Title:
Intelligent Speech for Information Systems (ISIS):
A Multi-modal, Trilingual, Distributed Conversational System with Combined Interaction and Delegation Dialogs

Speaker:
Helen Meng
Associate Professor and Director, Human-Computer Communications Laboratory
The Chinese University of Hong Kong

Abstract:
ISIS is a trilingual spoken dialog system in the stocks domain. It supports the three languages commonly used in Hong Kong (Cantonese, Putonghua and English), and serves as a test-bed for our research in various speech and language technologies. This talk presents the ISIS system with a focus on its several unique features. We use the CORBA middleware to implement a distributed architecture that supports interoperability across platforms. We incorporate KQML (Knowledge Query and Manipulation Language) software agents to handle delegation dialogs. We also implement a mixed-initiative dialog management strategy that combines online interaction with offline delegation. ISIS can automatically assimilate newly listed stock names into the system’s knowledge base. A recent enhancement supports multi-modal and mixed-modal input that suit the natural affordances of specific interactions in order to improve usability. Input modalities include speaking, typing or mouse-clicking. Output media include synthesized speech, text, tables and graphics.
ISIS: A Multi-modal, Trilingual, Distributed Conversational System with Combined Interaction and Delegation Dialogs

Helen Meng
Associate Professor and Director,
Human-Computer Communications Laboratory
The Chinese University of Hong Kong
Invited Lecture, ISCSLP 2002

Outline

- ISIS system overview
- Architecture – CORBA
- Delegation to KQML software agents
- Dialog management
  - Online interaction and offline delegation
- A Learning system
- Multimodality
- Conclusions and future work
ISIS System Overview

- Intelligent Speech for Information Systems (ISIS)
  - trilingual spoken dialog system [Meng et al., ICSLP-2000]
  - Cantonese, Mandarin, English (regional languages)
  - financial domain (real-time stock quotes, simulated personal portfolios, buy / sell transactions, etc.)

- Speaker verification
  - text-dependent, 16 Gaussian mixtures, hypothesis testing against a background speaker model [Wang, Chen & Chi, ICSLP-2000]

- Speech recognition
  - English: Dragon Naturally Speaking
  - Cantonese & Mandarin: HMM, Initial / Final models (bi-IF), word bigram [Choi et al., ICSLP-2000]

ISIS System Overview (cont)

- Natural language understanding
  - spoken & typed input
  - single framework for English and Chinese
  - concept parsing, Belief Network goal inference, out-of-domain (OOD) rejection [Meng et al., Eurospeech-99]
  - handle out-of-vocabulary words (Meng & Tsui, ISCSLP-00)

- Dialog and discourse
  - context inheritance with E-forms [Meng, ICSLP-2000]
  - mixed-initiative dialog with turn management table

- Response generation
  - English: FESTIVAL [Taylor et al., 1998]
  - both Chinese dialects: concatenative synthesis, CU VOCAL [Fung, Meng, et al., ICASSP-00, ISCSLP-02]
CORBA

Common Object Request Broker Architecture

CORBA (cont)

- Implementation
  - software, documentation, online demo
  - www.se.cuhk.edu.hk/~isis/download

- Location Transparency
  - remote server class invocations handled by Internet and Inter-ORB Protocol (IIOP)
  - only name of the server needs to be known
  - no need to manage host/port information

- Interoperability
  - across platforms over the network via the Interface Definition Language (IDL)

- Scalability
  - additional server classes inserted by adding interface definitions to the IDL

- Quality of Service
  - manages connection between client and server objects,
  - automatically rebind in case of communication problems
Delegation to Software Agents

- ISIS delegates to an agent the task of monitoring financial information **offline**
  - mismatch between the user’s ask / bid price and the real-time price quote
  - agent sends alert messages to the user when specified condition is met

- Knowledge Query and Manipulation Language
  - DARPA Knowledge Sharing Effort (JKQML)
  - message format and handling protocol for run-time information exchange and knowledge sharing
  - facilitator (software substrate with agent registry)
  - three-tiered structure

---

KQML Software Agents in ISIS

![Diagram of KQML software agents in ISIS](image-url)

- Requester Agent
  - request (non-blocking) from Dialog Manager
    - `<CORBA> ... </CORBA>`
    - (monitor HSBC dropping to 90 dollars)
  - respond to Dialog Manager
    - `<KQML> ... </KQML>`
    - (HSBC is at 90 dollars)

- Facilitator
  - request interpreted and stored
  - transmit request
  - trigger
  - alert message

- Alert Agent
  - monitor data feed

- Data Capture
Dialog involving Multiple Tasks

Register
Register current OI dialog state before switching to OD
Restores OI dialog state from Register

Online Interaction (OI): Mixed-Initiative Dialog
 Sys
 User
 Sys
 Status: subdialog complete
<Anchor>
- Notify user of alerts
- Offer option to switch to OD task

Offline Delegation (OD): Alert messages from KQML Agents
 Status: subdialog complete
 referential anchor (stock name)

A Learning System

- Expand knowledge base of spoken dialog system
  - newly listed stocks are out-of-vocabulary (OOV) words
  - e.g. Artel

- Assume speech recognizer can handle OOV
  - recognizable fragments, e.g. acronyms

- NLU detects OOV, hypothesizes semantic category
  - transformation rules [Brill 96]

OOV STOCK_NAME_OOV PREV_BIGRAM NEWS ABOUT

Query: i'd like to check the news about artel
Concepts1: <dummy> <check> <news> <about> <OOV>
Concepts2: <dummy> <check> <news> <about> <stock_name_OOV>
A Learning System (cont)

- Dialog strategy for OOV
  - dialog & discourse manager
  - detects <stock_name_OOV>ARTEL</stock_name_OOV>
  - find possible listing with substring matching
    0931.HK ARTEL GROUP
    1229.HK ARTFIELD GROUP
    ......

  - confirmation sub-dialog
  - call NLU server object to add grammar rules
    stock_name → ARTEL | ...
    stock_code → 0931

- Speech generation with OOV
  - English: letter-to-sound rules in FESTIVAL
  - Chinese: domain-independent syllable concatenation for
    Cantonese in CU VOCAL [Fung & Meng, ISCSLP-2002]

Multimodality

- User input modalities
  - speech, typed text, and mouse clicks (can switch freely)

- System output modalities
  - speech, text, graphics and tables

- Involves a set of java classes

- Example: multimodal scenario
  - I don’t know of the stock listed as Artel. This name is
    similar to the following list of stocks… Please speak or click
    on the correct listing…

- Example: mixed-modal scenario
  - Show the transaction details of these two stocks <click1>
    <click2> (on table listing)
  - mixed-modal events must occur within a short time window
  - client produces an integrated XML message for NLU
  - Show the transaction details of these two stocks, 0066.hk,
    1122.hk
Conclusions and Future Work

- Design and development of ISIS
- Trilingual spoken language interface
- Distributed architecture based on CORBA
- Mixed-initiative dialog management combines online interaction and offline delegation
- Learning system assimilates newly listed stocks
- Multimodal and mixed-modal input
- Migration towards a Web services model