the construction of the solar interface. In particular, the author would like to thank Sanjay Chitte, who wrote the Java code for the prototype speech browser and Barbara Szydlowski who built the web interface to the solar man interpreter. The author acknowledges the support provided by NSERC in the form of an individual research grant.

References

Frost, R. A. 1999a. Improving the efficiency of executable grammar objects. University of Windsor, Department of Computer Science Technical Report CS1999–01.

Frost, R. A. 1999b. SpeechNet: A network of hyperlinked speech-accessible objects. Proc. WECWIS '99: *International Workshop on Advanced Issues of E-Commerce and Web-Based Information Systems.* Santa Clara, CA. Apr. 1999.

Frost, R. A. 1995. Use of executable specifications in the construction of speech interfaces. *Proc. IJCAI Workshop on Developing AI Applications for the Disabled*, Montreal 1995.

Frost, R. A. 1994. W/AGE The Windsor Attribute Grammar Programming Environment. *Schloss Dagstuhl International Workshop on Functional programming in the Real World*.

Paaki, J. 1995. Attribute grammar paradigms — a high-level methodology in language implementation, *ACM Computing Surveys* 27(2) 196–255.

Dowty, D. R., Wall, R. E. and Peters, S. 1981. *Introduction to Montague Semantics*. D. Reidel Publishing Company, Dordrecht, Boston, Lancaster, Tokyo.