ISCA Archive
http://www.isca-speech.org/archive

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

NATURAL LANGUAGE PROCESSING FOR TAIWANESE SIGN LANGUAGE TO SPEECH CONVERSION

Chung-Hsien Wu, Yu-Hsien Chiu, and Chi-Shiang Guo

Department of Computer Science and Information Engineering
Cheng Kung University, Tainan, Taiwan

ABSTRACT

This paper addresses some natural language processing technologies in generating sentences from ill-formed Taiwanese Sign Language (TSL) for people with speech or hearing impairments. The design and development of the PC-based TSL Augmentative and Alternative Communication (AAC) system aims to improve the input rate and accuracy of communication aids. We demonstrate 1) developing an effective TSL virtual keyboard to assist users to efficiently input sign sequences, 2) investigating TSL prediction strategies for input rate enhancement, 3) integrating bottom-up parsing with top-down filtering strategy and variable N-gram language model for sentence generation and selection. The proposed system assists people with language defects in sentence formation and grammatical correction. To evaluate the performance of our approach, a pilot study for clinical evaluation and education training was undertaken. Evaluation results show that the generation rate and subjective satisfactory level for sentence construction have been significantly improved.

1. INTRODUCTION

People with communication disabilities range widely from cerebral palsy, stroke, aphasia, aging, development delay, articulation disorder and traumatic brain injury [1]. Since 1970, developed countries have designed extensively Augmentative and Alternative Communication (AAC) technology to provide language learning and speech correction aids for improving expressive communication abilities [2,3]. An AAC system intelligently provides syntactically and semantically correct sentences will greatly improve their communication in daily activities and independence [4,5,6]. Assistive communication technology and the corresponding AAC devices have been developed recently in developed countries for twenty years. The underlying principle is to facilitate the disabled with minimum efforts of alternative manual-sign input to generate comprehensive information output and greatly improve expressive communication abilities [7,8]. The implementation has been proved successfully in special education. Unfortunately the developed AAC technology cannot be directly applied to the people using Chinese language. In addition, the usage and characteristics of sign language varies with the time and the place of native circumstances.

Current research largely focused on the goals of developing more efficient user interface and extending the accessibility to people with communication impairments [9,10]. This has been largely applied by augmenting the virtual keyboard (VK) with additional processing [11]. The purpose of this paper is to investigate the methods for improving the input rate and accuracy of communication aid for Taiwanese Sign Language (TSL) AAC system. More specifically, the study focuses on: 1) developing an effective TSL virtual keyboard, 2) investigating TSL prediction strategies for input rate enhancement, 3) integrating bottom-up parsing with top-down filtering strategy and variable N-grams language model for sentence generation and selection. The significance of this paper is to provide people with communication problems an AAC system to become independent and therefore reduce the burden of family and the cost of the society.

2. SYSTEM DESIGN AND DEVELOPMENT

An AAC assistive device generally can be divided into three basic components, which are interface, natural language processing and output units. Figure 1 shows the structure of the proposed system.

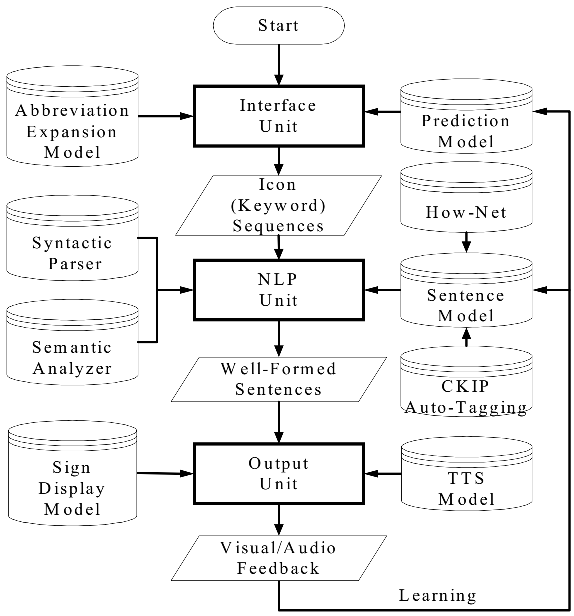

Figure 1 The structure of the system

2.1 Virtual Keyboard (VK)

The TSL-VK can be modeled with a language set (LS) and a user interface, which is functionally defined by the mapping the selected input to a sentence output. The interface unit is mainly used for human and machine interface. The users can input the sign sequences via the keyboard, mouse or touch screen. The choice of the interface technology will depend on individual physical capability. The LS contains the elements that can be selected by the user for expressing the idea, which can be phonetic symbols, Chinese characters, words, phrases, icons, pictures and so on. The LS is a structured collection of TSL signs corresponding to linguistic units. It is hierarchically categorized based on syntactic attributes and statistical n-grams information for the development of a large TSL vocabulary. The commonly used nouns are divided into many categories; each contains many most frequently used words. According to keystroke saving concept, the row-column scanning strategy is applied for the arrangement of the VK. In VK, the frequently used symbols are arranged on the upper-left corner to reduce the steps of scanning and makes system more efficiently accessible to the disabled by providing whole-word or phrase access via pictorial representations. The implementation of Taiwanese signs based VK is a highly individualized process. Therefore, a rehabilitation team must establish a relevant TSL vocabulary for the user and organize it in a way that is both efficient and relatively easy to learn and retain over time consideration.

2.2 TSL Prediction Strategies

For input rate enhancement, the TSL-VK has been endowed with the predictive capacities. The LS is designed as a dynamic tree structure in accordance with a sentence-template based grammar tree and a variable n-gram language model. In this tree-based data structure, the internal node, defined as keyword/syntactic slot, records the Part-of-Speech (POS) features, the frequency of reference word/symbol and bi-gram information. The external node records the frequency of the reference sentence templates, internal nodes, nouns, verbs, function words and n-gram information. The proposed sentence template-based grammar tree is constructed by using breadth-first search algorithm to match slot definitions.

The TSL icon sequences on the display are dynamically updated based on previous input and constrained with syntactic characteristics in Chinese grammar. The task of predicting the next keyword can be stated as attempting to estimate the probability function P:

$$P(POS_n \mid POS_{1\wedge} POS_{n-1}) = \frac{P(POS_{1\wedge} POS_n)}{P(POS_{1\wedge} POS_{n-1})} \quad \ldots(1)$$

where POS represents the part-of-speech feature. Figure 2 shows an example of the variable n-gram language model. Using a pre-processed TSL corpus and linguistic relationship, the keywords are grouped into semantic classes and used to reduce the value of n and the combination explosion problems for rule-based systems trained from large daily dialogue corpus.

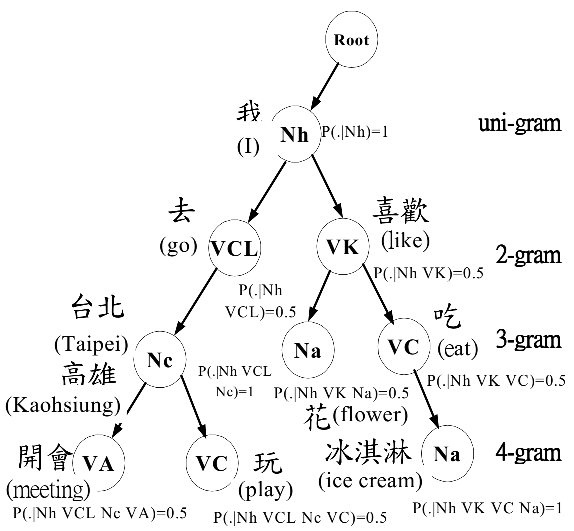

Fig 2 An example using variable N-gram prediction

2.3 Abbreviation Expansion

The development of abbreviation expansion method is to limit the number of selected inputs necessary to access a particular symbol while maintaining a large TSL bank. A Chinese Augmented Transition Network (ATN) is constructed based on Chinese phonetic characteristics. The Chinese phonetic abbreviation expansion method is proposed to fast search TSL icon from large vocabulary coverage with relatively few keystrokes using the ATN. This search mechanism can reduce the scanning number by using the ATN scheme shown in Figure 3, similar to English acronym. This scheme can be word-based or sentence-based.

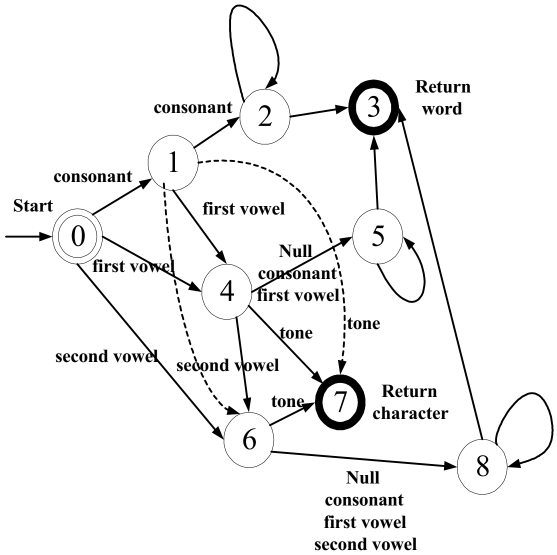

Fig. 3 ATN for abbreviation expansion and search modes

2.3 TSL to Sentence Translation

The Chinese sentences formed by people with hearing impairment are almost ill formed, such as wrong sentence types, the lack of adjectives, input disorder and redundant word problems. For accuracy enhancement, the bottom-up parsing with top-down filtering strategies and sentence template-based case-frame grammar are proposed to generate well-formed syntactic sentences from ill-formed TSL icon sequence inputs.

2.3.1 Bottom-Up Parsing Stage

Input keywords are parsed based on syntactic categories and partial-parsing rules (PPR). The PPR is categorized with syntactic properties of objects, time and place. Maintaining the various sets of keywords in a hierarchical structure allows users to reduce the search space and handle syntactically ill-formed sentences using the expansive mechanism of PPRs. In this stage, the PPRs are used to merge possible phrases and concatenate the adjectives and quantifiers into a phrase segment.

2.3.2 The Top-Down filtering stage

Phrase segments are filled into possible syntactic slots, except for functional words, time and place phrases. Then the keyword-matching criteria and PPR constraints are applied to retrieve candidate grammar paths. The preliminary matching criteria are defined as follows: 1) The number of verbs in the template is equal to the number of verbs in the phrase segment;

2). The number of nouns is less or equal to the number of nouns in the phrase segment; 3). The number of function words in the phrase segment should be less than 2. Figure 4 shows an example for matching phrase segment in the sentence templates.

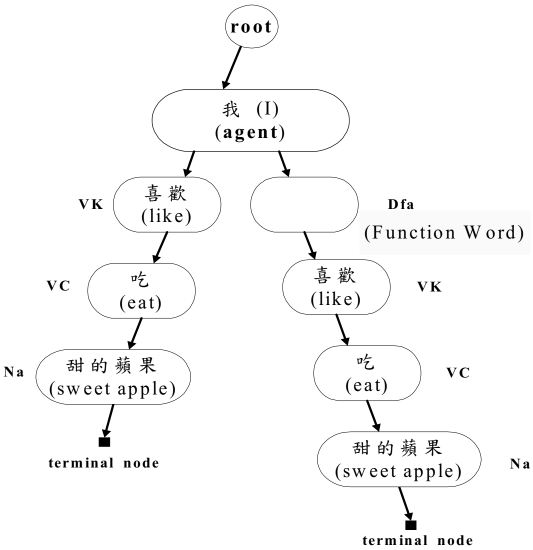

Figure 4 An example for illustrating phrase segment matching.

2.3.3 Semantic Parser and Sentence Generation

In a semantically unconstrained domain, a case-frame semantic parser is constructed based on the How-Net, a Chinese knowledge base similar to WordNet in English. The semantic parser is used to identify the agent and filter out unsuitable sentences by inferring the semantic role and relation in a phrase segment. This semantic parser also plays a useful role to improve word prediction. This allows phrase segments to mutually constrain each other and further reduce ambiguity and perplexity. Then the time and place are put into the final sentence templates. And the variable n-gram method is used to patch up function or redundant words and ranks candidate sentences.

2.4 Automatic Learning Mechanism

In addition, a statistical learning mechanism, for rate enhancement, is utilized to dynamically record personalized and frequently used words, phrases, and sentence patterns in the template-based tree. This can be used to reduce scanning numbers and remove infrequently used symbols in order to improve flexibility and search efficiency. Finally, candidate sentences, ranked by scoring variable n-gram information, are selected and used to generate speech output via a Chinese text-to-speech conversion system.

3. RESULTS AND DISCUSSION

In order to evaluate the performance of our approach, a pilot study for clinical evaluation and education training was undertaken. 1170 most often used TSL icons and their corresponding meanings were collected by a special educator. These icons were used to investigate the effect of message expressing communication for 20 students with speech-hearing impairments. 1000 Chinese sentences, in which the average length of sentences is 4.9 words, were selected to develop 553 sentence templates in grammar tree and utilized as the training and testing databases.

3.1 Evaluation of System Performance

In the practice evaluation, the usefulness of the proposed system is dependent on many factors, ranging from the user's cognition abilities to the interface design. Figure 5 shows this intelligent TSL to speech conversion system.

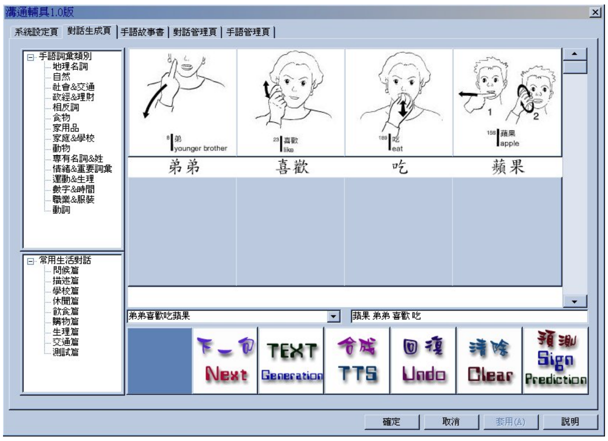

Figure 5 The TSL to speech conversion system

An evaluation of the potential effectiveness is developed based on an analysis of training corpus to estimate the selection saving over a comparable keyword-based system. An additional goal of the bottom-up parsing with top-down filtering strategies is to relieve the user from the burden of inputting function words such as the word categories of prepositions and stopping words in Chinese. Table 1 shows a summary of this analysis, a 26.25% keystroke (KS) saving rate for sentences with functional word deletion was achieved.

Table 1 Keystroke saving rate for functional word deletion

	KS numbers	KS saving rate
Whole sentence	2777	baseline
Keyword-based with functional word deletion	2048	26.25%

In this row-column scanning arrangement system, this estimation must consider several factors. The initial estimates are perhaps low for subjects because of effect in relation to the arrangement and scanning strategies. That is, the display mode has been tailored to match user cognition demand, such as the degree of attention and focus. For evaluating our proposed prediction strategies, the scanning rate enhancement using keyword prediction, sentence pattern prediction and Chinese phonetic abbreviation were 67.71%, 79.50%, and 96.87%, respectively. Table 2 shows the comparison of various prediction strategies in the evaluation of keyword selection.

Table 2 Comparison for scanning improvement rate with various prediction strategies.

	Average Scanning Steps	Improvement Rate
Without prediction	39.27	baseline
Word prediction	12.68	67.71%
Syntax prediction	8.05	79.50%
Abbreviation	1.23	96.87%

3.2 Case Study

For the assessment of practical communication aid, the special educator was asked to select suitable subjects meeting a predefined requirement in cognition and language capability. Through an integration of clinical and education evaluation, a single case with hearing-impairment was asked to conduct the experiments. After training, adaptation and evaluation phases, the sentence construction consistency from the subject compared to user transcription is 47.37%, 65.0% and 68.38%, respectively. The generation rate and subjective satisfactory level of sentence construction have been significantly improved. Table 3 shows the results of practice evaluation.

Table 3 Single case study for evaluating practice communication aids and system performance.

	Training Phase	Adaptation Phase	Evaluation Phase
Sentence Numbers	19	20	22
Average Keyword Numbers	6.1	5.5	4.2
Average Generation Time (sec/ sentence)	191.67	114.59	68.38
Successful Generation Numbers	9	13	15
Successful Generation Rate (%)	47.37	65.0	68.18
Subjective Satisfactory Level (5 levels)	3.6	4.2	4.4

For the accuracy of sentence generation, the iconic interface based system may be unable to generate exactly one sentence from the user's input. In this case, the system faces with multi-intention and complicated sentence type problems. An alternative approach is to display the original input sign sequence for selection. The subjective evaluation is used to judge whether the subjects understand their expressive intention or not.

4. CONCLUSIONS

The proposed system applies natural language processing technologies to the development of Chinese AAC system for Chinese people in need of communication aids. In addition to speeding up sentence generation rate for individuals, the proposed system has potential to assist people with language defects in sentence formation and grammatical correction. The HowNet-based semantic parser is also developed to handle unknown words by inferring the likely semantic role to construct a semantic representation for sentence generation. Evaluation results show that the generation rate and subjective satisfactory level for sentence construction have been significantly improved.

However, the current sentence generation model can only be used to generate simple sentence types. This is because the training corpus is a small-sized daily dialogue corpus. For the further development of language generation, our work will focus on analyzing the translation between TSL and Chinese and dealing with the complicated sentence structures, e.g. embedded clauses. According to the needs of rehabilitation technology and assistive devices, a computer-aided instruction (CAI) training program will be developed to improve the language learning and speech communication ability for communication impaired people. We will also aim to establish domestic AAC technologies to provide the users, clinicians and special educators with flexible and useful tools to design and use customized communication-aided devices. After the AAC system has been developed and fine turned, the system will be transferred to a palm-size platform and other multi-media applications.

5. ACKNOWLEDGEMENT

The authors would like to thank the National Science Council, Republic of China, for financial support of this work, under Contract No. NSC89-2614-H-006-003-F20.

6. REFERENCES

1. Reichle, Joe, "*Implementing Augmentative and Alternative Communication: strategies for learners with severe disabilities,*" Paul H.Bookes Publishing Co.,1992.
2. Webster, J.G., et. al., "*Electronic Devices for Rehabilitation*," John Wiley & Sons, Inc., 1985.
3. David, R. B. and Mirendan, Pat, "*Augmentative and Aternative Communication*," Paul H. Bookes Publishing Co., 1992.
4. Sara, Carmeli and Yechayahu, Shen, "Semantic Transparency and Translucency in Compound Blissymbols", *Augmentative and Alternative Communication*, Vol. 14, pp. 171-183, 1998.
5. Chang, S. K., et. al., " A Methodology for Iconic Language Design with Application to Augmentative Communication," *Preceedings of the 1992 IEEE Workshop on Visual Language*, pp.110-116, 1992.
6. Demasco, P., et. al., "Towards More Intelligent AAC Interfaces: The Use of Natural Language Processing," *The12th Annual Conference of RESNA*, pp. 27-30, 1989.
7. Demasco, P. and McCoy, K. F., "Generating Text from Compressed Input: An Intelligent Interface for Prople with Severe Motor Impairements," *Communication of the ACM*, Vol. 35, No. 5., pp.68-78, 1992.
8. Baker, B., "Minspeak," *Byte*, 186ff, 1982.
9. Simpson, R. C. and Koester, H. H., "Adaptive One-Switch Row-Column Scanning," *IEEE Transaction on Rehabilitation Engineering*, Vol. 7, No. 4, pp. 464-473, 1999.
10. Koester, H. H. and Levine, S. P., "Modeling the Speed of Text Entry with a Word Prediction Interface," *IEEE Transaction on Rehabilitation Engineering*, Vol. 2, No. 3., pp. 177-187, 1994.
11. Vanderheiden, G. C., "A High-Efficiency Flexible Keyboard Input Acceleration Techniques: SPEEDKEY," *Proceedings of the Second International Conference on Rehabilitation Engineering, RESNA*, pp. 353-354., 1984.