

Pertemuan ke-2

Tujuan, Komponen, dan Penggunaan Sistem Pengenalan Ucapan

R. Rizal Isnanto



Describe the purpose, components,
and use of speech recognition
systems

What Exactly is Speech Recognition?

- Speech Recognition is the process of translating spoken words into text words on the computer.
- Through a speech recognition program/application, the computer is able to process words you say and turn them into text on the screen just as if you had typed them on the keyboard.



Speech Recognition

There are different kinds of voice or speech "engines" that take the sounds of your voice and match it with words. The engine is software that is loaded into your computer. Each version has its advantages and disadvantages. Like everything else in computers, the key is to understand what you want to do with the program and then find the best match.

SINYAL SUARA DIGITAL

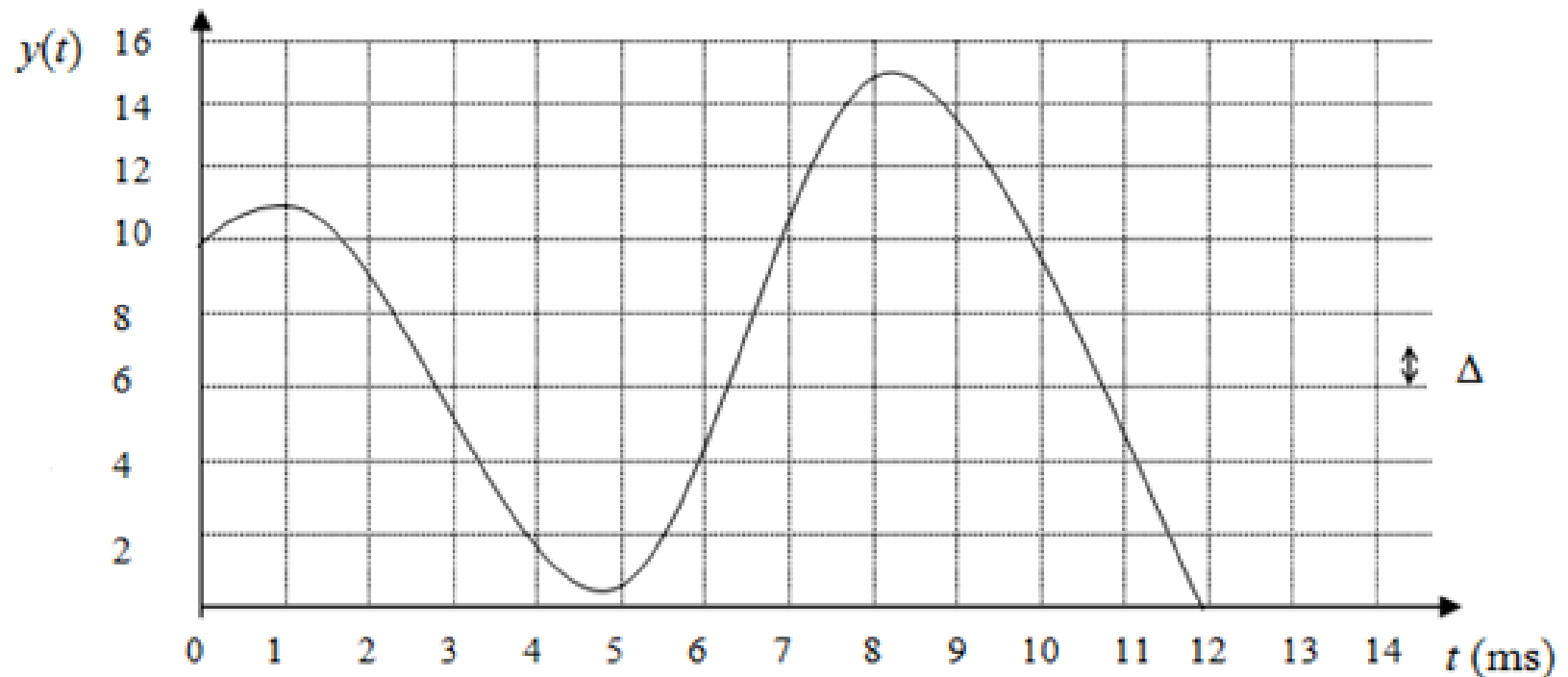
Untuk bisa disimpan dan kemudian dilakukan proses berikut (penyimpanan data secara massal, ekstraksi ciri, pelatihan, dan pengujian, dan pengenalan), suara asli harus melewati tahapan-tahapan sebagai berikut.

1. Akuisisi sinyal analog
2. Pencuplikan (*sampling*)
3. Kuantisasi/pembulatan (*quantizing*)
4. Digitalisasi/digitasi menjadi kode biner

Lihat contoh soal berikut. (*Jawaban lihat papan tulis*)

Jika sebuah sinyal suara pada gambar dicuplik tiap 1 ms dengan langkah kuantisasi $\Delta = 1$,

- bagaimana bentuk isyarat tercupliknya? Semua grafik harus menyertakan garis terputus-putus sebagai penjas.
- Bagaimana bentuk isyarat terkuantisasi dari isyarat tercuplik pada soal no 2?
- Bagaimanakah hasil pengkodean menjadi isyarat digital biner dari isyarat hasil terkuantisasi pada soal no. 3? Gunakan pula tabel konversi desimal ke biner sesuai nilai-nilai pada jawaban butir b untuk menjelaskan jawaban Saudara.
- Gambarkan grafik hasil konversi digital ke analog menggunakan metode *zero order hold* (ZOH).





How Speech Recognition Works

You speak into a microphone connected to your computer. The soundcard or multimedia chip and the speech engine processes your speech.

Each person who uses the speech program has their own unique "voice profile" that is separate from others. You choose your voice profile when you begin. Then the computer learns more about your voice each time you use it. Like a child, it will learn what words you said and can be corrected immediately.



How Speech Recognition Works

Most programs have an enrollment option where you speak a sample list of individual words or sentences into the microphone. This gives a base sample of how you make sounds. From there the computer "guesses" at other words that you are saying. The success rate depends on your computer's speed and how often you use it, but it is not unusual to have a 95% accuracy rate with some engines.



How Speech Recognition Works

Continuous speech recognition applications allow a user to dictate text fluently into the computer. These new applications can recognize speech at up to 200 words per minute. While these systems do give the user system control they are not yet hands-free.

Voice recognition uses a neural net to "learn" to recognize your voice. As you speak, the voice recognition software remembers the way you say each word, even though everyone speaks with varying accents and inflection.

Perbandingan Sistem Digital dibandingkan Sistem Analog pada Aplikasi Sinyal Suara (Audio) (1)

1. Teknologi Penyimpan Data Suara

Analog: terbatas, contoh: pita magnetik, piringan hitam

Digital: jenisnya sangat beragam: Flash Disk (*Flash Memory*), disket, hard disk, CD, dan sebagainya

2. Waktu Akuisisi Data sampai dengan tersimpan

Analog: lebih cepat, waktu nyata (*real-time*)

Digital: lebih lambat, ada proses kuantisasi dan digitasi

3. Proses penyuntingan (*editing*), *synthesizing*, dan *analyzing*

Analog: sulit

Digital: mudah

Perbandingan Sistem Digital dibandingkan Sistem Analog pada Aplikasi Sinyal Suara (Audio) (2)

4. Kompresi Data

Analog: (?)

Digital: sangat banyak algoritma, contoh **wav**, **mp3**, dsb.

5. Pemutar Suara

Analog: terbatas dan sudah kadaluarsa: tape recorder, gramafone

Digital: beragam: ipod, mp3 player, aplikasi multimedia

6. Aplikasi pengenalan suara otomatis

Analog: (?)

Digital: beragam dan terus berkembang, Text-To-Speech, Speech-To-Text, Google Voice, dan sebagainya



Discrete and Continuous Systems

Depending on the speed of your computer, there are continuous speech (normal speaking speed) and discrete speech (slight pauses between each word) versions available. Continuous speech is more natural but it typically requires a more powerful computer to process the information. Modern speech systems are continuous whereas older systems were discrete. Many people (especially those with learning disabilities) prefer discrete systems to the newer continuous speech.



More Memory is Better

Having more RAM memory (the amount of RAM determines how much the computer can "think about" at once without checking its "library" on the hard disk) is a plus. This is especially true if you are using other programs at the same time. As speed and memory of computers increase, speech recognition becomes more and more effective.



Memory and Processor Speed Requirements

Speech engines work on top of Windows and your other programs like your word processor. 128 MB of RAM is the minimum memory requirement for most speech recognition software. 256 is required for some professional versions. Generally a 400-500 mhz processor or better is recommended.



Training Requirements

All of the speech products require either training or a great deal of expertise to be effective. As with all higher end computer programs, the ability to use the speech recognition successfully involves learning to communicate with your speech recognition program as you see how it responds to your dictation and voice commands.



Dictation and Command Modes

Speech Recognition allows a user to use his/her voice as an input device. Voice recognition may be used to dictate text into the computer or to give commands to the computer (such as opening application programs, pulling down menus, or saving work).



Why Learn Speech Recognition?

- Increases productivity
- Helps avoid injury or overcome a handicap
- Improves writing skills
- Improves reading skills
- Improves speaking skills

Types of Microphones

Acceptable



**Boom
Microphone**

Good



**Single Ear
Headphone
Microphone**

Very
Good



**Dual Ear Headphone
Microphone with
volume and on/off
controls**

Acceptable



**Handheld
Microphone**

Good



**Dual Ear
Headphone
Microphone**

Extremely
Good



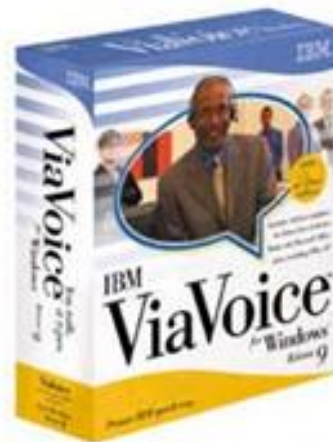
**USB Headphone
Microphone**

Speech Recognition Software

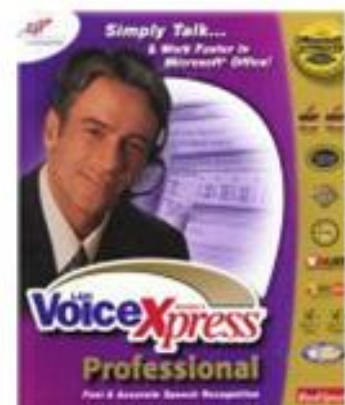
Microsoft Office XP or 2003 Speech Recognition



Dragon Naturally
Speaking 7.0 or 7.3



IBM ViaVoice



L&H Voice
Express



Speech Recognition System Requirements in Microsoft Office XP or 2003

- A high quality close-talk (headset) microphone (a universal serial bus (USB) microphone with gain adjustment support is recommended. Gain adjustment is a feature that modifies microphone amplification so the input sound level is appropriate for use by the system.)
- 400-megahertz (MHz) or faster computer
- 128 megabytes (MB) or more of RAM
- Microsoft Windows 98 or later or Microsoft Windows NT® 4.0, XP or later
- Microsoft Internet Explorer 5 or later