Computer-Aided Design and Applications

http://www.cadanda.com

Design by Talking with Computers

X. Y. Kou and S. T. Tan

The University of Hong Kong, kouxy@hku.hk, sttan@hku.hk

ABSTRACT

This paper describes our preliminary research on a voice-aided CAD system, where human voice is utilized to assist CAD modeling tasks. The goal is to provide a novel type of CAD system, in which users can design by talking with computers. Voice is regarded as an alternative interaction medium to boost design efficiency, to promote productivity and to improve ease of use. The proposed system accepts users to utter voices such as "Draw a circle, radius 12", "Extrude it to 40", "Delete it" etc, and the system will execute the modeling commands if the required inputs are valid and complete. A prototype voice-aided CAD system is presented and its framework is outlined. Major challenges and suggested solutions are presented; and preliminary results are also provided.

Keywords: voice, voice user interface, speech recognition, CAD, modeling, design by talking.
DOI: 10.3722/cadaps.2008.266-277

1. INTRODUCTION

In 2004, a science fiction film "I, Robot" [1] made a splash all over the world. The film takes place in 2035, Chicago, where intelligent humanoid robots violated human's commands and even killed their inventors. The audiences were amazed at the technologies demonstrated in the film, for instance, robots can fluently communicate with human beings in natural voices, and they can even display emotions such as anger and fear. Although this fancy scene is still a fiction, it does represent human's strong desire and expectation to conduct natural communications with computers/robots. Such natural interactions have manifested many unique advantages over the current human-machine (e.g. mouse/keyboard based) interactions, for instance, they are natural, flexible and efficient.
Speech or human voice has aroused tremendous interest in the field of human-computer interactions. With the recent development in speech synthesis and recognition technologies [2-4], voice based interaction has demonstrated excellent flexibility and wide applicability in a variety of areas, such as ***word processing*** (e.g. Microsoft Word [5]), ***medical*** (e.g. Naturally Speaking Medical [6]), ***biomedical*** [7] and ***telephony*** [8] applications. Using voice input, a predefined operation can be efficiently executed through speech recognitions, as opposed to the cumbersome command searching through traditional Graphical User Interfaces (GUI).

This paper is motivated to leverage human voice and integrate it into CAD modeling systems. The goal is to provide a novel type of CAD system, in which users can design by talking with computers. Voice is regarded as an alternative interaction medium to boost design ***efficiency***, to promote ***productivity*** and to improve ***ease of use***. To achieve such goals, we have conducted a preliminary study on a voice-aided CAD system, and this paper reports the technical challenges we encountered as well as our solutions to these problems.

The subsequent sections of this paper are logically structured as follows: Section 2 reviews related work and the technical challenges; and the motivations of this project are discussed. Section 3 presents our solutions to these challenges and detailed paradigms are described. Section 4 offers some preliminary results of this pilot investigation; the efficacy of the proposed voice assisted CAD system is demonstrated. Finally, discussions and future work are described and the paper is concluded in Section 5.

2. RELATED WORK AND MOTIVATIONS

2.1 Novel Technologies for Human-Computer Interactions

Human beings rely heavily on acoustic, haptic, visual, gustatory and olfactory senses to acquire knowledge from the outside world and to generate feedbacks, reactions or to exert controls. Among these means of communication, the mouse and keyboard based interactions have been maturely used in Human-Computer Interactions (HCI) and are generally termed as direct manipulations [4, 7]. These traditional interactions are well suited for applications where the hands have easy access to the keyboards or mouse devices.

There are, however, many scenarios in which conventional direct manipulations are inconvenient or additional hand-free manipulations are expected. In word processing, voice based input prove to be efficient and easy to use. It is reported that most people can speak about five times faster than they can type [9]. Microsoft [10] has fully supported voice input in the latest Windows Vista operating system, and users can not only input literal words in word processors such as Notepad and Microsoft Word, they can also easily execute commands (e.g. open files, copy and paste etc.) via voices. Nuance Communications Inc. also takes advantages of speech recognition in medical applications. With the speech recognition module in Naturally Speaking Medical [6], healthcare professionals are greatly freed from the tedious manual transcription of patient notes which traditionally results in high cost and long turnaround time. At research institutes, voice based interactions are also widely investigated. Guy et al [11] developed a computer program called “Math Speak & write” to read and hear mathematical input using human voices. The recognized words are converted to math symbols for quick creation of mathematical expressions, which substituted the traditional input using tedious keyboard strokes. Reddy et al [12] proposed a Voice Operated Information System (VOIS) for drivers information and guidance where the hands of a driver are occupied in driving tasks. By using voice as the medium, increased road transport safety can be obtained since the driver's attention is no longer distracted from traditional hand-busy cases. Victor Zue [9] devised voice based interaction systems for weather forecasts and flight schedules, and users can talk to computers, raise questions and get responses.

In addition to voice based human-computer interaction, many novel types of interactions have also emerged in recent years. In the latest release of Microsoft Math 3.0 [13], a featured “Ink Handwriting” module is proposed, which enables users to input mathematical formulae using Tablet and Ultra-Mobile PCs by means of handwriting recognitions; improved ease of use and better efficiencies are obtained. Shwetak and Gregory [14] devised a hands-free input scheme by directly blowing at computer screens . The approach “leverages the amplitude and phase responses produced by different wind patterns” [14], and the target region a user blows at is determined from distinct response signatures.

2.2 Novel Human-Computer Interactions in CAD/CAM Systems

Apart from the aforementioned schemes, novel human-computer interactions have also been integrated into CAD/CAM applications in recent years. Liu et al [15, 16] proposed to use haptic devices to aid the surface and solid modeling in their virtual DesignWorks system; designers can use a haptic interface to touch a “native B-rep CAD model” for direct surface manipulations [16], which traditionally, is carried out using mouse drag-and-drop. Zhu and Lee [17] presented a haptic-aided 5-axis pencil-cut tool path generation method and haptic interface are integrated to calculate the force-torque feedbacks. Zou and Lee [18, 19] targeted at reconstructing 3D models from inaccurate 2D sketches and proposed several algorithms to eliminate inconsistent constraints and generate valid 3D polyhedral models. Diniz and Branco [20] proposed a direct freehand drawing system, which tracks the movement of two LED lights attached to the designer’s hands. The 3D positions and the paths of each light/hand are taken as the truly 3D input in surface construction. These novel human-computer interaction schemes, to certain degrees, either enhanced the ease of use or improved design efficiencies of CAD/CAM systems.

2.3 Motivations of Voice-Aided CAD System

Despite the many novel interaction schemes used in CAD/CAM systems, voice, “the most natural and intuitive communication medium between people” [4] appears to have not been well exploited in the context of CAD/CAM applications. Unlike such applications as word processing, where the recognized words can be directly mapped to literal words, it seems that voice based interactions can offer only limited benefits to CAD designers, as most elements of interest are of graphical natures. In fact, speech or voice user interface has a number of unique characteristics and advantages for use in the CAD modeling environment, some of the most important ones are described as follows.

- First, it is natural and easy to use. We know how to speak before we know how to read and write. Normal people have a natural familiarization with speech; there is no need for special trainings; non-specialists and kids can equally evoke the same operation as CAD professionals via voices. In contrast, traditional keyboard and mouse based interactions tend to block the design ideas, since they require users to be familiar with the interface designed by software developers. Therefore designer usually focus on the software (CAD modeler), instead of the design itself.
- Second, speech is efficient and flexible. This is especially useful to improve the design efficiencies. It is known that contemporary CAD systems are getting more and more powerful, and at the same time it is also true that they are far more complex than ever. Even for a simple CAD system, it is not surprising to see hundreds of menu items and toolbar buttons squeezed to limited screen regions. Users are frequently compelled to toggle on and off certain modules, traverse across a number of menu hierarchies or conjecture among many button icons, just to get a desired command executed. Fig. 1 shows parts of toolbar buttons in SolidWorks 2007 CAD modeler [21], and it is not difficult to imagine how complicated it will be to select a right button from the 500+ candidates! Voice based interaction, however, can alleviate such cumbersome problems and promote the overall productivities.
- Thirdly, speech is omnidirectional and can communicate with multiple users [7]. Using the voice as the medium, cooperative CAD modeling can be carried out both locally or remotely via networks. Notably and uniquely, such communications are both understandable to computers as well as design partners. Real time modifications can be immediately generated, approved or rejected with intuitive and natural voice controls (such as "Good", "I love this", "I dislike it" etc.), which may greatly increase the cooperative efficiency, the ease of use and user satisfactions.
- Last but not least, due to the inherent complexity of CAD design and the long time use of keyboard and mouse, designers usually suffer from repetitive strain injuries (RSI) [22]. Voice-aided CAD system can offer an alternative approach to keep the design in progress without sacrifices of their health. Moreover, physically handicapped people who have no/limited control over mouse/keyboard can also benefit from this novel scheme.

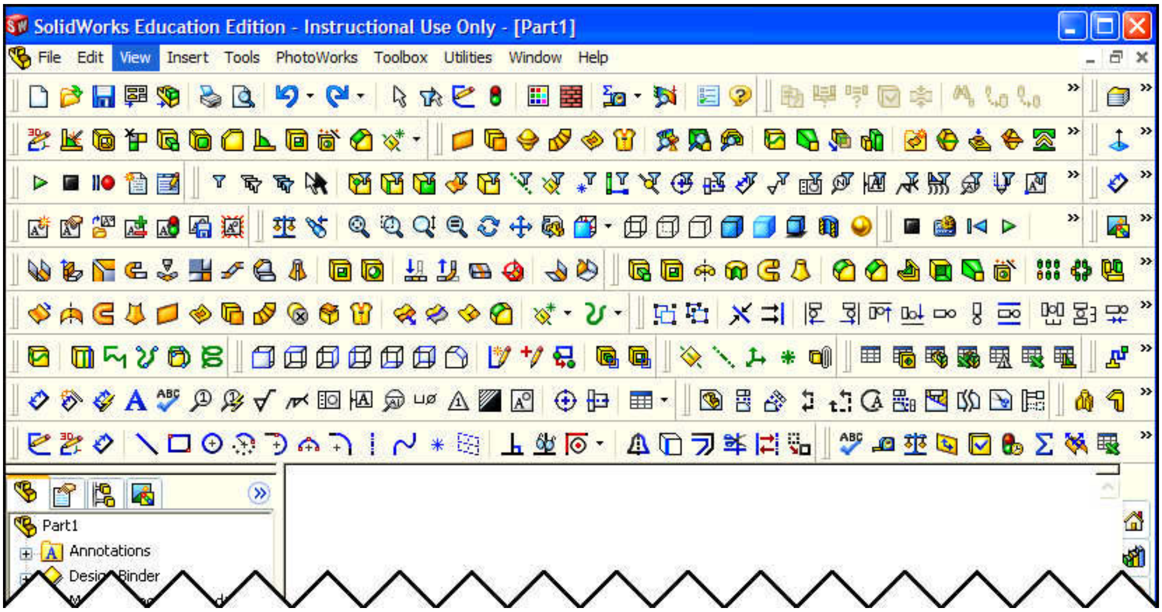

Fig. 1: Parts of toolbar buttons in SolidWorks 2006.

2.4 Challenges

Speech technology has gained considerable advancement in recent decades. IBM, Intel, Microsoft, Nuance and many other famous companies have devoted considerable resources to the R&D of speech recognitions and synthesis. In market, there are also many speech related products ready for use. However, there are still many challenges to integrate voice into CAD systems, some of the most important ones are briefly listed below:

- **Accuracy and speed**: although existing software toolkits claimed to have high speech recognition accuracies and speeds, however the results are actually far from satisfactions [3, 8]. According to the investigations reported in [4], current states of art for generating and interpreting speech are "still a travesty of what a young child can achieve with ease". Moreover, when speech recognition from a large vocabulary is undertaken or complex grammars are involved, the performance will be even worse.
- **Robustness**: most voice assisted schemes are rather sensitive to speaker's voices, tones, pitches or rhythms, resulting in serious robust problems. To illustrate, Tab. 1 lists some misrecognized words from human and computer synthesized voice input. These results are recognized based on a general vocabulary with 20, 000+ words commonly used in daily English.
- **Interaction unfriendliness**: due to the above limitations, speech recognition of voice input is error prone. In the case of misrecognitions or wrong operations, the available error correction mechanism is very limited and ineffective, which directly worsens the user satisfactions.
- **Lack of professional knowledge in action interpretation**: the off-the-shelf, general-purpose speech recognition toolkits are mostly targeted to deal with textual or literal words, for instance, Microsoft Words translates the recognized words to texts [5], Naturally Speaking Medical converts the voice input to medical prescriptions[6]. These toolkits can hardly understand the technical semantics extensively used in CAD communities. They are not fully competent for CAD jargon interpretation and the subsequent mapping of words to the non-textual modeling actions.

Human voice	Recognition result
Circle	So cool
3D sketch	Three key sketch
Plane	Lane
Extruded Boss	Extruded balls
Mirror Feature	Near future

(a)

Computer synthesized voice	Recognition result
Edit Component	Added component
Mate	Rate
Exploded View	Exploded the you
Join	Joint
Change Transparency	Chains transparency

(b)

Tab. 1: Misrecognized words from human and computer synthesized voice.

Motivated to tackle the above challenges and to realize the goal of "design by talking with computers", in this paper, some solutions are presented and the proposed methodologies are detailed as follows.

3. DESIGN BY TALKING WITH COMPUTERS: THE PROPOSED SOLUTIONS

As is discussed in Section 2, one of the key challenges in integrating human voice into CAD systems is to increase the speech recognition accuracy and speed. Our preliminary study shows that the general-purpose vocabulary and grammar for the purpose of word dictation do not adequately fulfill such requirements. This is mostly due to the large number of candidate words to be matched which accounts for the speed issues, and the many similarly pronounced words, which accounts for the accuracy penalties.

3.1 CAD Targeted Lexicon and Grammar

A natural idea to improve the speech recognition accuracy is to limit the candidate words to be recognized within strictly confined scopes so that only those words of interest will be considered in the recognition. Since the objective is to integrate voice into CAD systems, a CAD targeted lexicon can be constructed for such purposes. This helps decimate a large number of irrelevant words for general (e.g. dictation) purposes and increase the robustness of speech recognition.

A typical CAD targeted lexicon may includes extensively used CAD jargons or keywords, for instance V_{CAD}= […, *assembly*, *chamfer*, *circle*, *continuity*, *dimension*, *draft*, *extrude*, *feature*, *from*, *left*, *line*, *mate*, *normal*, *pan*, *part*, *rotate*, *select*, *shade*, *sketch*, *stretch*, *view*, *wireframe*, *zoom*, …]. Take for the recognition of the word "Circle" in Tab. 1 (A) as an example, since the expression "so cool" is not a CAD related term; therefore it is excluded from the CAD targeted vocabularies. This naturally eliminated the chance of misrecognition, as the expression “So cool” in Tab. 1 (A) is out of the scope of the vocabularies.

The idea of CAD targeted grammar is a similar but more effective approach to enhance the recognition accuracies. It does not only limit the candidate words, but also constrains the patterns/rules that will be “listened” or “processed” in speech recognition. A CAD targeted grammar G can be defined as a collection of rules or patterns:

$$G_{CAD} = \{R_1, R_2, R_i, \cdots R_m\} \quad (3.1)$$

where *m* is the number of rules in the grammar. Each rule R_i is a collection of predefined expressions and the sequence of the words follows certain patterns:

$$R_i = \{[W_1] \rightarrow [W_2] \rightarrow \cdots W_j \cdots \rightarrow [W_k] \cdots \rightarrow \cdots [W_n] \big| W_l \in V_{CAD}, 1 \leq l \leq n\} \quad (3.2)$$

where [X] indicates that the word X is optional, “ $\rightarrow$ ” indicates the order/sequence of the words; *n* denotes the number of words (including the optional ones) in the expression, and W_l ($1 \leq l \leq n$) is an element (word) in the CAD targeted vocabulary.

Fig. 2 illustrates a simple CAD targeted grammar which consists of a single rule *R*. The valid patterns regulated by *R* include such word sequences as “View from left”, “View left” and “View isometric” etc., and only such word combinations will be processed by the speech recognition engines.

As is seen CAD targeted grammar imposes stiffer constraints and the benefit is that highly accurate speech recognitions can be obtained. To verify this, we have conducted several speech recognition experiments using both the general grammar (containing commonly used 20,000+ words) and a simple CAD targeted grammar (containing hundreds of CAD jargons), and the detailed results are presented in Section 4.

3.2 Context-aware Inference

CAD targeted lexicon and grammar can increase the speech recognition accuracies by filtering irrelevant words, however if the words in the targeted vocabulary have strong vocal similarities, it is still difficult for computers to recognize. This problem is especially significant when background noise such as hiss, hum (e.g. from air conditioners), clicks and pops (e.g. from wireless microphones) are existent. Actually, the same phenomenon also happens in human-human communications when one is not sure about what he has heard. Under such circumstances, one may say “Pardon?” to confirm the listened words. Similarly, computer may prompt users to clarify the word if it is not certain about the possible words under recognition. This can be easily accomplished by the Text-To-Speech (TTS) technology [23, 24], by which computer synthesizes human voices to interact with designers.

However, frequently saying ‘Pardon?’ may be frustrating and degrade user satisfactions. Most experienced adults, however, solve this problem by conjecturing other people’s words from the context; their knowledge or past experiences might be utilized in inferring the words and semantics. In addition, one may frequently encounter such situations: even if you have not finished the entire sentence, the one you talk to may already have caught what you wish to convey precisely. Why and how do this work? The answer is: **context**. Experienced people can intelligently infer the words/meanings of the conversation from the topics and the scenarios.

Enlightened by this mechanism, we propose to introduce a novel type of Context-Aware Inference (CAI) to improve the ease of use and the speech recognition accuracy. The idea is to simulate the human’s inference by means of context awareness. We monitor the CAD modeling contexts such as object *selection*, file open/close status or the entity’s dimensionality etc., and we take advantage of such knowledge to infer/guess the possible words to be spoken or recognized. To make this more understandable, let us have a look at the following example.

Consider when a user wishes to stretch a circle to alter its radius. He may utter ‘Stretch circle’ to the voice–aided CAD system. Because of the pronunciation similarity of ‘stretch’ and ‘sketch’ and the background noise, it is very likely that the computer takes it as ‘sketch a circle’. To tell which word best matches the user’s design intent, the computer may check the design context and make the decision accordingly, as conceptually demonstrated in Fig. 3.

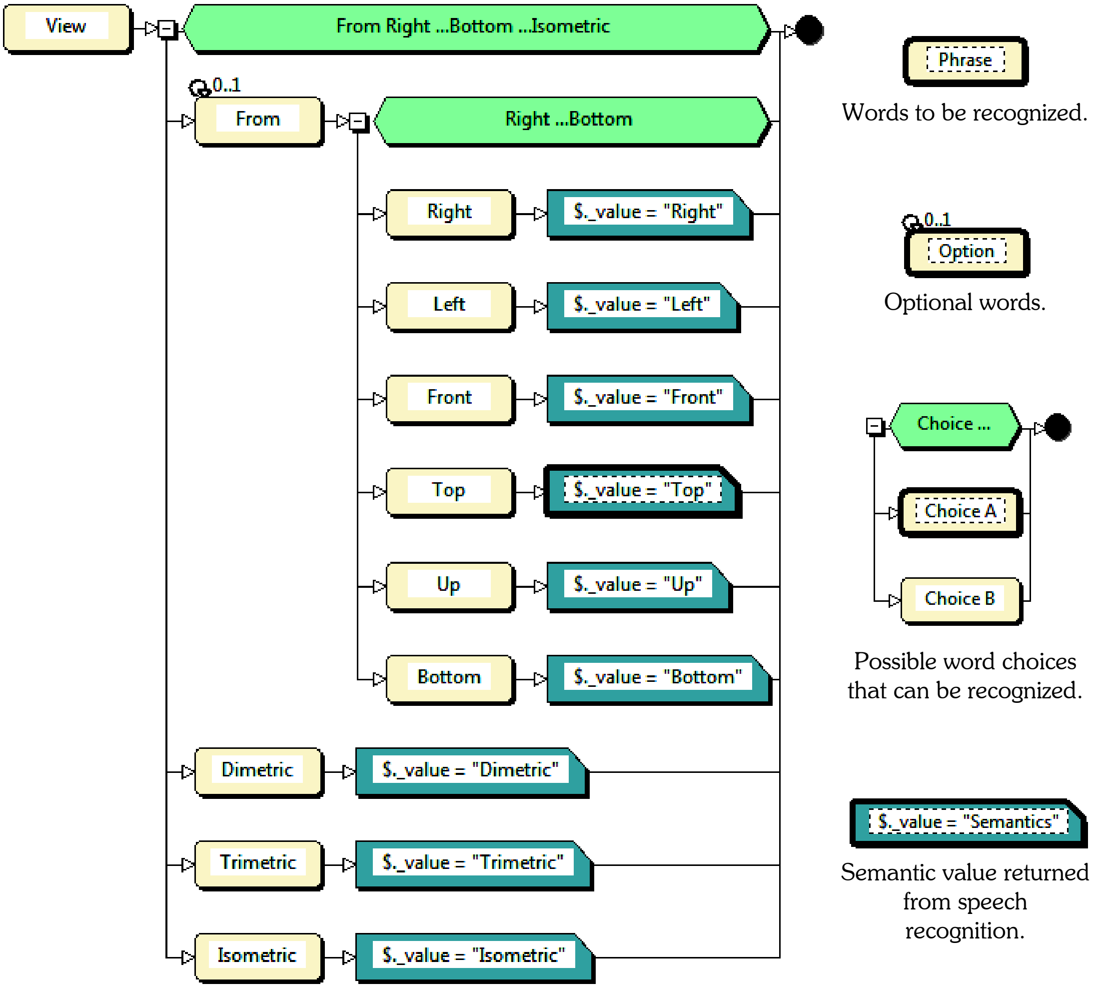

Fig. 2: An example of CAD targeted grammar composed of a single rule.

In this example, when the computer fails to distinguish “sketch” from “stretch”, it first checks if there is any entities available in the active CAD model: if there is no entities, it means there is nothing to stretch; therefore the word “stretch” can be excluded and the result is output as “**sketch**”; otherwise, the computer checks if there is a reference plane selected, if yes, it is very likely that the designer wishes to create a **sketch** and draw a circle thereon. On the contrary if there are stretchable entities (e.g. circles, lines or spline curves) currently selected, then it is most likely that the designer will edit them via **stretching** operations. Finally, if no explicit judgment can be obtained from all of the above interrogations, then as the last resort, the computer will ask the user: “Pardon?” or “Do you mean …” to assure correct word recognitions.

In this paper, a general algorithm is proposed to implement the context-aware inference (CAI), and the pseudo code is illustrated in Fig. 4. In the algorithm, the context-aware inference will be performed only if there is more than one candidate in the recognized word collection, as is seen in Fig. 4 (L2). According to the probability or the recognition confidence, a quick sort (L4, Fig. 4) is then executed to arrange the hypothetical words in a descending order. The purpose of such quick sort is to avoid the brute-force data traversal: the most likely word is first processed and once it is successfully validated through the CAI rules, the reminder of words will no longer be processed.

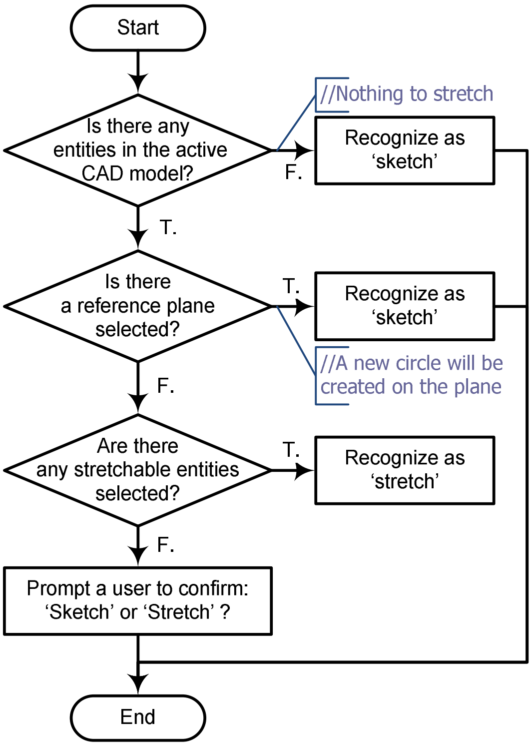

Fig. 3: A Context-Aware Inference (CAI) example.

```
L1    String PerfomCAI ( StringArray CandidateCollection )
L2    if( CandidateCollection.Count >= 2 )
L3    {
L4            QuickSortByConfidence( CandidateCollection );
L5            foreach( WordCandidate wCurrent in CandidateCollection )
L6            {
L7                    WordCandiate wNext=wCurrent.Next;
L8                    if( wNext != null && abs( wCurrent.Confidence - wNext.Confidence ) <= Threshold )
L9                    {
L10                           bool bSuccess=Search_CAI_Rule_Library( wCurrent, wNext, InferredResult );
L11                           if ( bSuccess == true )
L12                                   return InferredResult;
L13                   }
L14                   else if ( wCurrent.Confidence >= MinimumConfidence && wCrurrent.Previous==null )
L15                           return wCurrent;
L16           }
L17           PromptViaTTS("Pardon? Do you mean ... or … ? ");
L18           return null;
L19   }
```

Fig. 4: The pseudo code of the Context-Aware Inference (CAI) algorithm.

For each word in the candidate collections, we first check if the difference of recognition confidence of adjacent candidates is below the preset threshold (L8, Fig. 4), and if the answer is "Yes", it indicates that two candidates are very similar in pronunciation, then a context-aware inference will be conducted by searching within the CAI rule library (L10 - L12, Fig. 4). The CAI rule library contains a collection of rules and each of which is similarly represented as shown in Fig. 3. Otherwise, if the current word's recognition confidence is far higher than the next word (L14, Fig. 4),

then it is relatively safe to output the current word as the result; in this case, the first word in the sorted collection may be returned, however, we also need to make sure that the recognition confidence of the first word is above the preset minimum confidence threshold (L14, Fig. 4). Finally, similar to the case in the previous example, if the context-aware inference fails to provide definite answers, the computer will interact with the user with TTS prompts (L17, Fig. 4).
As is seen from the above algorithm, context-aware inference helps intelligently guess the candidate spoken words and increase the success rate of speech recognitions. This intelligence is especially beneficial to increase the user satisfactions.

4. IMPLEMENTATIUONS AND RESULTS

To verify the feasibility of the proposed scheme and test its performance, we have developed a prototype voice-aided CAD system and named it as "NaturalCAD". NaturalCAD is developed with VisualStudio.Net in Windows Vista, on a PC with 2.8GHz CPU and 2 GB memory. The commercial CAD system Solidworks is used to conduct CAD modeling operations.

The components of NaturalCAD are illustrated in Fig. 5. A USB microphone is used to capture the input voice signal. The analog signal is digitized by the sound card and further transferred to the computer memories. The digital data is then passed to the Speech Recognition Engine (managed SAPI in DotNet Framework 3.5).
The proposed CAD targeted grammar is used to increase the speech recognition speed and accuracies whenever possible. To further distinguish similarly pronounced words, context based inference is used to intelligently conjecture the candidate spoken words. In case the speech recognition engine fails to understand the meaning of the human's voice input, then the general dictation grammar will be used. The recognized words are then mapped to voice commands (e.g. sketch a circle) and are then fed to the Solidworks Addin module. Solidworks conducts corresponding CAD modeling operations and the generated CAD model is then rendered for visual feedbacks. Voice feedbacks generated with TTS engines can be also sent to notify the operation success or prompt the user about the reasons of failure.

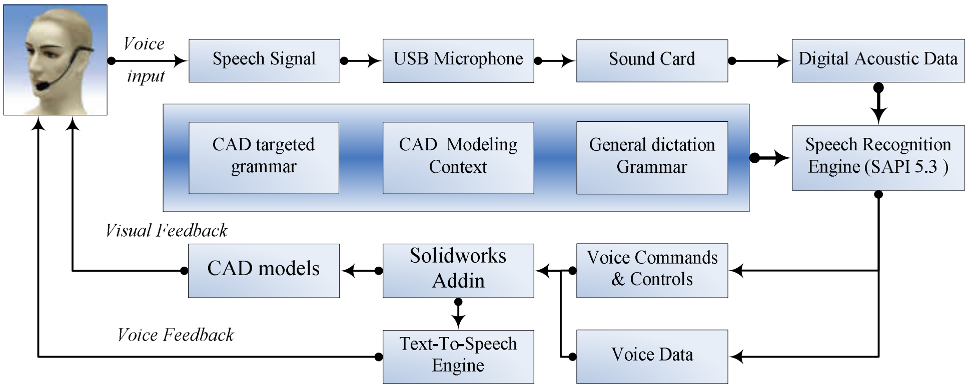

Fig. 5: System components of the proposed Voice-aided CAD system.

To test the performance of the CAD targeted grammar, we have conducted several speech recognition experiments using both the general dictation grammar (by Microsoft) and simple CAD targeted grammars (by ourselves). To test the efficacy in the noise-free environment, we used the text-to-speech engine (Microsoft Anna) to synthesize the input voice signal. Some commonly used CAD commands (e.g. the words in Fig. 2) are spoken by TTS engines, with 2 seconds breaks in between. The general dictation grammar and the proposed CAD targeted grammar are then used in speech recognitions. We tried different TTS engines in the above experiment and the recognition success rates are listed in Tab. 2. In this experiment, the TTS engines A_i (i=1,2,3), B_j (j=1,2), C_k (k=1,2) are products from the same vendor. As can be clearly seen from the results, the performance of CAD targeted grammars is much better than the general dictation grammar.

Using the proposed methods, a trial of "design by talking with the computer" is carried out. A demonstrated conversation is shown in Fig. 6. The user starts by fooling the computer: "Good morning" and in fact the conversation took place in the afternoon. NaturalCAD checks the computer system time and tells the user it should be "afternoon". After the user corrected this, the voice-aided modeling begins. As there is nothing selected in the active document, therefore when the user requests the computer to stretch something, NaturalCAD notified the user with synthesized voice: "*There is nothing to stretch now. Select the last created entity?*" When the user accepted the suggestion, a stretching operation began, and now when the mouse moves, the circle's radius changes accordingly. After completing this simple operation, the user terminated the conversation by saying "See you", and NaturalCAD politely replied, "Bye Bye."

Input TTS Engine	A_1	A_2	A_3	B_1	B_2	C_1	C_2
General Grammar	15.4%	7.7%	7.7%	30.7%	61.5%	53.8%	53.8%
CAD Targeted Grammar	92.3%	84.6%	84.6%	100%	100%	100%	100%

Tab. 2: Speech recognition success rate using different TTS engine and different grammars.

User:	**Good morning.**
NaturalCAD:	*GOOG MORNING? Haha, you should say good afternoon.*
User:	**Good afternoon.**
NaturalCAD:	*Good afternoon. What can I do for you?*
User:	**Stretch. (now no circle is selected)**
NaturalCAD:	*There is nothing to stretch now. Select the last created entity?*
User:	**Yes, please do that.**
NaturalCAD:	*Roger,done.*
User:	**See you.**
NaturalCAD:	*Bye Bye.*

Fig. 6: A conversation between a user and the computer.

Fig. 7 illustrates a more complex CAD model designed with the proposed approach. To demonstrate the efficacy, we have intentionally disabled **all** the toolbar commands and the model is constructed using only mouse and voice inputs. For demonstration purposes, the words spoken by the designer (David Kou, the 1st author) and the computer synthesized feedbacks are also rendered in text output in the "History Log", as shown in Fig. 7.

The detailed modeling process is illustrated in Fig. 8 and Fig. 9. Fig. 8 shows the history log of human-computer interactions as well as the corresponding visual output of the modeling result. To be concise, the remainder voice commands are presented in textual format only, see Fig. 9. For those who are interested to get more details, they can watch the flash animation of the entire process at the first author's web webpage http://web.hku.hk/~kouxy/VoiceCAD/Animations/demo1/demo1.html.

5. CONCLUSIONS AND DISCUSSIONS

This paper describes our preliminary research on a voice-aided CAD system, where human voice is utilized to assist CAD modeling tasks. The objective of this voice-aided CAD system is to allow CAD software users to design by talking with computers. The key technologies that enable this functionality are speech recognition and computer speech synthesis.

In order to increase the success rate of speech recognition without sacrifices of speed, the paper employs novel CAD-targeted vocabulary as well as CAD-targeted grammars, which greatly enhances the practical applicability of the voice-aided CAD system. A novel type of Context-Aware Inference (CAI) is also proposed to simulate human's intelligent inference by means of context awareness. CAD modeling contexts such as object *selection* status or the entity's dimensionality etc. is harnessed to infer possible words from ambiguous candidates. The pseudo code of the Context-Aware Inference (CAI) algorithm is presented.

The proposed voice-aided CAD system is easy to use and efficient in carrying out predefined CAD modeling commands such as 2D/3D sketching and feature construction etc. Newbie users do not need to have much familiarization with the graphical user interfaces of the CAD system and can enjoy the ease of use; professional CAD designers can take the extreme of the functionalities by directly uttering the most relevant voice commands, therefore benefit from productivity improvements.

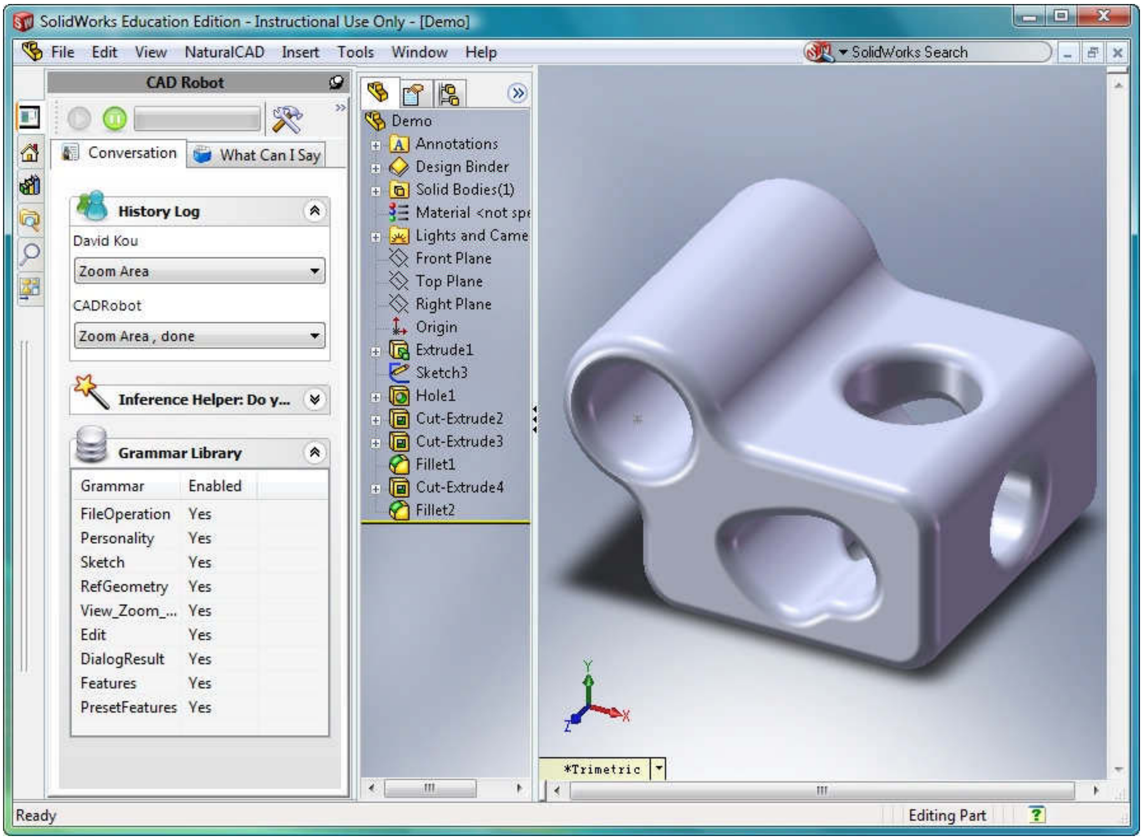

Fig. 7: A design example by talking with computer.

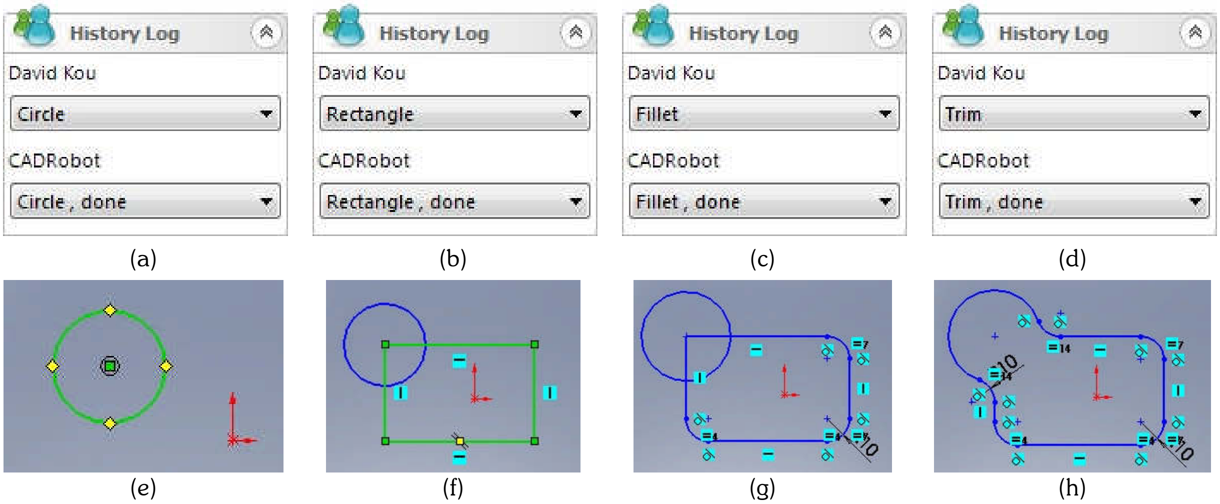

Fig. 8: The history log of human-computer voice interactions and the corresponding results.

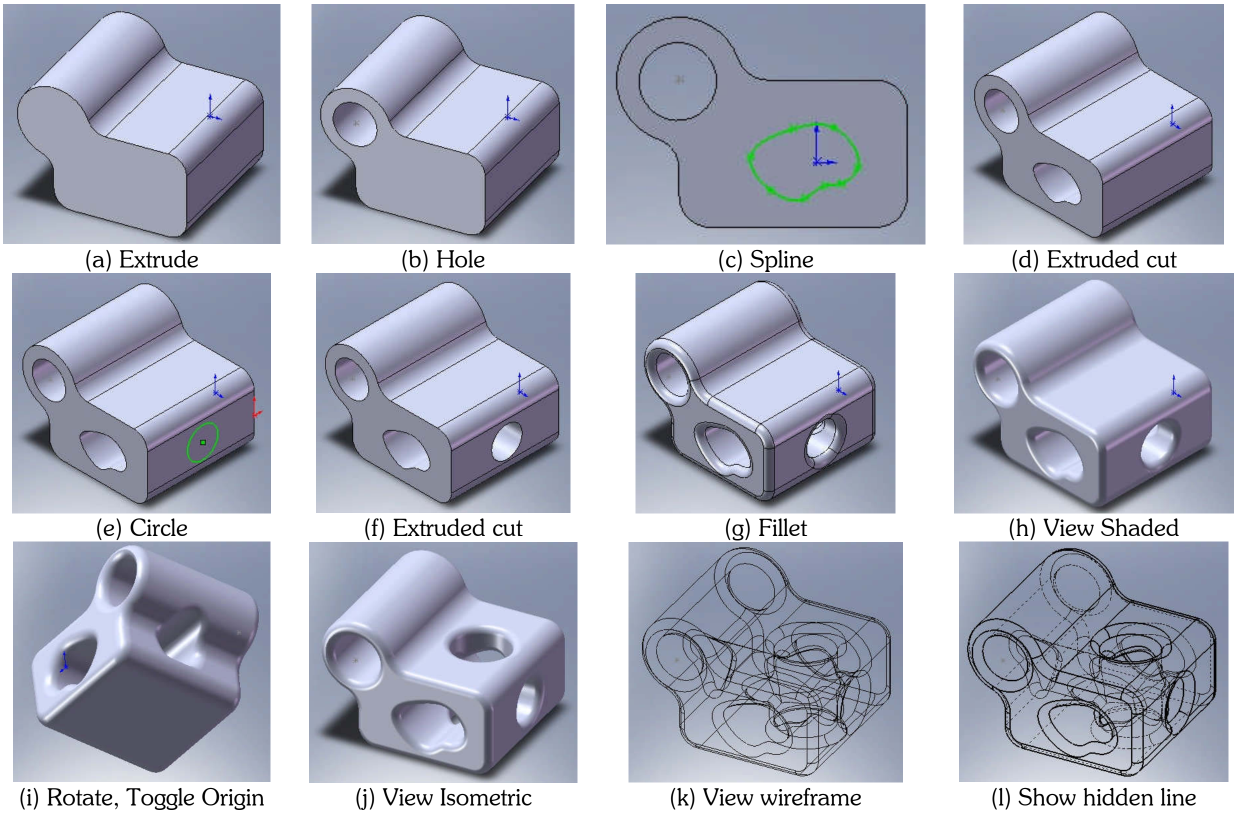

Fig. 9: A design example by talking with computer.

The proposed voice-aided CAD system is easy to use and efficient in carrying out predefined CAD modeling commands such as 2D/3D sketching and feature construction etc. Newbie users do not need to have much familiarization with the graphical user interfaces of the CAD system and can enjoy the ease of use; professional CAD designers can take the extreme of the functionalities by directly uttering the most relevant voice commands, therefore benefit from productivity improvements.

Despite the many merits of this voice-aided CAD system, our research is still in its pilot stage. It could be enhanced in many aspects in terms of the design powerfulness, user friendlessness and algorithmic elegance. In its current status, we are using simple generic data structures to hold the candidate lexicon items as well as the grammar rules, and they are best to be archived into databases, so that easier data isolation and maintenance can be realized. The current vocabularies and the grammar rules are also manually selected from the most widely used CAD jargons; and in our future work, a more systematic and automated lexicon/grammar construction scheme will be investigated. We hope to automate this process by a detailed analysis from the help manuals of CAD software packages. This will help to seamlessly and precisely hook specific voice commands with an existing CAD modeler.

6. ACKNOWLEDGEMENT

This work is the formative parts of a project for the Hong Kong CERG grant. The authors would like to thank the Department of Mechanical Engineering and the Committee on Research and Conference Grants of the University of Hong Kong for supporting this project (Project No. 200707176061).

7. REFERENCES

[1] I, Robot, http://www.imdb.com/title/tt0343818/.
[2] Becchetti, C.: Speech recognition: theory and C++ implementation, Chichester, Wiley, 1998.

[3] Schroeder, M. R.: Computer speech: recognition, compression, synthesis; with introductions to hearing and signal analysis and a glossary of speech and computer terms, 2nd ed. Berlin, New York, Springer, 2004.
[4] Holmes, J. N.: Speech synthesis and recognition, 2nd ed. London, New York, Taylor & Francis, 2001.
[5] Microsoft Word, http://www.microsoft.com/word/.
[6] Dragon NaturallySpeaking Medical, http://www.nuance.com/naturallyspeaking/medical/.
[7] Michael, A. G.; David, S. E.; Timothy, W. F.: The integrality of speech in multimodal interfaces, 5, ACM Press, 1998.
[8] Peacocke, R. D.; Graf D. H.: An Introduction to Speech and Speaker Recognition, Computer, 23(8), 1990, 26-33.
[9] Zue, V.: The Oxygen Project: Talking with your computer, Scientific American, 281, 1999, 56-57.
[10] Microsoft, http://www.microsoft.com/en/us/default.aspx.
[11] Guy, C.; Jurka, M.; Stanek, S.; Fateman, R.: Math Speak & Write, National Science Foundation (NSF) Research project (CCCR-9901933) of USA, available at http://www.cs.berkeley.edu/~fateman/msw/msw.html, 2004.
[12] Reddy, P. D. V. G.; Kitamura, R.; Jovanis P. P.: Voice Operated Information System (VOIS) for driver's route guidance, Mathematical and Computer Modeling, 22(4-7), 1995, 269-278.
[13] Microsoft Math, http://www.microsoft.com/math/productDetails.aspx?pid=001&active_tab=Features#InkHandwritingSupport.
[14] Patel, S.; Abowd, G.: BLUI: Low-cost Localized Blowable User Interfaces, Proceedings from UIST '07: ACM Symposium on User interface Software and Technology, 2007, 217-220.
[15] Liu, X.; Dodds, G.; McCartney, J.; Hinds, B. K.: Virtual DesignWorks - Designing 3D CAD models via haptic interaction, Computer Aided Design, 36(12), 2004, 1129-1140.
[16] Liu, X.; Dodds, G.; McCartney, J.; Hinds, B. K.: Manipulation of CAD surface models with haptics based on shape control functions, Computer Aided Design, 37(14), 2005, 1447-1458.
[17] Zhu, W.; Lee, Y. S.: Five-axis pencil-cut planning and virtual prototyping with 5-DOF haptic interface, Computer Aided Design, 2004, 36(13), 1295-1307.
[18] Zou, H. L.; Lee, Y. T.: Skewed rotational symmetry detection from a 2D line drawing of a 3D polyhedral object, Computer-Aided Design, 38(12), 2006, 1224-1232.
[19] Zou, H. L.; Lee, Y. T.: Constraint-based beautification and dimensioning of 3D polyhedral models reconstructed from 2D sketches, Computer-Aided Design, 39(11), 1025-1036, 2007.
[20] Diniz N.; Branco, C.: An approach to 3D digital design free hand form generation, in Proceedings of the International Conference on Information Visualization, London, 2004, 239-244.
[21] SolidWorks, http://www.solidworks.com/.
[22] Keller, K.; Corbett, J.; Nichols, D.: Repetitive strain injury in computer keyboard users: Pathomechanics and treatment principles in individual and group intervention, Journal of Hand Therapy, 11(1), 1998, 9-26.
[23] Multilingual text-to-speech synthesis: the Bell Labs approach, Dordrecht, Boston, Kluwer, 1998.
[24] Narayanan, S.; Alwan, A.: Text to speech synthesis: new paradigms and advances, Prentice Hall Professional Technical Reference, 2005.