CU FOREX: A Bilingual Spoken Dialog System for Foreign Exchange Enquiries

Helen M. Meng[1], Steven Lee[2] and Carmen Wai[1]
[1]Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China
[2]SpeechWorks International Ltd.
695 Atlantic Avenue, Boston, MA 0211, USA
hmmeng@se.cuhk.edu.hk, sclee@speechworks.com, cmwai@se.cuhk.edu.hk

ABSTRACT

This paper reports on the design, development and evaluation of a bilingual telephone-based spoken language system for foreign exchange inquiries – the CU FOREX system. The telephone-based system supports Cantonese and English in a single spoken language interface to access real-time foreign exchange information. The specific domain covers information on the exchange rates between foreign currencies, as well as deposit interest rates of various time durations for a specified currency. Overall our system achieves a performance of 0.95 (kappa-statistic), with a typical transaction duration of 2 minutes on average.

1. INTRODUCTION

This paper reports on the design, development and evaluation of a bilingual telephone-based spoken language system for foreign exchange inquiries – the CU FOREX system. The telephone-based system supports Cantonese (a dialect of Chinese)[1] and English in a single spoken language interface. Real-time foreign exchange information is captured directly from a dedicated Reuters satellite feed. The specific domain covers information on the exchange rates between foreign currencies, as well as deposit interest rates of various time durations for a specified currency.

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 [2,3,4,5,6,7].

In this work, we have chosen an application domain that is well-suited for Hong Kong. The region has a trilingual populace speaking Cantonese, Mandarin and English. It is one of the largest foreign exchange trading centers in the world. Both landline and mobile phone penetrations are high – the former is near saturation and the latter is over 50%. Our ultimate goal is to explore the research issues involved in the development of a trilingual conversational system. As an initial step, we developed a bilingual system for the foreign exchange domain. At this stage, we focused on the following issues:

- Bilinguality – we began with Cantonese and English, the two predominant languages used in the region.
- Affordance of the dialog design – we aim to support effective interaction of both novice and expert users.
- Evaluation – we conducted a series of user trials and usability surveys to evaluate the end-to-end system.

Discourse handling is planned as a next step. Nevertheless, our current system prompts for disambiguation if the user merely specifies a currency name which may refer to the currency in multiple countries, e.g. the Franc (Belgian Franc, French Franc and Swiss Franc); or the Krone (Danish Krone and Norwegian Krone).[2]

2. SYSTEM ARCHITECTURE

Figure 1 shows the overall architecture of CU FOREX system. A satellite dish is mounted on our rooftop to receive the real-time datafeed from Reuters. We developed a data-capture component that continuously updates the data in a relational database. The core engines include:

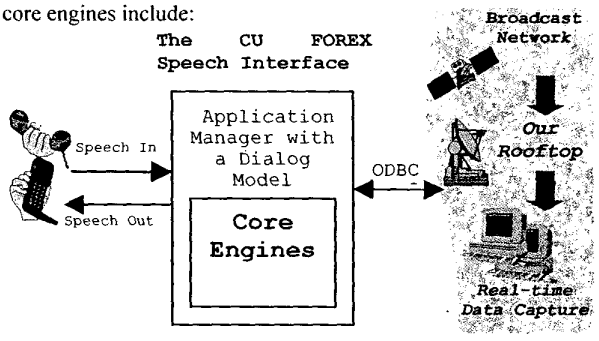

Figure 1: The CU FOREX System – overall architecture.

bilingual speech recognition, an optional natural language processing component, and a concatenative speech synthesis component. These are integrated with the application manager, which includes a dialog model. The interface is developed on a SpeechWorks 4.0 and InterVoice InVision platform, running on

[1] Cantonese is spoken by tens of millions of people in Hong Kong, Southern China and many overseas Chinese communities. The dialect is monosyllabic and tonal, with approximately 1,800 distinct syllables and nine lexical tones [1].

[2] Our knowledge domain covers 32 currencies provided by our Reuters feed. We have also included the Euro, since it became a trading currency in 1999.

0-7803-6293-4/00/$10.00 ©2000 IEEE.

a Pentium II machine (300MHz) with 64M RAM. Communication between the speech interface and the relational database (SQL server) is effected by the ODBC.

3. TASK PROFILE AND SYSTEM FUNCTIONALITIES

Our system is designed to support 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 and presented to the user in spoken form via concatenative synthesis. As illustrated, the system starts a session with a bilingual welcome message, followed by a language prompt. Selection of the preferred language by the user determines the language used in subsequent prompts for the session. However, the user is able to change to the other language at any time, by uttering "*change language*" or "轉換語言".

Our speech recognition component handles both Cantonese and English. Our Cantonese transcription is based on the LSHK standard [8], 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*". The cellular phone service providers in Hong Kong adopt a number of standards, including PCS, GSM and CDMA [9]. Hence our speech recognizer needs to handle these various types of data. In addition, it may select a lexical item from the vocabulary list during an interaction with a novice user, or perform continuous speech recognition on a natural language query during an interaction with an expert user. We have also experimented with system configurations whereby the user is free to speak either language at any time; or requested to make a language selection at the start of the session, and subsequently constrained to speak the selected language. While the former configuration provides greater flexibility upon active use, it is inherently more complex. The latter configuration eases subsequent processing (e.g. parsing and response generation), leading to a higher task completion rate.

For natural language processing, we have developed parallel context-free grammars, as English and Chinese have rather 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. Both directed dialog and natural language interaction produce a semantic frame to invoke database access. The response is generated by template-based word concatenation.

As regards dialog control in CU FOREX, we provide a directed dialog interaction for novice users, and natural language 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.[3] The dialog is also integrated with confidence-based rejection from speech recognition. Rejection is subsequented by prompting the user to repeat the utterace, or reconfirm based on the *N*-best recognition outputs (*N*=3). 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

The natural language interaction is tailored for expert users who may wish to expedite the enquiry session by uttering a full query, and traverse the entire session within one interaction. Hence the query may carry multiple attributes for database retrieval, as opposed to a single attribute per utterance in the directed dialog case. Should the utterance be unrecognized, the system prompts the user for the *entire* query, as well as offer to backoff to a directed dialog interaction. When discourse handling is incorporated in the future, we will be able to support correction of specific attributes in the query. The integration of the directed dialog and the natural language interface provides the scalability required to support of user diversity – from the novice to the expert.

4. EVALUATION

The CU FOREX was made available to the public for experimentation since August 1999. Since then, we have received a few hundred calls (around 800) per month on average. During a three-week period in September 1999, we recruited 89 subjects[4] 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 [10], 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, with a breakdown tabulated in Table 3.

4.1 The Kappa Statistic

We adopted the PARADISE framework [11] for our evaluation. The 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

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

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'
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.

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 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).

$P(A)$ 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.

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

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

$$P(A) = \frac{\sum_{i=1}^{n} M(i,i)}{T} \quad \text{.....(2)} \qquad P(E) = \sum_{i=1}^{n} \left(\frac{t_i}{T}\right)^2 \quad \text{....(3)}$$

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 US dollar and the HK dollar 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.

Directed Dialog Queries 277				**Natural Language Queries** 146			
Cantonese 112		**English** 165		**Cantonese** 56		**English** 90	
Ex.	**Int.**	**Ex.**	**Int.**	**Ex.**	**Int.**	**Ex.**	**Int.**
76	36	88	77	33	23	53	37

Table 3. Breakdown of queries from our evaluators. Ex. and Int. stands for Exchange Rate and Interest Rate queries respectively.

Task Completion	**Directed Dialog Interaction**	**Natural Language Interaction**
Kappa statistic	0.938	0.876
Average Transaction Time	2.2 min	1.9 min

Table 4. Comparison between two interaction styles in terms of success rate and transaction times for task completion.

4.2 Comparison between Interaction Styles

We compared between the interaction styles of directed dialog (DD) and natural language (NL). We expect that the recognition performance of DD is better, leading to a higher task completion performance and kappa value. Results are shown in Table 4.

From Table 4 we see that the kappa success rate for DD is higher than NL. Analysis shows that the inferior performance of NL is due to (i) the higher difficulty in recognizing full queries versus utterances each with a single keyword. In addition, system re-confirmation is absent in the NL interaction. (ii) parse failures of the recognized query – we will elaborate on this point later.

Task failures are mostly caused by recognition errors, out-of-domain queries (e.g. the "Finland Markka") and parse failures from the NL interaction.

Average transaction time for the DD interaction is 12 seconds longer than that of the NL interaction, which implies that NL expedites the transaction to some extent. The difference is smaller than expected, mostly due to the greater latency in recognizing a full query compared to a short, keyword-based utterance.

4.3 Comparison between Subtasks

We have also compared the subtasks of exchange rate and interest rate enquiries. Results are shown in Table 5.

	Exchange Rate Enquiries	Interest Rate Enquiries
DD, kappa value	0.939	0.930
DD, duration	2.1 min	2.2 min
NL, kappa value	0.877	0.869
NL, duration	1.9 min	1.9 min

Table 5. Comparing the task completion success rates and durations between exchange rate and interest rate enquiries.

While the transaction times are comparable between the two subtasks, the kappa values for the exchange rate enquiries are higher than those for interest rate enquiries. We believe this is due to a more confusable vocabulary (i.e. the lexical items pertaining to the specification of time durations).

4.4 Comparison across Languages

Comparison between the Cantonese query set and English query set yields the results as displayed in Table 6.

	Cantonese	English
DD, kappa value	0.950	0.926
DD, duration	2.0 min	2.2 min
NL, kappa value	0.812	0.914
NL, duration	1.7 min	2.1 min

Table 6. Comparison of success rates and durations for task completion between the use of Cantonese versus English.

The choice of language does not seem to affect task completion duration, and the kappa values based on the DD interactions are close between the two languages. However, the NL interaction using Cantonese is noticeably worse than that of English. Data for NL interaction in Cantonese is relatively sparse, but analysis shows that we received some input phrasal structures which were not anticipated during grammar writing, e.g.

請問日圓兌加元而家係幾多
一蚊港銀兌幾多
我想問澳元存款利率一年
日圓兌澳元,唔該

The situation should improve as we extend our grammar.

5. Conclusions and Future Work

In this paper, we have reported on the design, development and evaluation of the first version of our bilingual spoken dialog system in the foreign exchange domain – CU FOREX. System performance is measured based on the kappa measure and time durations for task completions, and the effects of interaction styles (directed dialog versus natural language), choice of language, and difference in subtasks were all considered. Subsequent developments include improving the speech recognition performance based on the data collected by the system, refining our grammar for natural language processing, incorporating the capability of discourse handling, including Putonghua to achieve trilinguality, and expanding of the knowledge domain to cover other types of financial information. We will also research into strategies for dialog designs, which should be scalable in the comprehensibility, predicability and controllability of the interface, to fit the needs of novice, knowledgeable and expert users.

ACKNOWLEDGMENTS

CU FOREX is developed in collaboration with SpeechWorks International Ltd. and IVRS (International) Ltd. We thank Reuters for providing the real-time data feed via satellite.

REFERENCES

[1] Wong, Y. W. et al., "Acoustic Modeling and Language Modeling for Cantonese LVCSR," Proceedings of Eurospeech, pp. 1091-1094, 1999.

[2] P. Price, "Evaluation of Spoken Language Systems: the ATIS Domain," *Proceedings of the DARPA Speech and Natural Language Workshop*, pp.91-95, 1990.

[3] den Os et al., "Overview of the Arise Project," Proceedings of Eurospeech, 1999.

[4] 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.

[5] Jeanrenaud, P. et al., "A Multimodal, Multilingual Telephone Application: The Wildfire Electronic Assistant," Proceedings of Eurospeech, 1999.

[6] Deviller, L. and H. Bonneau-Maynard, "Evaluation of Dialog Strategies for a Tourist Information Retrieval System," Proceedings of Eurospeech, 1999.

[7] 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.

[8] Lingustic Society of Hong Kong, Hong Kong Jyut Ping Character Table, Linguistic Society of Hong Kong Press, 1997.

[9] Office of the Telecommunications Authority, Hong Kong SAR Government, http://www.ofta.gov.hk.

[10] http://www.se.cuhk.edu.hk/hccl/demos/cu_forex

[11] Walker, M. et al., "PARADISE: A General Framework for Evaluating Spoken Dialog Agents," ACL/EACL 1997.