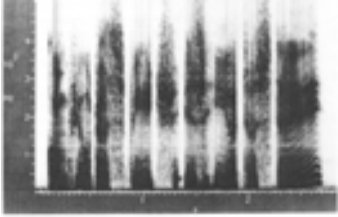


## Slajd 1

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Analiza i synteza mowy - wprowadzenie**

---



Spektrogram wyrażenia: „computer speech”

---

Zygmunt Ciota 1/4/2005 Slide 1

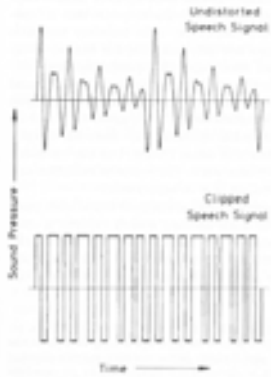
Detailed description: This slide displays a spectrogram of the phrase "computer speech". The spectrogram is a rectangular plot with a dark background and lighter, vertical, irregular bands of varying intensity, representing the frequency components of the speech signal over time. The title "Analiza i synteza mowy - wprowadzenie" is underlined. The footer contains the author's name "Zygmunt Ciota", the date "1/4/2005", and "Slide 1".

## Slajd 2

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Analiza i synteza mowy - wprowadzenie**

---



Sound Pressure

Time

Undistorted Speech Signal

Clipped Speech Signal

---

Zygmunt Ciota 1/4/2005 Slide 2

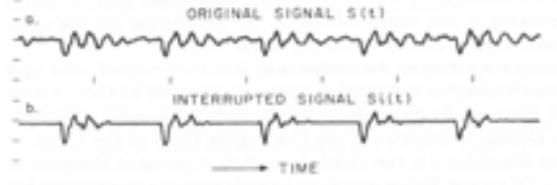
Detailed description: This slide compares two speech signals. The top graph, labeled "Undistorted Speech Signal", shows a smooth, periodic waveform. The bottom graph, labeled "Clipped Speech Signal", shows a similar waveform but with the peaks flattened into a square wave, indicating signal clipping. The vertical axis is labeled "Sound Pressure" and the horizontal axis is labeled "Time". The title "Analiza i synteza mowy - wprowadzenie" is underlined. The footer contains the author's name "Zygmunt Ciota", the date "1/4/2005", and "Slide 2".

### Slajd 3

Rozpoznawanie i synteza mowy w systemach multimedialnych

#### Analiza i synteza mowy - wprowadzenie

---



The image shows two waveforms. The top waveform is labeled 'ORIGINAL SIGNAL  $S(t)$ ' and shows a continuous, irregular wave. The bottom waveform is labeled 'INTERRUPTED SIGNAL  $S_i(t)$ ' and shows the same wave with several gaps where the signal is zero. A horizontal arrow at the bottom is labeled 'TIME'.

---

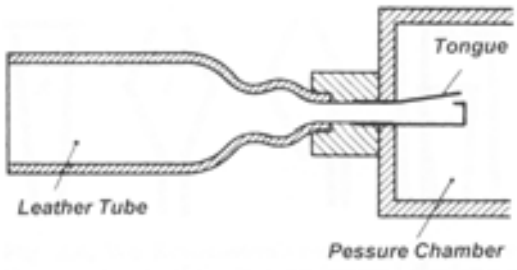
Zygmunt Ciota 1/4/2005 Slide 3

### Slajd 4

Rozpoznawanie i synteza mowy w systemach multimedialnych

#### Pierwsze próby syntezy mowy

---



The diagram shows a cross-section of a mechanical device. On the left is a 'Leather Tube' that tapers towards the right. On the right, a 'Tongue' is attached to the end of the tube. The tongue is positioned inside a 'Pressure Chamber' which has a rectangular cross-section. The tongue is shown in a position that would allow air to pass through it, creating a specific sound.

Maszyna wymawiająca samogłoski z 1769 roku zbudowana przez Wolfganga Rittera von Kempelen (rekonstrukcja)

---

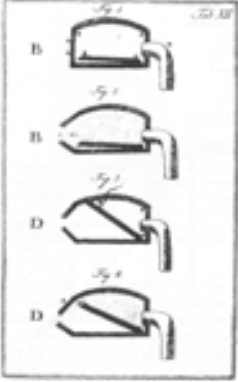
Zygmunt Ciota 1/4/2005 Slide 4

## Slajd 5

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Pierwsze próby syntezy mowy**

---



Urządzenia do wymawiania spółgłosek „plozyjnych”: „b” i „d” (Wolfgang Ritter)

---


Zygmunt Ciota 1/4/2005 Slide 5

## Slajd 6

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Pierwsze próby syntezy mowy**

---



Rezonatory (wibrujące piszczałki) Kratzensteina - 1779

---

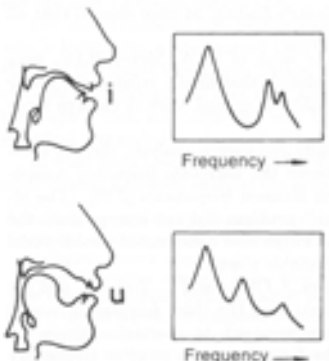
Zygmunt Ciota 1/4/2005 Slide 6

## Slajd 7

Rozpoznawanie i synteza mowy w systemach multimedialnych

### Pierwsze próby syntezy mowy

---



Przekroje kanału głosowego przy wymawianiu samogłosek „i” oraz „u” i odpowiadające im formanty (rezonanse)

---

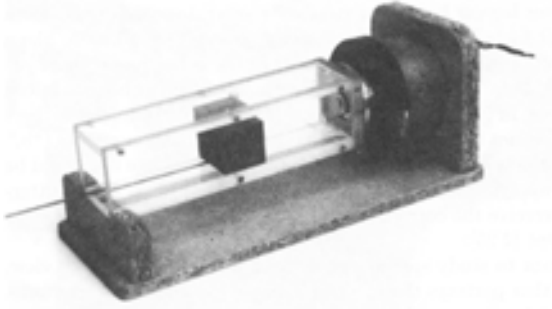
Zygmunt Ciota 1/4/2005 Slide 7

## Slajd 8

Rozpoznawanie i synteza mowy w systemach multimedialnych

### Pierwsze próby syntezy mowy

---



Najprostszy model kanału głosowego

---

Zygmunt Ciota 1/4/2005 Slide 8

## Slajd 9

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Rozpoznawanie mowy oraz identyfikacja głosu**

---



Pomiar parametrów fizycznych kanału głosowego

---

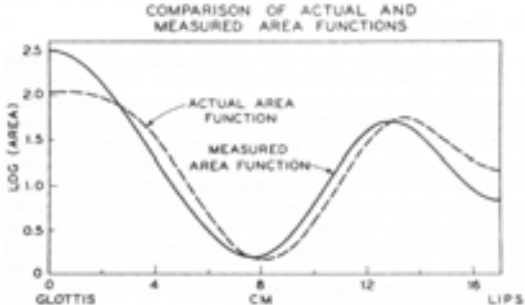
Zygmunt Ciota 1/4/2005 Slide 9

## Slajd 10

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Rozpoznawanie mowy oraz identyfikacja głosu**

---

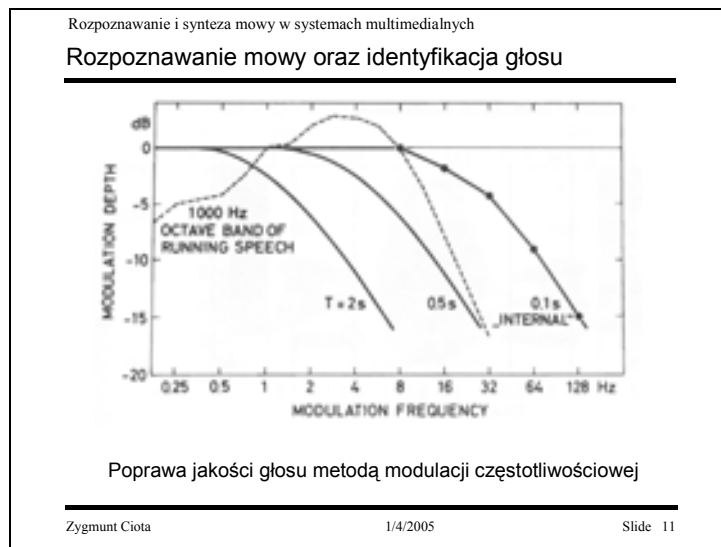


Pomiar parametrów geometrycznych kanału głosowego

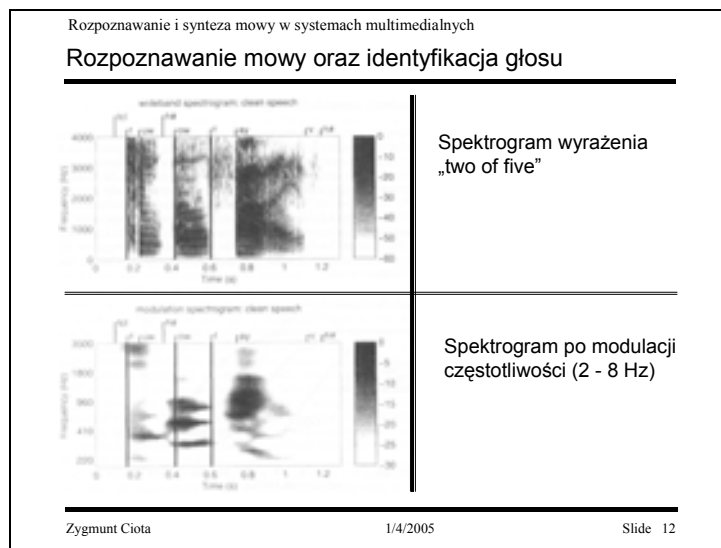
---

Zygmunt Ciota 1/4/2005 Slide 10

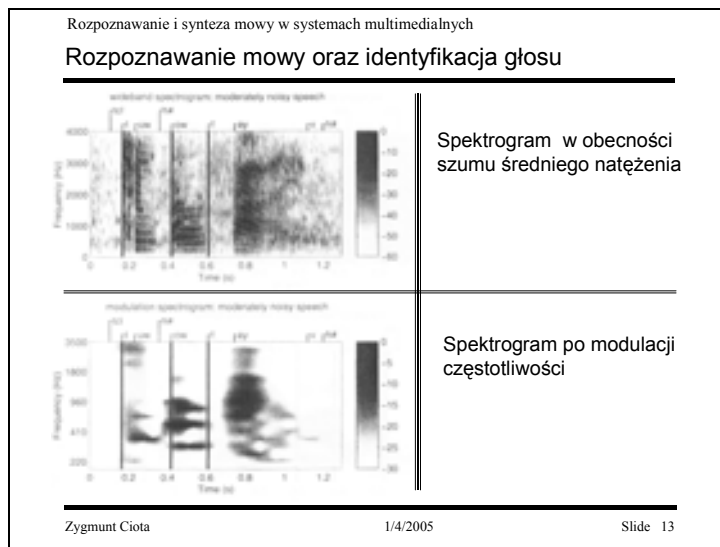
## Slajd 11



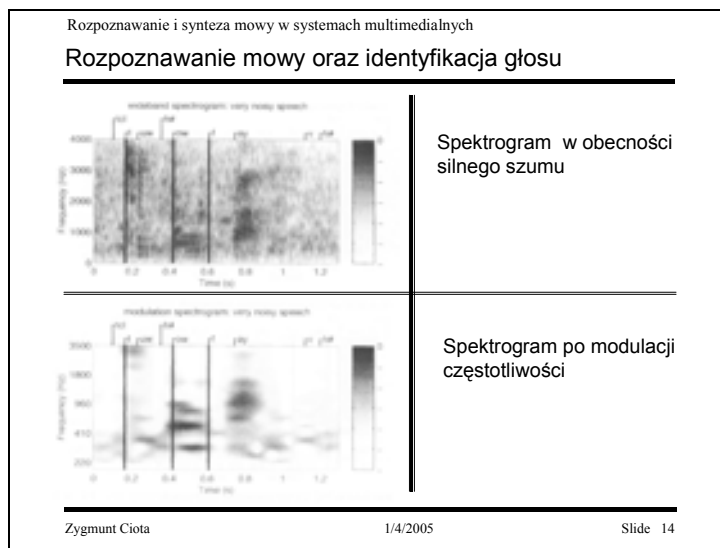
## Slajd 12



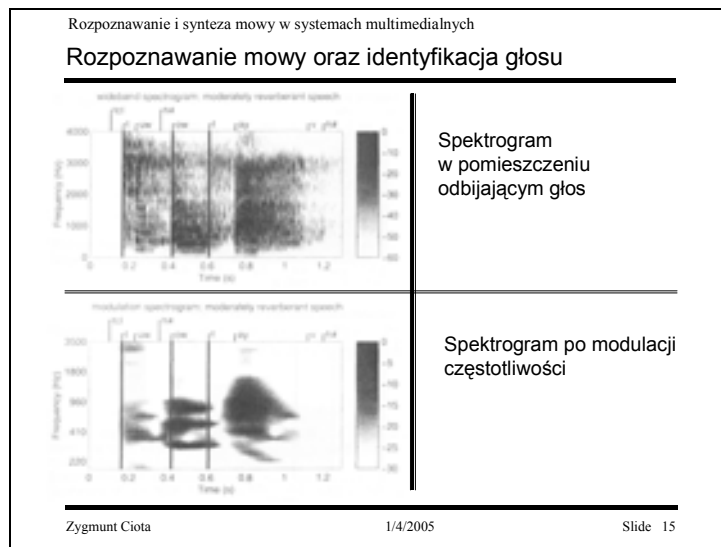
## Slajd 13



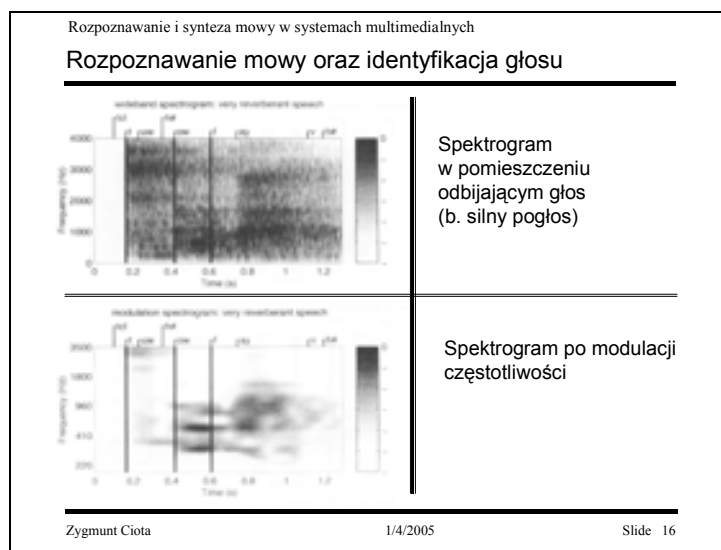
## Slajd 14



## Slajd 15



## Slajd 16



## Slajd 17

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Rozpoznawanie mowy**

---

**Reguły Bayesa: wyznaczanie prawdopodobieństw a posteriori w funkcji prawdopodobieństw a priori**

$$p(W|A) = p(A|W) \cdot p(W) / p(A)$$

---

Zygmunt Ciota 1/4/2005 Slide 17

## Slajd 18

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Kompresja mowy**

---

**Predykcja liniowa**

$$s_n = -a_1 s_{n-1} - a_2 s_{n-2} - \dots - a_p s_{n-p} + e_n \quad (1)$$

---

$$E = \sum_{n=1}^N e_n^2 = \sum_{n=1}^N \left( \sum_{k=0}^p a_k s_{n-k} \right)^2 \quad a_0 = 1$$

---

Zygmunt Ciota 1/4/2005 Slide 18

## Slajd 19

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Rozpoznawanie mowy**

---

$$\frac{\partial E}{\partial a_m} = E = \sum_{n=1}^N 2s_{n-m} \sum_{k=0}^p a_k s_{n-k} = 0$$

**Odwracając porządek sumowania**

$$\sum_{k=0}^p r_{mk} a_k = 0 \quad \text{gdzie} \quad r_{mk} = \sum_{n=1}^N s_{n-m} s_{n-k}$$

---

Zygmunt Ciota 1/4/2005 Slide 19

## Slajd 20

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Rozpoznawanie mowy**

---

**Pamiętaj, że  $a_0=1$**

$$\sum_{k=1}^p r_{mk} a_k = -r_{m0}$$

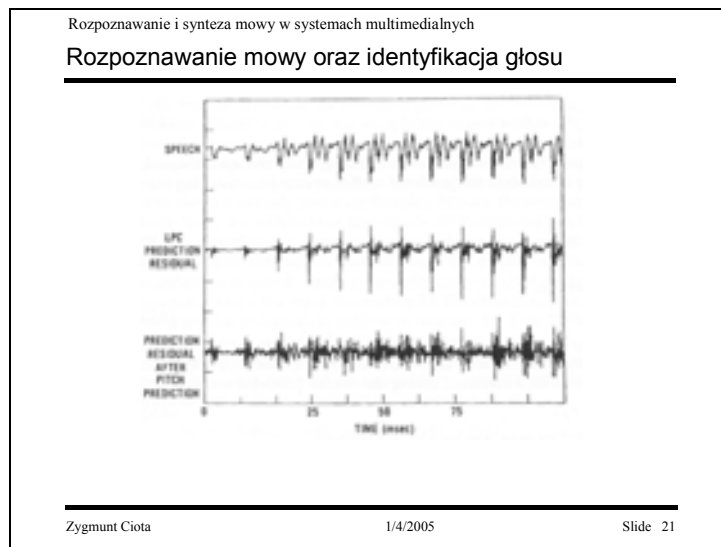
**lub w postaci macierzowej**

$$\mathbf{R} \cdot \mathbf{a} = -\mathbf{r}_0 \quad \text{czyli} \quad \mathbf{a} = \mathbf{R}^{-1} \cdot \mathbf{r}_0$$

---

Zygmunt Ciota 1/4/2005 Slide 20

## Slajd 21



## Slajd 22

Rozpoznawanie i synteza mowy w systemach multimedialnych

### Rozpoznawanie mowy

---

**Filtr predykcyjny**

**Filtr FIR (Finite Impulse Response):**

$$H(z) = \sum_{k=0}^N a_k z^{-k}$$

**Odpowiedź impulsowa filtru FIR:**

$1, a_1, a_2, \dots, a_p$

---

Zygmunt Ciota 1/4/2005 Slide 22

## Slajd 23

Rozpoznawanie i synteza mowy w systemach multimedialnych

### Rozpoznawanie mowy

---

$$s_n \xrightarrow{\text{yellow}} s(z) \qquad e_n \xrightarrow{\text{yellow}} e(z)$$

**Przyjmując:**

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_p z^{-p}$$

**Z równania (1):**

$$e_n = s_n + a_1 s_{n-1} + a_2 s_{n-2} + \dots + a_p s_{n-p}$$

**czyli:**

$$s(z) \cdot A(z) = e(z) \qquad s(z) = \frac{e(z)}{A(z)}$$

---

Zygmunt Ciota 1/4/2005 Slide 23

## Slajd 24

Rozpoznawanie i synteza mowy w systemach multimedialnych

### Rozpoznawanie mowy

---

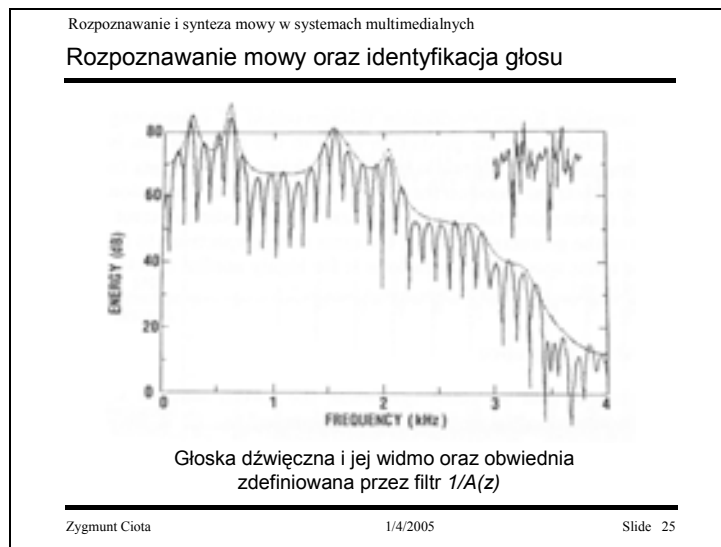
$$A(z) = \left(1 - \frac{z_1}{z}\right) \cdot \left(1 - \frac{z_2}{z}\right) \cdots \quad z_1, z_2 \cdots \text{zera } A(z)$$

$$H_p(z) = \frac{1}{A(z)} \quad \text{all-pole filter}$$

---

Zygmunt Ciota 1/4/2005 Slide 24

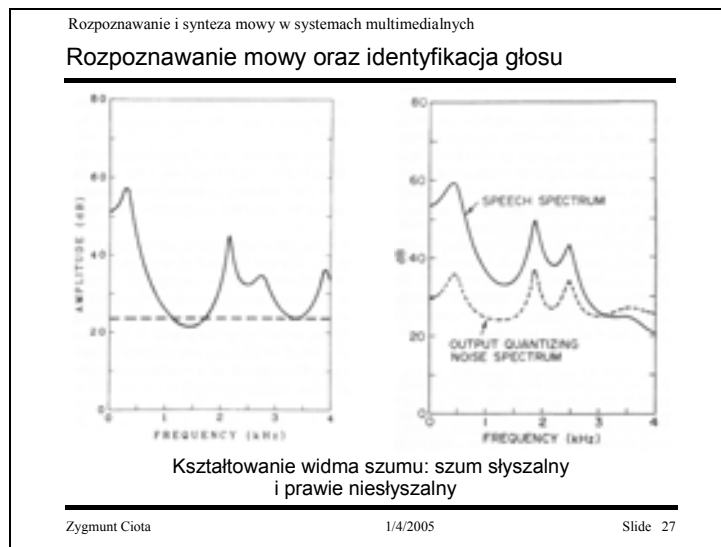
Slajd 25



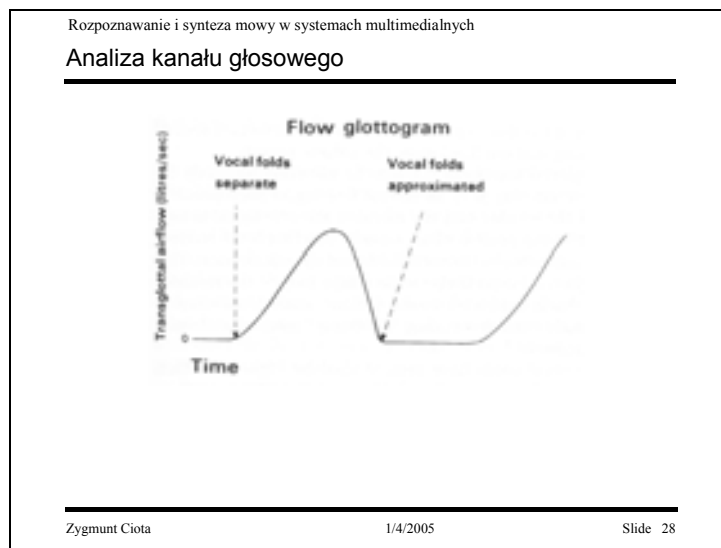
Slajd 26



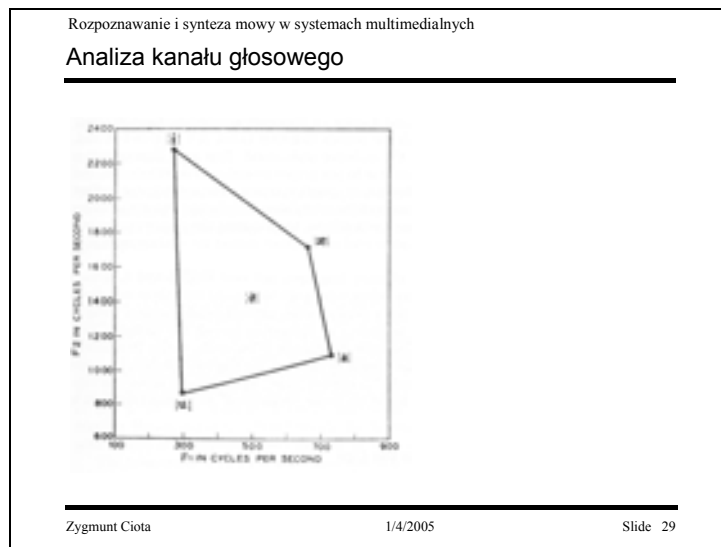
## Slajd 27



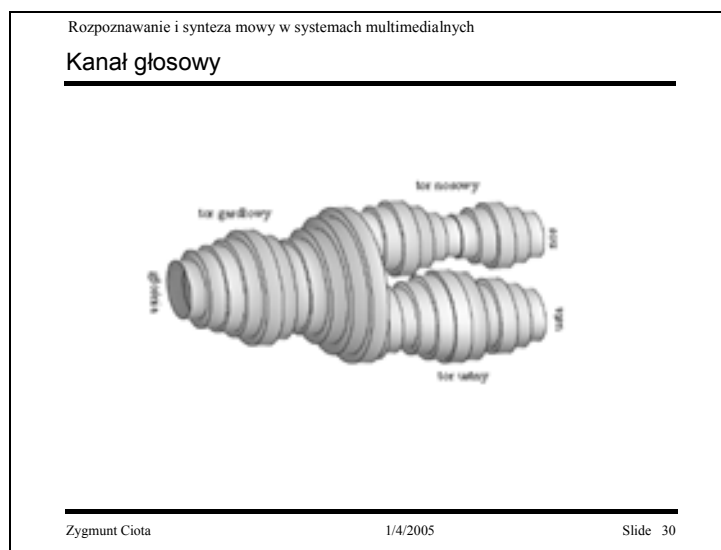
## Slajd 28



## Slajd 29



## Slajd 30



## Slajd 31

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Kanał głosowy**

---

**Równanie Webster'a, czyli równanie natężenia (ciśnienia) fali dźwiękowej  $p(x,t)$  wzdłuż kanału:**

$$\frac{\partial^2 p}{\partial x^2} + \frac{1}{A} \frac{dA}{dx} \frac{\partial p}{\partial x} = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2}$$

**$A(x)$  jest przekrojem kanału głosowego między głośnią ( $x=0$ ) a wargami  $x=L \approx 170$  mm)**

---

Zygmunt Ciota 1/4/2005 Slide 31

## Slajd 32

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Lokalizacja położenia źródła dźwięku**

---

$$\Delta\alpha = 180^\circ \frac{c}{d} \Delta t$$

**Prawo pierwszej fali frontowej (Joseph Henry) i efekt Haas'a**

---

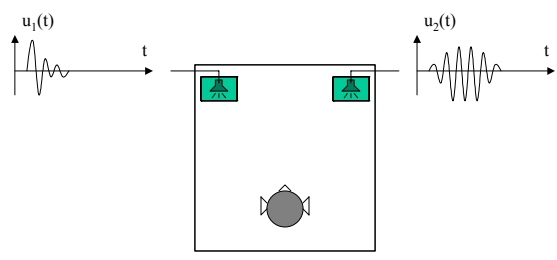
Zygmunt Ciota 1/4/2005 Slide 32

## Slajd 33

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Lokalizacja położenia źródła dźwięku**

---



**Continuity effect**

---

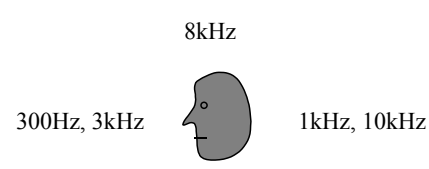
Zygmunt Ciota 1/4/2005 Slide 33

## Slajd 34

Rozpoznawanie i synteza mowy w systemach multimedialnych

**Lokalizacja położenia źródła dźwięku**

---



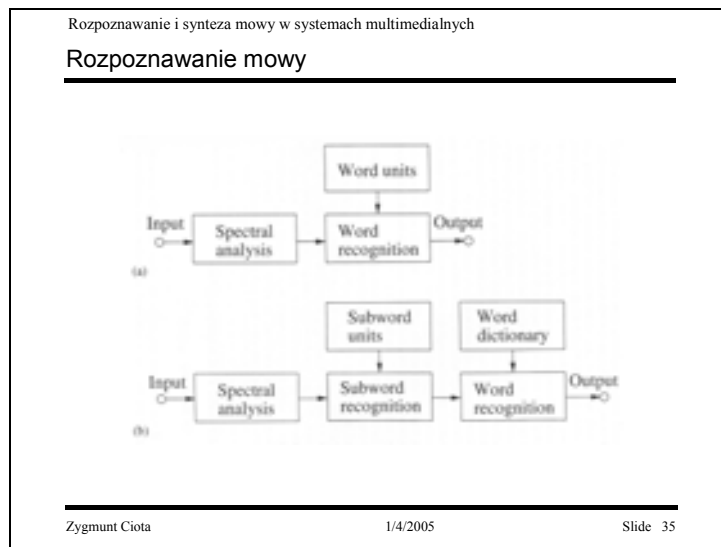
8kHz

300Hz, 3kHz 1kHz, 10kHz

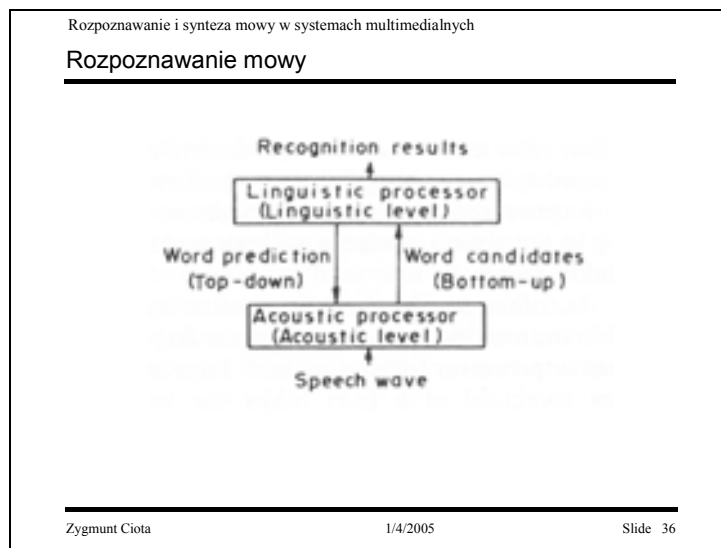
---

Zygmunt Ciota 1/4/2005 Slide 34

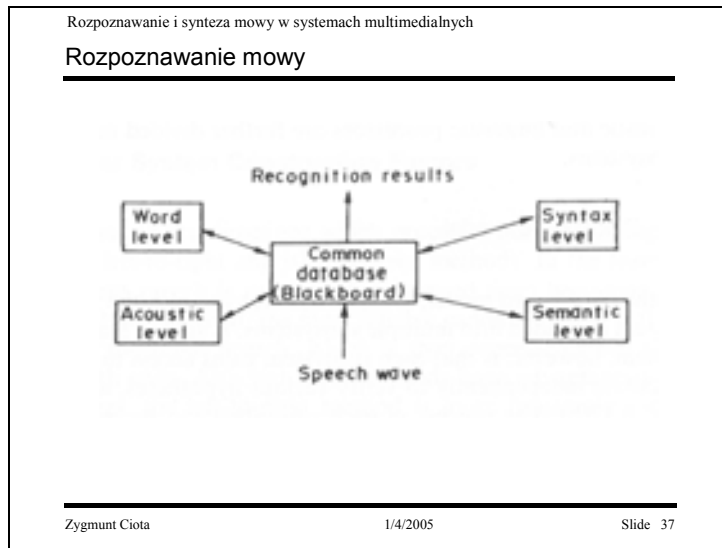
## Slajd 35



## Slajd 36



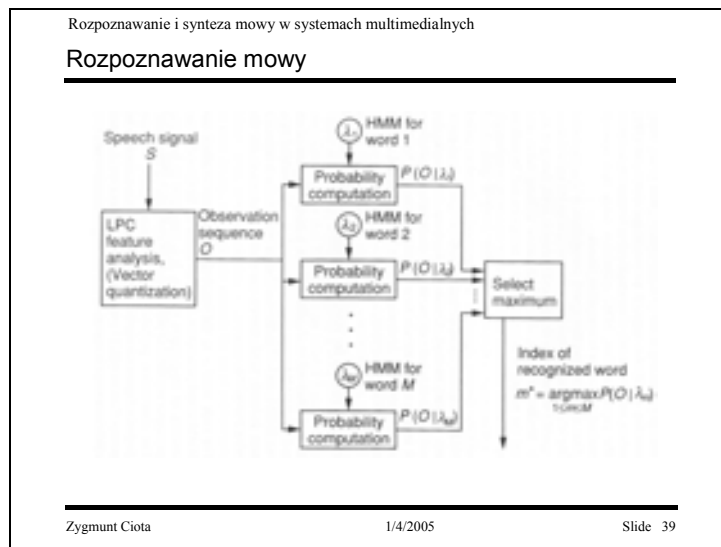
# Slajd 37



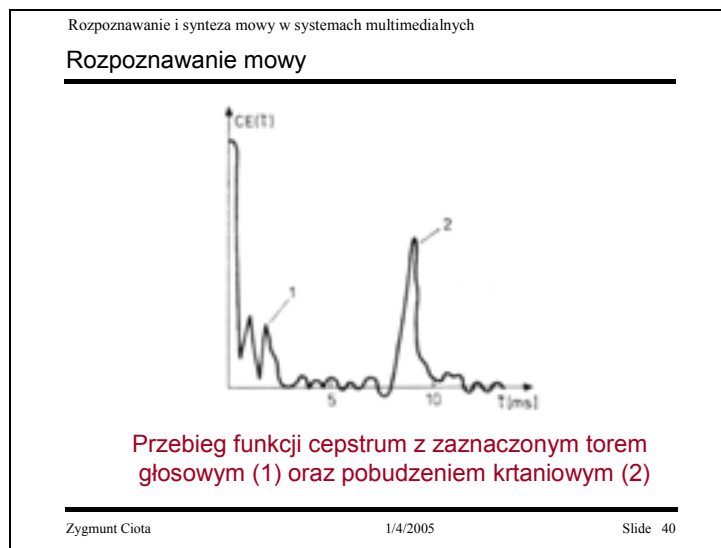
# Slajd 38



## Slajd 39



## Slajd 40



## Slajd 41

Rozpoznawanie i synteza mowy w systemach multimedialnych

---

**Formaty dźwiękowe**

**LPCM (Linear Pulse Code Modulation): udoskonalony PCM**

Częstotliwość próbkowania	Rozdzielczość	Ilość kanałów	Strumień max.
48 kHz	16 bitów	8	6,144 Mb/s
48 kHz	20 bitów	6	5,76 Mb/s
48 kHz	24 bity	5	5,76 Mb/s
96 kHz	16 bitów	4	6,144 Mb/s
96 kHz	20 bitów	3	5,76 Mb/s
96 kHz	24 bity	2	4,608 Mb/s

---

Zygmunt Ciota 1/4/2005 Slide 41

## Slajd 42

Rozpoznawanie i synteza mowy w systemach multimedialnych

---

**MPEG – AUDIO**

**A. MPEG-1 audio, opisuje trzy warstwy kodowania dźwięku o następujących właściwościach:**

- 1 lub 2 kanały dźwiękowe
- Częstotliwość próbkowania 32kHz, 44.1kHz lub 48kHz
- Przepustowość od 32kbps do 448kbps

---

Zygmunt Ciota 1/4/2005 Slide 42

## Slajd 43

Rozpoznawanie i synteza mowy w systemach multimedialnych

---

**B. MPEG-2 audio występuje w dwóch odmianach**

- **MPEG-2/LSF**
  - 1 lub 2 kanały dźwiękowe
  - Dwa razy mniejszą częstotliwość próbkowania
  - Przepustowość od 8 do 256 kbs
- **MPEG-2/Multichannel**
  - Do 5 pełno-zakresowych kanałów dźwiękowych
  - Częstotliwość próbkowania taką samą jak w MPEG-1
  - Rozdzielczość 16 bitów
  - Większe przepustowości dochodzące do około 1 Mb/s do systemu 5+1 kanałów

---

Zygmunt Ciota 1/4/2005 Slide 43

## Slajd 44

Rozpoznawanie i synteza mowy w systemach multimedialnych

---

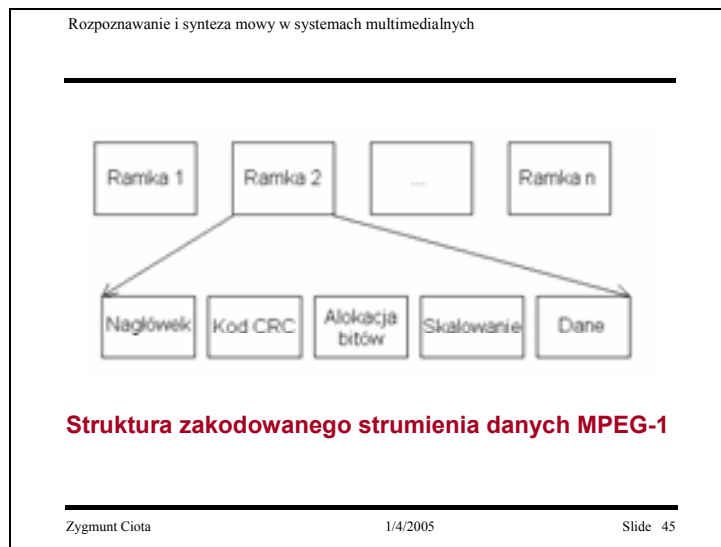
```
graph LR; PCM --> Bank[Bank 32 filtrów pasmowych]; Bank --> Skalowanie[Skalowanie próbek]; Skalowanie --> Kwantyzacja[Kwantyzacja + Kodowanie]; Kwantyzacja --> Multiplexing[Multiplexing]; Multiplexing --> Strumien[Strumień danych MPEG-1]; Model[Model psychoakustyczny] -- SMR --> Kwantyzacja; Bank --> Model;
```

**Schemat blokowy układu kodera standardu MPEG-1**

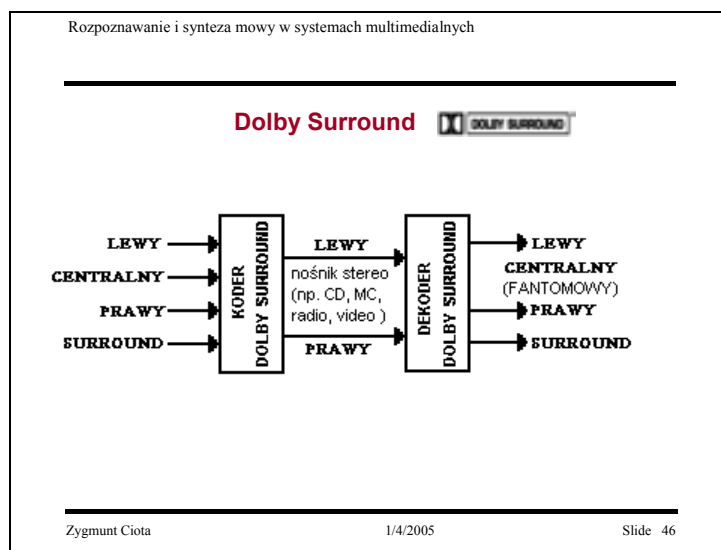
---

Zygmunt Ciota 1/4/2005 Slide 44

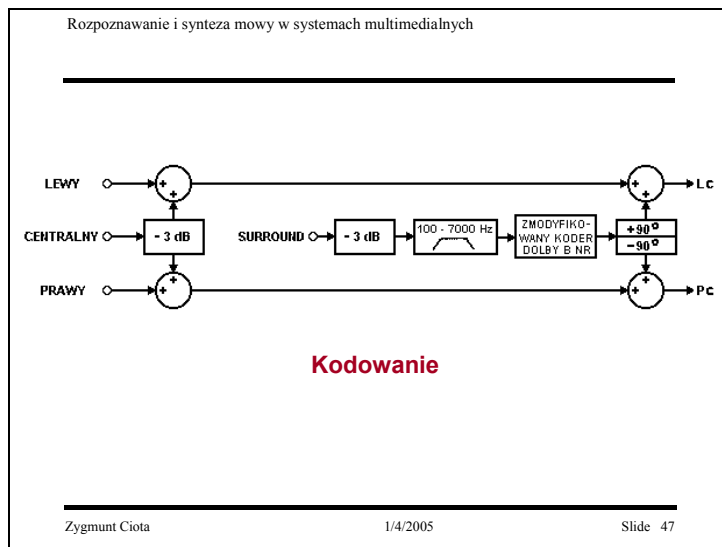
## Slajd 45



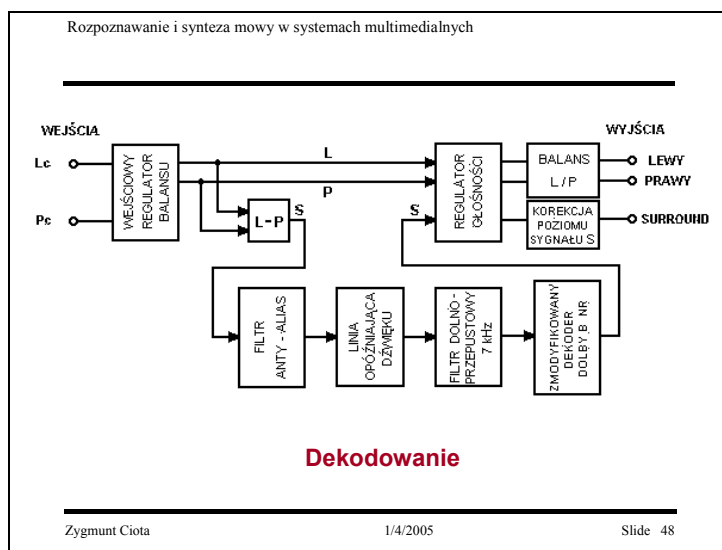
## Slajd 46



Slajd 47




Slajd 48

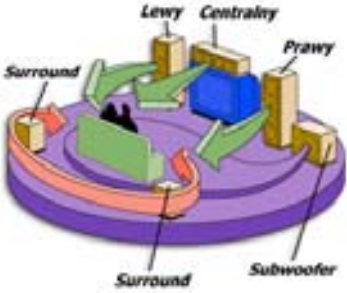


# Slajd 49

Rozpoznawanie i synteza mowy w systemach multimedialnych

---

**Dolby Pro Logic** 



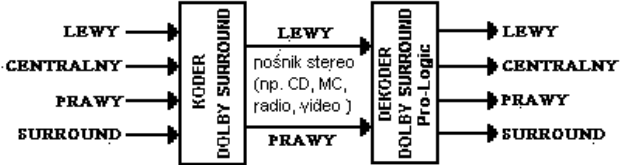
---

Zygmunt Ciota 1/4/2005 Slide 49

# Slajd 50

Rozpoznawanie i synteza mowy w systemach multimedialnych

---

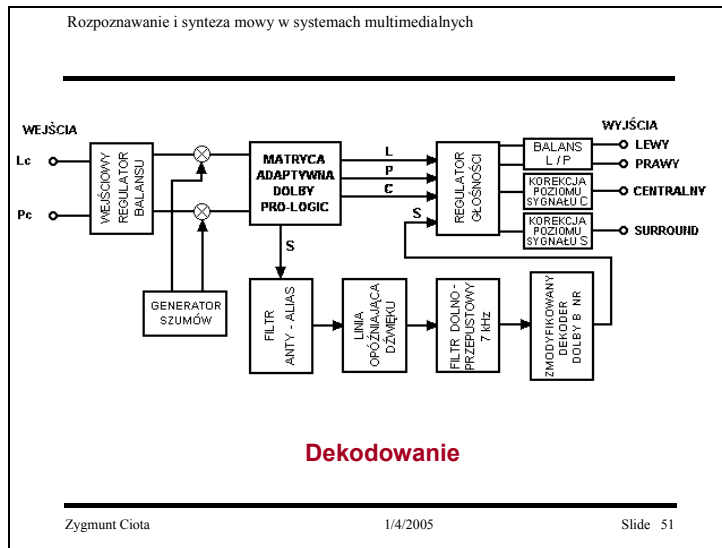


**Kodek**

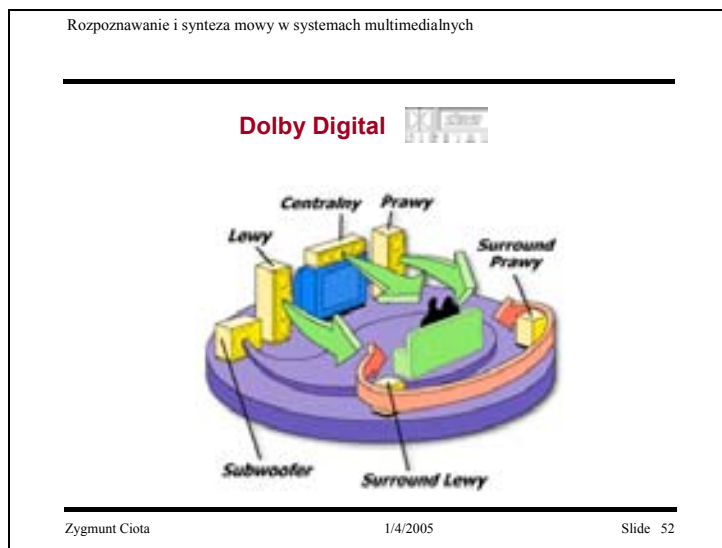
---

Zygmunt Ciota 1/4/2005 Slide 50

Slajd 51



Slajd 52





## Slajd 55

Rozpoznawanie i synteza mowy w systemach multimedialnych

---

**DC** **DVD**

---

Zygmunt Ciota 1/4/2005 Slide 55

## Slajd 56

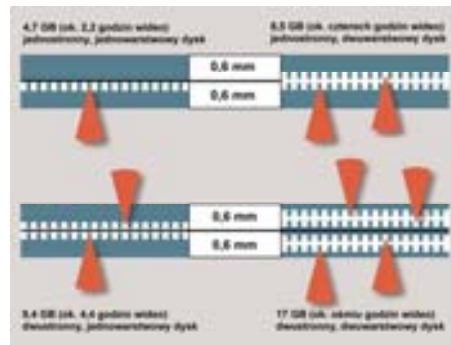
Rozpoznawanie i synteza mowy w systemach multimedialnych

---

<b>Standard:</b>	<b>Pojemność:</b>	<b>Rodzaj płyty:</b>
DVD 5	4.7 GB	jednostronna – jednowarstwowa
DVD 9	8.5 GB	jednostronna – dwuwarstwowa
DVD 10	9.4 GB	dwuwarstwowa – jednostronna
DVD 18	17 GB	dwuwarstwowa – dwustronna

---

Zygmunt Ciota 1/4/2005 Slide 56



**Struktura warstw płyty DVD w zależności od standardu**