Intelligent Speech for Information Systems: Towards Biliteracy and Trilingualism

Helen M. Meng
Human-Computer Communications Laboratory
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong SAR, China
+852 2609 8327
hmmeng@se.cuhk.edu.hk

ABSTRACT
This paper reports on our research and development effort in human-computer spoken language interfaces, capable of processing English and Chinese, including two dialects for Chinese (Cantonese and Putonghua). This is the language environment in Hong Kong, and in order to develop human-computer spoken language interfaces that can be used by almost *anybody* in the region, we strive to develop speech and language technologies capable of handling biliteracy and trilingualism. The context of use is in accessing real-time information in the financial domain.

Keywords
Speech interfaces, spoken language systems, multilingual

INTRODUCTION

This paper reports on our research and development effort in human-computer spoken language interfaces, capable of processing English and Chinese, including two dialects for Chinese. This is the language environment in Hong Kong, where the official documents are written in English as well as Chinese; and the populace speaks Cantonese, Putonghua and English. Hence in order to develop human-computer spoken language interfaces which can be used by almost *anybody* in Hong Kong, we strive to develop speech and language technologies capable of handling biliteracy and trilingualism. The context of use is in accessing real-time information in the financial domain.

Spoken language systems have previously been developed to support mixed-initiative dialog interaction in a multitude of application domains, which characteristically have several task-specific user goals and constraints. Examples include air travel, railway information, restaurant guide (BeRP), ferry timetables (WAXHOLM), weather (JUPITER), electronic automobile classifieds, electronic assistants and tourist information. The languages concerned include English and a number of European languages. A few systems have also been developed for Mandarin Chinese. [1,2,3,4,5,6]

For our work, we have chosen the financial domain due to the abundance of real-time and dynamic information. Furthermore, the context of use is well-suited to Hong Kong, a financial capital in Asia. The financial domain provides many opportunities for information-critical applications. A spoken language interface that allows users to access financial information simply by asking questions should benefit both the computer-savvy and computer-naïve alike.

In the following, we will describe a foreign exchange inquiry system and its component technologies.

CU FOREX: A FOREIGN EXCHANGE INQUIRY SYSTEM

CU FOREX is a system which supports users' inquiries about foreign exchange, including the bid / ask exchange rates between two currencies, and deposit interest rates for a particular currency at various time durations (twenty four hours, one week, one month, two months... one year). Real-time financial information is retrieved from the Reuters satellite feed by a data capture process and stored in a relational database. Figure 1 illustrates the overall architecture of the system.

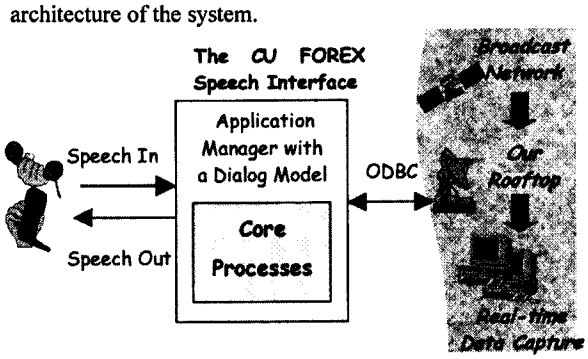

Figure 1: The CU FOREX System – overall architecture.

The core processes include:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CUU '00 Arlington VA USA
Copyright ACM 2000 1-58113-314-6/00/11...$5.00

- Bilingual speech recognition – currently the recognizer handles spoken Cantonese and English, but we are extending it to cover Putonghua as well.
- Natural language processing– this transforms the user's natural language query into a semantic frame (a meaning representation).
- Concatenative speech synthesis – this automatically generates a spoken presentation of the relevant raw data (codes and numeric expressions from Reuters) in either Cantonese or English.

These processes are integrated in the application manager, together with a dialog model. The interface is developed on a SpeechWorks 4.0 and InterVoice InVision platform, running on a Pentium II machine (300MHz) with 64M RAM. Communication between the speech interface and the relational database (SQL server) is effected by ODBC.

Speech Recognition

Our speech recognition component handles both Cantonese and English. Our Cantonese transcription is based on the LSHK standard [7], while our English transcription adopts the ARPABET phonetic labels. Our vocabulary has approximately 500 entries, covering country and currency names in the foreign exchange domain, as well as their (colloquial) variations, e.g. the "*German Mark*" and "*D-Mark*" both refer to the "*Deutsche Mark*". Similar examples include "馬仔" for "馬克", and the "*greenback*" for "*US Dollar*". To increase flexibility, users can call with either landline or mobile phones. The cellular phone service providers in Hong Kong adopt a number of standards, including PCS, GSM and CDMA [8]. Hence our speech recognizer needs to handle these various types of speech data.

Natural Language Processing

We have developed parallel context-free grammars for natural language processing, one for English and the other for Chinese. This is necessary as English and Chinese have different word orders. As an illustration, consider the following query pairs:

English: *What is the exchange rate between the US dollar and the Hong Kong dollar?*

Chinese: 我想問美金兌港幣 o 既匯率 o 係幾多

Approximate Translation for the Chinese query:

<US DOLLAR><VERSUS><HK DOLLAR><EXCHANGE RATE>

Natural language is bypassed should the user select a directed dialog interaction. In a directed dialog interaction, the user inputs only one informational attribute per dialog turn.

Dialog Modeling

The system provides a directed dialog interaction for *novice* users, and natural language shortcut interaction for *expert* users. The dialog structure is consistent across the two languages. Example dialogs are shown in Tables 1 and 2. A directed dialog guides the user through a session, and elucidates what may be said at various stages of the interaction. User feedback is enriched by the use of audio icons to distinguish among the various menu choices.[1] Novice users may also take control and self-navigate around the dialog tree by the use of meta-commands, which include:

HELP 求助: context-sensitive help instructions

REPEAT 重覆: repeat the last generated response

START AGAIN 重新開始: return to the main menu

CHANGE LANGUAGE 轉換語言: switch to the other language and start again from the main menu

GOODBYE 拜拜: end the session

Domain-specific intelligence is also integrated into the directed dialog model. For example, if the user's input is "francs", the system prompts for disambiguation between the French Franc, Swiss Franc or Belgian Franc. Similar prompting applies for other currencies with the same name, e.g. Danish Krone, Norwegian Krone or Swedish Krona.

System	Welcome to CU FOREX Hotline. Which language would you prefer, Cantonese or English? 歡迎致電寰宇之聲.請講出你想用既語言? '廣東話' 定 o 係 '英文'?
Caller	**English**
System	Main Menu. Would you like 'exchange rates' or 'interest rate'?
Caller	**Interest rate.**
System	Interest rate menu. Which currency are you interested in?
Caller	<UNRECOGNIZED>
System	Sorry, I didn't understand you. Please say the currency name or the country name again.
Caller	**Australian Dollar.**
System	I think you said Australian Dollar. Is that correct? Please say 'yes' or 'no'.
Caller	**Yes.**
System	Would you like deposit durations for twenty four hours, one month.... or one year? Or, you can say say 'all of them' to hear all the interest rates quotes.
Caller	**One month.**
System	The quotes we provide are for reference only. Interest rate, Australian dollar for one month is four point nine. Main menu. Please select 'exchange rate' or 'interest rate'.
Caller	**Change language.**
System	請問你想查詢邊種服務, '外幣匯率' 定係 '存款利率'

[1] The idea is contributed by Brian Lawrence from SpeechWorks.

Caller	外幣滙率
System	請問你想查詢 o 既匯率是兌港幣還是兌其他外幣?
Caller	兌港幣
System	請問你想查詢邊種貨幣 o 既匯率?
Caller	丹麥克郎
System	以下資料只供參考. 丹麥克郎兌港幣,買入價係 1.087,賣出價係 1.086.

Table 1. Example of a Directed Dialog Interaction.

The natural language interaction is tailored for expert users who may wish to expedite the inquiry session by uttering a full query, and traverse the entire session within one interaction. Hence the query may carry multiple attributes for database retrieval, in contrast with a single attribute per utterance in the case of a directed dialog.

System	Welcome to CU FOREX Hotline. Which language would you prefer, Cantonese or English? 歡迎致電寰宇之聲.請講出你想用既語言? '廣東話' 定 o 係 '英文'?
Caller	**English**
System	This system can handle a full question for your enquiry. Main Menu. What kind of currency information are you interested in?
Caller	**I'd like to know the exchange rate between the greenback and the german mark please.**
System	The quotes we provide are for reference only. Exchange rate. US Dollar to HK Dollar. The buying rate is seven point seven seven. The selling rate is seven point seven seven one. What kind of currency information are you interested in?
Caller	**Change language.**
System	請講出你所查詢既外滙資料.
Caller	澳幣 一星期 o 既存款利率
System	以下資料只供參考 澳幣一星期存款利率 o 係 4.82

Table 2. Example Dialog of a Natural Language Interaction.

Both directed dialog and natural language interaction produce a semantic frame to invoke database access. The response is generated by template-based word concatenation.

Current Status

CU FOREX is currently made available to the public for experimentation. We run two phone lines constantly throughout the day – one for the directed dialog interaction (+852.2603.7884) and the other for natural language shortcut interaction (+852.2603.7830). Calls are recorded at certain times and the data acquired is used for usability studies.

We have recruited 89 subjects[2] over a three-week period to conduct an evaluation of the system. All our subjects were interacting with a spoken language system for the first time. They were asked to refer to the system's homepage on the Web [9], to obtain some brief information about our system. Each evaluator was asked to formulate several queries related to foreign exchange prior to calling the system. Our analysis is based on system logs, as well as questionnaires returned by our evaluators. We received a total of 423 foreign exchange queries in all. Based, on this corpus, we adopted the PARADISE framework [10] for system evaluation.

The PARADISE framework offers a way to evaluate task completion with considerations in task complexity. We organized our evaluation data into Attribute Value Matrices (AVMs), where the columns are reference values to the task attributes, and rows are hypothesized values to the task attributes. Our attributes include language, EXCHANGE_RATE, INTEREST_RATE, CURRENCY_TO_BUY, CURRENCY_TO_SELL, CURRENCY_FOR_DEPOSIT and TIME_DURATION, and their values include bilingual lexical items. For a given confusion matrix M with total count T, the kappa coefficient (κ) measures the rate of actual agreement between the reference and hypothesized values, P(A), normalized by the rate of agreement by chance, as shown in Equation (1).

$$K = \frac{P(A) - P(E)}{1 - P(E)} \text{.....(1)}$$

$P(A)$ and $P(E)$ are computed according to Equations (2) and (3), and t_i is the sum of counts in column i of the AVM.

$$P(A) = \frac{\sum_{i=1}^{n} M(i,i)}{T} \text{.....(2)}$$

$$P(E) = \sum_{i=1}^{n} (\frac{t_i}{T})^2 \text{....(3)}$$

[2] Our evaluators are students from the Chinese University of Hong Kong.

Overall, our system achieves a kappa-coefficient of 0.938 for directed dialog interaction, and 0.876 for natural language shortcuts. The average transaction time for directed dialog is longer (2.2 minutes), compared to that of natural language shortcuts (1.9 minutes). Details are provided in [11].

TOWARDS BILITERACY AND TRILINGUALITY

We are extending our system to cover Putonghua as well, thereby becoming a trilingual system. It is observed that a multilingual system may be beneficial for this application context – users generally know the set of the globally traded currencies, but they may not know all the country/currency names in a single language. Hence multilinguality offers enhanced flexibility to support the user's inquiries.

Additionally, we aim to scale up our system to application domains of higher complexities, e.g. the stocks domain and the financial news domain. Figure 2 illustrates the architecture of such a system (named ISIS). It is enhanced with a speaker verification component to secure access to private or personal financial information. The speech recognition and speech generation components are *trilingual*, while the language understanding component is *biliteral* for handling English / Chinese text coming from the recognizers. The remaining components in the system remain *language independent.*

Biliteral Language Understanding

We have devised a methodology that can semi-automatically induce a context-free grammar from un-annotated text corpora [12]. The methodology has been demonstrated to work for both English and Chinese textual queries. Grammar induction is an agglomerative word clustering procedure which can capture semantic categories as well as (syntactic) phrasal structures. The induction algorithm is amenable to prior human knowledge injection to catalyze the induction process. Moreover, the induced grammars are amenable to hand refinement as a post-process. The induced grammars can then couple with a parser to analyze natural language input and extract meaning from the user's query. The semi-automatic nature of the approach enhances portability across domains and languages. Our experiments have shown encouraging results when the induced grammar is compared with a handcrafted grammar for understanding [12].

Trilingual Speech Recognition and Generation

We are integrating three monolingual recognizers to achieve trilingual speech recognition. As regards speech generation, the objective is to generate a spoken presentation of the raw data (numbers and codes) for the user, and maximizing the degrees of intelligibility and naturalness in the output acoustics. We have integrated the FESTIVAL [13] speech synthesizer for English synthesis. For Chinese, we have developed our own corpus-based concatenative synthesis system [14], and applied the same technique to *both* Cantonese and Putonghua. We have chosen the syllable as our basic unit for synthesis, since the Chinese language is monosyllabic in nature. Each syllable unit is appended with two digits to encode the distinctive features in its left and right co-articulatory contexts. Our concatenative synthesis technique aims to maximize the intelligibility and naturalness of the generated acoustics within the scope of the domain. A listening test based on 12 subjects showed that this concatenative approach compares favorably with a domain-independent PSOLA synthesizer based on intelligibility and naturalness.

ACKNOWLEDGMENTS

We acknowledge our collaborators, SpeechWorks International Ltd. and IVRS (International) Ltd., on the development of CU FOREX system. We also acknowledge our collaborator, National Key Laboratory for Machine Perception in Peking University, on the joint development of the ISIS system. We thank Reuters Hong Kong for donating the real-time data feed via satellite. The work reported here involves projects conducted by various members in the CUHK Human-Computer Communications Laboratory.

REFERENCES

1. P. Price, "Evaluation of Spoken Language Systems: the ATIS Domain," *Proceedings of the DARPA Speech and Natural Language Workshop*, pp.91-95, 1990.
2. Os et al., "Overview of the Arise Project," Proceedings of Eurospeech, 1999.
3. Zue V. et al., "From Interface to Content: Translinngual Access and Delivery of On-Line Information," *Proceedings of the European Conference on Speech Communication and Technology*, pp. 2227-2230, 1997.
4. Jeanrenaud, P. et al., "A Multimodal, Multilingual Telephone Application: The Wildfire Electronic Assistant," Proceedings of Eurospeech, 1999.
5. Deviller, L. and H. Bonneau-Maynard, "Evaluation of Dialog Strategies for a Tourist Information Retrieval System," Proceedings of Eurospeech, 1999.
6. Yang, Y. and L. S. Lee, "A Syllable-Based Chinese Spoken Dialogue System for Telephone Directory Services Primarily Trained with A Corpus," Proceedings of ICSLP, 1998.
7. Lingustic Society of Hong Kong, Hong Kong Jyut Ping Character Table, Linguistic Society of Hong Kong Press, 1997.
8. Office of the Telecommunications Authority, Hong Kong SAR Government, http://www.ofta.gov.hk.
9. http://www.se.cuhk.edu.hk/hccl/demos/cu_forex/.
10. Walker, M. et al., "PARADISE: A General Framework for Evaluating Spoken Dialog Agents," ACL/EACL 1997.

11. Meng, H., S. Lee and C. Wai, "CU FOREX: A Bilingual Spoken Dialog System for Foreign Exchange Inquiries," Proceedings of ICASSP, 2000.

12. Siu, K. C. and H. Meng, "Semi-Automatic Acquisition of Domain-Specific Semantic Structures," Proceedings of Eurospeech, 1999.

13. Taylor, P. et al., "The Architecture of the FESTIVAL speech synthesis system," Proceedings of the Third ESCA Workshop on Speech Synthesis.

14. Fung, T. Y. and H. Meng, "Concatenating Syllables for Response Generation in Spoken Language Applications," Proceedings of ICASSP, 2000.

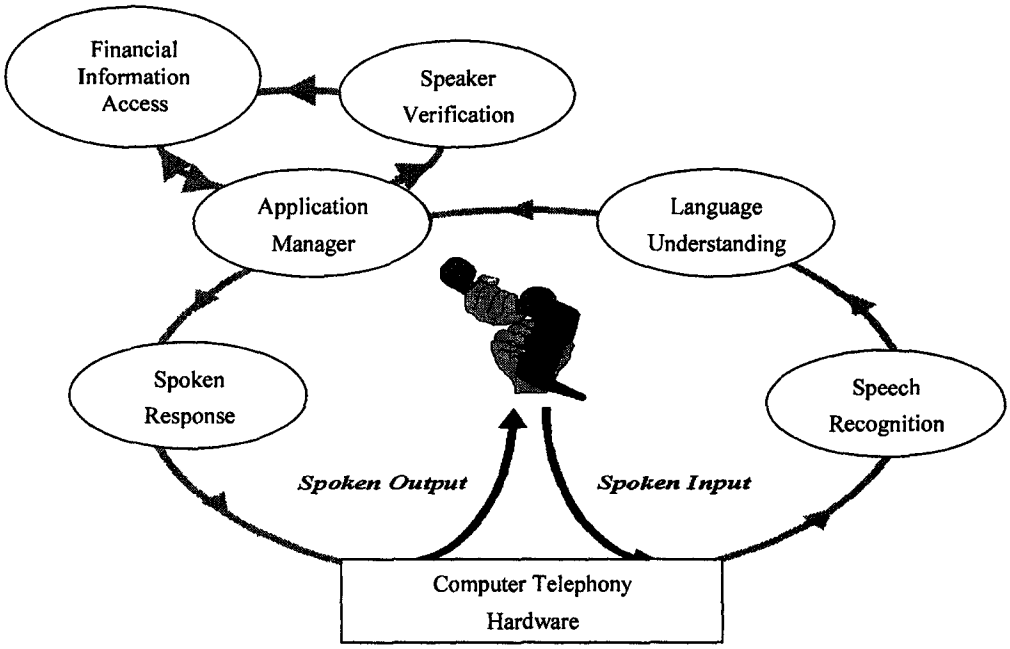

Figure 2. System Architecture of a Spoken Language Interface for Financial Information Access.