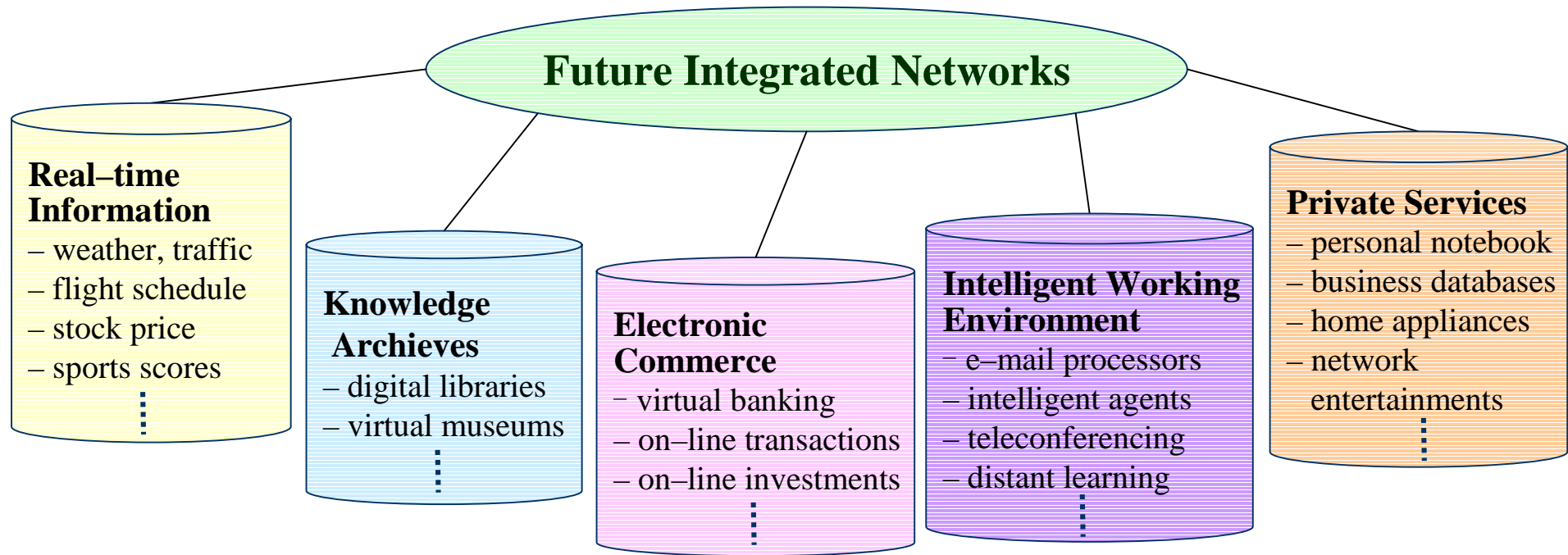


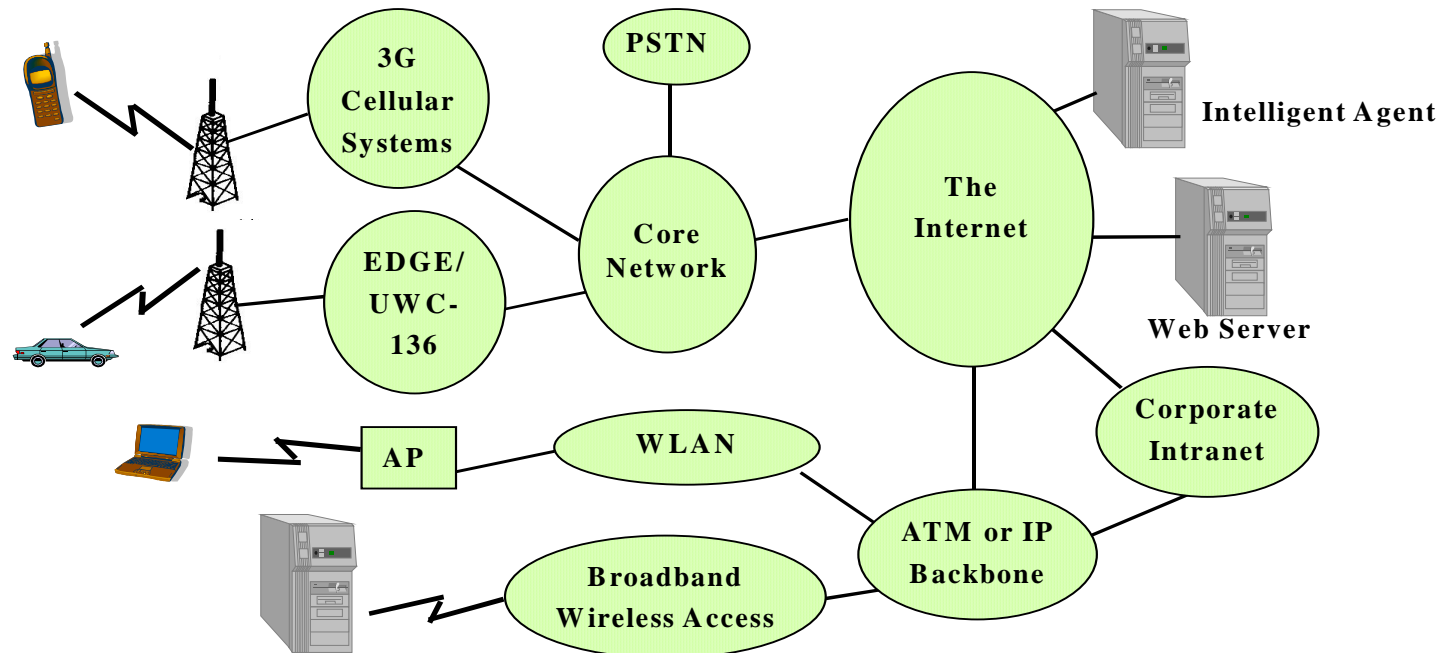
Human – Network Voice Interface in A Wireless Era

Information-related Activities, Applications and Services in Future Network Era



- **Multi-media, Multi-lingual, Multi-functionalities**
- **Cross-cultures, Cross-domains, Cross-regions**
- **Integrating All Knowledge Systems and Information-related Activities and Services Globally**
- **Multiple User Terminals**
 - telephone set, hand set, PDA, vehicular electronics, home appliance, personal computer, etc.

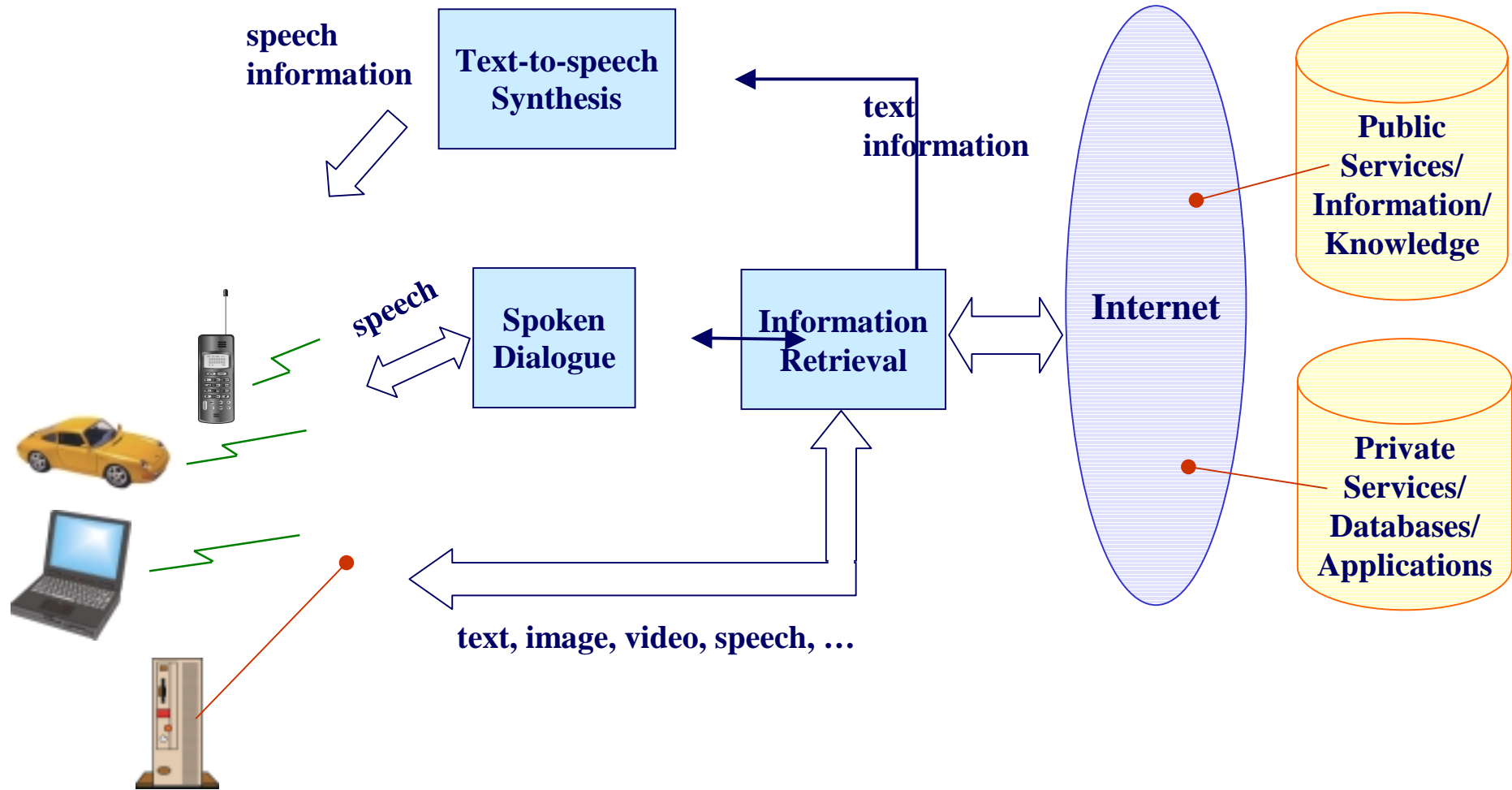
Wireless Access of Global Multi-media Information



- **At Any Time, from Anywhere**
- **As Handset Size Shrinks While Required Functionalities Grows and the User Environment Changes, Voice Interface will be Useful for all User Terminals**
- **Examples**
 - voice retrieval, voice browser, voice portal, voice web
 - spoken dialogue based access to intelligent agents

Scenario for Network Information

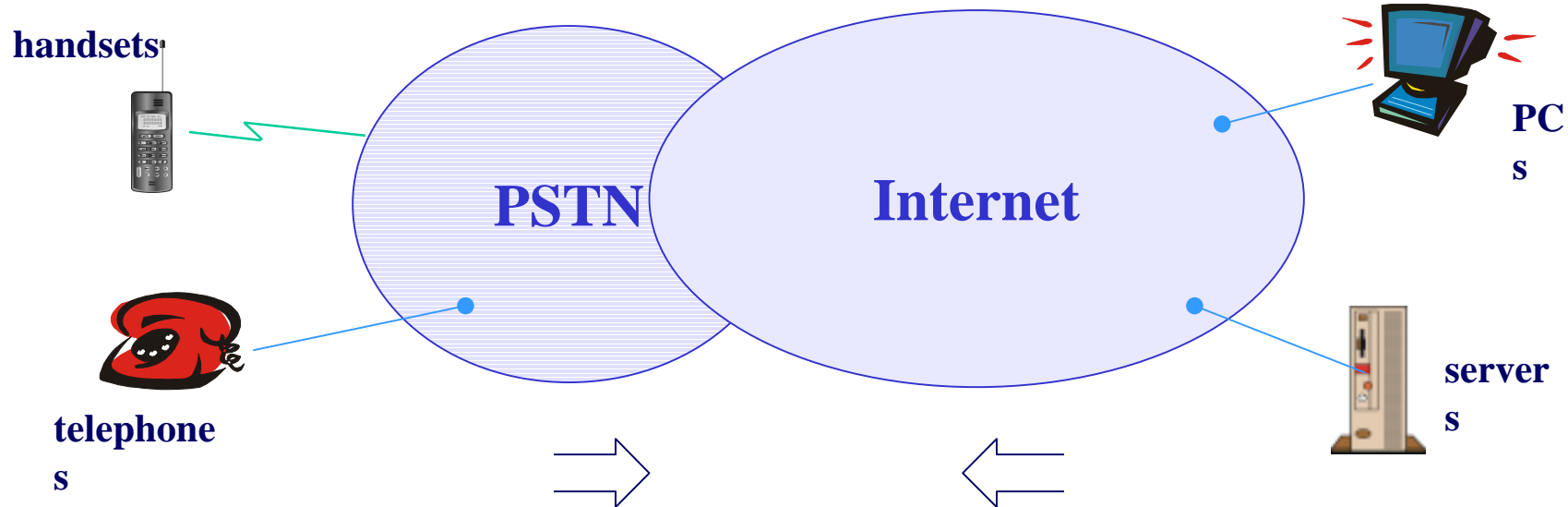
Access



Convergence of PSTN and

Internet

- PSTN(for Voice) and Internet(for Data and Multi-media Contents) are Converging



- **Driving Force for the Convergence**

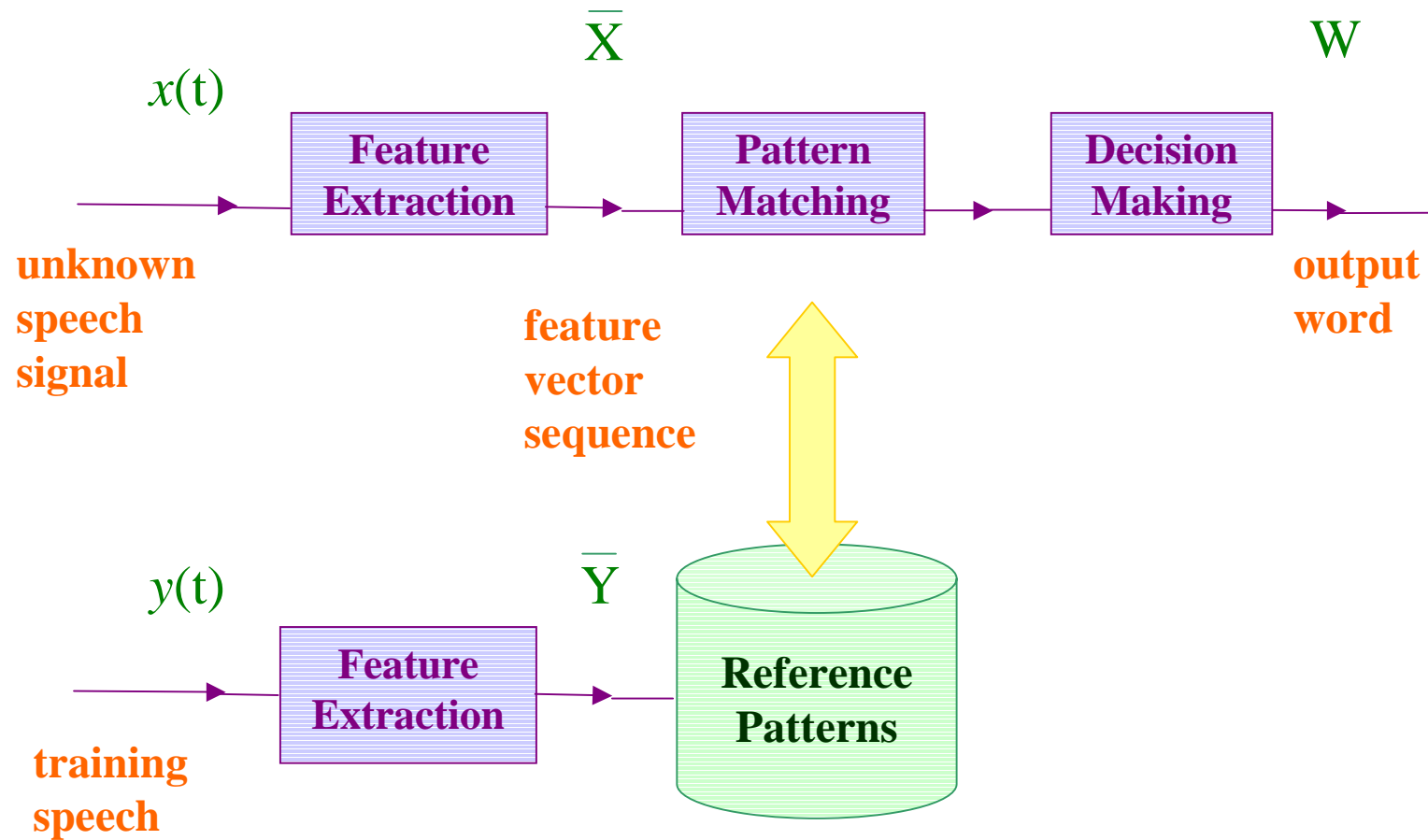
- “anywhere, any time” of wireless services
- voice provides the most convenient and natural interaction interface
- attractive contents over the Internet
- contents(human information) are why the Internet is attractive, while voice directly carries human information
- Speech-enabled Access of Web-based Applications

Voice Interface for Human-network Interaction

- huge volumes of data disseminated across the globe by optical fiber networks
- any time, from anywhere by wireless terminals
- vehicular electronics, PDA, handset, home appliance, etc.
new platforms accessing the global network information/services
- traditional keyboard/mouse not adequate any longer
size shrinkage, different user environment, etc.
desired functionalities/human–network interactions increasing
- voice interface will be one out of the few most important, natural, user friendly, attractive interface
- examples: voice retrieval, voice browser, voice portal, voice web
voice–based web–user interaction
voice–based web tools/Application Interfaces, etc.
- voice interface is the only major “missing link” in the “semi–mature” technology chain

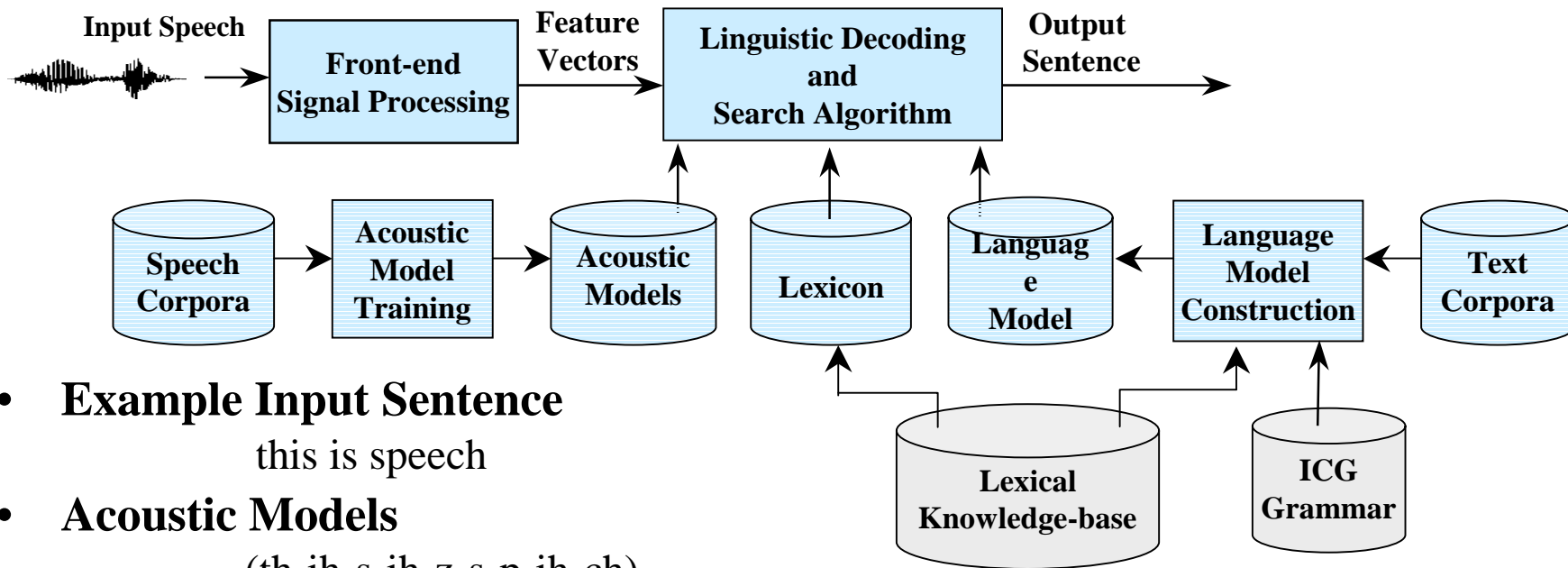
**Core Technologies /
Functionalities for Voice Interface**

Speech Recognition as a pattern recognition problem



Basic Approach for Large Vocabulary Speech Recognition

- **A Simplified Block Diagram**



- **Example Input Sentence**

this is speech

- **Acoustic Models**

(th-ih-s-ih-z-s-p-ih-ch)

- **Lexicon** (th-ih-s) → this

(ih-z) → is

(s-p-iy-ch) → speech

- **Language Model** (this) – (is) – (speech)

$P(\text{this}) P(\text{is} \mid \text{this}) P(\text{speech} \mid \text{this is})$

$P(w_i \mid w_{i-1})$ bi-gram language model

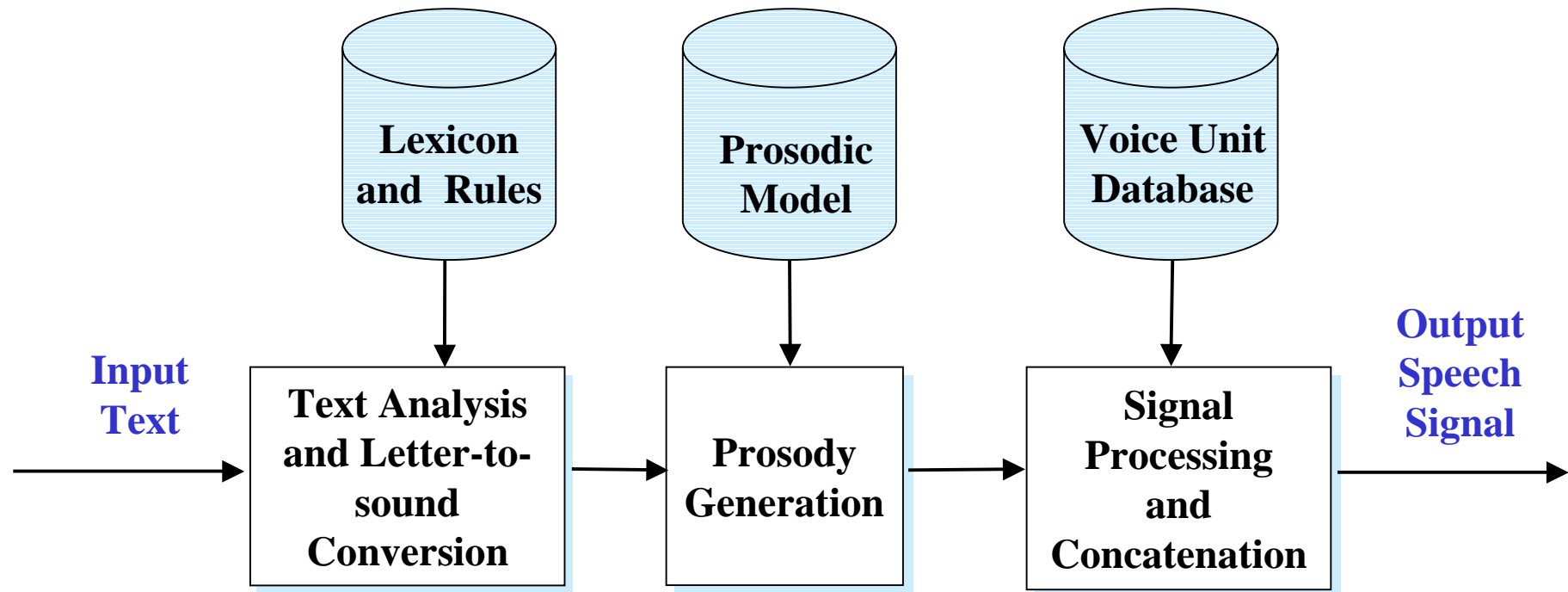
$P(w_i \mid w_{i-1}, w_{i-2})$ tri-gram language model, etc

Speech Recognition Technologies, Applications and Problems

- **Word Recognition**
 - voice command/instructions
- **Keyword Spotting**
 - identifying the keywords out of a pre-defined keyword set from input voice utterances
- **Large Vocabulary Continuous Speech Recognition**
 - entering longer texts
 - remote dictation
- **Speaker Dependent/Independent/Adaptive**
- **Acoustic Reception/Background Noise/Channel Distortion**
- **Read/Spontaneous/Conversational Speech**

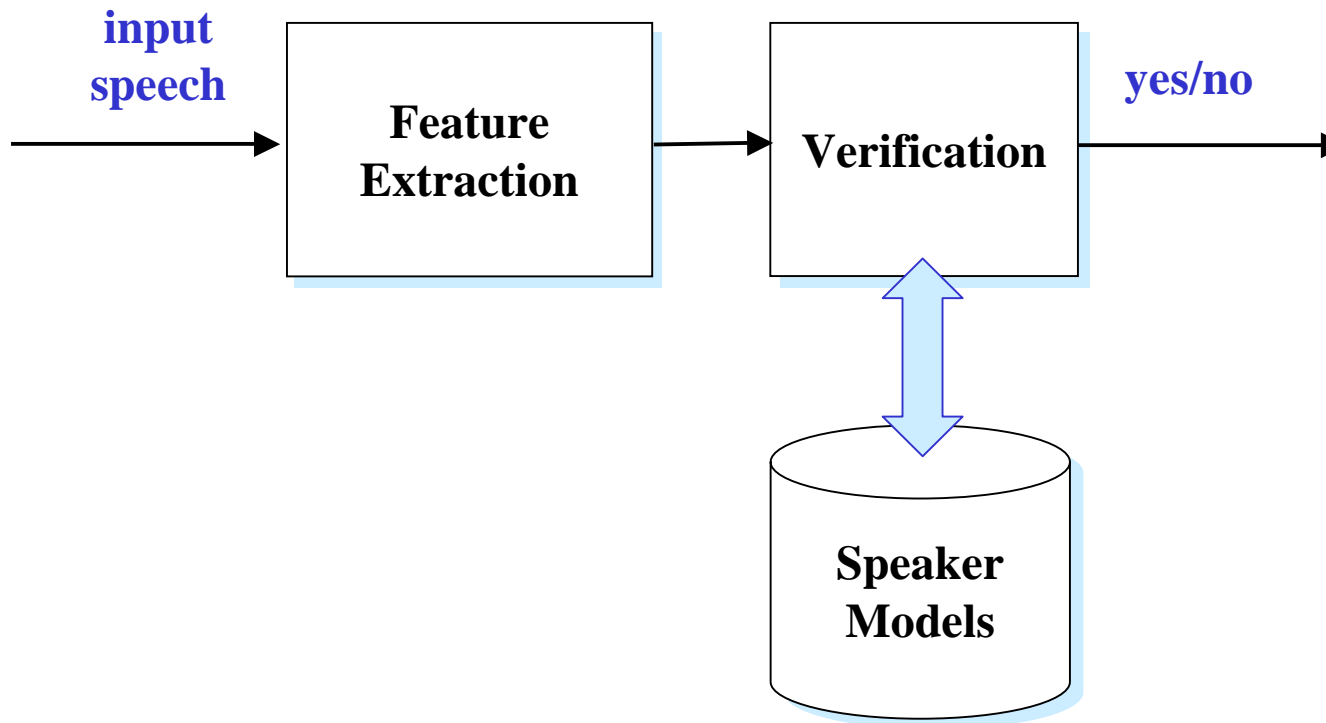
Text-to-speech Synthesis

- Transforming any input text into corresponding speech signals
- E-mail/Web page reading
- Prosodic modeling
- Basic voice units/rule-based, non-uniform units/corpus-based



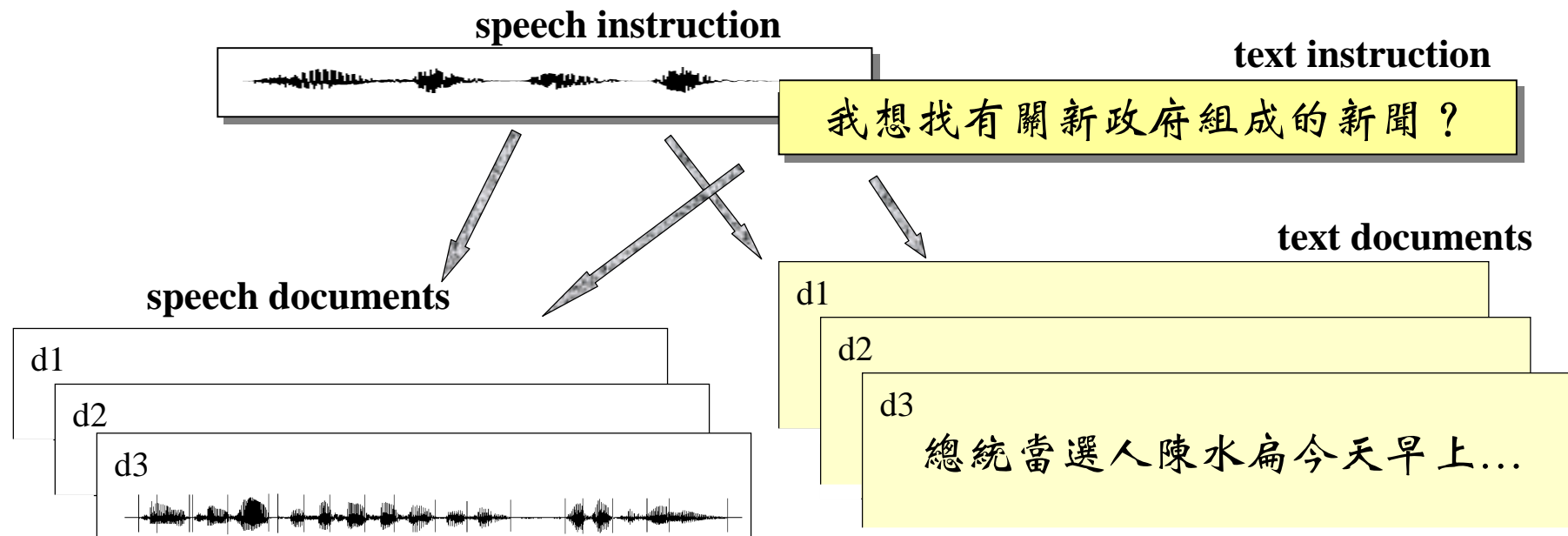
Speaker Verification

- Verifying the speaker as claimed
- Applications requiring verification
- Text dependent/independent
- Integrated with other verification schemes



Information Retrieval Including Voice

- Text Documents/Instructions
- Speech Documents/Instructions
- Voice Personal Notebook/Private Database



Multi-lingual Functionalities

- **Code-Switching Problem**

- English words/phrases inserted in Spoken Chinese sentences

人人都用Computers，家家都上Internet

- the whole sentence switched to English

準備好了嗎？Let's go!

- **Cross-language Network Information Processing**

- globalized network with multi-lingual content/users

- cross-language network information processing with spoken Chinese language input as an example

- **Chinese Dialects/Accents**

- Taiwanese, Cantonese, Shanghainese, etc.

- hundreds of Chinese dialects

- code-switching problem—dialects mixed with Mandarin(or plus English)

- Mandarin with a variety of strong accents

- **Language Dependent/Independent Technologies**

Spoken Dialogue Systems

- Almost all human-network interactions can be made by spoken dialogue
- Speech understanding
- System/user/mixed initiatives
- Reliability/efficiency, dialogue modeling/flow control

