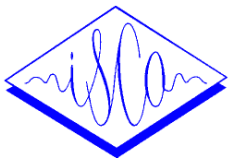

Speech and Language Technology
for Disabled Persons
Stockholm, Sweden
May 31 - June 2, 1993

ALTERNATIVE STRATEGIES FOR CREATING AUTOCAD DRAWINGS

Corine Bickley[1,2], Sheri Hunnicutt[2], and Lori Lamel[3]

[1]Research Lab. of Electronics, Massachusetts Inst. of Technology
Room 36-521, Cambridge, MA 02139, USA, Fax 1-617-258-7864
[2]Department of Speech Communication and Music Acoustics, KTH
Box 70014, S-100 44 Stockholm, Sweden, Fax 46-8-790-7854
[3]CNRS LIMSI, bat. 508, b.p. 133
91403 Orsay cedex, France, Fax 33-1-69-85-80-88

ABSTRACT

Three different approaches for creating AutoCAD drawings with voice input have been examined: isolated-word speech recognition, connected-speech recognition, and word prediction (followed by translation into drawings). Commercially available isolated-word recognizers are limited by a small vocabulary size and by training requirements. Connected-speech recognizers permit faster and more natural interaction. Word prediction can be used to specify textual rules which facilitates modification of drawings.

INTRODUCTION

Access to software through alternatives to a standard computer keyboard is necessary to enhance the vocational options for persons with motoric disabilities. One example of vocational software is AutoCAD, a commercially-available computer-aided design package: speech, joystick, and head pointing have been documented as input methods to this system (Burnett et al., 1991; Bickley et al., 1993). While these projects have used isolated-word speech recognizers, in this research we extend the set of input strategies to include connected-speech recognition and prediction. We hope that by demonstrating accessibility using these new alternatives for input to AutoCAD, we will make available a broader range of input choices for vocational software for all persons, particularly persons with motoric disabilities.

AutoCAD drawings can be generated either by creating each line, arc, and label in terms of endpoints, centers, radii, and such, or alternatively, by writing rules which describe the types and sizes of each entity and the spatial relationships among them. Each approach has certain advantages and disadvantages, and can be the appropriate choice depending on the needs of the user and the characteristics of the application. Drawings that are needed just once, or that will be in a continual process of modification, can be created line by line, i.e., by drawing each component. In contrast, drawings that will be reused frequently, possibly with some minor modifications, can more conveniently be represented by a set of rules in text. These rules can then be translated and the drawing generated automatically by a computer program. This technology has potential application for persons with motoric disabilities: the process of creating the rule-based drawings by voice might be more effective than by other means because various technologies for text creation already exist.

USE OF ISOLATED-WORD RECOGNITION WITH AUTOCAD

Most commercially available speech recognizers are speaker-dependent, isolated word systems with relatively small vocabulary sizes (a few hundred words), which are based on pattern matching algorithms, such as dynamic time-warping. Advantages of these sorts of systems are low cost and reasonably high recognition accuracy. While higher recognition accuracies are typically obtained with speaker-dependent systems, this has the disadvantage that each user must train the system. The training process usually consists of pronouncing each of the items in the vocabulary several times. When the vocabulary is modified, the new words must be trained as well.

The BUG system (developed by Command Corp., Atlanta, GA, USA), which was used in this project, enhances recognition accuracy by the use of command sets. The vocabulary is divided by the user into a number of separate sets, each of which can be chosen to contain acoustically distinct words. Voice triggers are used to switch between command sets. A judicious selection of the words in each set, as well as the number of command sets (total and currently active), and the organization of words into separate sets which reflects the functionality of AutoCAD is needed to balance the tradeoff between recognition accuracy and ease of use. Training consists of prompting the user to speak each word three times. If any error is made in the training of a word, the word must be retrained.

Excerpts from a portion of a session of AutoCAD controlled by the BUG recognizer are shown in Figure 1. Each quoted word in the figure represents one voice command; these 14 commands would be used to create two lines.

```
"modify"  "offset"  "four"  "feet"
"drawline"  "tools"  "endpoint"
"draw"  "ten"  "feet"  "six"  "inches"  "comma"  "zero"
```

Figure 1. *BUG voice commands to create lines*

USE OF CONNECTED-SPEECH RECOGNITION WITH AUTOCAD

Another approach to controlling AutoCAD by voice is to use phrases of connected words instead of isolated words. Laboratory research systems are capable of recognizing continuous speech with 5000-word vocabularies, without speaker-specific training data, with a word accuracy of better than 94% (Pallett et al., 1993). However, such systems are neither real-time nor commercially available. There are some more limited continuous speech recognizers which *are* commercially available. We used one such system, DATAVOX (marketed by the Vecsys Company, Bievres, France) which can recognize continuous speech with vocabularies of 300 - 500 words in a speaker-dependent mode. This recognizer uses a grammar to guide the recognition strategy. This approach has two potential advantages: greater speed of entry of voice commands and increased accuracy of word recognition. An increase in speed can result due to the possibility of the user entering several words as a phrase, for instance, "three feet six inches comma zero." Highly robust recognition can be achieved because the grammar restricts the possible set of target words at each stage of command entry.

The DATAVOX system uses a specialized VLSI processor for Dynamic Programming matching developed by LIMSI and VECSYS in collaboration with the BULL Company to produce a spectral analysis of the speech signal. The recognizer processes acoustic units of words, disyllables, and noises, and adapts robustly to the presence of noise and to changes in speaking rate. A software package aids the user in defining the vocabulary, the syntax, and the sequence of characters generated as a result of the output of the speech recognizer. The training consists of two phases in which the vocabulary items are spoken in isolation and in context, where the training prompts are automatically generated in compliance with the specified grammar. Since the program which generates these sentences attempts to minimize the total number of sentences to be spoken, the sentences can become quite long if permitted by the grammer.

Excerpts from a portion of a VECSYS-controlled AutoCAD session which creates two lines are shown in Figure 2; only three phrases are needed with this connected-speech recognizer. But more importantly, the connected-speech recognizer permits the user to speak in a more natural fashion.

```
"offset four feet"
"line from endpoint"
"to ten feet six inches comma zero"
```

Figure 2. *VECSYS voice commands to create line of Figure 1*

USE OF PREDICTION IN RULE-WRITING WITH AUTOCAD

A third alternative is to create rules by text entry which can then be translated automatically into AutoCAD drawings. In this project, we investigated a way to create textual rules which is known to be appropriate for persons with motoric disabilities (Magnuson and Hunnicutt, 1992). A method of achieving the text coverage of a large-vocabulary speech recognizer while using an inexpensive smaller vocabulary system is to employ word prediction. The input vocabulary can be a set of maximally distinctive code words for each letter, number and set of special commands. Prediction of possible words may begin as soon as a first letter is recognized. In this way, almost all words can be accessed by speaking one to three code words plus an acceptance command. A lexicon of word pairs also provides the facility of predicting possible second words with no additional input after a first word has been accepted.

Figure 3 shows a set of rules which when translated by a program such as The ICAD System (marketed by ICAD, Inc., Cambridge, MA, USA) would result in two lines. Thirty-two voice commands (or any kind of ``keystroke" equivalent means of input) would be needed to create these lines.

```
side-wall type wall
        thickness centimeters 15
        length-of-wall meters 3
        position right 0
```

Figure 3. *Textual rules created by Profet prediction program*

DISCUSSION OF THE THREE STRATEGIES

In the AutoCAD application, there are some commands which lend themselves well to control by an isolated-word recognition system, but others for which connected speech would be preferable. An obvious contrast is the entry of single-word commands such as "trim" and the entry of coordinates such as "three feet six inches comma zero." The command "trim" must be followed by some sort of graphic input (such as pointing with a mouse, joystick, or headpointer), and is an example of a case for which an isolated-word recognition system works quite adequately. In the case of specifying coordinates such as those given above, a connected-speech recognizer would obviously be preferable.

The approach of using rule-based commands to describe the sizes and positions of each entity in the drawing, for which prediction holds much promise, addresses the specification of drawings in an entirely different manner. The number of ``keystrokes" needed to create the rules which specify a drawing for the first time is greater than the number of AutoCAD commands needed to create the drawing directly. In cases in which the drawing will be modified (e.g., to change thickness), after the second modification the rule-based approach gives a savings in input commands over the isolated-word recognizer, and after several more modifications the rule-based technology is faster than even the connected-speech recognizer.

ACKNOWLEDGEMENTS

Partial funding for this work was supplied by the U.S. Department of Education Rehabilitation Services Administration Grant #H128A01011, by "Money directed to excellent areas," KTH, Stockholm, and by a LIMSI "Action Incitative." We appreciate the support and use of software from ICAD Inc. and the software support and DATAVOX system provided by the VECSYS Company. We thank James E. Mszanski for his efforts in using the speech recognition systems during the trials.

REFERENCES

Bickley, C., H. Tan, and D. Horowitz (1993): "Voice input for graphics and text creation: a case study," in *Proceedings of CSUN'93, Technology and Persons with Disabilities.*

Burnett, J.K., C.R. Klabunde, and C.W. Brittell (1991): "Voice and head pointer operated electronics computer assisted design workstations for individuals with severe upper extremity impairments," pp. 48-49 in *RESNA '91, Proceedings of the 14th Annual Conference*, RESNAPRESS, Washington, D.C.

Gauvain, J.L. (1990): "Le systeme de reconnaisance AMADEUS: Principe et algorithmes," *LIMSI Report.*

Lamel, L.F. and J.L. Gauvain (1992): "Large-vocabulary speech recognition at LIMSI," in *Proceedings of the final review of the DARPA Artificial Neural Network Technology (ANNT) Speech Program*, Stanford, CA.

Magnuson, T., and Hunnicutt, S. (1992): "Swedish Word Prediction: A Follow-Up Investigation," *AAC*, **8**:2, p. 151 (1992 ISAAC Conference Abstracts).