Speech Recognition

語音辨識

Berlin Chen, 陳柏琳

berlin@csie.ntnu.edu.tw
http://berlin.csie.ntnu.edu.tw

Course Contents

- Both the theoretical and practical issues for spoken language processing will be considered
- Technology for **Automatic Speech Recognition** (ASR) will be further emphasized
- Topics to be covered
  - Statistical Modeling Paradigms
    - Spoken Language Structure
    - Hidden Markov Models
    - Speech Signal Analysis and Feature Extraction
    - Acoustic and Language Modeling
    - Search/Decoding Algorithms
  - Systems and Applications
    - Keyword Spotting, Dictation, Speaker Recognition, Spoken Dialogue, Speech-based Information Retrieval etc.

Textbook and References (1/2)

- Textbook
  - X. Huang, A. Acero, H. Hon. Spoken Language Processing, Prentice Hall, 2001
  - C. Manning and H. Schutze. Foundations of Statistical Natural Language Processing. MIT Press, 1999
- References books
  - T. F. Quatieri. Discrete-Time Speech Signal Processing - Principles and Practice. Prentice Hall, 2002
  - J. R. Deller, J. H. L. Hansen, J. G. Proakis. Discrete-Time Processing of Speech Signals. IEEE Press, 2000
  - F. Jelinek. Statistical Methods for Speech Recognition. MIT Press, 1999
  - S. Young et al.. The HTK Book. Version 3.0, 2000 "http://htk.eng.cam.ac.uk"
  - L. Rabiner, B.H. Juang. Fundamentals of Speech Recognition. Prentice Hall, 1993
  - 王小川教授， 語音訊號處理， 全華圖書 2004

Textbook and References (2/2)

- Reference papers

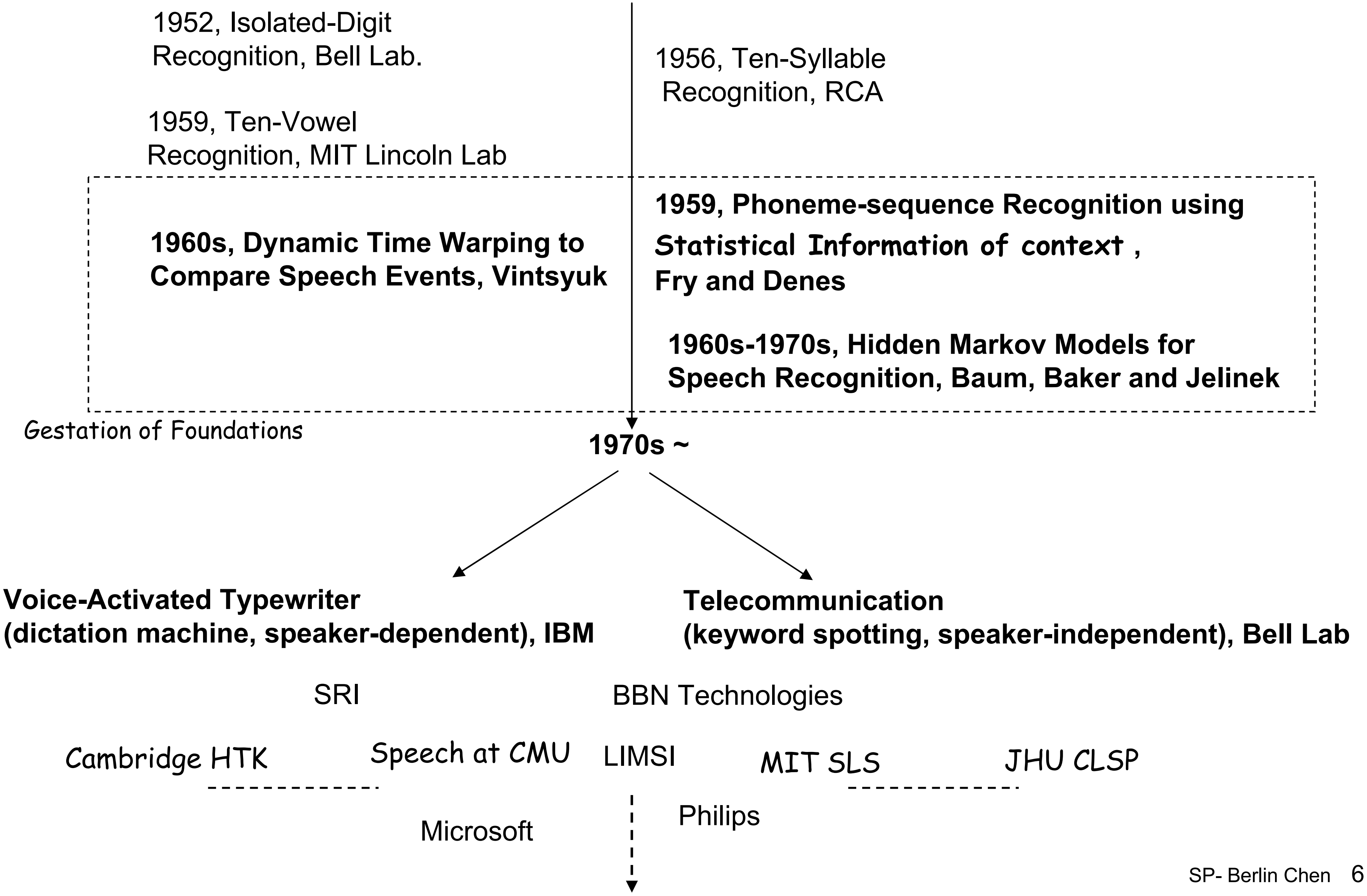
1. L. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition,” Proceedings of the IEEE, vol. 77, No. 2, February 1989
2. A. Dempster, N. Laird, and D. Rubin, "*Maximum likelihood from incomplete data via the EM algorithm*," J. Royal Star. Soc., Series B, vol. 39, pp. 1-38, 1977
3. Jeff A. Bilmes "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," U.C. Berkeley TR-97-021
4. J. W. Picone, “Signal modeling techniques in speech recognition,” proceedings of the IEEE, September 1993, pp. 1215-1247
5. R. Rosenfeld, ”Two Decades of Statistical Language Modeling: Where Do We Go from Here?,” Proceedings of IEEE, August, 2000
6. H. Ney, “Progress in Dynamic Programming Search for LVCSR,” Proceedings of the IEEE, August 2000
7. H. Hermansky, "Should Recognizers Have Ears?", Speech Communication, 25(1-3), 1998

Introduction

References:

1. B. H. Juang and S. Furui, "Automatic Recognition and Understanding of Spoken Language - A First Step Toward Natural Human-Machine Communication,“ Proceedings of IEEE, August, 2000
2. I. Marsic, Member, A. Medl, And J. Flanagan, “Natural Communication with Information Systems,“ Proceedings of IEEE, August, 2000

Historical Review



1952, Isolated-Digit Recognition, Bell Lab.

1956, Ten-Syllable Recognition, RCA

1959, Ten-Vowel Recognition, MIT Lincoln Lab

1960s, Dynamic Time Warping to Compare Speech Events, Vintsyuk

1959, Phoneme-sequence Recognition using *Statistical Information of context*, Fry and Denes

1960s-1970s, Hidden Markov Models for Speech Recognition, Baum, Baker and Jelinek

Gestation of Foundations

1970s ~

Voice-Activated Typewriter (dictation machine, speaker-dependent), IBM

Telecommunication (keyword spotting, speaker-independent), Bell Lab

SRI

BBN Technologies

Cambridge HTK

Speech at CMU

LIMSI

MIT SLS

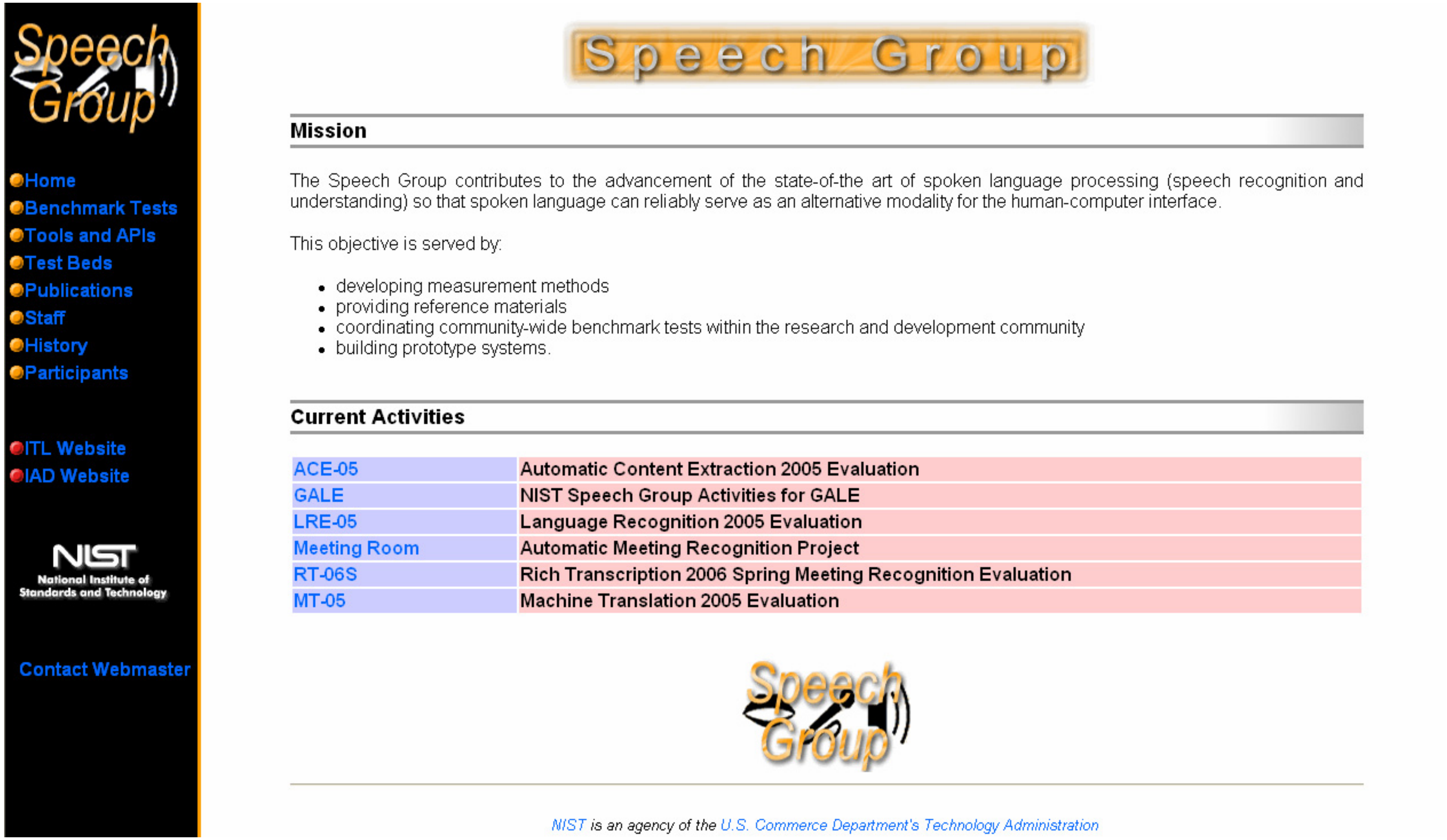
JHU CLSP

Microsoft

Philips

Progress of Technology (1/6)

- US. National Institute of Standards and Technology (NIST)



Mission

The Speech Group contributes to the advancement of the state-of-the art of spoken language processing (speech recognition and understanding) so that spoken language can reliably serve as an alternative modality for the human-computer interface.

This objective is served by:

- developing measurement methods
- providing reference materials
- coordinating community-wide benchmark tests within the research and development community
- building prototype systems.

Current Activities

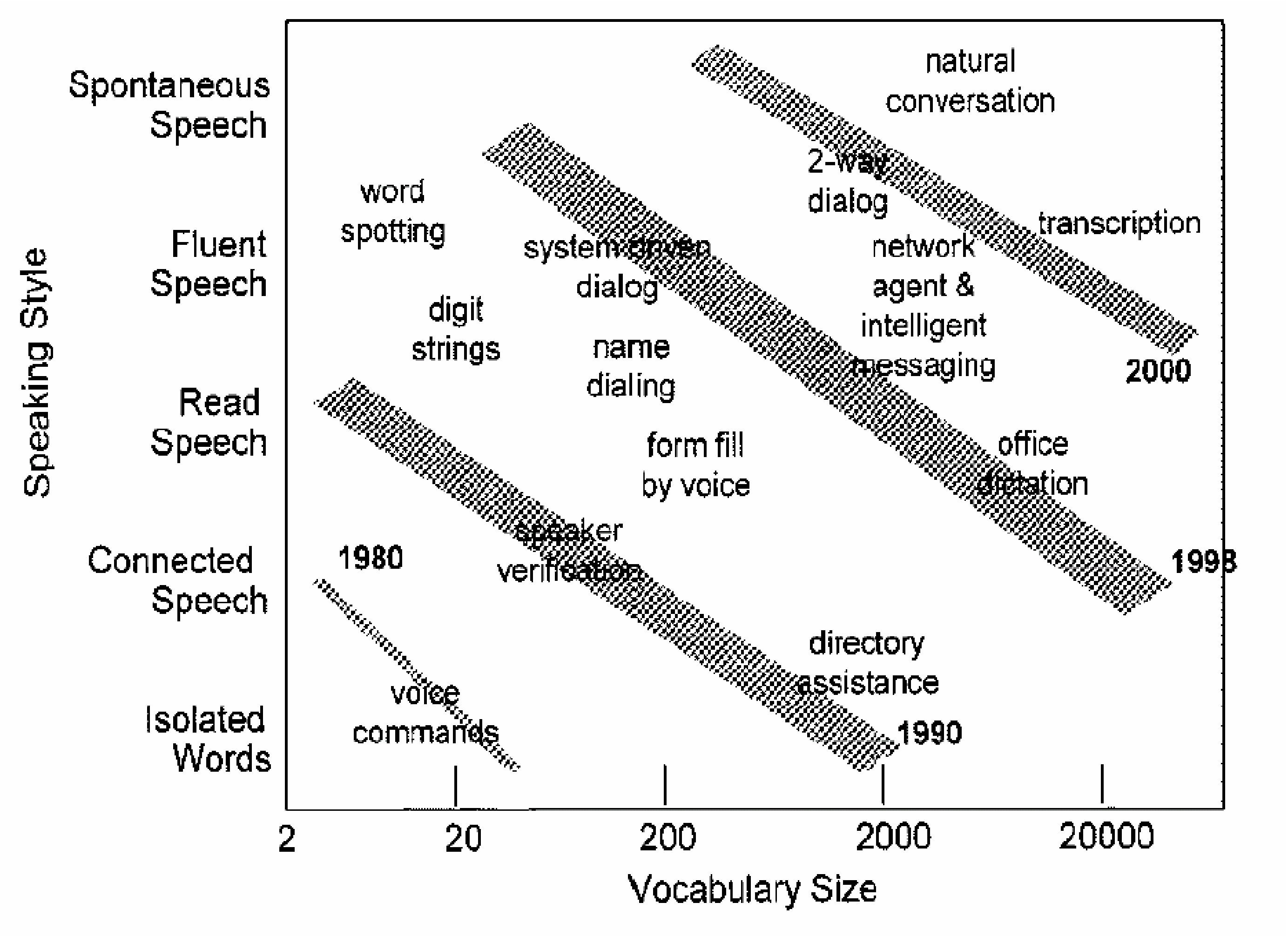
ACE-05	Automatic Content Extraction 2005 Evaluation
GALE	NIST Speech Group Activities for GALE
LRE-05	Language Recognition 2005 Evaluation
Meeting Room	Automatic Meeting Recognition Project
RT-06S	Rich Transcription 2006 Spring Meeting Recognition Evaluation
MT-05	Machine Translation 2005 Evaluation

NIST is an agency of the U.S. Commerce Department's Technology Administration

http://www.nist.gov/speech/

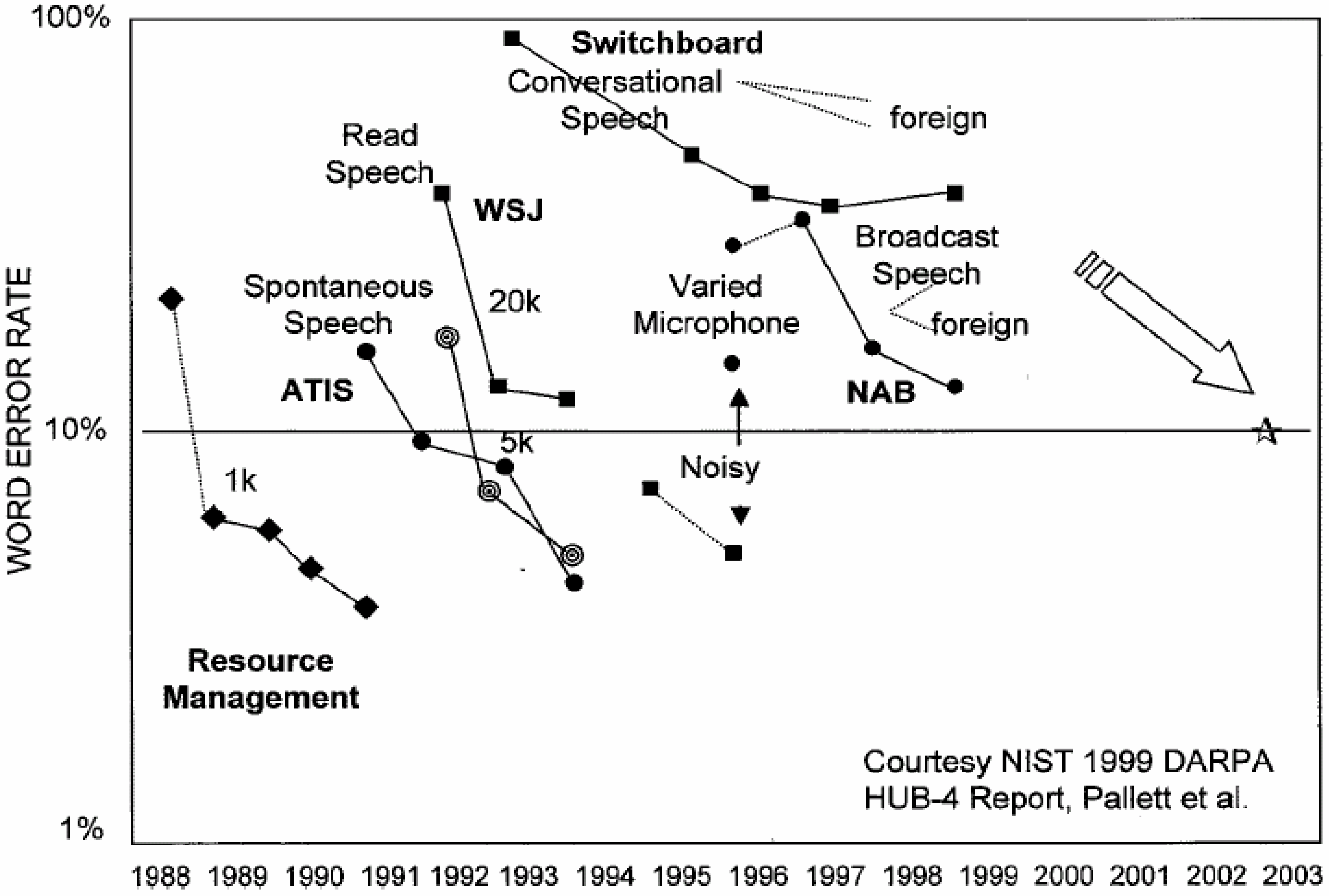
Progress of Technology (2/6)

- Generic Application Areas (vocabulary vs. speaking style)



Progress of Technology (3/6)

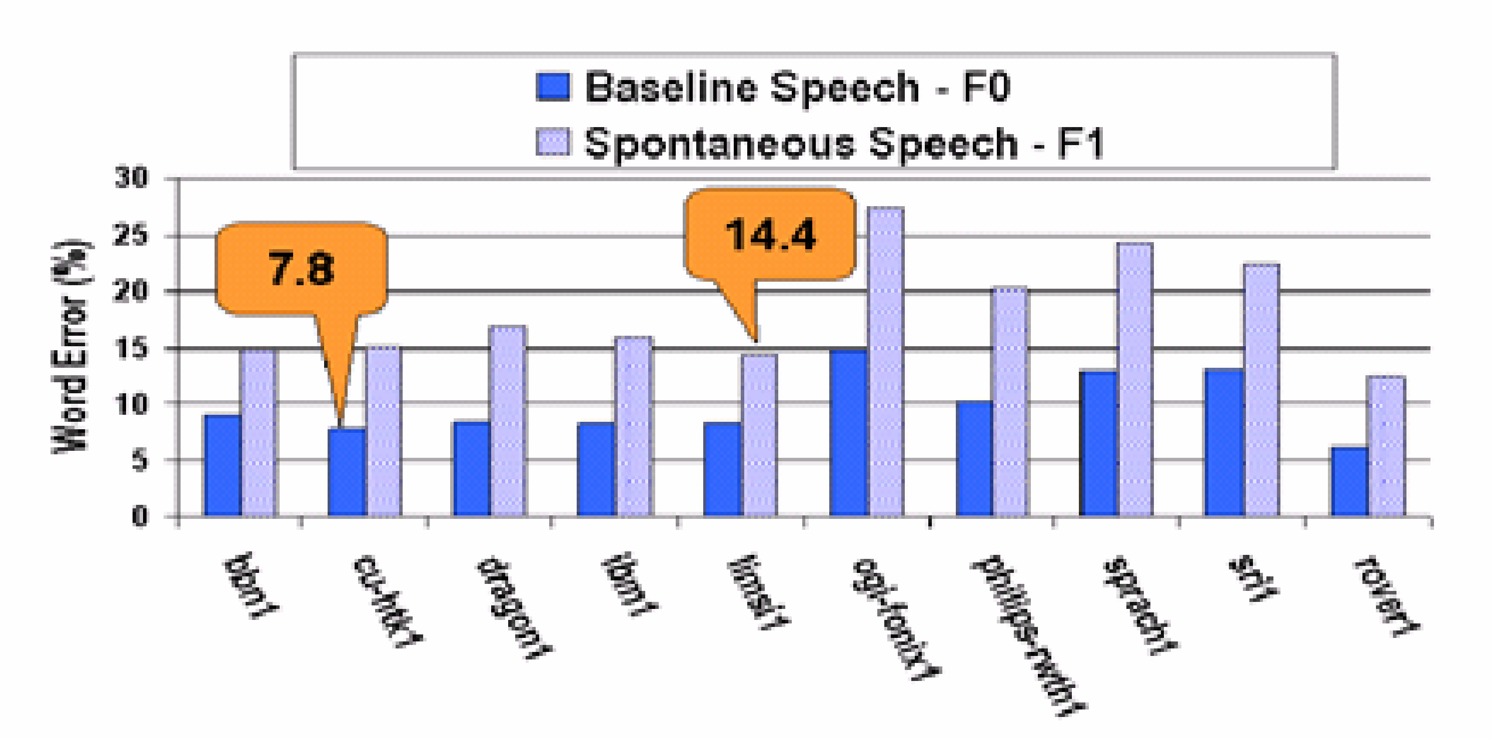
- Benchmarks of ASR performance: Overview



Courtesy NIST 1999 DARPA HUB-4 Report, Pallett et al.

Progress of Technology (4/6)

- Benchmarks of ASR performance: Broadcast News Speech



Progress of Technology (5/6)

- Benchmarks of ASR performance: Conversational Speech

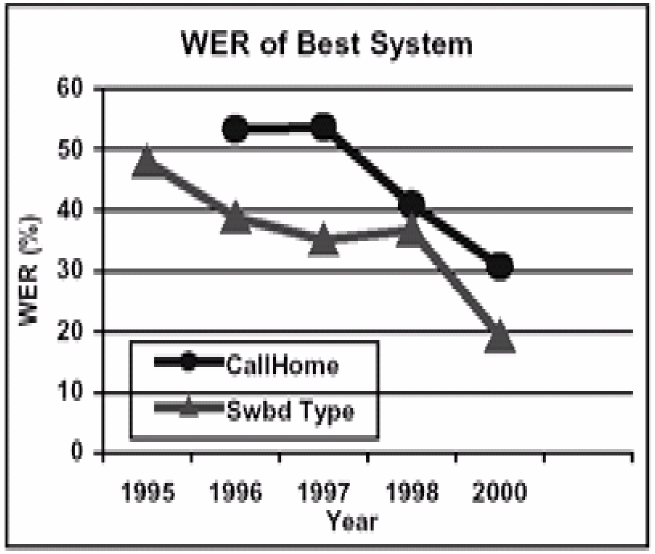


Figure 4 History of lowest word error rates (WER) obtained in NIST conversational speech evaluations on Switchboad and CallHome type conversations in English [26].

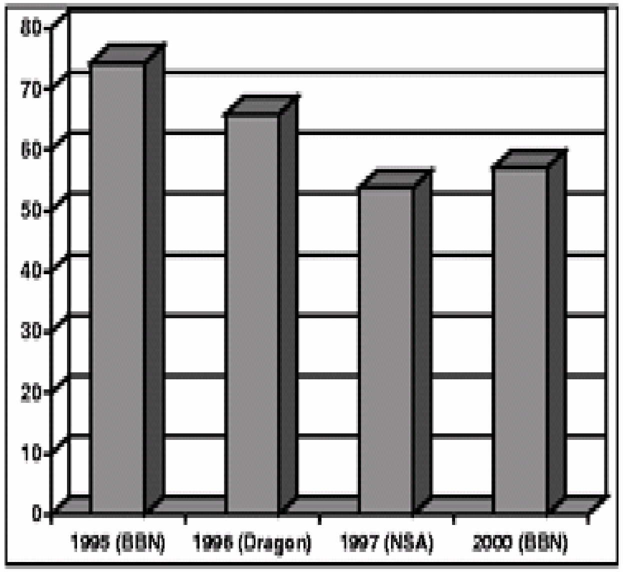
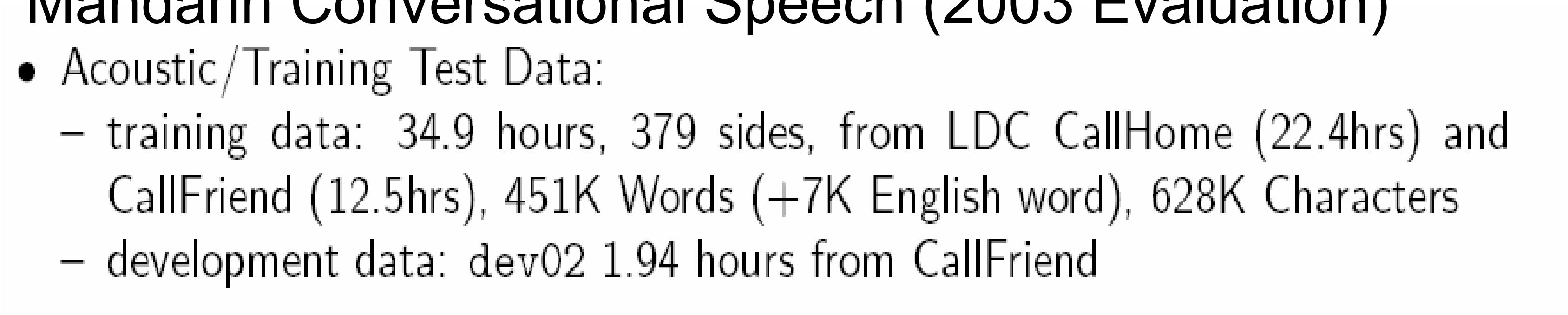


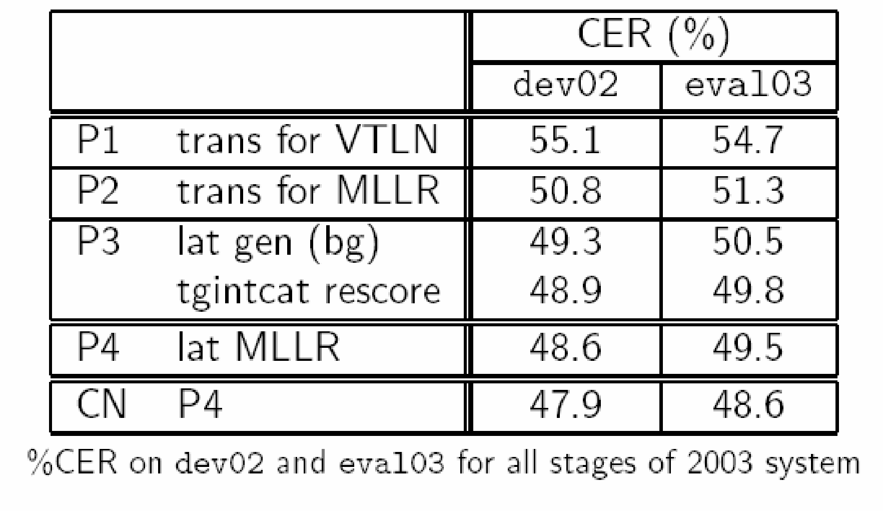
Figure 5 Chinese Character error rates of the best performing evaluation system in NIST Mandarin conversational speech evaluations 1995-2000 [26].

Progress of Technology (6/6)

- Mandarin Conversational Speech (2003 Evaluation)



- Acoustic/Training Test Data:
  - training data: 34.9 hours, 379 sides, from LDC CallHome (22.4hrs) and CallFriend (12.5hrs), 451K Words (+7K English word), 628K Characters
  - development data: dev02 1.94 hours from CallFriend



		CER (%)	
		dev02	eval03
P1	trans for VTLN	55.1	54.7
P2	trans for MLLR	50.8	51.3
P3	lat gen (bg)	49.3	50.5
	tgintcat rescore	48.9	49.8
P4	lat MLLR	48.6	49.5
CN	P4	47.9	48.6

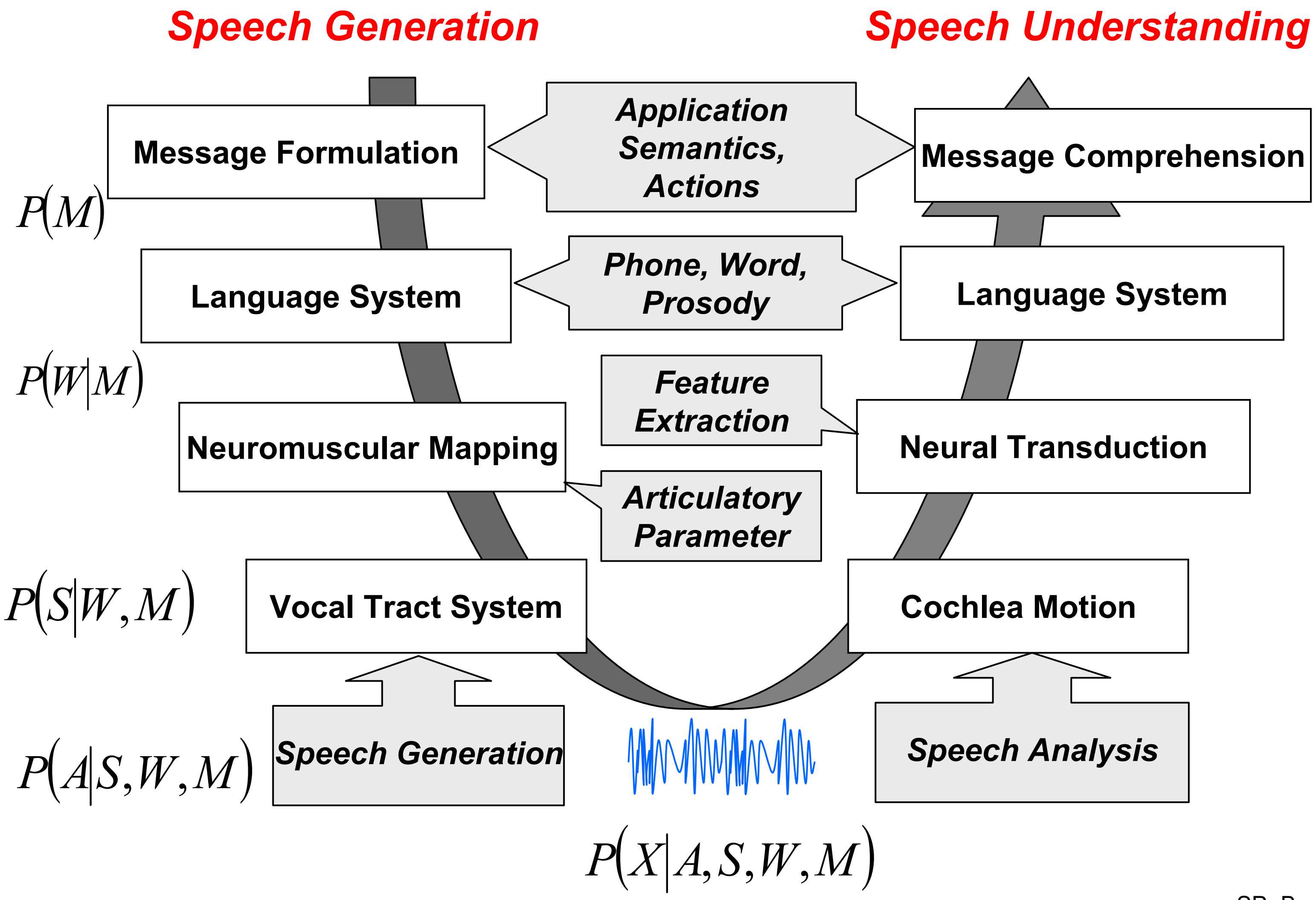
%CER on dev02 and eval03 for all stages of 2003 system

  - Adopted from

Cambridge University
Engineering Department

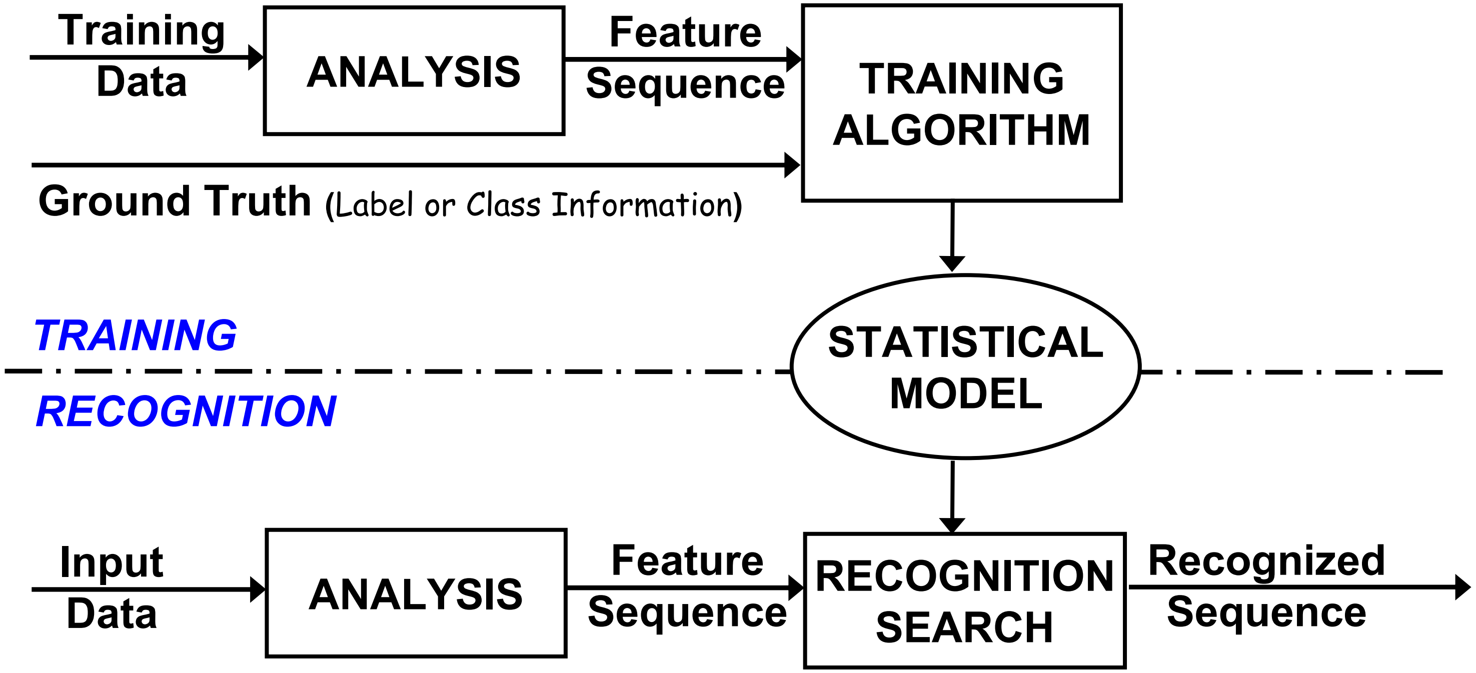
Rich Transcription Workshop 2003

Determinants of Speech Communication



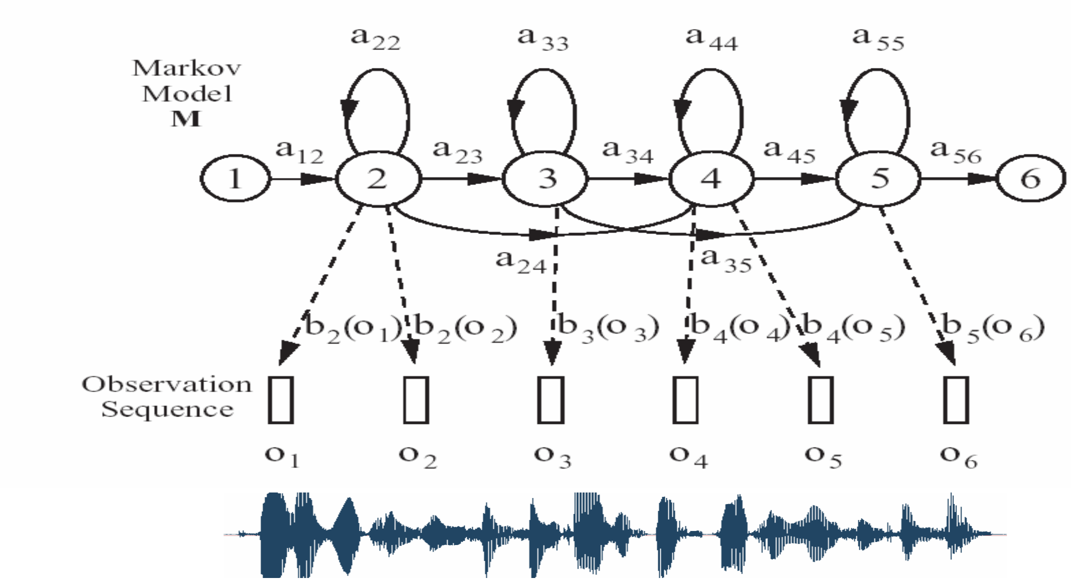
Statistical Modeling Paradigm (1/2)

- The statistical modeling paradigm used in speech and language processing

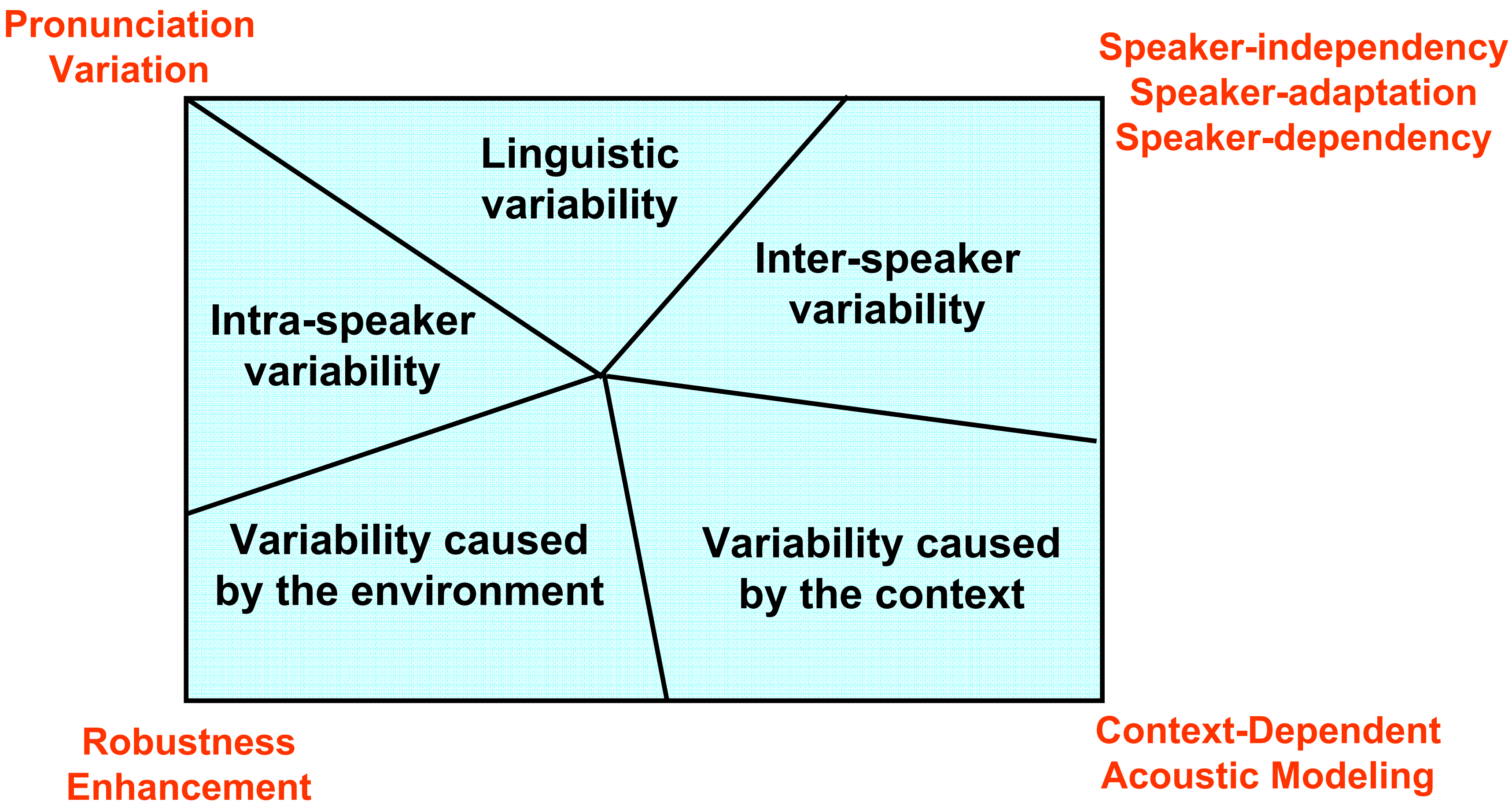


Statistical Modeling Paradigm (2/2)

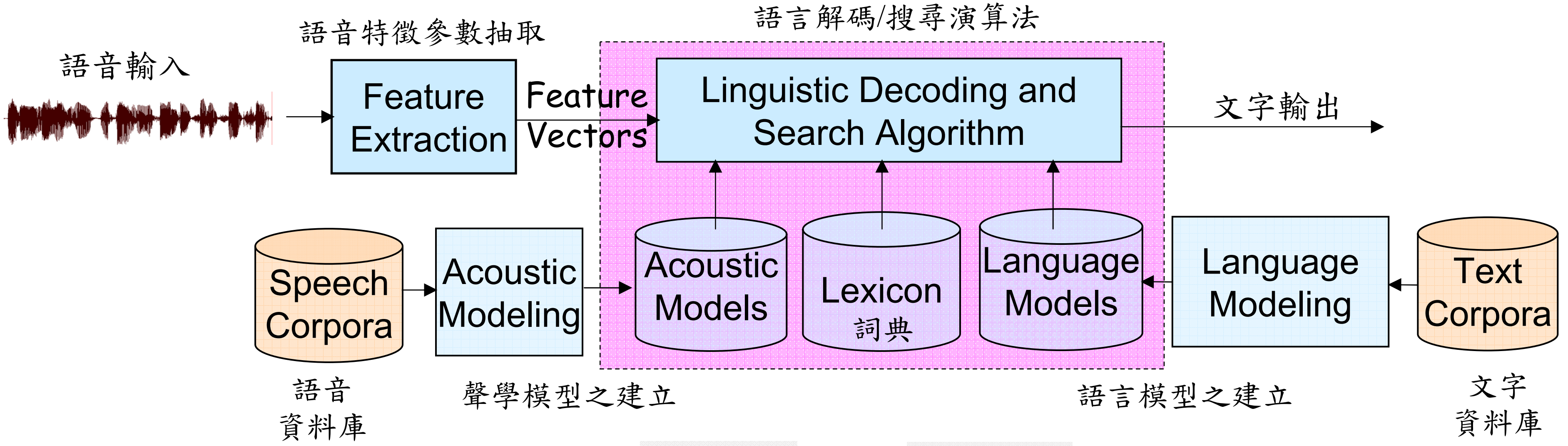
- Approaches based on Hidden Markov Models (HMMs) dominate the area of speech recognition
  - HMMs are based on rigorous mathematical theory built on several decades of mathematical results developed in other fields
  - HMMs are generated by the process of training on a large corpus of real speech data



Difficulties: Speech Variability


Pronunciation Variation
Speaker-independency
Speaker-adaptation
Speaker-dependency
Linguistic variability
Inter-speaker variability
Intra-speaker variability
Variability caused by the environment
Variability caused by the context
Robustness Enhancement
Context-Dependent Acoustic Modeling

Large Vocabulary Continuous Speech Recognition (LVCSR) (1/2)



語音輸入

語音特徵參數抽取

Feature Extraction

Feature Vectors

語言解碼/搜尋演算法

Linguistic Decoding and Search Algorithm

文字輸出

Speech Corpora

語音資料庫

Acoustic Modeling

聲學模型之建立

Acoustic Models

Lexicon 詞典

Language Models

Language Modeling

語言模型之建立

Text Corpora

文字資料庫

可能詞句　　語音輸入

$$\hat{\mathbf{W}} = \arg\max_{\mathbf{W}} P(\mathbf{W}|\mathbf{X})$$

貝氏定理

$$= \arg\max_{\mathbf{W}} \frac{P(\mathbf{X}|\mathbf{W})P(\mathbf{W})}{P(\mathbf{X})}$$

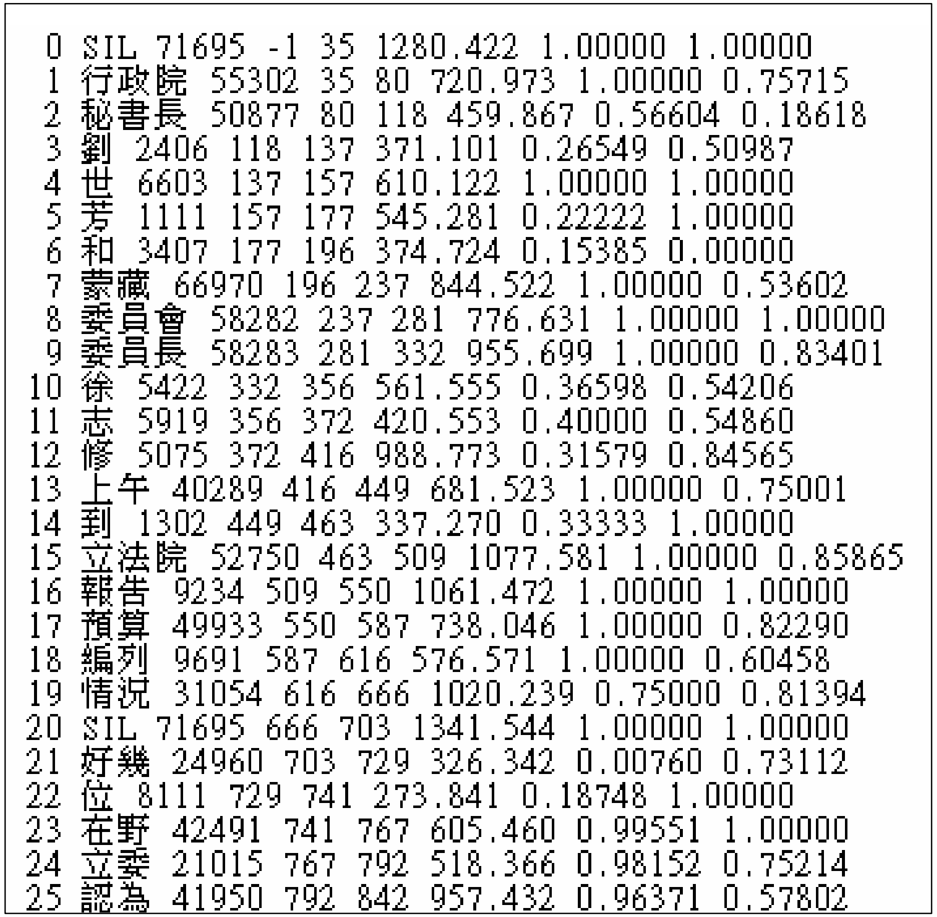
$$= \boxed{\arg\max_{\mathbf{W}} P(\mathbf{X}|\mathbf{W})P(\mathbf{W})}$$

詞彙網路搜尋

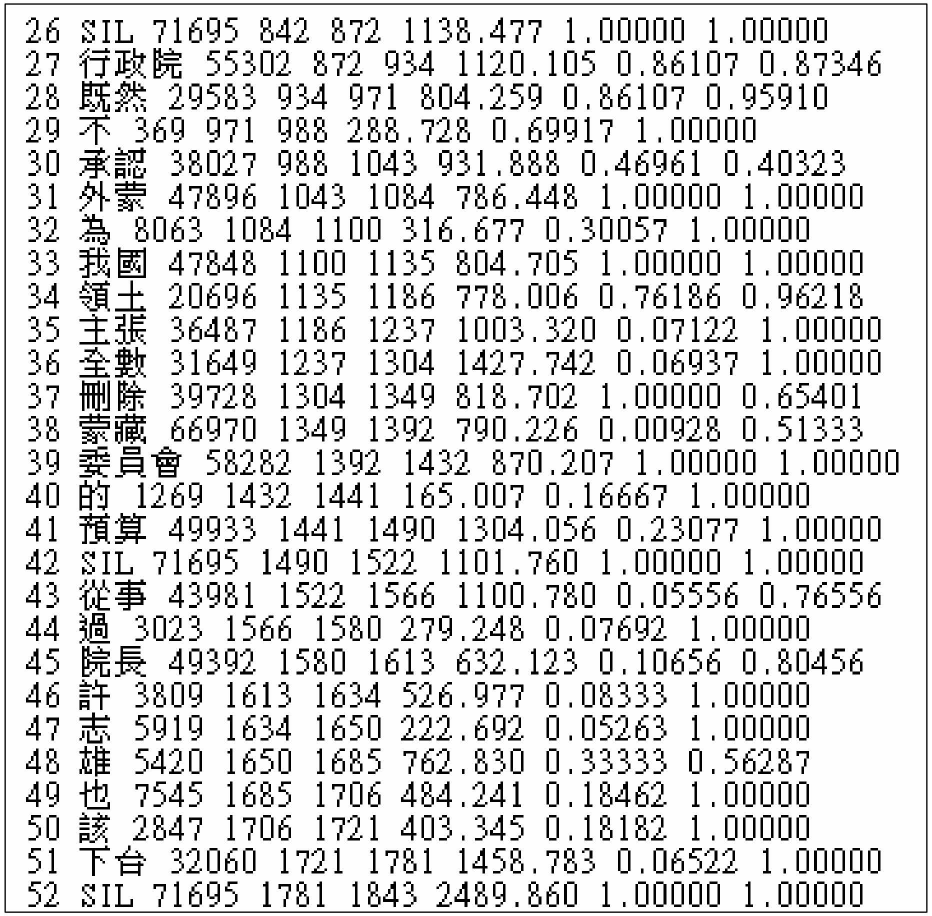
聲學模型機率　　語言模型機率

Large Vocabulary Continuous Speech Recognition (cont.) (2/2)

- Transcription of Broadcast News Speech



```
 0 SIL 71695 -1 35 1280.422 1.00000 1.00000
 1 行政院 55302 35 80 720.973 1.00000 0.75715
 2 秘書長 50877 80 118 459.867 0.56604 0.18618
 3 劉 2406 118 137 371.101 0.26549 0.50987
 4 世 6603 137 157 610.122 1.00000 1.00000
 5 芳 1111 157 177 545.281 0.22222 1.00000
 6 和 3407 177 196 374.724 0.15385 0.00000
 7 蒙藏 66970 196 237 844.522 1.00000 0.53602
 8 委員會 58282 237 281 776.631 1.00000 1.00000
 9 委員長 58283 281 332 955.699 1.00000 0.83401
10 徐 5422 332 356 561.555 0.36598 0.54206
11 志 5919 356 372 420.553 0.40000 0.54860
12 修 5075 372 416 988.773 0.31579 0.84565
13 上午 40289 416 449 681.523 1.00000 0.75001
14 到 1302 449 463 337.270 0.33333 1.00000
15 立法院 52750 463 509 1077.581 1.00000 0.85865
16 報告 9234 509 550 1061.472 1.00000 1.00000
17 預算 49933 550 587 738.046 1.00000 0.82290
18 編列 9691 587 616 576.571 1.00000 0.60458
19 情況 31054 616 666 1020.239 0.75000 0.81394
20 SIL 71695 666 703 1341.544 1.00000 1.00000
21 好幾 24960 703 729 326.342 0.00760 0.73112
22 位 8111 729 741 273.841 0.18748 1.00000
23 在野 42491 741 767 605.460 0.99551 1.00000
24 立委 21015 767 792 518.366 0.98152 0.75214
25 認為 41950 792 842 957.432 0.96371 0.57802
```



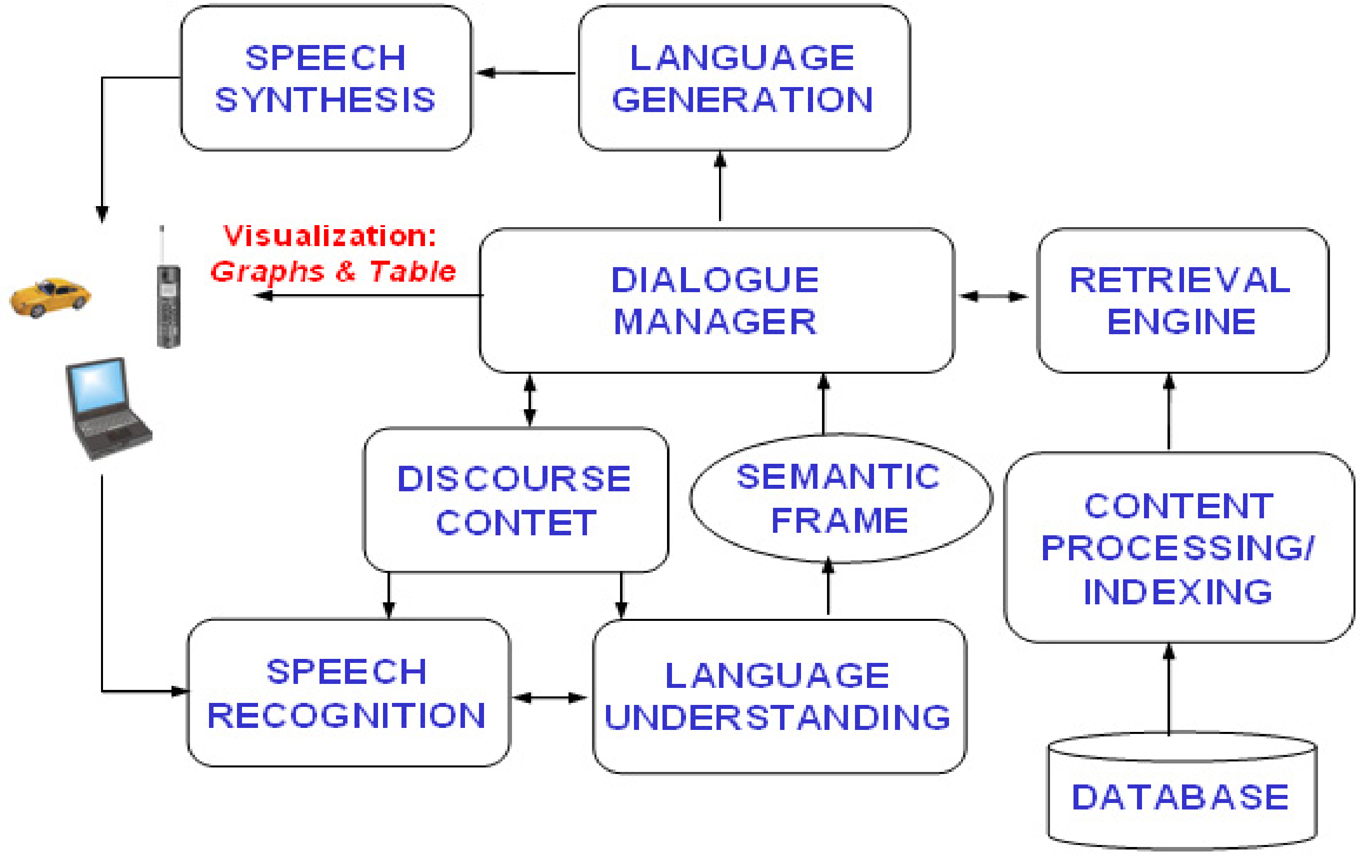
```
26 SIL 71695 842 872 1138.477 1.00000 1.00000
27 行政院 55302 872 934 1120.105 0.86107 0.87346
28 既然 29583 934 971 804.259 0.86107 0.95910
29 不 369 971 988 288.728 0.69917 1.00000
30 承認 38027 988 1043 931.888 0.46961 0.40323
31 外蒙 47896 1043 1084 786.448 1.00000 1.00000
32 為 8063 1084 1100 316.677 0.30057 1.00000
33 我國 47848 1100 1135 804.705 1.00000 1.00000
34 領土 20696 1135 1186 778.006 0.76186 0.96218
35 主張 36487 1186 1237 1003.320 0.07122 1.00000
36 全數 31649 1237 1304 1427.742 0.06937 1.00000
37 刪除 39728 1304 1349 818.702 1.00000 0.65401
38 蒙藏 66970 1349 1392 790.226 0.00928 0.51333
39 委員會 58282 1392 1432 870.207 1.00000 1.00000
40 的 1269 1432 1441 165.007 0.16667 1.00000
41 預算 49933 1441 1490 1304.056 0.23077 1.00000
42 SIL 71695 1490 1522 1101.760 1.00000 1.00000
43 從事 43981 1522 1566 1100.780 0.05556 0.76556
44 過 3023 1566 1580 279.248 0.07692 1.00000
45 院長 49392 1580 1613 632.123 0.10656 0.80456
46 許 3809 1613 1634 526.977 0.08333 1.00000
47 志 5919 1634 1650 222.692 0.05263 1.00000
48 雄 5420 1650 1685 762.830 0.33333 0.56287
49 也 7545 1685 1706 484.241 0.18462 1.00000
50 該 2847 1706 1721 403.345 0.18182 1.00000
51 下台 32060 1721 1781 1458.783 0.06522 1.00000
52 SIL 71695 1781 1843 2489.860 1.00000 1.00000
```

Spoken Dialogue (1/5)

- Spoken language is attractive because it is the most natural, convenient and inexpensive means of exchanging information for humans
- In mobilizing situations, using keystrokes and mouse clicks could be impractical for rapid information access through small handheld devices like PDAs, cellular phones, etc.

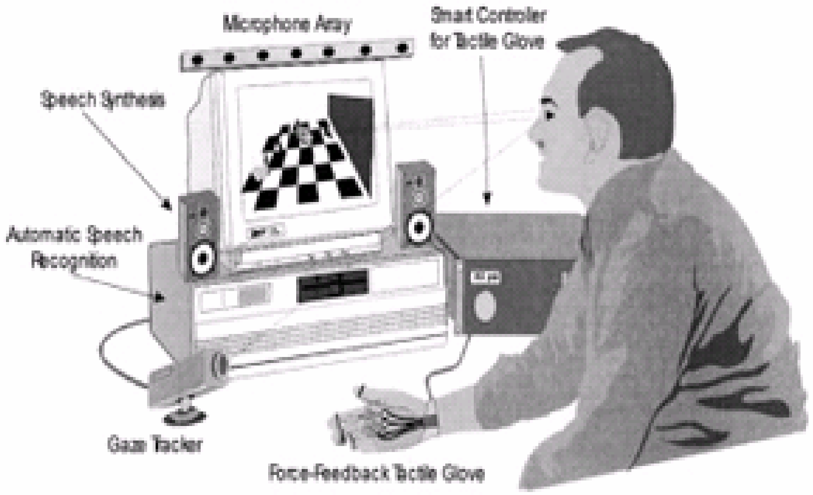
Spoken Dialogue (2/5)

- Flowchart



Spoken Dialogue (3/5)

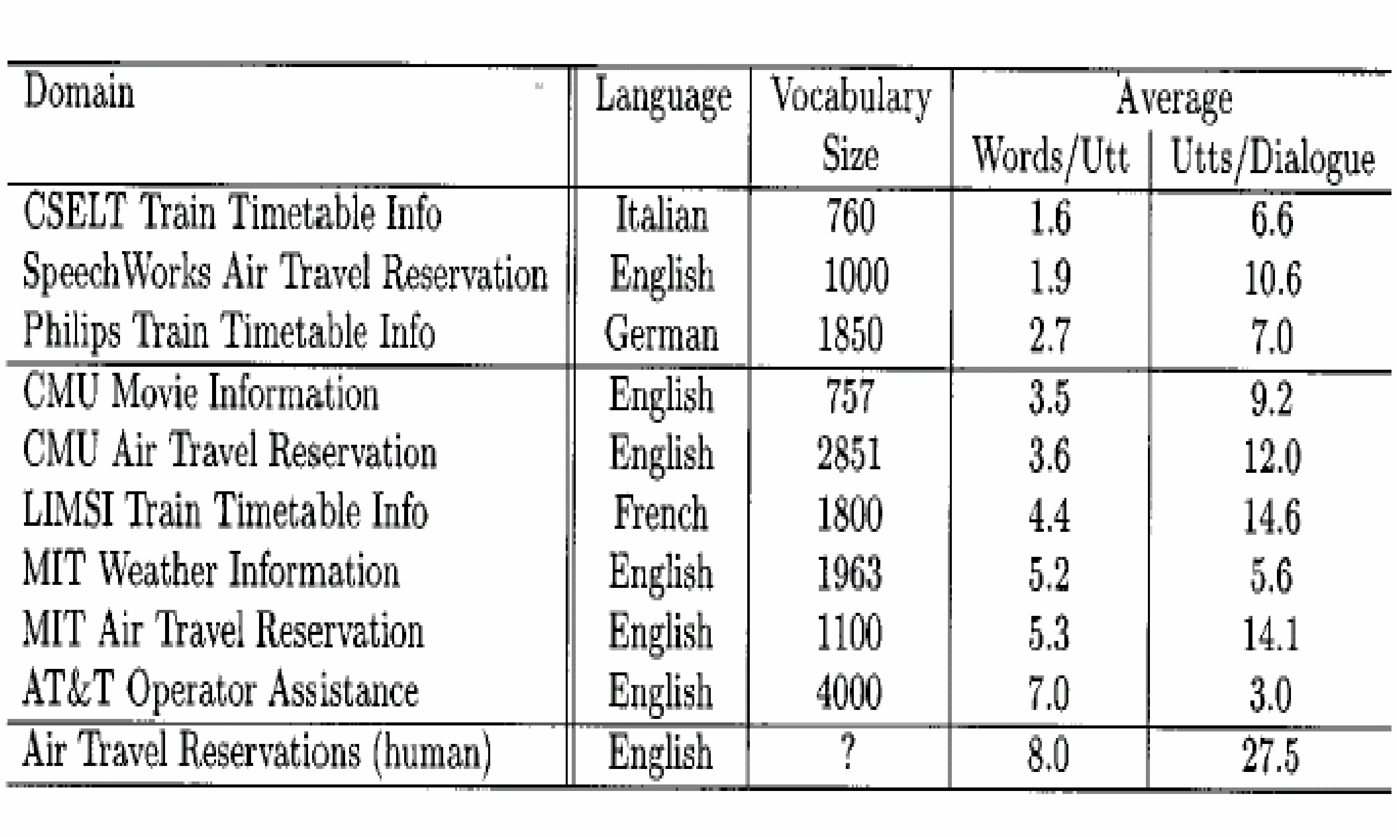
- Multimodality of Input and Output


Microphone Array
Smart Controller for Tactile Glove
Speech Synthesis
Automatic Speech Recognition
Gaze Tracker
Force-Feedback Tactile Glove

Experimental client workstation incorporating sight, sound, and touch modalities for human/machine communication. The eye tracker provides a gaze-controlled cursor for indicating objects in the display. The tactile force-feedback glove allows displayed objects to be grasped, “felt,” and moved. Hands-free speech recognition and synthesis provides natural conversational interaction [7].

Spoken Dialogue (4/5)

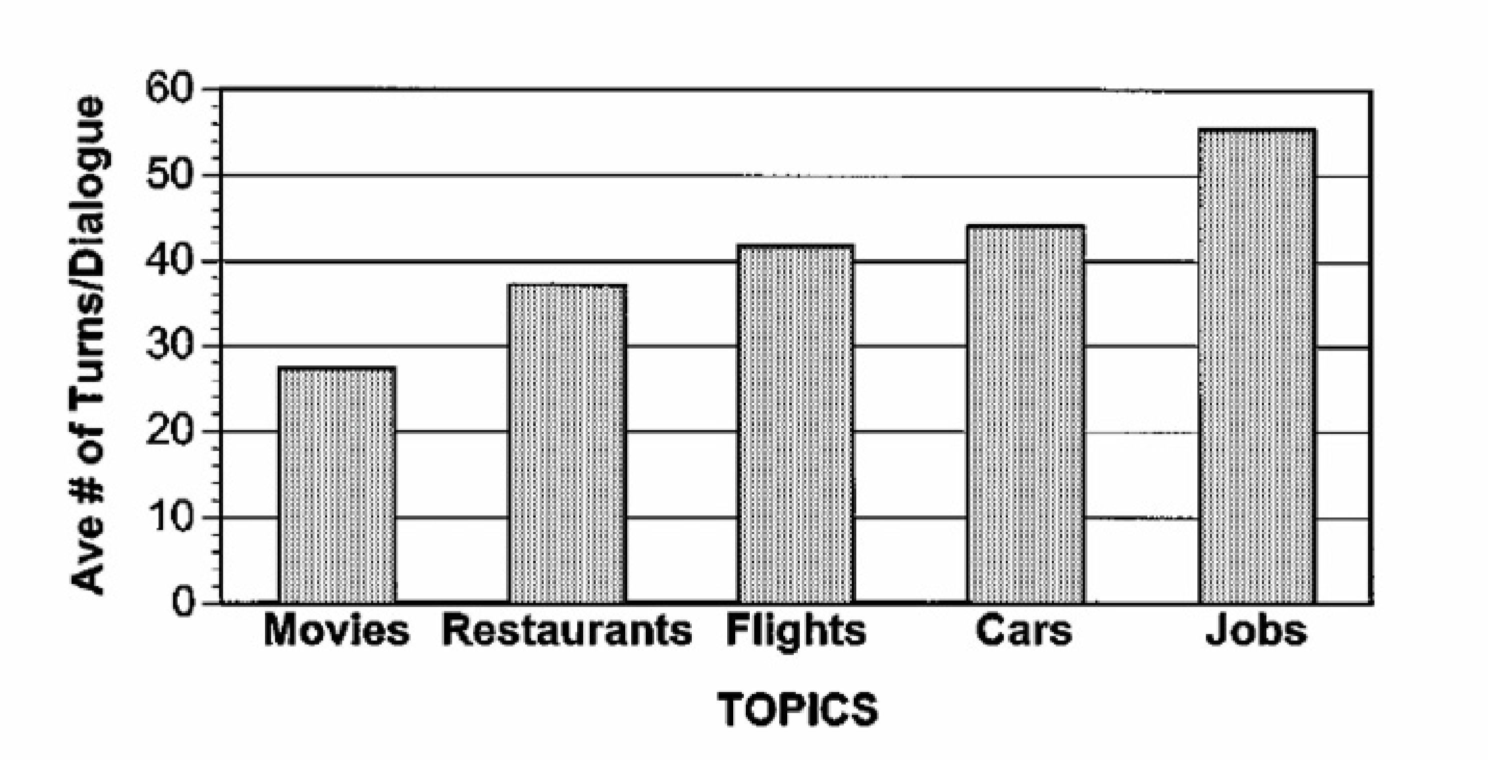
- Deployed Dialogue Systems



Domain	Language	Vocabulary Size	Average Words/Utt	Average Utts/Dialogue
CSELT Train Timetable Info	Italian	760	1.6	6.6
SpeechWorks Air Travel Reservation	English	1000	1.9	10.6
Philips Train Timetable Info	German	1850	2.7	7.0
CMU Movie Information	English	757	3.5	9.2
CMU Air Travel Reservation	English	2851	3.6	12.0
LIMSI Train Timetable Info	French	1800	4.4	14.6
MIT Weather Information	English	1963	5.2	5.6
MIT Air Travel Reservation	English	1100	5.3	14.1
AT&T Operator Assistance	English	4000	7.0	3.0
Air Travel Reservations (human)	English	?	8.0	27.5

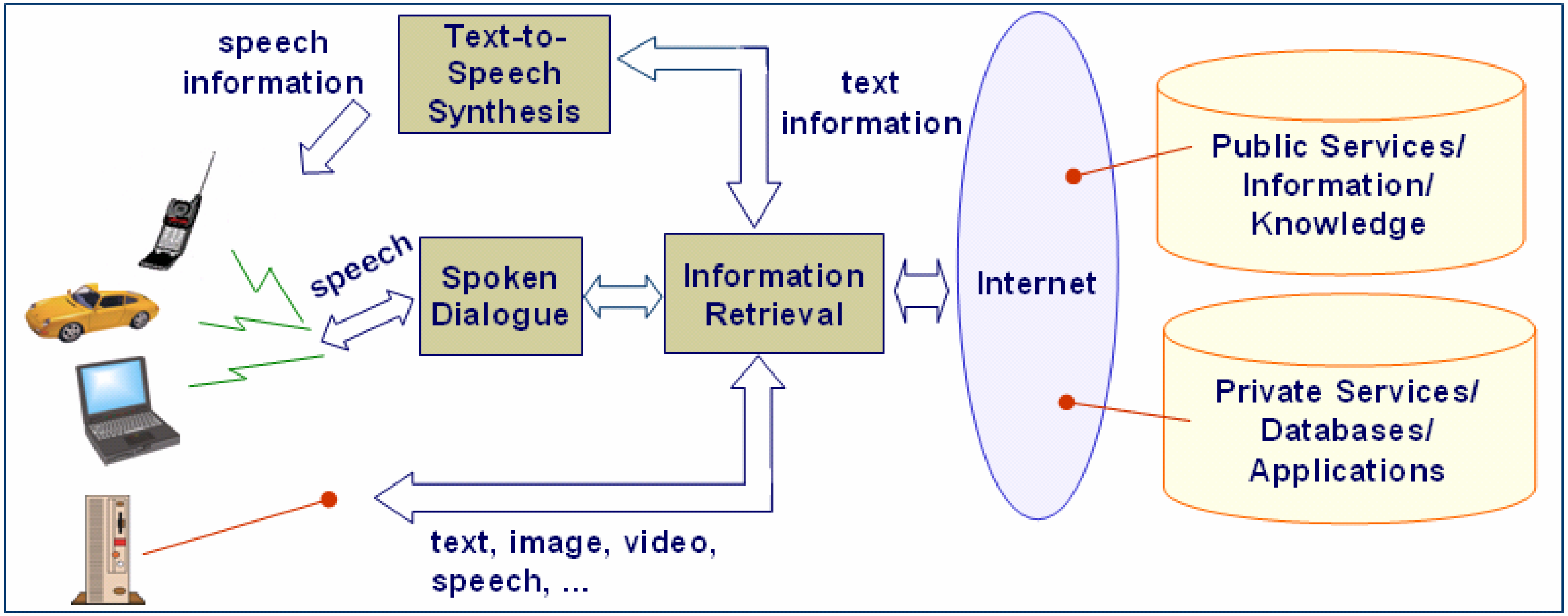
Spoken Dialogue (5/5)

- Topics vs. Dialogue Terms

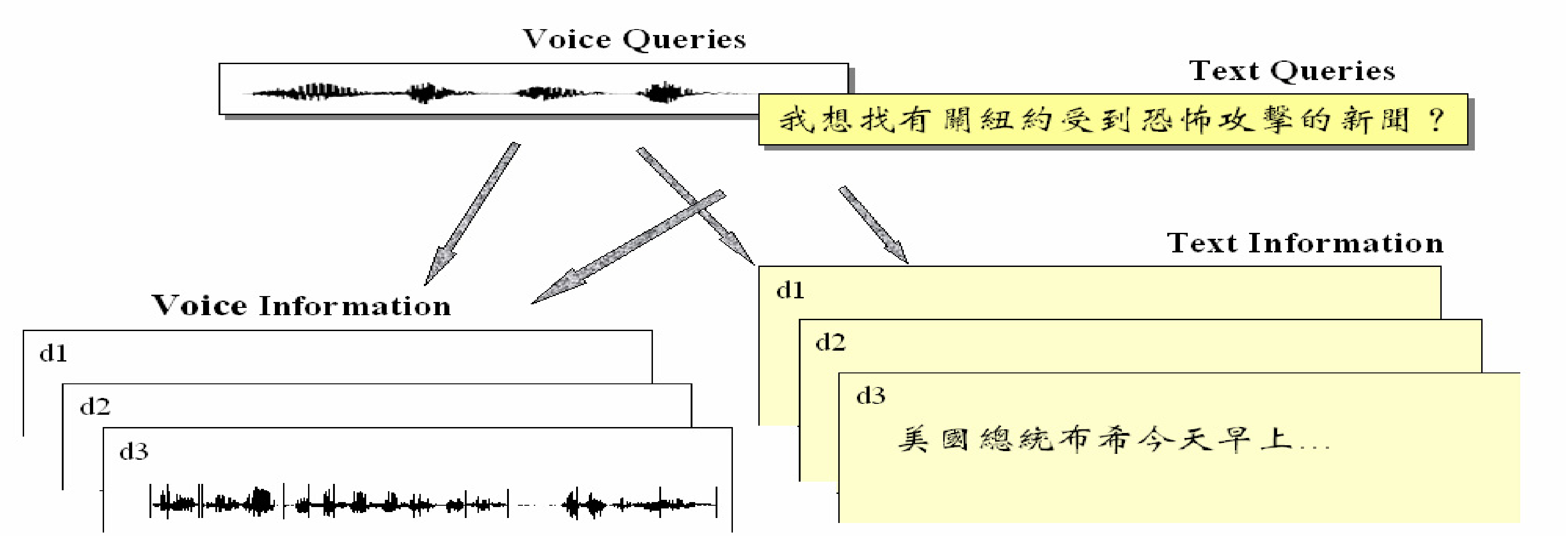


Speech-based Information Retrieval (1/5)

- Task :
  - Automatically indexing a collection of spoken documents with speech recognition techniques
  - Retrieving relevant documents in response to a text/speech query


speech
information
Text-to-Speech Synthesis
text
information
speech
Spoken Dialogue
Information Retrieval
Internet
Public Services/ Information/ Knowledge
Private Services/ Databases/ Applications
text, image, video, speech, ...

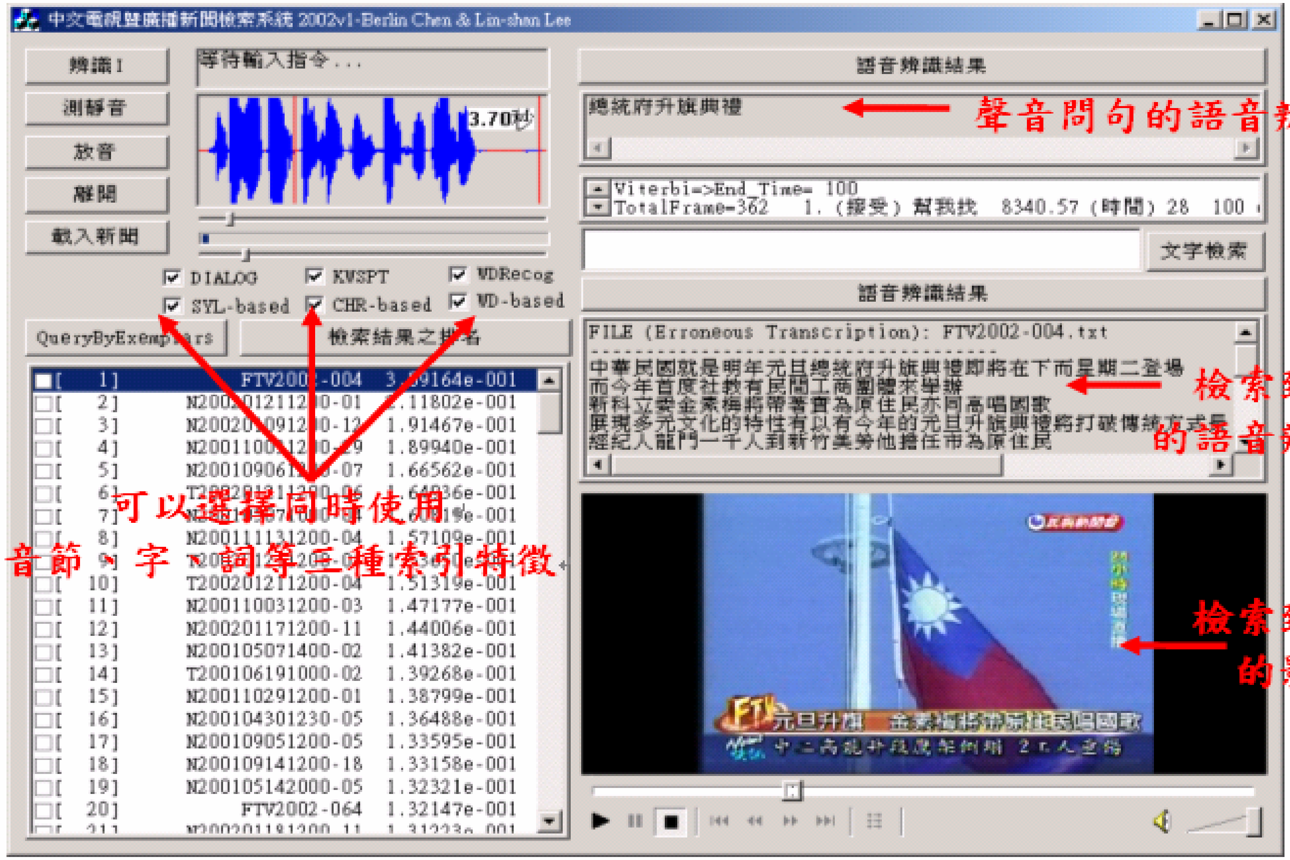
Speech-based Information Retrieval (2/5)



在四種不同時機下的資訊檢索過程。使用聲音問句(VQ，Voice Queries)或文字問句(TQ，Text Queries)去檢索聲音資訊(VI，Voice Information)或者是傳統的文字資訊(TI，Text Information)。

Speech-based Information Retrieval (3/5)

輸入聲音問句："請幫我查總統府升旗典禮"



中文電視暨廣播新聞檢索系統 2002v1-Berlin Chen & Lin-shan Lee

等待輸入指令...

☑ DIALOG ☑ KWSPT ☑ WDRecog
☑ SYL-based ☑ CHR-based ☑ WD-based

QueryByExemplars　檢索結果之排名

```
□[  1]      FTV2002-004  3.[illegible]9164e-001
□[  2] N2002012112[illegible]0-01  [illegible].11802e-001
□[  3] N20020[illegible]0912[illegible]0-12  1.91467e-001
□[  4] N20011[illegible]12[illegible]0-29  1.89940e-001
□[  5] N20010906[illegible]-07  1.66562e-001
□[  6] T2002012[illegible]1200-06  1.64[illegible]36e-001
□[  7] [illegible]  1.6[illegible]e-001
□[  8] N200111131200-04  1.57109e-001
□[  9] [illegible]  [illegible]
□[ 10] T200201211200-04  1.51319e-001
□[ 11] N200110031200-03  1.47177e-001
□[ 12] N200201171200-11  1.44006e-001
□[ 13] N200105071400-02  1.41382e-001
□[ 14] T200106191000-02  1.39268e-001
□[ 15] N200110291200-01  1.38799e-001
□[ 16] N200104301230-05  1.36488e-001
□[ 17] N200109051200-05  1.33595e-001
□[ 18] N200109141200-18  1.33158e-001
□[ 19] N200105142000-05  1.32321e-001
□[ 20]      FTV2002-064  1.32147e-001
□[ 21] N200201181200-11  1.31223e-001
```

語音辨識結果

總統府升旗典禮

```
Viterbi=>End_Time= 100
TotalFrame=362    1.（接受）幫我找  8340.57（時間）28   100
```

語音辨識結果

```
FILE (Erroneous Transcription): FTV2002-004.txt
-----------------------------------------
中華民國就是明年元旦總統府升旗典禮即將在下而星期二登場
而今年首度社教有民間工商團體來舉辦
新科立委金素梅將帶著會為原住民亦同高唱國歌
展現多元文化的特性有以有今年的元旦升旗典禮將打破傳統方式是
經紀人龍門一千人到新竹美勞他擔任市為原住民
```

聲音問句的語音辨識結果

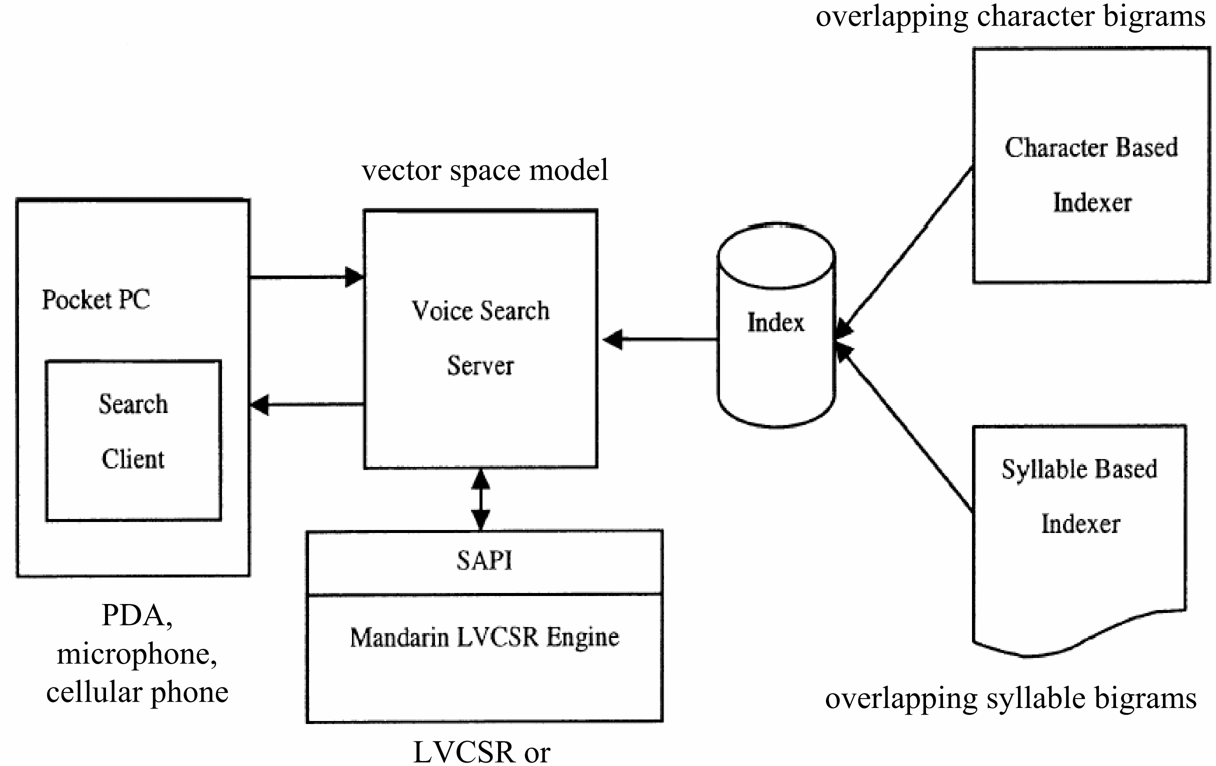
檢索到新聞的語音辨識結果

可以選擇同時使用音節、字、詞等三種索引特徵

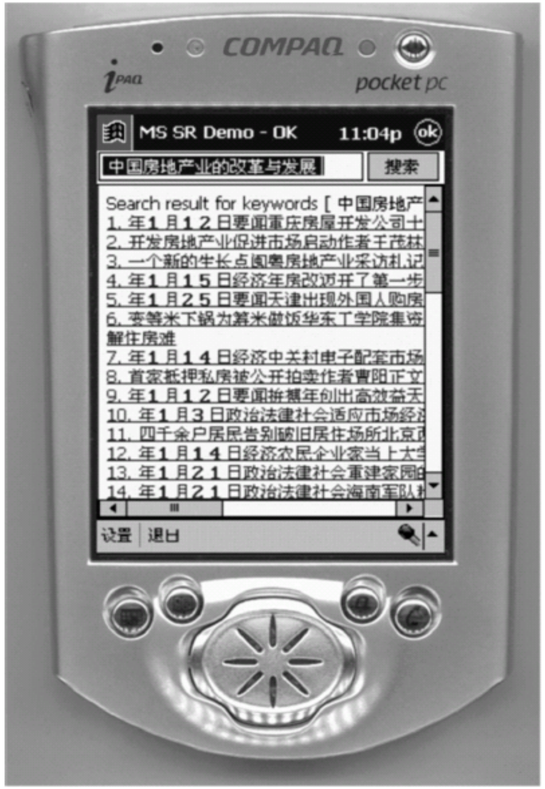
檢索到新聞的影音

中文語音資訊檢索雛形展示系統。

Speech-based Information Retrieval (3/5)



LVCSR or
syllable decoding



MS SR Demo - OK 11:04p ok

中国房地产业的改革与发展 搜索

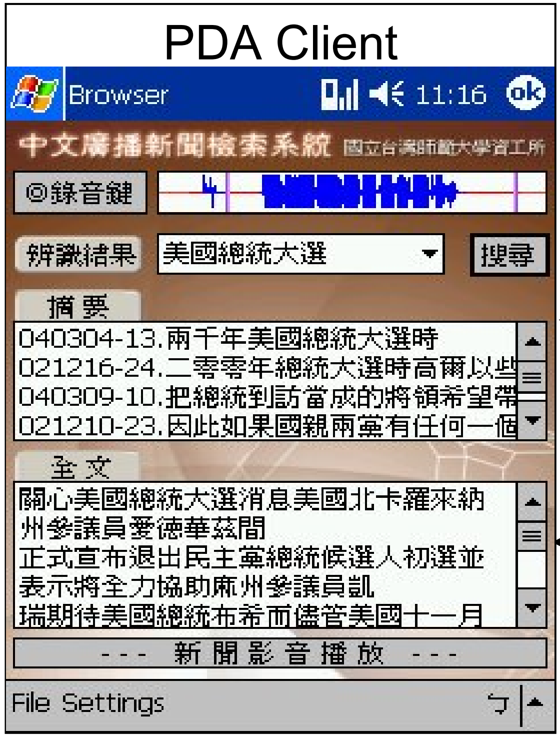
Search result for keywords [中国房地产

1. 年１月１２日要闻重庆房屋开发公司十
2. 开发房地产业促进市场启动作者王茂林
3. 一个新的生长点闽粤房地产业采访札记
4. 年１月１５日经济年房改迈开了第一步
5. 年１月２５日要闻天津出现外国人购房
6. 变等米下锅为筹米做饭华东工学院集资解住房难
7. 年１月１４日经济中关村电子配套市场
8. 首家抵押私房被公开拍卖作者曹阳正文
9. 年１月１２日要闻拚搏年创出高效益天
10. 年１月３日政治法律社会适应市场经
11. 四千余户居民告别破旧居住场所北京
12. 年１月１４日经济农民企业家当上大
13. 年１月２１日政治法律社会重建家园
14. 年１月２１日政治法律社会海南军队

设置 退出

Speech-based Information Retrieval (4/5)

- PDA-based IR system for Mandarin broadcast news



中文廣播新聞檢索系統 國立台灣師範大學資工所

◎錄音鍵

辨識結果 美國總統大選 搜尋

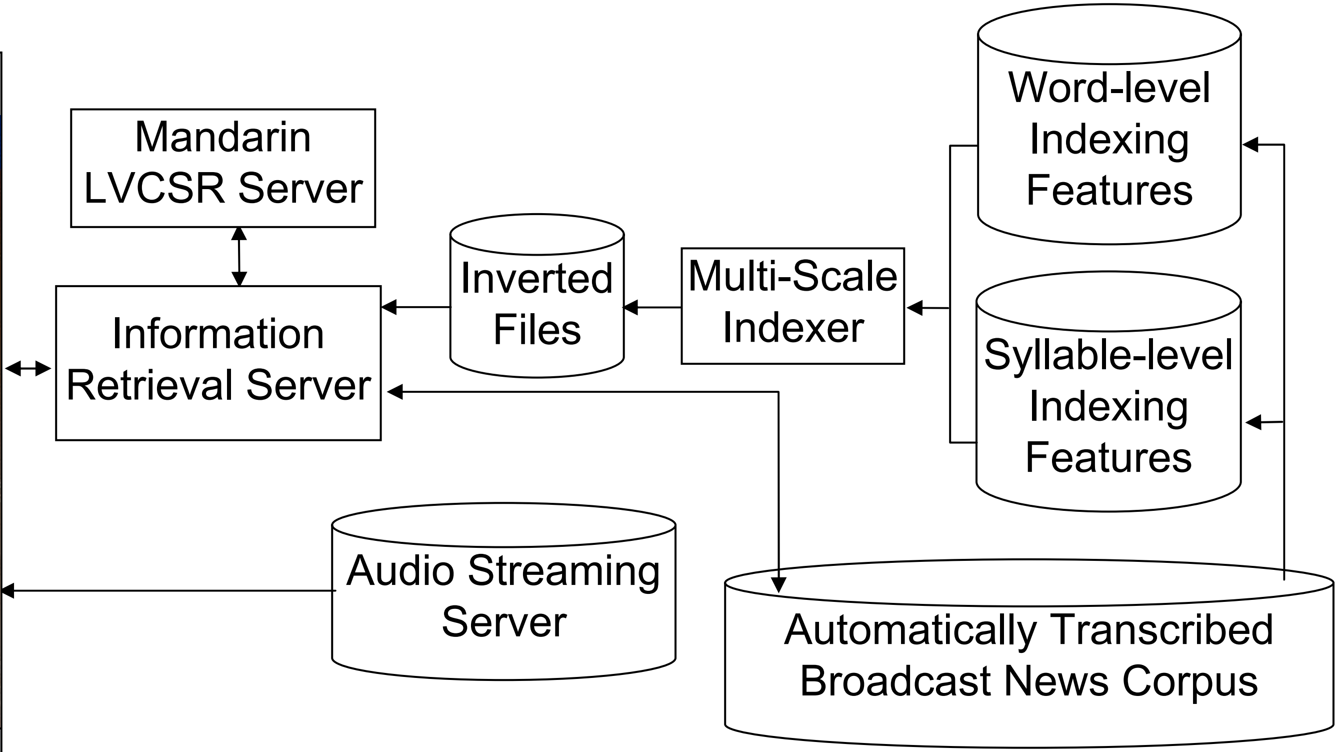
摘要

040304-13.兩千年美國總統大選時
021216-24.二零零年總統大選時高爾以些
040309-10.把總統到訪當成的將領希望帶
021210-23.因此如果國親兩黨有任何一個

全文

關心美國總統大選消息美國北卡羅來納
州參議員愛德華茲間
正式宣布退出民主黨總統候選人初選並
表示將全力協助麻州參議員凱
瑞期待美國總統布希而儘管美國十一月

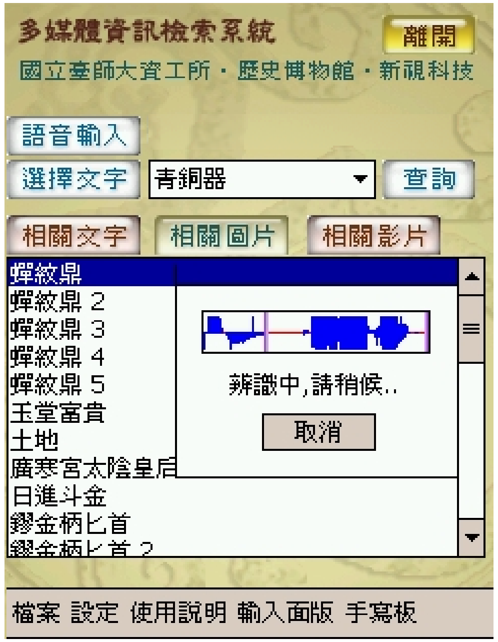
--- 新聞影音播放 ---



Mandarin LVCSR Server
Information Retrieval Server
Inverted Files
Multi-Scale Indexer
Word-level Indexing Features
Syllable-level Indexing Features
Audio Streaming Server
Automatically Transcribed Broadcast News Corpus

Speech-based Information Retrieval (5/5)

- PDA-based IR system for digital archives
  - Current deployed at National Museum of History, Taipei



多媒體資訊檢索系統 離開

國立臺師大資工所・歷史博物館・新視科技

語音輸入

選擇文字 青銅器 查詢

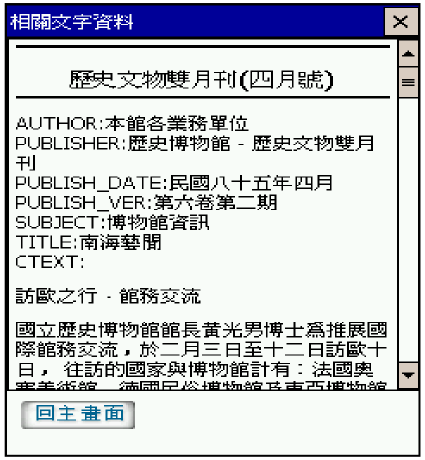
相關文字 相關圖片 相關影片

蟬紋鼎
蟬紋鼎 2
蟬紋鼎 3
蟬紋鼎 4
蟬紋鼎 5
玉堂富貴
土地
廣寒宮太陰皇后
日進斗金
鎏金柄匕首
鎏金柄匕首 2

辨識中,請稍候..

取消

檔案 設定 使用說明 輸入面版 手寫板



相關文字資料

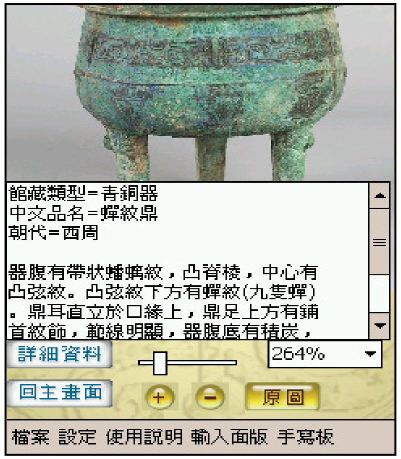
歷史文物雙月刊(四月號)

AUTHOR:本館各業務單位
PUBLISHER:歷史博物館 - 歷史文物雙月刊
PUBLISH_DATE:民國八十五年四月
PUBLISH_VER:第六卷第二期
SUBJECT:博物館資訊
TITLE:南海藝聞
CTEXT:

訪歐之行 - 館務交流

國立歷史博物館館長黃光男博士為推展國際館務交流，於二月三日至十二日訪歐十日， 往訪的國家與博物館計有：法國奧

回主畫面



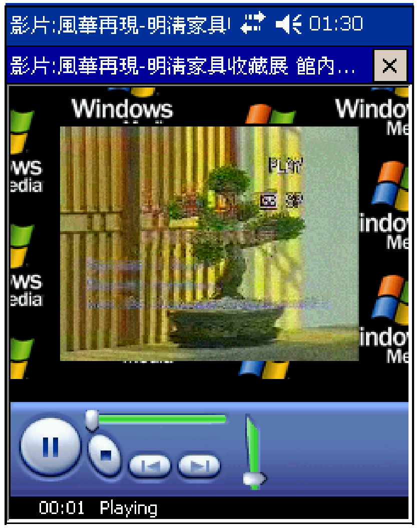
館藏類型=青銅器
中文品名=蟬紋鼎
朝代=西周

器腹有帶狀蟠螭紋，凸脊棱，中心有凸弦紋。凸弦紋下方有蟬紋(九隻蟬)。鼎耳直立於口緣上，鼎足上方有鋪首紋飾，範線明顯，器腹底有積炭，

詳細資料 264%

回主畫面 + − 原圖

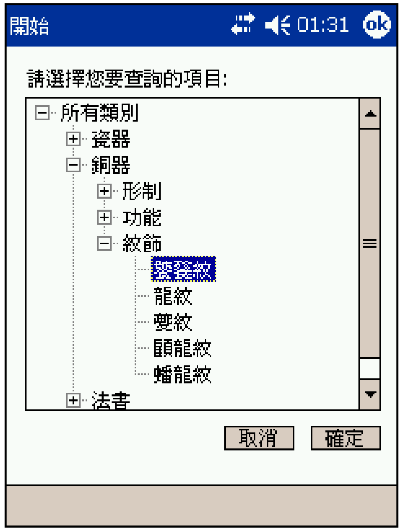
檔案 設定 使用說明 輸入面版 手寫板



影片:風華再現-明清家具! 01:30

影片:風華再現-明清家具收藏展 館內...

00:01 Playing



開始 01:31 ok

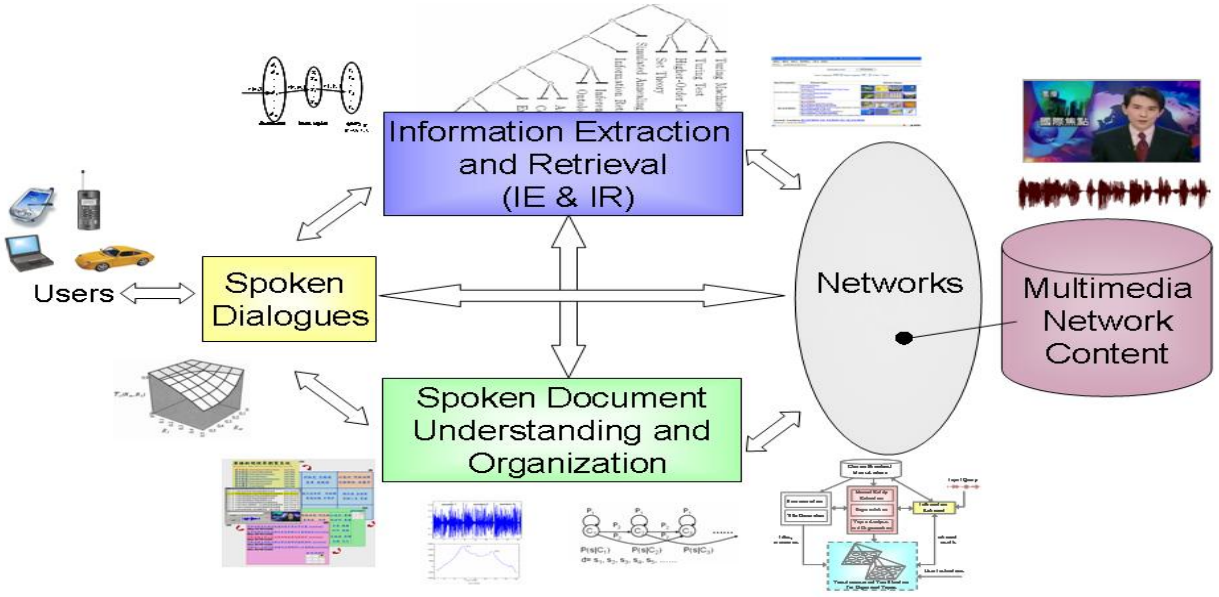
請選擇您要查詢的項目:

- 所有類別
  - 瓷器
  - 銅器
    - 形制
    - 功能
    - 紋飾
      - 饕餮紋
      - 龍紋
      - 夔紋
      - 顧龍紋
      - 蟠龍紋
  - 法書

取消 確定

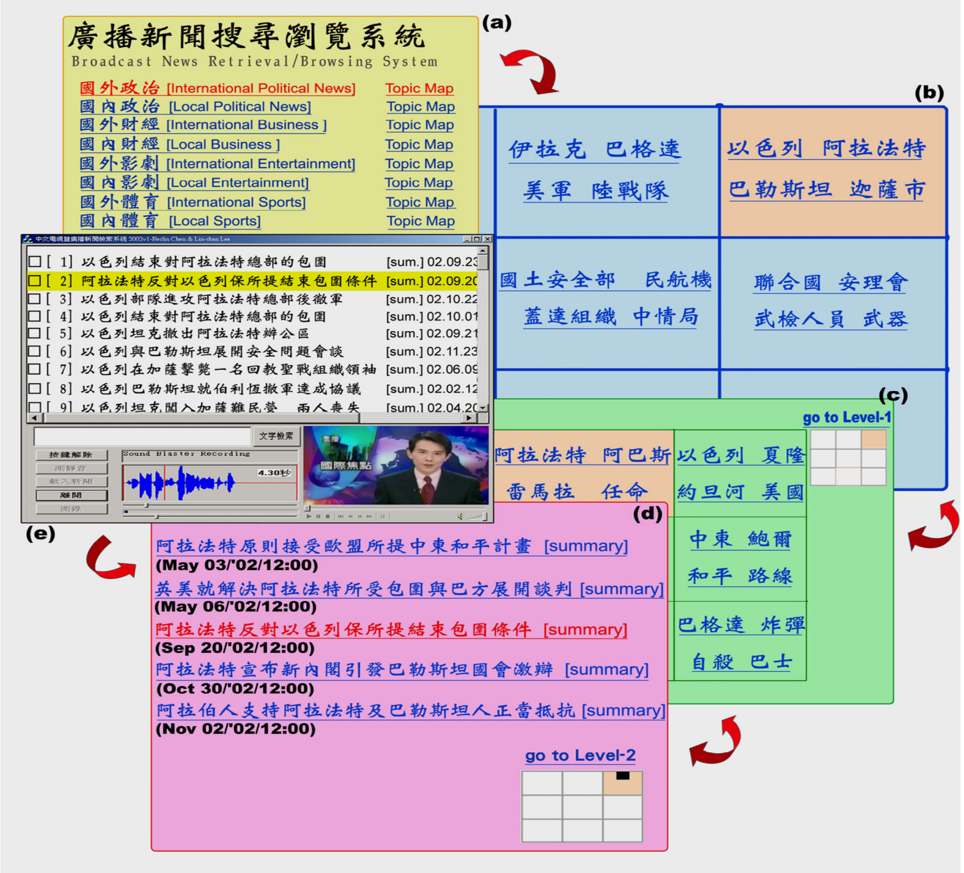
Spoken Document Organization and Understanding (1/2)

- Problem
  - The content of multimedia documents very often described by the associated speech information
  - Unlike text documents with paragraphs/titles easy to look through at a glance, multimedia/spoken documents are unstructured and difficult to retrieve/browse


Information Extraction and Retrieval (IE & IR)
Users
Spoken Dialogues
Networks
Multimedia Network Content
Spoken Document Understanding and Organization

Spoken Document Organization and Understanding (2/2)

- For example, spoken documents can be clustered by the latent topics and organized in a two-dimensional tree structure, or a two-layer map



廣播新聞搜尋瀏覽系統
Broadcast News Retrieval/Browsing System

國外政治 [International Political News] Topic Map
國內政治 [Local Political News] Topic Map
國外財經 [International Business] Topic Map
國內財經 [Local Business] Topic Map
國外影劇 [International Entertainment] Topic Map
國內影劇 [Local Entertainment] Topic Map
國外體育 [International Sports] Topic Map
國內體育 [Local Sports] Topic Map

中文電視暨廣播新聞檢索系統 2002v1-Berlin Chen & Lin-shan Lee

□ [1] 以色列結束對阿拉法特總部的包圍 [sum.] 02.09.23
□ [2] 阿拉法特反對以色列保所提結束包圍條件 [sum.] 02.09.20
□ [3] 以色列部隊進攻阿拉法特總部後撤軍 [sum.] 02.10.22
□ [4] 以色列結束對阿拉法特總部的包圍 [sum.] 02.10.01
□ [5] 以色列坦克撤出阿拉法特辦公區 [sum.] 02.09.21
□ [6] 以色列與巴勒斯坦展開安全問題會談 [sum.] 02.11.23
□ [7] 以色列在加薩擊斃一名回教聖戰組織領袖 [sum.] 02.06.09
□ [8] 以色列巴勒斯坦就伯利恆撤軍達成協議 [sum.] 02.02.12
□ [9] 以色列坦克闖入加薩難民營　兩人喪失 [sum.] 02.04.20

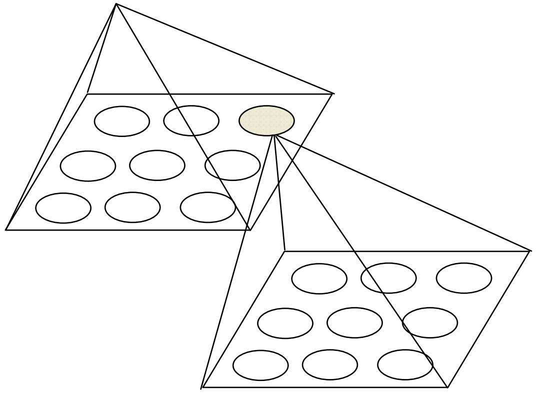
伊拉克 巴格達 美軍 陸戰隊
以色列 阿拉法特 巴勒斯坦 迦薩市
國土安全部 民航機 蓋達組織 中情局
聯合國 安理會 武檢人員 武器

go to Level-1

阿拉法特 阿巴斯 雷馬拉 任命
以色列 夏隆 約旦河 美國
中東 鮑爾 和平 路線
巴格達 炸彈 自殺 巴士

阿拉法特原則接受歐盟所提中東和平計畫 [summary]
(May 03/'02/12:00)
英美就解決阿拉法特所受包圍與巴方展開談判 [summary]
(May 06/'02/12:00)
阿拉法特反對以色列保所提結束包圍條件 [summary]
(Sep 20/'02/12:00)
阿拉法特宣布新內閣引發巴勒斯坦國會激辯 [summary]
(Oct 30/'02/12:00)
阿拉伯人支持阿拉法特及巴勒斯坦人正當抵抗 [summary]
(Nov 02/'02/12:00)

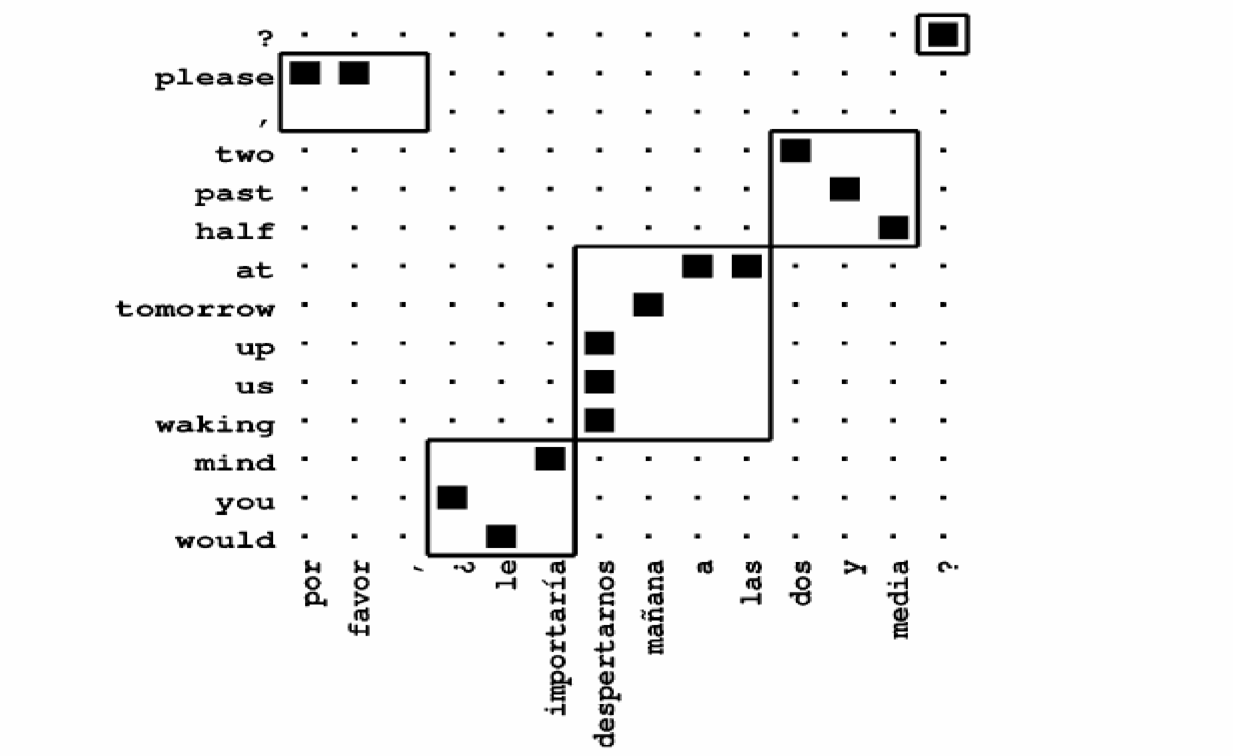
go to Level-2



Two-dimensional Tree Structure for Organized Topics

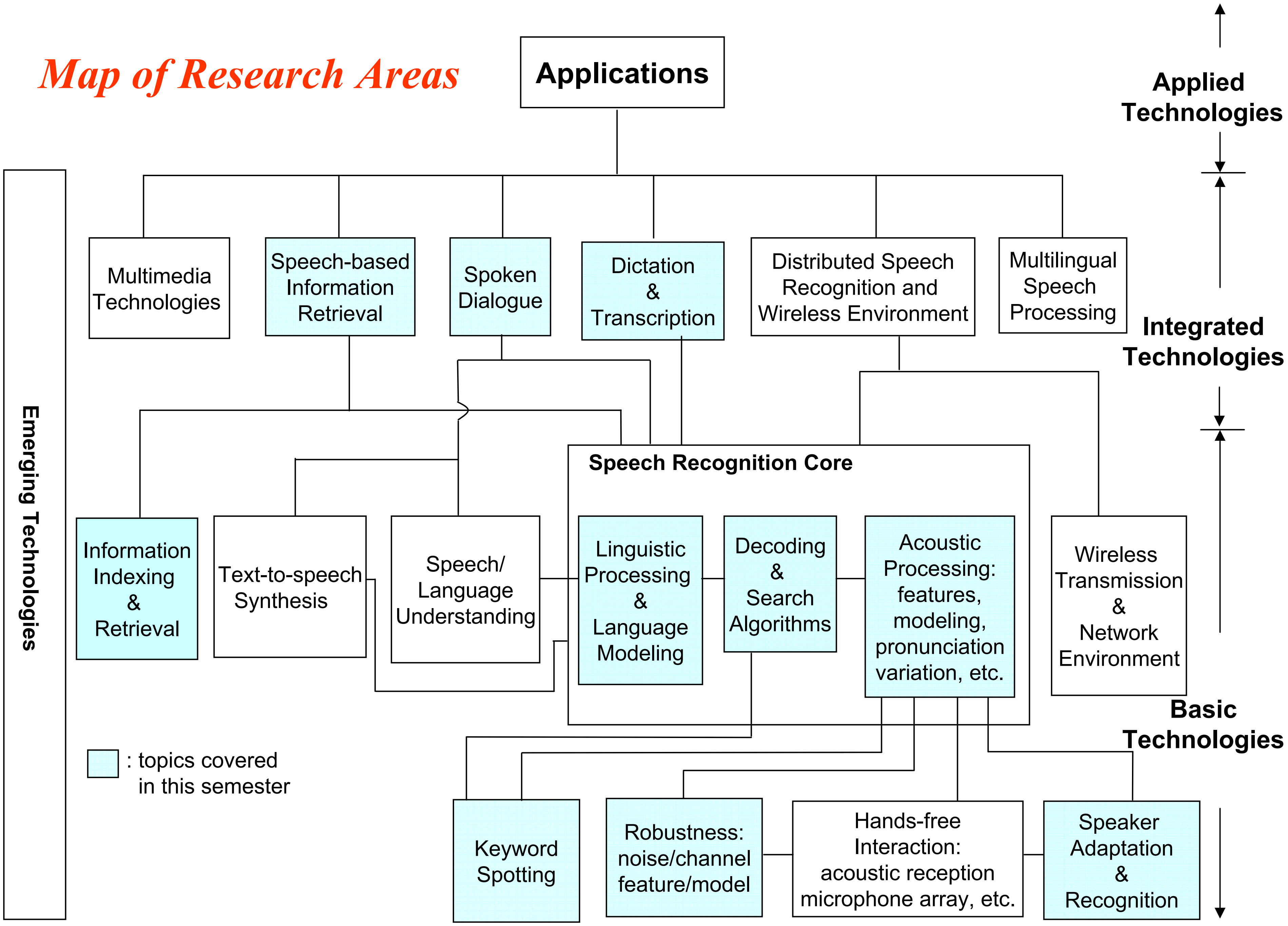
Speech-to-Speech Translation

- Multilingual interactive speech translation
  - Aims at the achievement of a communication system for precise recognition and translation of spoken utterances for several conversational topics and environments by using human language knowledge synthetically (adopted form ATR-SLT)



Example of a word alignment and of extracted alignment templates.

Map of Research Areas



Applications

Applied Technologies

Multimedia Technologies

Speech-based Information Retrieval

Spoken Dialogue

Dictation & Transcription

Distributed Speech Recognition and Wireless Environment

Multilingual Speech Processing

Integrated Technologies

Emerging Technologies

Speech Recognition Core

Information Indexing & Retrieval

Text-to-speech Synthesis

Speech/ Language Understanding

Linguistic Processing & Language Modeling

Decoding & Search Algorithms

Acoustic Processing: features, modeling, pronunciation variation, etc.

Wireless Transmission & Network Environment

Basic Technologies

Keyword Spotting

Robustness: noise/channel feature/model

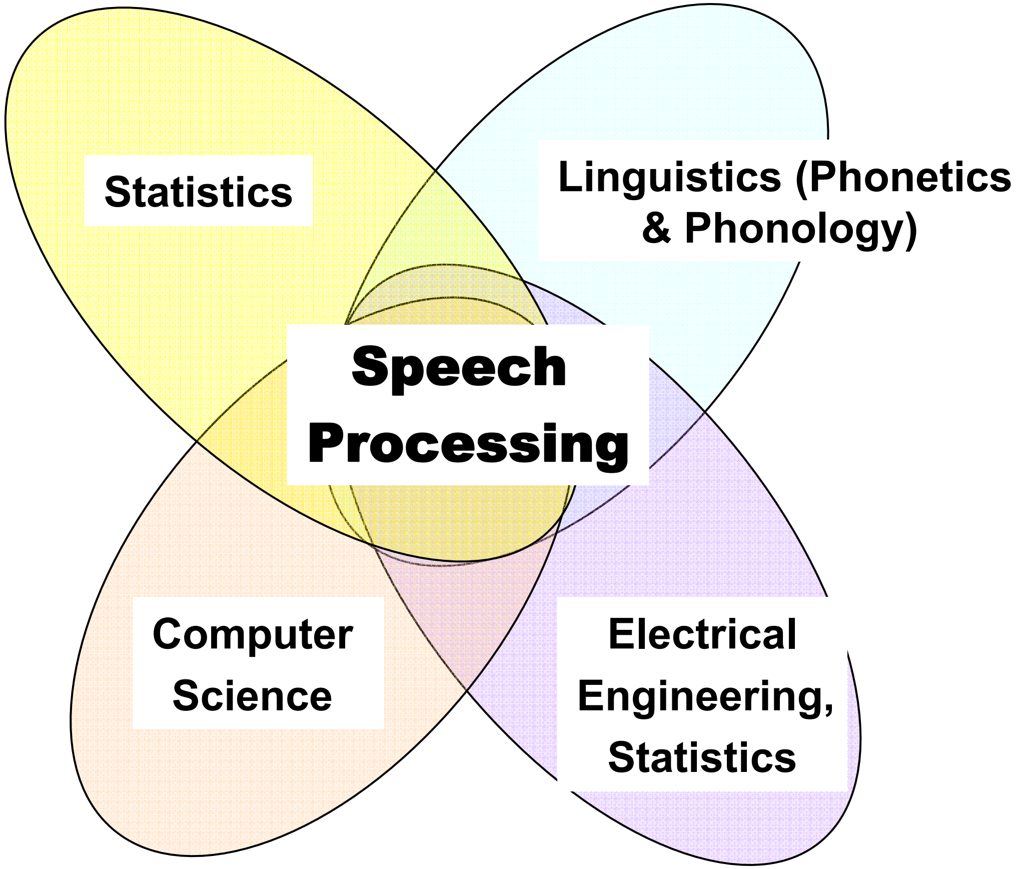
Hands-free Interaction: acoustic reception microphone array, etc.

Speaker Adaptation & Recognition

□ : topics covered in this semester

Adapted from Prof. Lin-shan Lee

Different Academic Disciplines

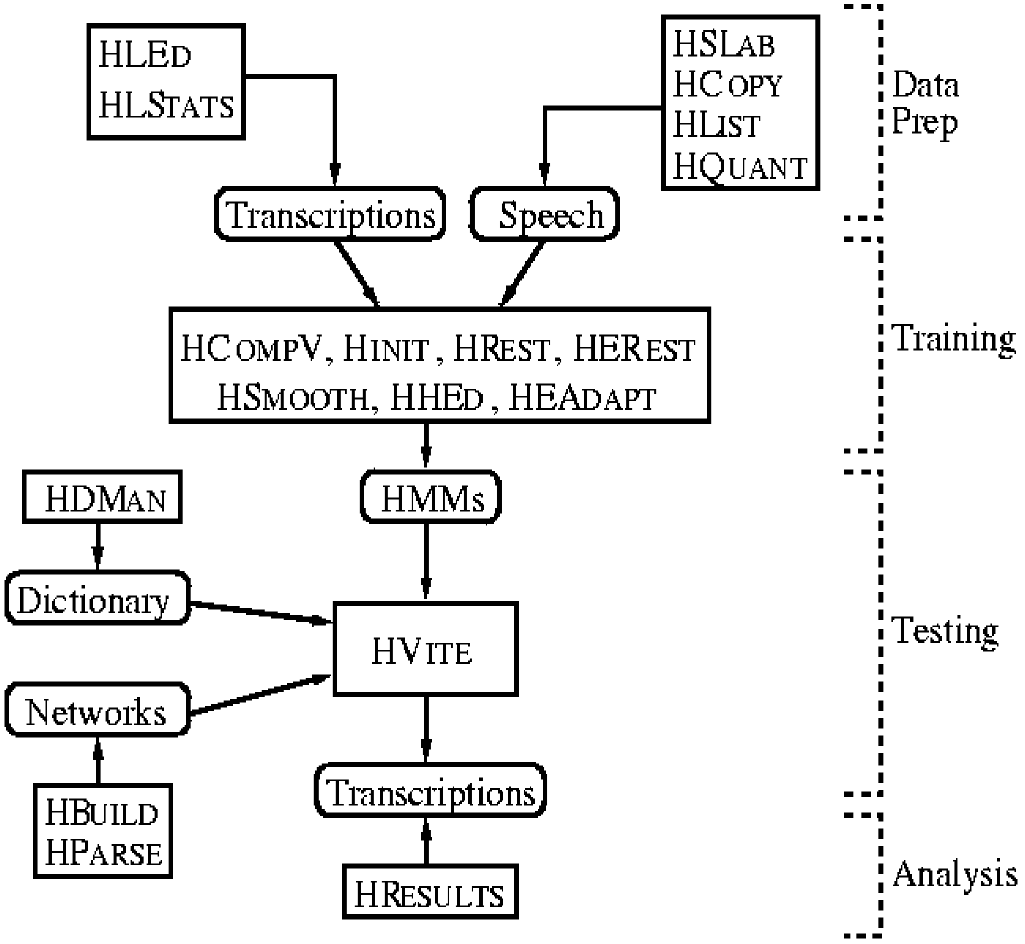

Statistics
Linguistics (Phonetics & Phonology)
Speech Processing
Computer Science
Electrical Engineering, Statistics

Speech Processing Toolkit (1/2)

- HTK (**H**idden Markov Model **T**ool**K**it)
  - A toolkit for building Hidden Markov Models (HMMs)
  - The HMM can be used to model any time series and the core of HTK is similarly general-purpose
  - In particular, for the acoustic feature extraction, HMM-based acoustic model training and HMM network decoding

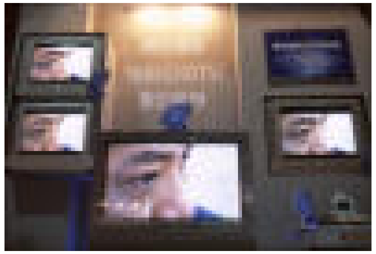
Speech Processing Toolkit (2/2)

- HTK (**H**idden Markov Model **T**ool**K**it)



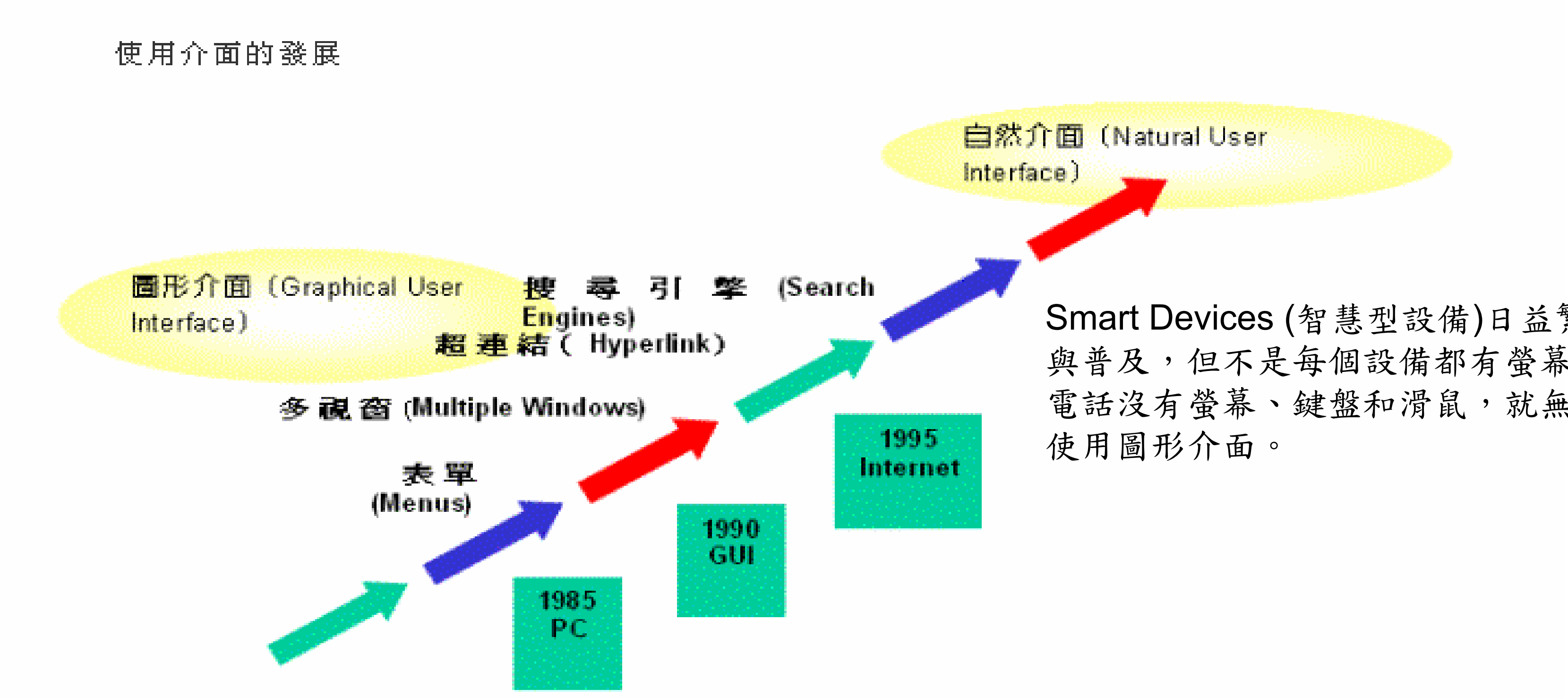
Speech Industry (1/3)

- Telecommunication
- Information Appliance
- Interactive Voice Response
- Voice Portal
- Multimedia Database
- Education
- …..



Speech Industry (2/3)

- Microsoft: Smart Device/Natural UI


使用介面的發展
自然介面（Natural User Interface）
圖形介面（Graphical User Interface）
搜 尋 引 擎 (Search Engines)
超 連 結（ Hyperlink）
多 視 窗 (Multiple Windows)
表 單 (Menus)
1985 PC
1990 GUI
1995 Internet

Source：微軟自然互動服務產品部門 (NISD)副總裁李開複博士講稿， 2003/04

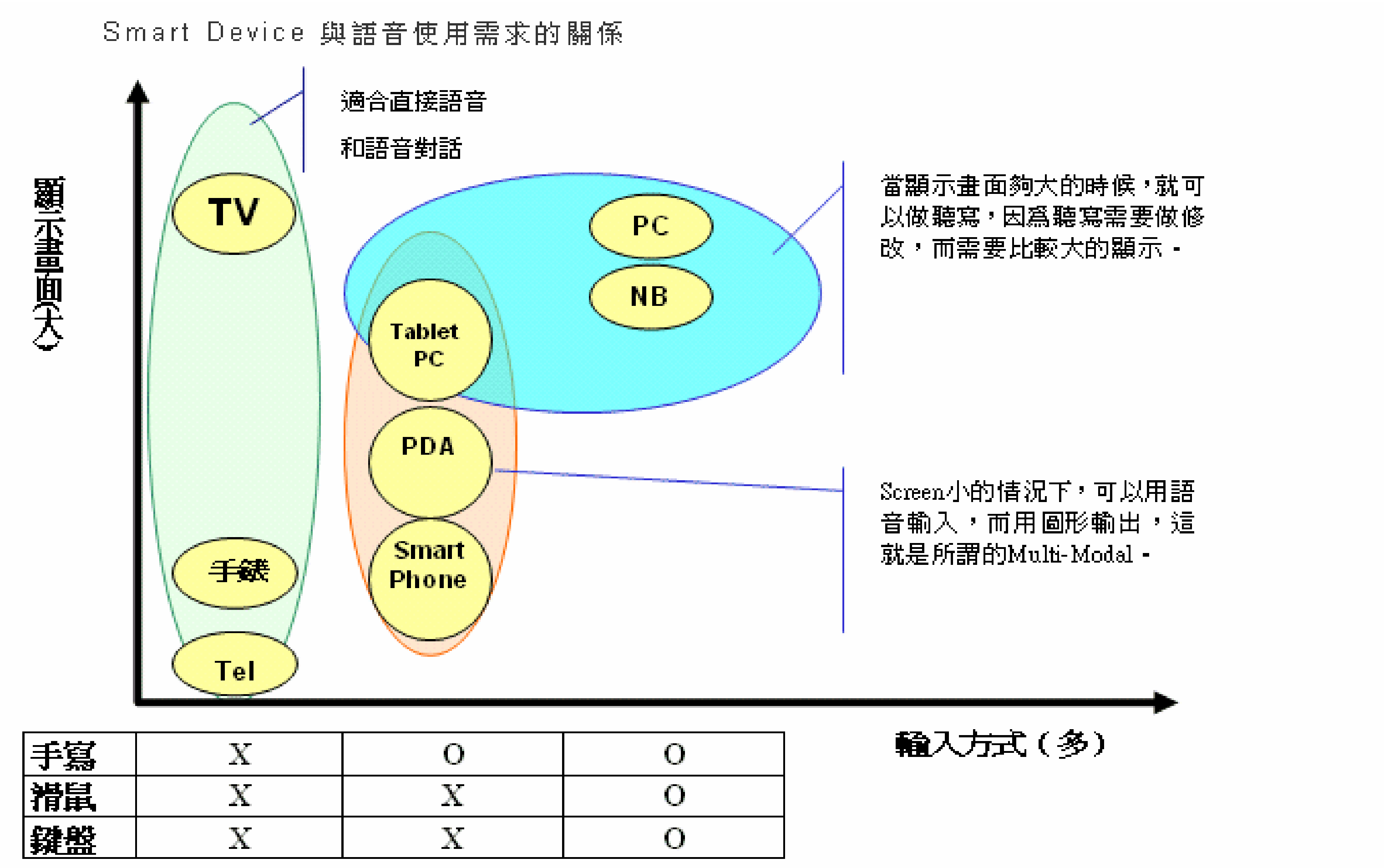
Smart Devices (智慧型設備)日益繁多與普及，但不是每個設備都有螢幕，例如：電話沒有螢幕、鍵盤和滑鼠，就無法使用圖形介面。

.NET 的最初構想，以符合人類需求的自然介面，其包括 –

- 語音合成
- 語音辨識技術
- 結合XML為基礎的網路服務

Speech Industry (3/3)

- Microsoft: Smart Device/Natural UI



Smart Device 與語音使用需求的關係

適合直接語音和語音對話

當顯示畫面夠大的時候，就可以做聽寫，因爲聽寫需要做修改，而需要比較大的顯示。

Screen小的情況下，可以用語音輸入，而用圖形輸出，這就是所謂的Multi-Modal。

手寫	X	O	O
滑鼠	X	X	O
鍵盤	X	X	O

Journals & Conferences

- Journals
  - IEEE Transactions on Speech and Audio Processing
  - Computer Speech and Language
  - Speech Communication
  - Proceedings of the IEEE
  - IEEE Signal Processing Magazine
  - ACM Transactions on Asian Language Information Processing
  - ACM Transactions on Speech and Language Processing
  - …
- Conferences
  - IEEE Int. Conf. Acoustics, Speech, Signal processing (ICASSP)
  - Int. Conf. on Spoken Language Processing (ICSLP)
  - European Conference on Speech Communication and Technology (Eurospeech)

Tentative Schedule

Date	**Tentative Topic List**
9/13	Overview & Introduction
9/20, 9/27	Hidden Markov Models
10/04	Spoken Language Structure
10/11, 10/18	Acoustic Modeling & HTK Toolkit
10/25, 11/01	Statistical Language Modeling & SRI LM Toolkit
11/08	**Midterm**
11/15, 11/22	Speech Signal Processing
11/29, 12/06	Speech Signal Representations
12/13	Model Training and Adaptation Techniques
12/20	Digit Recognition, Word Recognition and Keyword Spotting
12/27	Large Vocabulary Continuous Speech Recognition
01/03	Speech Enhancement and Robustness
01/10	**FINAL**