

XỬ LÝ TIẾNG NÓI

Trịnh Văn Loan – ĐHBK Hà Nội

1

1

TÀI LIỆU THAM KHẢO

- ✖ La parole et son traitement automatique
Calliope, Masson, 1989
- ✖ Traitement de la parole
Rene Boite et Murat Kunt, Presse Polytechnique Romandes, 1987
- ✖ Introduction to Digital Speech Processing
Lawrence R. Rabiner, Ronald W. Schafer, Now.2007
- ✖ Fundamentals of Speech Signal Processing
Saito S., Nakata K. , Academic Press, 1985
- ✖ Digital Processing of Speech Signals
Lawrence R. Rabiner, Ronald W. Schafer, Prentice-Hall .1978
- ✖ Discrete-Time Processing of Speech Signals
John R. Deller, John G. Proakis, Hansen John H. L.. IEEE Press, 2000
- ✖ Tiếng Việt hiện đại (Ngữ âm, ngữ pháp, phong cách)
Nguyễn Hữu Quỳnh, Hà Nội, 1994
- ✖ Dẫn luận Ngôn ngữ học
Nguyễn Thiện Giáp, Đoàn Thiện Thuật, Nguyễn Minh Thuyết, Hà Nội, 1994

2

2

1

NỘI DUNG

1. Một số khái niệm cơ bản
2. Xử lý tín hiệu tiếng nói
3. Mã hóa tiếng nói
4. Tổng hợp tiếng nói
5. Nhận dạng tiếng nói

3

3

1. MỘT SỐ KHÁI NIỆM CƠ BẢN

What is speech processing?

- The study of speech signals and their processing methods
- Speech processing encompasses a number of related areas
 - **Speech recognition:** extracting the linguistic content of the speech signal
 - **Speaker recognition:** recognizing the identity of speakers by their voice
 - **Speech coding:** compression of speech signals for telecommunication
 - **Speech synthesis:** computer-generated speech (e.g., from text)
 - **Speech enhancement:** improving intelligibility or perceptual quality of speech signals

4

4

Applications of speech processing

- Human computer interfaces (e.g., speech I/O, affective)
- Telecommunication (e.g., speech enhancement, translation)
- Assistive technologies (e.g., blindness/deafness, language learning)
- Audio mining (e.g., diarization, tagging)
- Security (e.g., biometrics, forensics)

Related disciplines

- Digital signal processing
- Natural language processing
- Machine learning
- Phonetics
- Human computer interaction
- Perceptual psychology

5

5

1. MỘT SỐ KHÁI NIỆM CƠ BẢN

- ✖ Xử lý thông tin chứa trong tín hiệu tiếng nói nhằm truyền, lưu trữ tín hiệu này hoặc tổng hợp, nhận dạng tiếng nói.
- ✖ Các nghiên cứu được tiến hành để xử lý tiếng nói yêu cầu những hiểu biết trên nhiều lĩnh vực ngày càng đa dạng: từ ngữ âm và ngôn ngữ học cho đến xử lý tín hiệu...

6

6

MỤC ĐÍCH

- ✖ Mã hóa một cách có hiệu quả tín hiệu tiếng nói để truyền và lưu trữ tiếng nói.
- ✖ Tổng hợp và nhận dạng tiếng nói tiến tới giao tiếp người-máy bằng tiếng nói.
- ✖ Tất cả các ứng dụng của xử lý tiếng nói đều cần phải dựa trên các kết quả của phân tích tiếng nói

7

7

MỘT SỐ KHÁI NIỆM CƠ BẢN

- ✖ Phân biệt tiếng nói và âm thanh

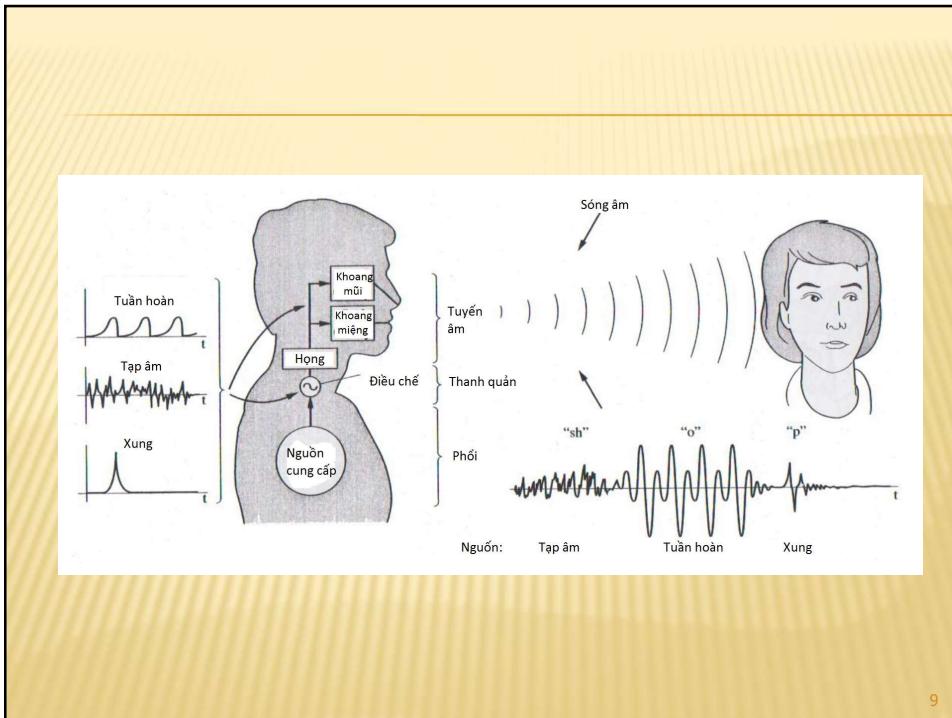
Tiếng nói được phân biệt với các âm thanh khác bởi các đặc tính âm học có nguồn gốc từ cơ chế tạo tiếng nói.

- ✖ Có các nguồn âm

- + Tuần hoàn (dây thanh rung)
- + Tạp âm (dây thanh không rung)
- + Xung

8

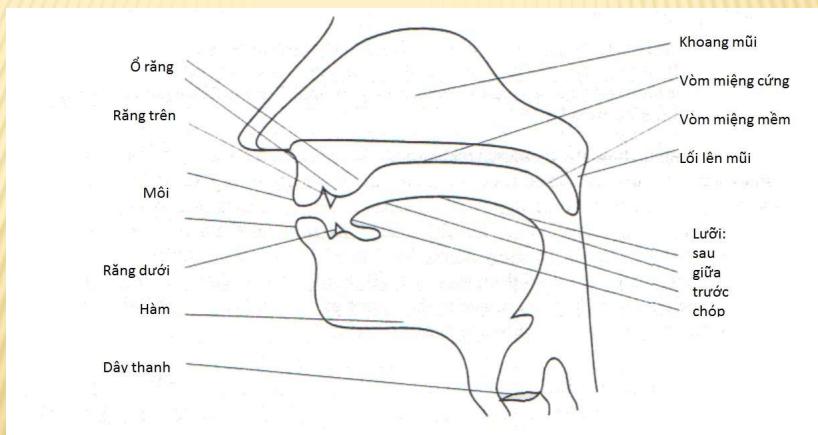
8



9

9

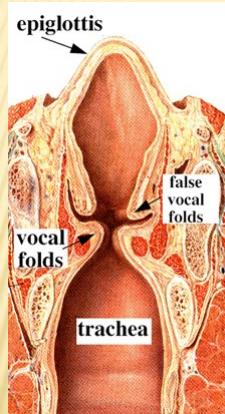
BỘ MÁY PHÁT ÂM



10

10

BỘ MÁY PHÁT ÂM

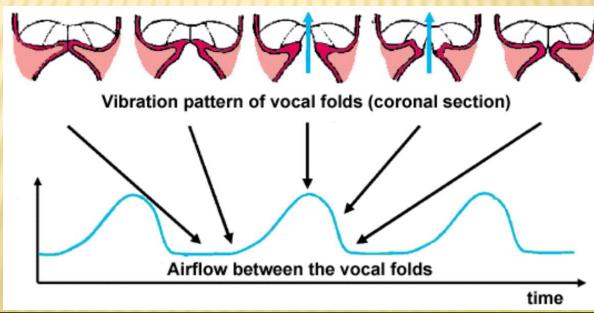
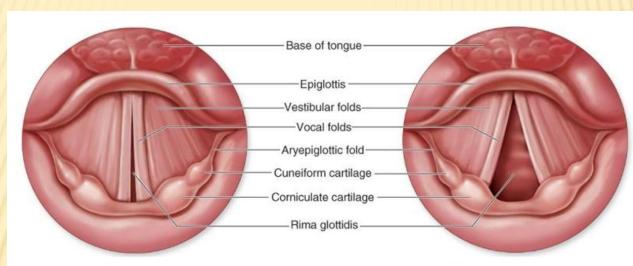


NASAL CAVITY: Khoang mũi
 SOFT PALATE: Vòm miệng mềm
 EPIGLOTTIS: Nắp thanh quản
 VOCAL FOLDS (CORDS): Dây thanh
 OESOPHAGUS: Thực quản
 TRACHEA: Khí quản
 PHARYNX: Họng

11

11

DÂY THANH VÀ THANH MÔN



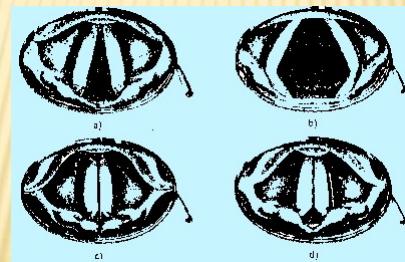
12

12

1. Một số khái niệm cơ bản

THANH MÔN

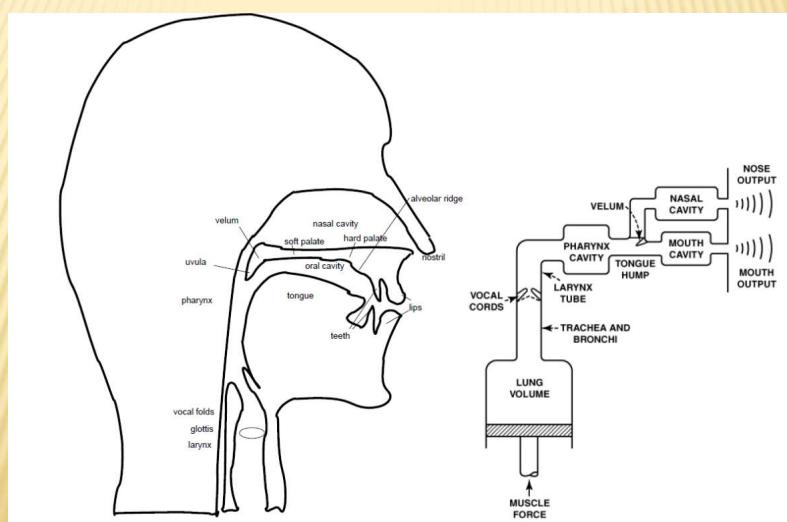
✗ Ở các vị trí hít, thở, phát âm, nói thì thào



13

13

SƠ ĐỒ KHỐI BỘ MÁY PHÁT ÂM



14

14

HỆ THỐNG THÍNH GIÁC

✗ Hệ thống thính giác có 2 thành phần quan trọng:

+ Cơ quan thính giác ngoại vi (tai)

✗ Biến đổi áp suất âm thanh thành dao động cơ học
kích thích tế bào thần kinh

+ Hệ thống thần kinh thính giác (não)

✗ Trích xuất các thông tin cảm nhận được ở mức độ khác nhau

15

15

HỆ THỐNG THÍNH GIÁC

✗ Tai có thể được phân chia

+ Tai ngoài:

✗ Bao gồm loa tai, ống tai ngoài và màng nhĩ
✗ Biến đổi áp suất âm thanh thành rung động

+ Tai giữa

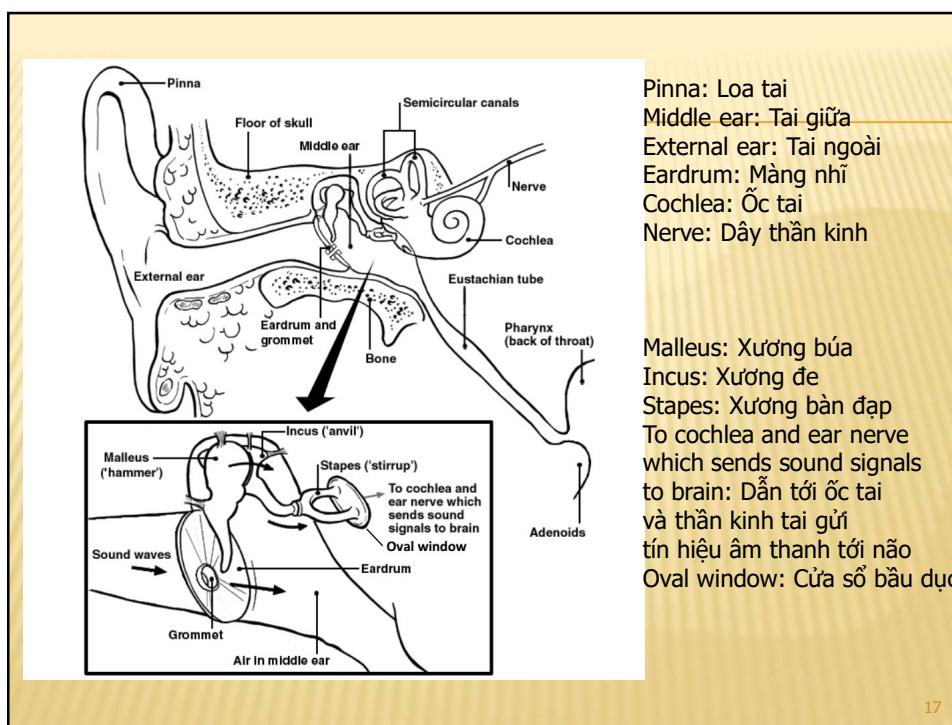
✗ Gồm các xương: xương búa, xương đe và xương bàn đạp
✗ Vận chuyển rung động màng nhĩ vào tai trong

+ Tai trong:

✗ Gồm ống tai
✗ Biến đổi các rung động thành các xung kích thích màng đáy
✗ Màng đáy có thể được mô hình hóa như băng bộ lọc

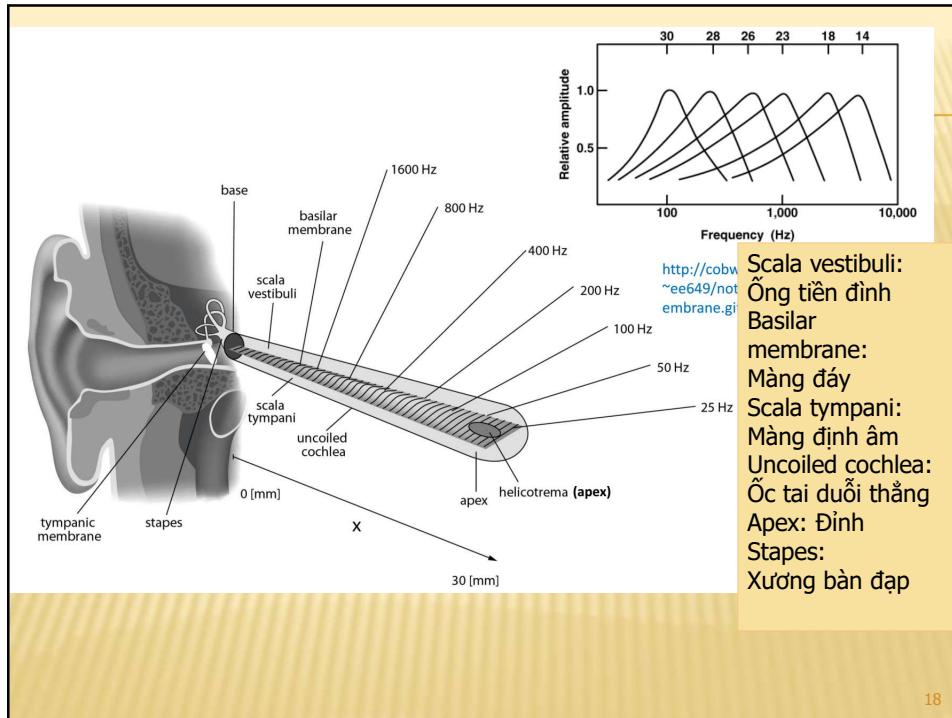
16

16



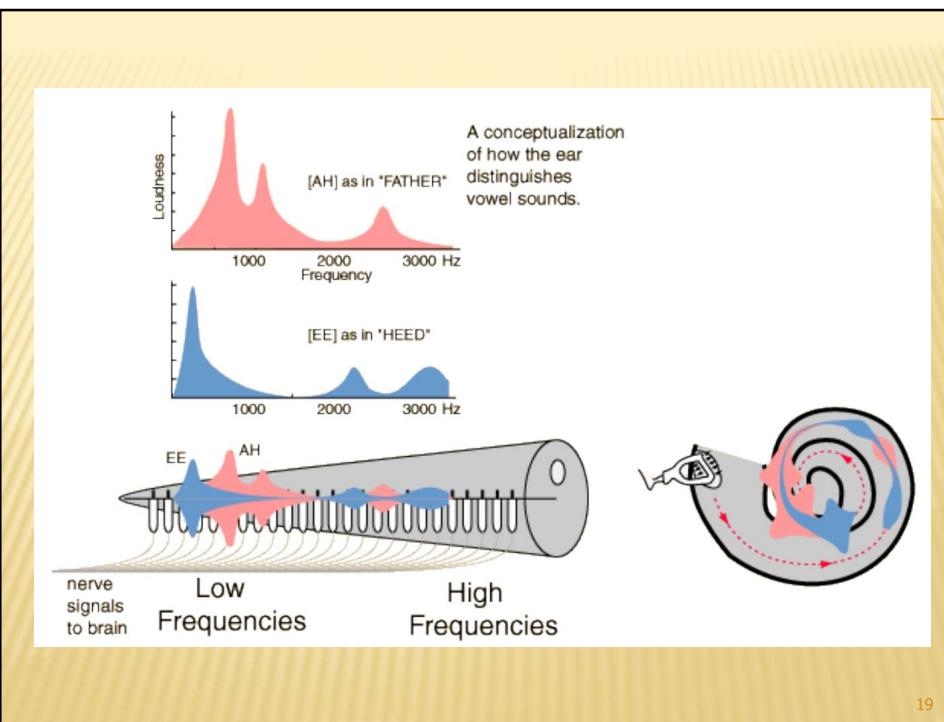
17

17



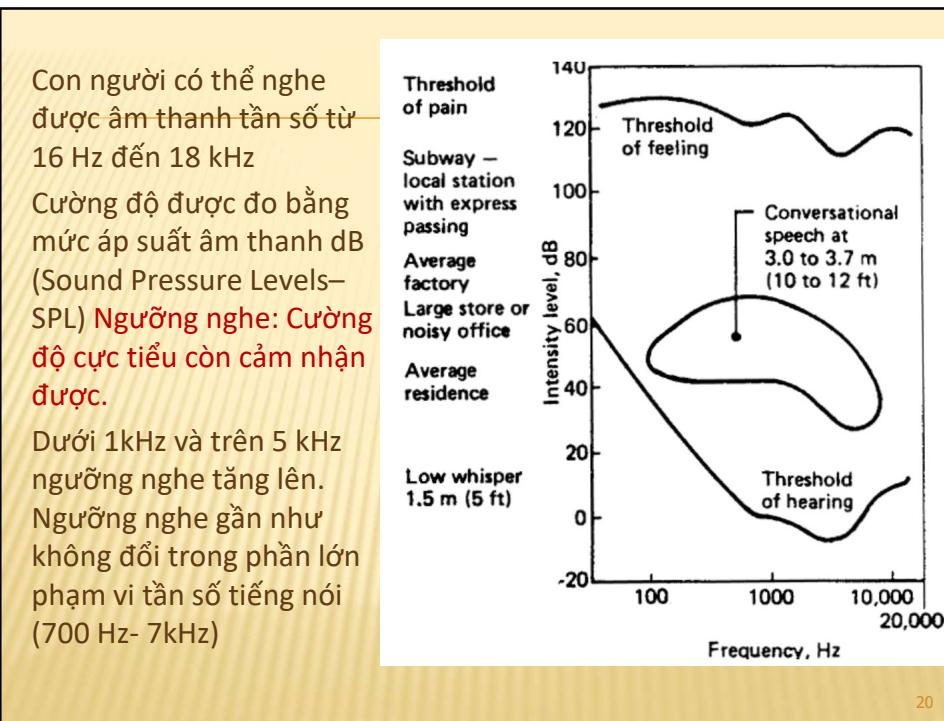
18

18



19

19

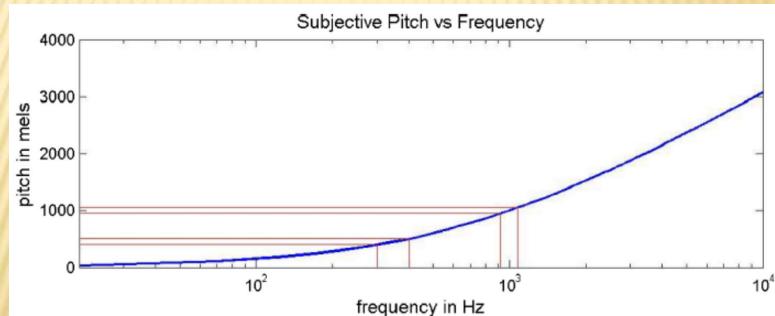


20

20

CẢM NHẬN CAO ĐỘ (PITCH)

- ✖ Cao độ là F0 (tần số cơ bản) được cảm nhận, mang tính chủ quan
- ✖ Quan hệ giữa Pitch và F0 là phi tuyến, có thể được mô tả theo thang Mel



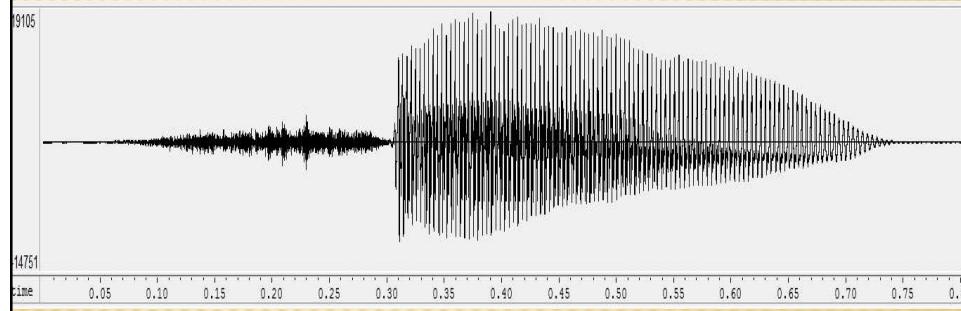
$$m = 2595 \log_{10}(1 + f/700)$$

21

21

BIỂU DIỄN TÍN HIỆU TIẾNG NÓI

- ✖ Dạng sóng theo thời gian



22

22

FILE WAV

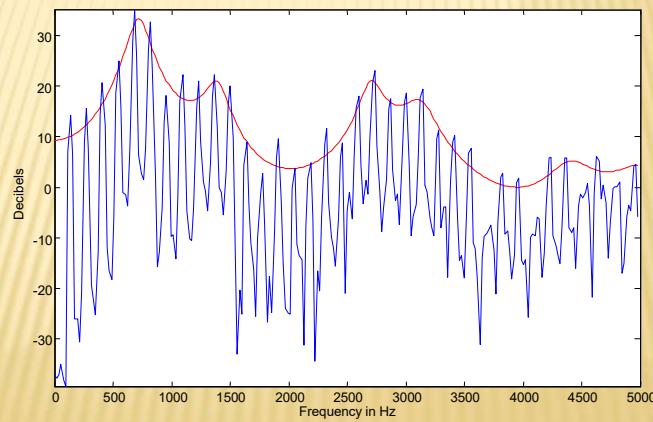
- ✖ Tần số lấy mẫu: 8kHz, F1= 11025 Hz, 2F1, 4F1
(16kHz, 10kHz)
- ✖ Số bit/mẫu: 8,16
- ✖ Mono, Stereo

23

23

BIỂU ĐIỂN TÍN HIỆU TIẾNG NÓI

- ✖ Phổ tín hiệu tiếng nói

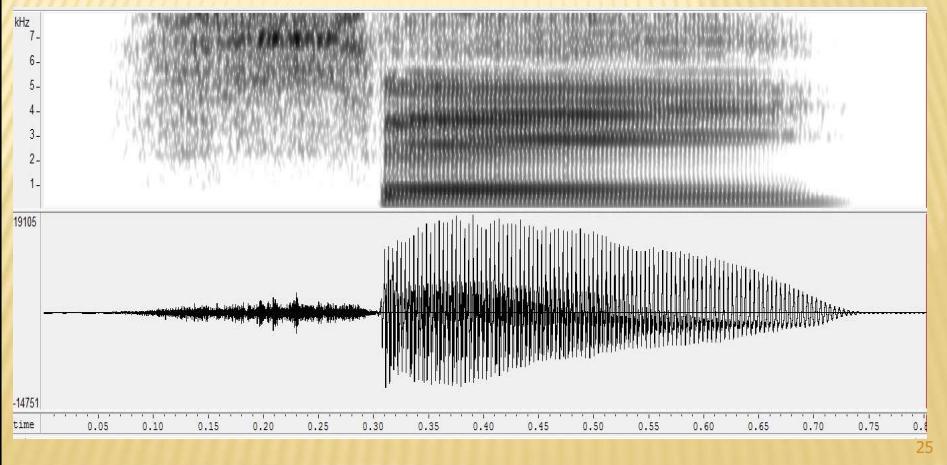


24

24

BIỂU ĐIỂN TÍN HIỆU TIẾNG NÓI

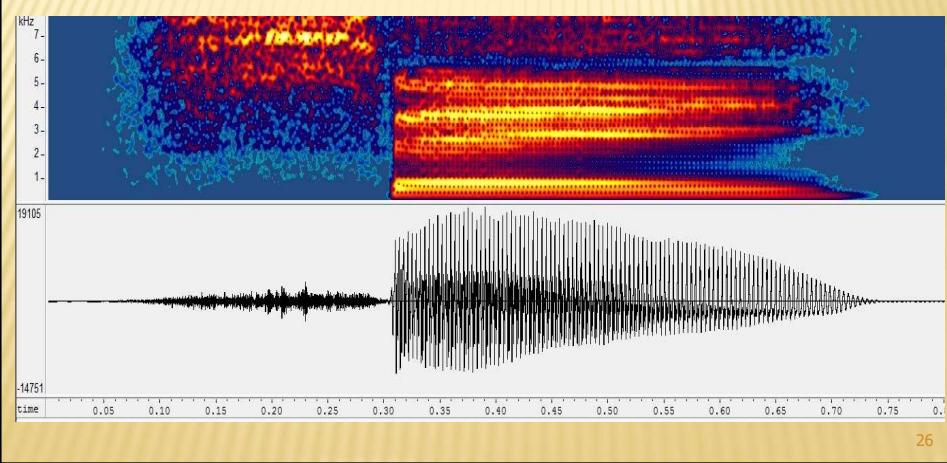
❖ Spectrogram (Sonagram)



25

BIỂU ĐIỂN TÍN HIỆU TIẾNG NÓI

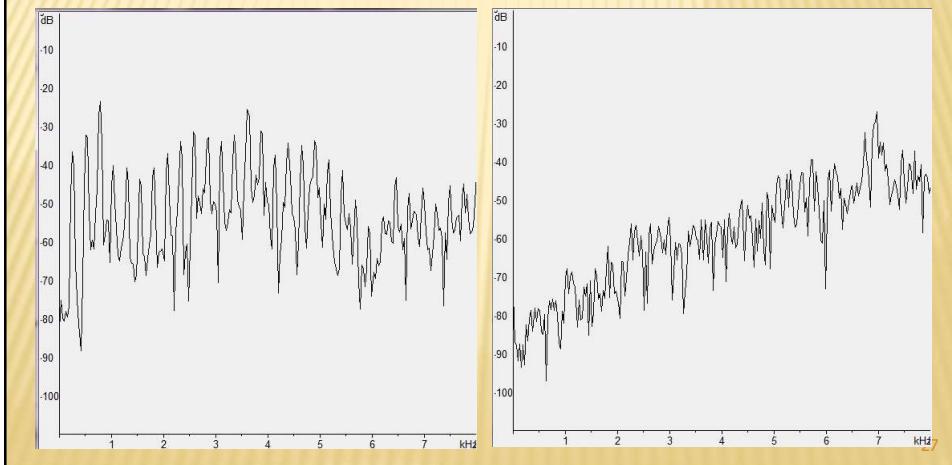
❖ Spectrogram (Sonagram)



26

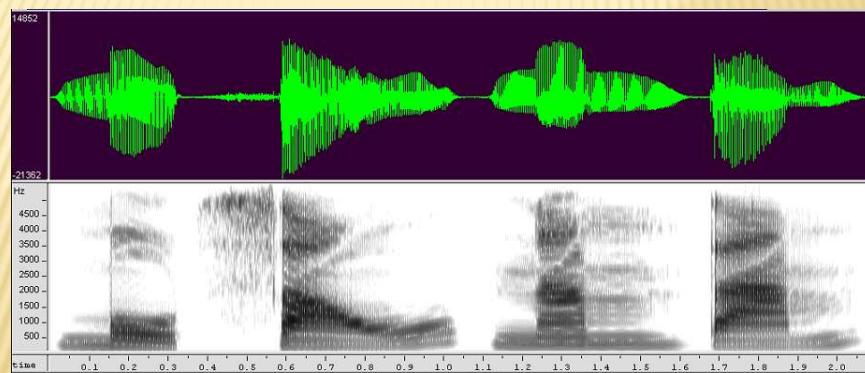
BIỂU ĐIỂN TÍN HIỆU TIẾNG NÓI

✖ Spectrogram (Sonagram)



27

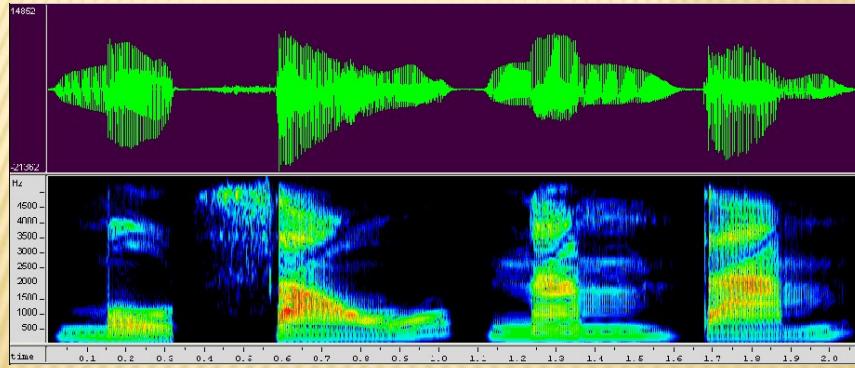
BIỂU ĐIỂN TÍN HIỆU TIẾNG NÓI



28

28

BIỂU ĐIỂN TÍN HIỆU TIẾNG NÓI

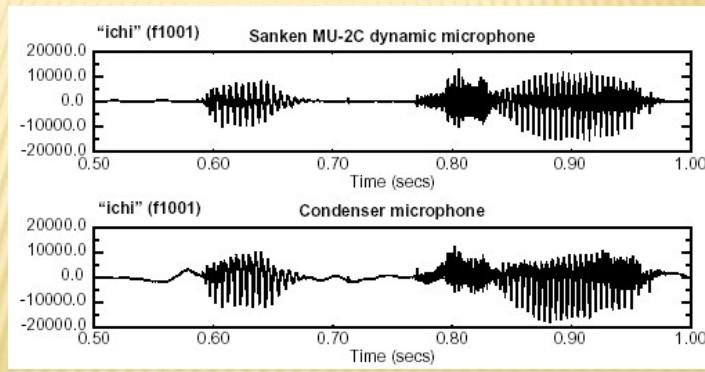


29

29

BIỂU ĐIỂN TÍN HIỆU TIẾNG NÓI

- ✖ Thu bằng micro khác loại

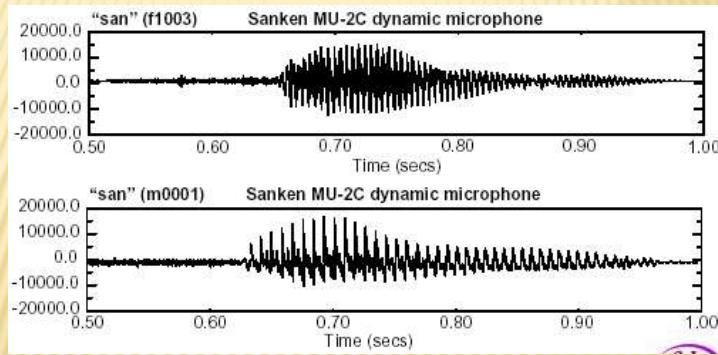


30

30

BIỂU DIỄN TÍN HIỆU TIẾNG NÓI

- ✖ Hai giọng khác nhau cho cùng một âm

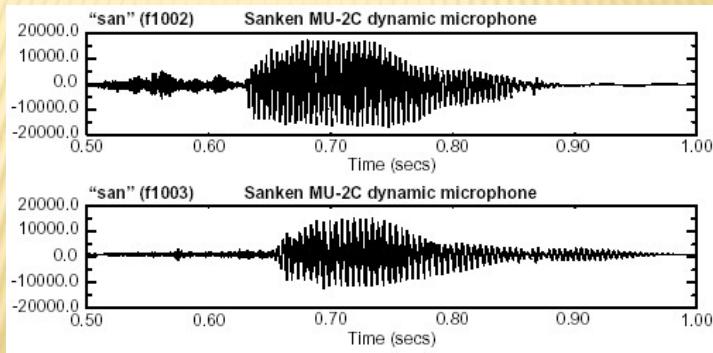


31

31

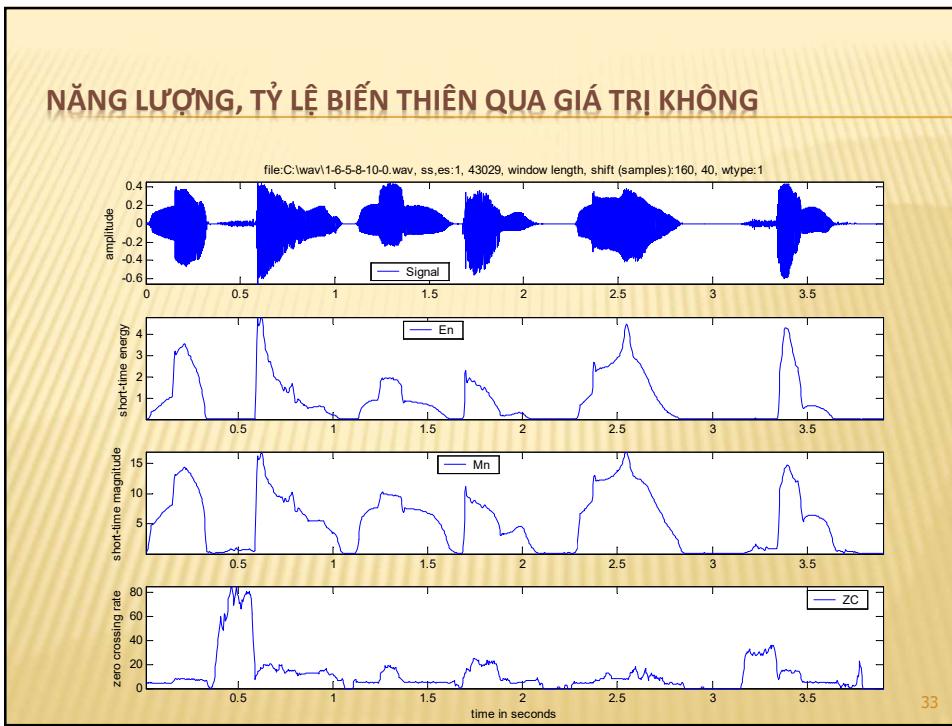
BIỂU DIỄN TÍN HIỆU TIẾNG NÓI

- ✖ Cùng người nói, cùng một âm

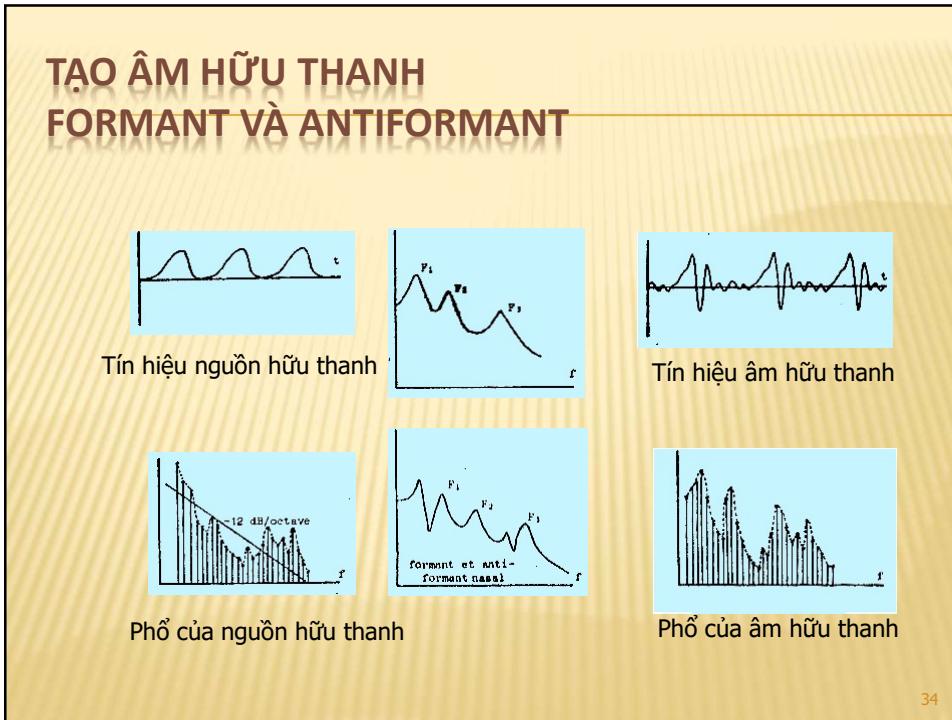


32

32

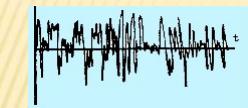


33

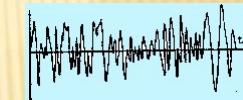


34

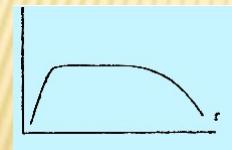
TẠO ÂM VÔ THANH



Tín hiệu nguồn vô thanh



Tín hiệu âm vô thanh



Phổ của nguồn vô thanh

35

35

MỘT SỐ ĐẶC ĐIỂM NGỮ ÂM TIẾNG VIỆT

- ✖ Đơn âm tiết
- ✖ Có thanh điệu (6), biến đổi thanh điệu kèm theo biến đổi nghĩa
- ✖ Không biến đổi hình thái

36

36

MỘT SỐ ĐẶC ĐIỂM NGỮ ÂM TIẾNG VIỆT

- ❖ Hệ thống âm vị: 14 nguyên âm (11 nguyên âm đơn, 3 nguyên âm đôi, 22 phụ âm)

1	i,y	ý chí
2	ê	é chè
3	e	e dè
4	a	a ha
5	ă	mắt
6	ơ	bơ phờ
7	â	ân cần
8	ư	tùtù
9	ô	ôtô
10	o	co ro
11	u	lù mù

1	ia,yê,ya,iê (đọc ia, yê)	kia kia, yêu kiều, khuya, tiên tiến
2	ua,uô (đọc ua)	tua tua, luôn
3	ưa,ươ (đọc ưa)	lưa thưa, lượt

37

37

MỘT SỐ ĐẶC ĐIỂM NGỮ ÂM TIẾNG VIỆT

- ❖ Hệ thống âm vị: 22 phụ âm

1	b	bồng bènh
2	p	óp ép
3	v	vẫn vơ
4	ph	phôi pha
5	m	mơ màng
6	đ	đất đai
7	t	tin tưởng
8	th	thơ thẩn
9	d,gi	duyên, giũ
10	n	nóng
11	l	long lanh

12	tr	trồng
13	s	sinh viên
14	r	rừng
15	ch	chông
16	nh	nhọc
17	ng,ngh	ngô nghê
18	c,k,q	con,kết,qua
19	kh	khúc
20	g,gh	gó ghề
21	h	hả hê
22	x	xa xôi

38

38

MỘT SỐ ĐẶC ĐIỂM NGỮ ÂM TIẾNG VIỆT

- Phân loại nguyên âm theo độ nâng của lưỡi và chuyển động của lưỡi

<i>Độ nâng Hàng</i>	<i>cao</i>	<i>trung bình</i>	<i>thấp</i>
trước	i e	e	
giữa	ư	ơ â	a ă
sau	u ô	o	

39

39

MỘT SỐ ĐẶC ĐIỂM NGỮ ÂM TIẾNG VIỆT

- Phân loại nguyên âm theo độ mở của miệng và chuyển động của lưỡi

<i>Hàng Độ mở</i>	<i>hang trước</i>	<i>hang sau không tròn môi</i>	<i>hang sau tròn môi</i>
hép	i ia,yê,ya,iê	ư ưa	u ua
hơi hép	ê	ơ â	ô
hơi rộng	e		o
rộng		a ă	

40

40

MỘT SỐ ĐẶC ĐIỂM NGỮ ÂM TIẾNG VIỆT

- Âm tắc: tiếng nổ, phát sinh do luồng khí từ phổi đi ra bị cản trở hoàn toàn, phải phá vỡ sự cản trở đó để thoát ra.
- Âm xát: tiếng cọ xát, phát sinh do luồng không khí đi ra bị cản trở không hoàn toàn (chỉ bị khó khăn), phải lách qua một khe hở nhỏ và trong khi thoát ra như vậy phải cọ xát vào thành của bộ máy phát âm.
- Phụ âm bên: đầu lưỡi tiếp xúc với lợi chặn lối thoát của không khí, buộc nó phải lách qua khe hở ở hai bên cạnh lưỡi tiếp giáp với má mà ra ngoài tạo nên tiếng xát nhẹ (l).
- Luồng không khí thoát ra ngoài bị cản trở, tạo nên tiếng xát hay tiếng nổ, dạng tín hiệu không tuần hoàn gọi là tiếng động (ồn).
- Trong khi phát âm một số phụ âm, dây thanh cũng hoạt động đồng thời tạo nên tiếng thanh.
- Phụ âm có tỉ lệ tiếng động lớn hơn gọi là phụ âm ồn.
- Phụ âm có tỉ lệ tiếng thanh lớn hơn gọi là phụ âm vang.

41

41

MỘT SỐ ĐẶC ĐIỂM NGỮ ÂM TIẾNG VIỆT

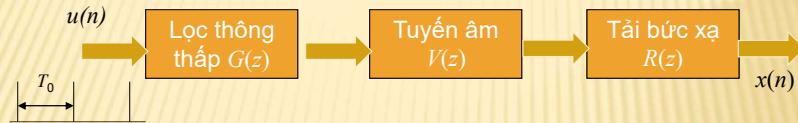
- Phân loại phụ âm theo tắc hay xát, hữu thanh hay vô thanh, mũi hóa

Phương thức câu âm		Vị trí câu âm		Môi	Đầu lưỡi		Mặt lưỡi	Cuối lưỡi	Họng
		Bật hơi	Vô thanh		Răng	Vòm miệng			
Tắc	Ôn	Không bật hơi	p	t	tr	ch	c,k,qu		
			b	đ					
	Xát	Vang mũi	m	n		nh	ng,ngh		
		Vô thanh	ph	x	s		kh	h	
		Hữu thanh	v	d,gi	r		g		
		Vang bên	l						

42

42

MÔ HÌNH TẠO TIẾNG NÓI (FANT-1960)



$$G(z) = \frac{A}{(1 + \alpha z^{-1})(1 + \beta z^{-1})}$$

$$R(z) = C(1 - z^{-1})$$

$$V(z) = \frac{B}{\prod_{k=1}^K (1 + b_{1k}z^{-1} + b_{2k}z^{-2})}$$

43

43

MÔ HÌNH TOÀN ĐIỂM CỰC (AR)

$$T(z) = G(z)V(z)R(z) = \frac{\sigma}{A(z)}$$

✖ $A(z)$: Hàm truyền đạt của bộ lọc đảo

$$T(z) = \frac{\sigma}{A(z)}$$

$$A(z) = 1 + \sum_{i=1}^{2K+1} a_i z^{-i} \quad A(z) = \sum_{i=0}^p a_i z^{-i} \quad a_0 = 1$$

$$x(n) + \sum_{i=1}^p a_i x(n-i) = \sigma u(n)$$

$$P = 2K+1$$

44

44

MÔ HÌNH ARMA (AUTOREGRESSIVE MOVING AVERAGE)

$$T(z) = \frac{\sigma_1}{A_1(z)} + \frac{\sigma_2}{A_2(z)} = \sigma \frac{C(z)}{A(z)}$$

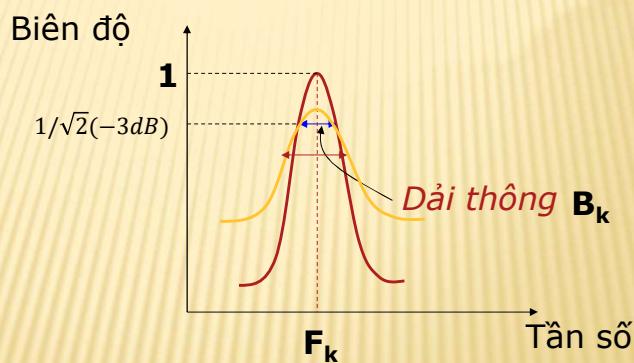
$$C(z) = \sum_{i=0}^q c_i z^{-i} \quad c_0 = 1$$

$$x(n) + \sum_{i=1}^p a_i x(n-i) = \sigma \sum_{l=0}^q c_l u(n-l)$$

45

45

DẢI THÔNG

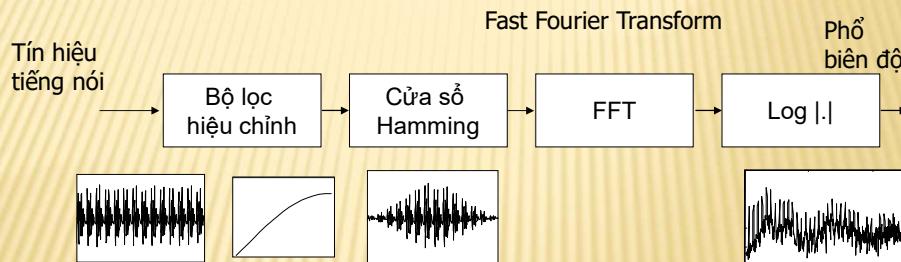


46

46

2. XỬ LÝ TÍN HIỆU TIẾNG NÓI

✖ Phân tích phổ



+ Bộ lọc hiệu chỉnh $H(z) = 1 - az^{-1}$, $a = 0,95..0,98$

47

47

✖ Lọc hiệu chỉnh $s_1(n) = s(n) - a \cdot s(n-1)$

✖ Sau cửa sổ Hamming $w_H(n)$

$$s_H(n) = s_1(n) \cdot w_H(n)$$

✖ Tính biến đổi Fourier

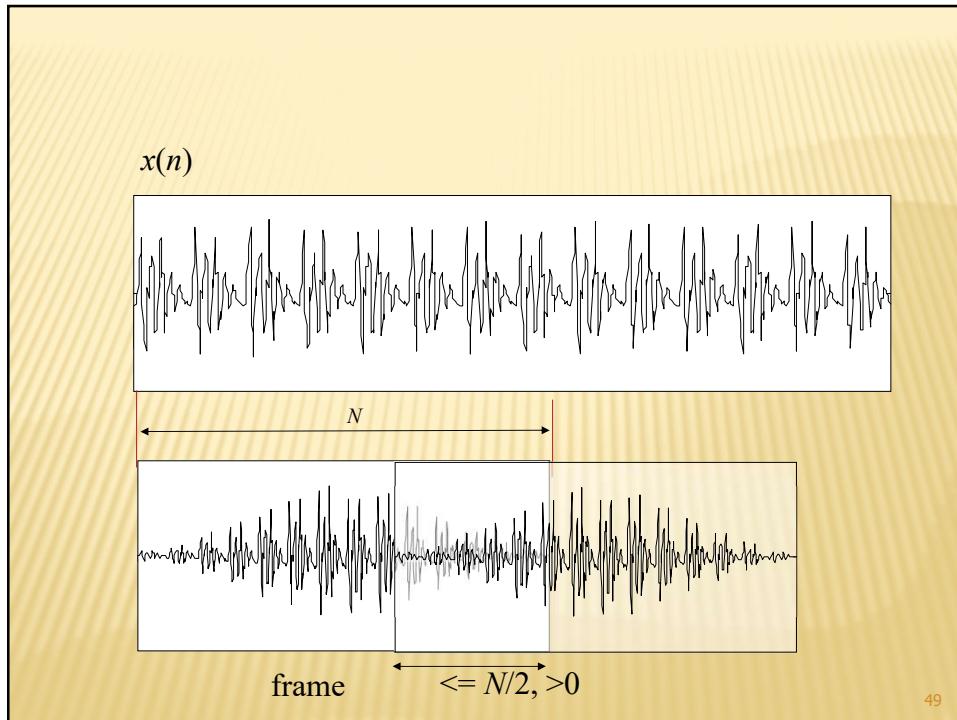
$$S(k) = \sum_{n=0}^{N-1} s_H(n) \cdot e^{-j \frac{2\pi}{N} kn}$$

✖ Phổ biên độ theo dB

$$S_{dB}(k) = -20 \lg \frac{|S(k)|}{|S(k)|_{max}}$$

48

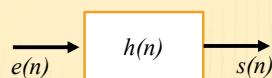
48



49

49

XỬ LÝ ĐỒNG HÌNH (HOMOMORPHIC)

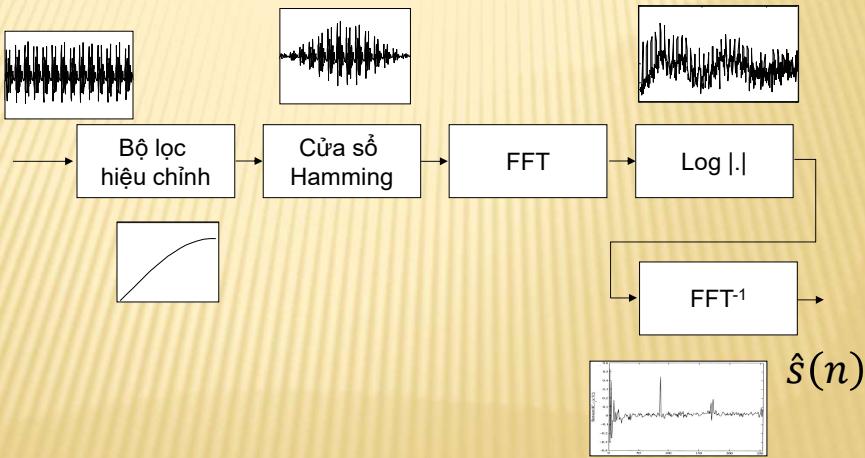


- ✖ $s(n) = h(n) * e(n) \rightarrow S(\omega) = H(\omega)E(\omega)$
- ✖ $\log S(\omega) = \log H(\omega) + \log E(\omega)$
- ✖ $\mathbb{F}^{-1}\{\log S(\omega)\} = \mathbb{F}^{-1}\{\log H(\omega)\} + \mathbb{F}^{-1}\{\log E(\omega)\}$
- ✖ $\hat{s}(n) = \hat{h}(n) + \hat{e}(n)$

50

50

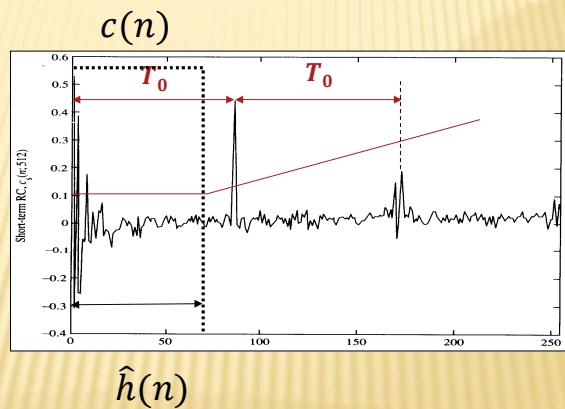
SƠ ĐỒ KHỐI XỬ LÝ ĐỒNG HÌNH



51

51

VÍ DỤ



52

52

TIÊN ĐOÁN TUYẾN TÍNH (LINEAR PREDICTION CODING)

✖ Mô hình AR

$$x(n) + \sum_{i=1}^p a_i x(n-i) = \sigma u(n)$$

Tiên đoán

$$\hat{x}(n) = - \sum_{i=1}^p \hat{a}_i x(n-i)$$

Sai số tiên đoán

$$e(n) = x(n) - \hat{x}(n)$$

Sai số bình phương toàn phần

$$E = \sum_n e^2(n)$$

Tối thiểu hóa sai số

$$\frac{\partial E}{\partial \hat{a}_i} = 0, i = 1, 2, \dots, p$$

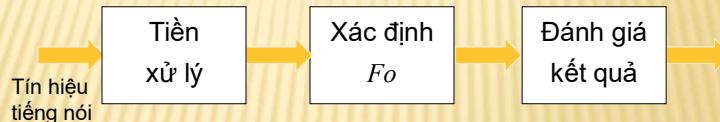
53

53

XÁC ĐỊNH TẦN SỐ CƠ BẢN

✖ Giá trị F_0 phụ thuộc vào giới tính và lứa tuổi

- + Giọng nam: 80..250 Hz
- + Giọng nữ: 150..500 Hz



54

54

MỘT SỐ PHƯƠNG PHÁP XÁC ĐỊNH F_o

- ✖ Dựa vào hàm tự tương quan
- ✖ Dựa vào hàm vi sai biên độ trung bình
- ✖ Dùng bộ lọc đảo và hàm tự tương quan
- ✖ Xử lý đồng hình

55

55

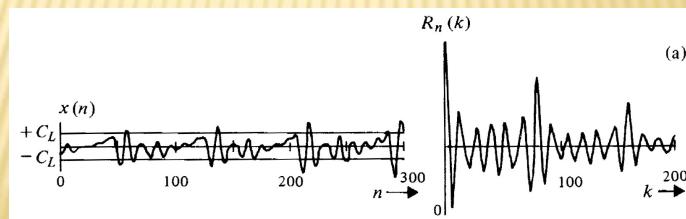
DỰA VÀO HÀM TỰ TƯƠNG QUAN

- ✖ Tính hàm tự tương quan $R(k)$ của tín hiệu tiếng nói $x(n)$

$$R(k) = \sum_{n=0}^{N-1-k} x(n)x(n+k) \quad k = 0, 1, \dots, K$$

$F_s = 10 \text{ kHz}, N = 300, K = 150.$

Tìm cực đại trong khoảng $(0, K)$

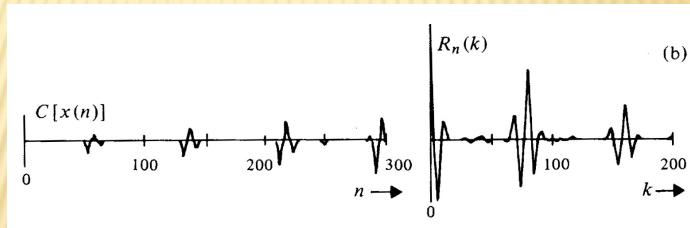


56

56

PHƯƠNG PHÁP TỰ TƯƠNG QUAN CÓ CẢI TIẾN

- ❖ Hạn chế, loại bỏ $|x| < C_L$



57

57

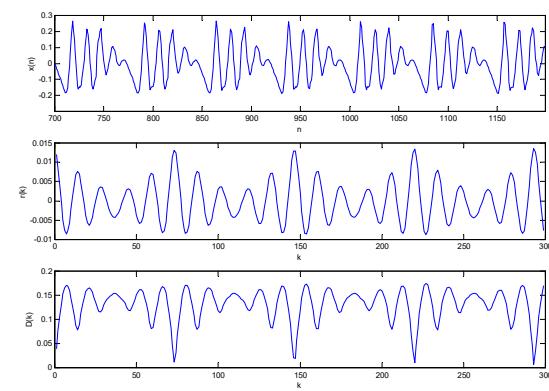
DỰA VÀO HÀM VI SAI BIÊN ĐỘ TRUNG BÌNH (AMDF- AVERAGE MAGNITUDE DIFFERENCE FUNCTION)

$$\begin{aligned}
 D(k) &= \sum_{m=0}^{N-k-1} |x(n+m) - x(n+m+k)| \\
 k &= 0, 1, \dots, K \\
 D(iP) &= 0, \quad i = 0, 1, \dots \quad \frac{1}{N} \sum_{n=0}^{N-1} |u(n)| \leq \left[\frac{1}{N} \sum_{n=0}^{N-1} u^2(n) \right]^{1/2} \\
 D(k) &= \lambda \left\{ \frac{1}{N} \sum_{m=0}^{N-1} [x(n+m) - x(n+m-k)]^2 \right\}^{1/2} \\
 &= \lambda \left\{ \frac{1}{N} [2r(0) - 2r(k)] \right\}^{1/2} \quad k = 0, 1, \dots, K \\
 \text{với } \lambda &< 1
 \end{aligned}$$

58

58

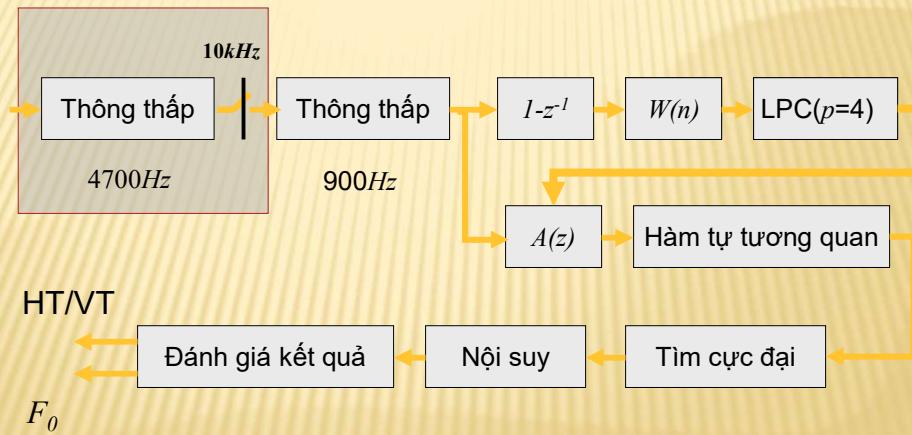
VÍ DỤ



59

59

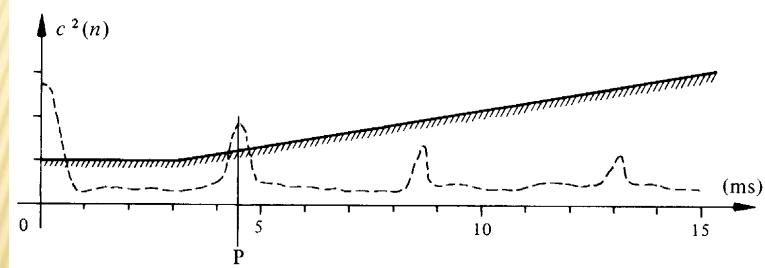
DÙNG BỘ LỌC ĐẢO (SIFT - SIMPLIFIED INVERSE FILTER TRACKING)



60

60

XỬ LÝ ĐỒNG HÌNH



61

61

XÁC ĐỊNH FORMANT

✖ Tham số cần xác định

- + Formant F_k

- + Dải thông B_k

✖ Phương pháp

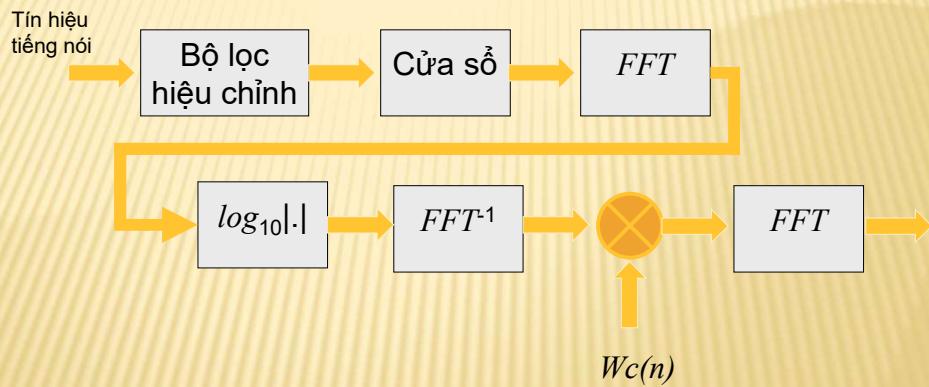
- + Xử lý đồng hình

- + LPC

62

62

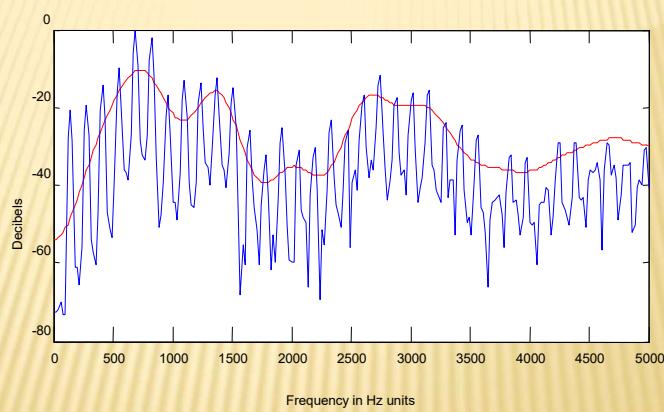
XỬ LÝ ĐỒNG HÌNH



63

63

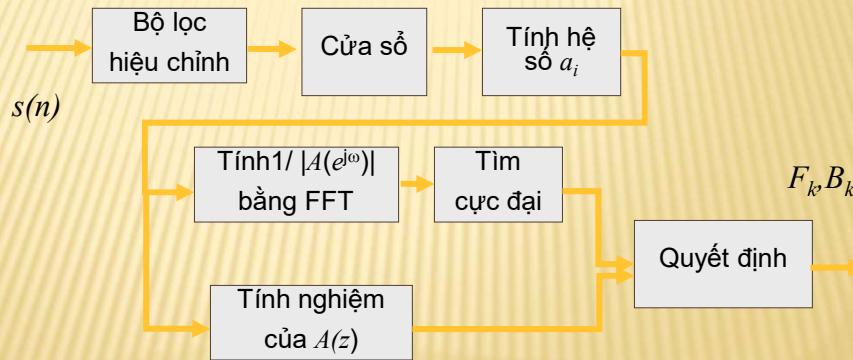
XỬ LÝ ĐỒNG HÌNH



64

64

PHƯƠNG PHÁP LPC

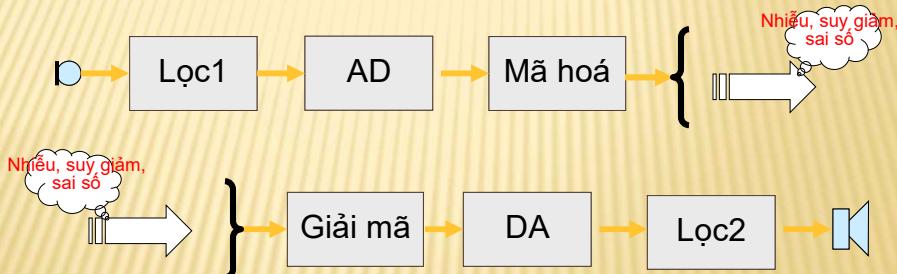


65

65

3. MÃ HÓA TIẾNG NÓI

- ❖ Dãy thao tác mã hóa và giải mã



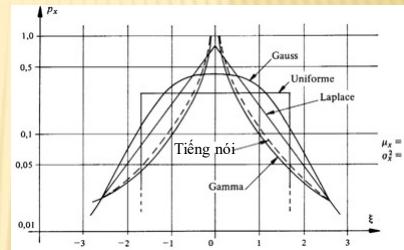
66

66

MỘT SỐ TÍNH CHẤT THỐNG KÊ CỦA TÍN HIỆU TIẾNG NÓI

✖ Mật độ xác suất

N_ξ : số lượng mẫu $x(n)$
có biên độ trong
khoảng $[\xi - \Delta\xi/2, \xi + \Delta\xi/2]$
 $n \in [-N, \dots, N]$
 x ergodic và dừng



$$p_x(\xi) = \lim_{\substack{N \rightarrow \infty \\ \Delta\xi \rightarrow 0}} [N_\xi / (2N + 1)]$$

67

67

GIÁ TRỊ TRUNG BÌNH VÀ PHƯƠNG SAI

✖ Giá trị trung bình của tín hiệu dừng

$$\mu_x = \int_{-\infty}^{\infty} \xi p_x(\xi) d\xi = \lim_{N \rightarrow \infty} \frac{1}{2N + 1} \sum_{n=-N}^N x(n)$$

với tín hiệu tiếng nói $\mu_x = 0$

✖ Phương sai

$$\sigma_x^2 = \int_{-\infty}^{\infty} \xi^2 p_x(\xi) d\xi = \lim_{N \rightarrow \infty} \frac{1}{2N + 1} \sum_{n=-N}^N x^2(n)$$

68

68

LƯỢNG TỬ TỨC THỜI (KHÔNG NHỚ)

- ✖ Luật lượng tử $y = Q(x)$ được định nghĩa:
 - + $(L + 1)$ mức tín hiệu $x(0), x(1), \dots, x(L)$
 - + L mức lượng tử hoá
- ✖ Mỗi mức lượng tử hoá biểu diễn bằng từ b bit

$$L = 2^b.$$
- ✖ Sai số lượng tử (tập âm lượng tử) $e_q = Q(x) - x$
- ✖ Bước lượng tử : hiệu 2 mức tín hiệu kề nhau

$$\Delta(i) = x(i) - x(i - 1)$$
- ✖ Thông lượng $I = bFs$ (bit/s). Fs : tần số lấy mẫu

69

69

THÔNG LƯỢNG

- ✖ Tín hiệu lượng tử 8 bit (256 mức), $Fs = 8 \text{ kHz} \rightarrow$
Thông lượng = 64 kbit/s
- ✖ Tín hiệu lượng tử 16 bit (65536 mức),
 $Fs = 16 \text{ kHz} \rightarrow$ Thông lượng = 256 kbit/s ,
1 giờ tiếng nói $\sim 100 \text{ Mbyte}$
- ✖ Cần phải mã hoá tín hiệu tiếng nói (**MPEG**, **GSM**, **G723**, ...) để
truyền tiếng nói trên mạng hoặc lưu trữ

70

70

THÔNG LƯỢNG

Tần số lấy mẫu (kHz)	Số bit cho 1 mẫu	Thông lượng kbit/s	Dung lượng / phút (kbyte)	Lĩnh vực
48	16	768	11520	Ghi âm chuyên nghiệp
44,1	16	705,6	10584	CD Audio
32	16	512	7680	Radio FM
22	12	264	3960	Radio AM
8	8	64	960	Điện thoại

71

71

LƯỢNG TỬ ĐỀU

- ✖ Tổng quát, bước lượng tử là hàm của biên độ tín hiệu x (lượng tử không đều) → đơn giản nhất là lượng tử đều.
- ✖ Lượng tử đơn cực: Tín hiệu tương tự biến thiên từ 0 von đến một giá trị dương nào đó.
- ✖ Lượng tử lưỡng cực: Tín hiệu tương tự biến thiên từ giá trị âm đến giá trị dương nào đó.
- ✖ x_{max}, x_{min} : giá trị cực đại và cực tiểu của tín hiệu tương tự x

72

72

LƯỢNG TỬ ĐỀU

- ✖ L : Số mức lượng tử, b : số bit cho một mức lượng tử dùng trong ADC. $L = 2^b$
- ✖ Bước lượng tử $\Delta = (x_{max} - x_{min})/L$
- ✖ i : chỉ số tương ứng với mã nhị phân
 $i = \text{round}((x - x_{min})/\Delta)$
- ✖ x_q : mức lượng tử
 $x_q = x_{min} + i\Delta, i = 0, 1, \dots, L - 1$
- ✖ e_q : sai số lượng tử $e_q = x_q - x$

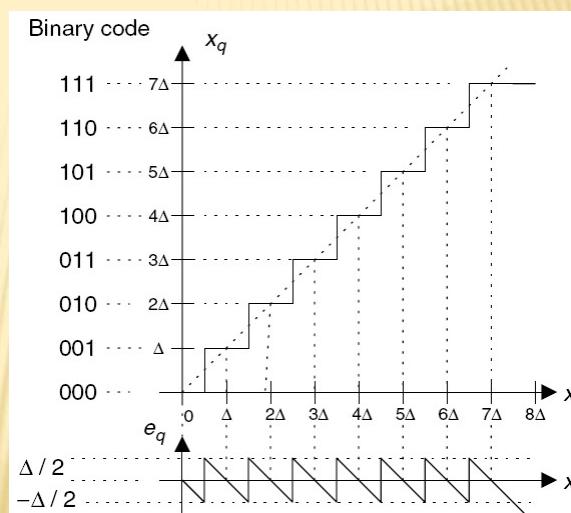
73

73

LƯỢNG TỬ ĐỀU

- ✖ Đơn vị c

$$\begin{aligned} x_{min} &= 0 \\ x_{max} &= 8\Delta \\ b &= 3 \\ L &= 8 \\ x_q &= 0 + i\Delta, \\ i &= 0, 1, \dots, L - 1 \\ -\Delta/2 \leq e_q &\leq \Delta/2 \end{aligned}$$



74

74

LƯỢNG TỬ ĐỀU

Bảng lượng tử của bộ lượng tử đơn cực 3 bit, $x_{min} = 0$

x_{max} = giá trị điện áp cực đại



Binary Code	Quantization Level x_q (V)	Input Signal Subrange (V)
0 0 0	0	$0 \leq x < 0.5\Delta$
0 0 1	Δ	$0.5\Delta \leq x < 1.5\Delta$
0 1 0	2Δ	$1.5\Delta \leq x < 2.5\Delta$
0 1 1	3Δ	$2.5\Delta \leq x < 3.5\Delta$
1 0 0	4Δ	$3.5\Delta \leq x < 4.5\Delta$
1 0 1	5Δ	$4.5\Delta \leq x < 5.5\Delta$
1 1 0	6Δ	$5.5\Delta \leq x < 6.5\Delta$
1 1 1	7Δ	$6.5\Delta \leq x < 7.5\Delta$

75

75

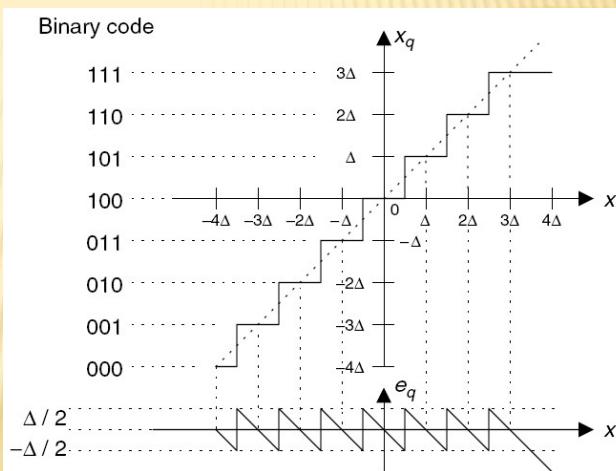
LƯỢNG TỬ ĐỀU

❖ Lưỡng cực

$x_{min} = -4\Delta$

$x_{max} = 4\Delta$

$b = 3$



76

76

LƯỢNG TỬ ĐỀU

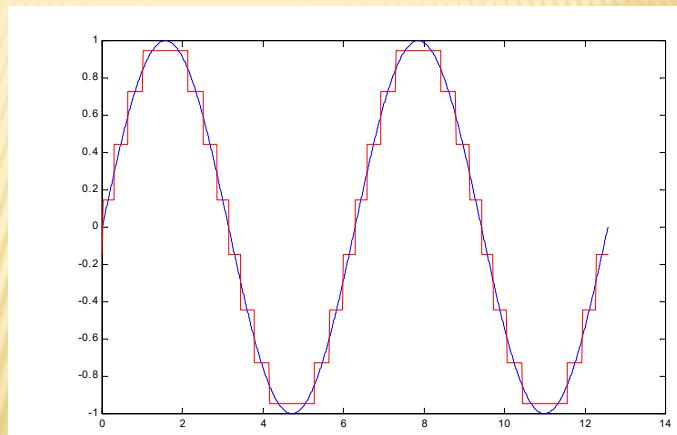
- Bảng lượng tử của bộ lượng tử lưỡng cực 3 bit,
 x_{max} = giá trị điện áp cực đại, $x_{min} = -x_{max}$

Binary Code	Quantization Level x_q (V)	Input Signal Subrange (V)
000	-4Δ	$-4\Delta \leq x < -3.5\Delta$
001	-3Δ	$-3.5\Delta \leq x < -2.5\Delta$
010	-2Δ	$-2.5\Delta \leq x < -1.5\Delta$
011	$-\Delta$	$-1.5\Delta \leq x < -0.5\Delta$
100	0	$-0.5\Delta \leq x < 0.5\Delta$
101	Δ	$0.5\Delta \leq x < 1.5\Delta$
110	2Δ	$1.5\Delta \leq x < 2.5\Delta$
111	3Δ	$2.5\Delta \leq x < 3.5\Delta$

77

77

LƯỢNG TỬ ĐỀU

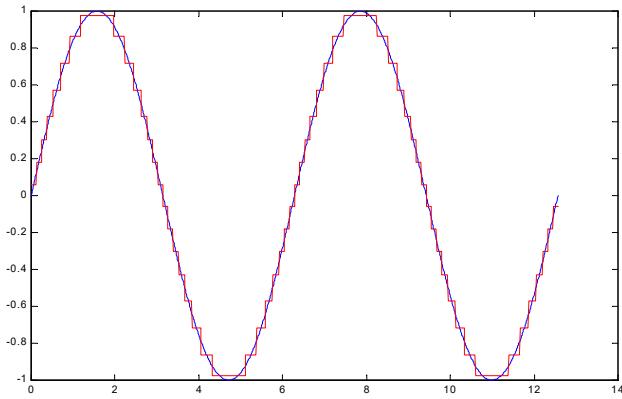


78

78

LƯỢNG TỬ ĐẦU

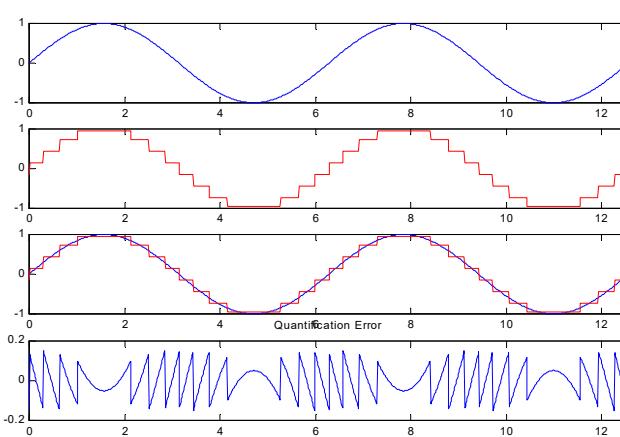
✖ $L = 16$



79

79

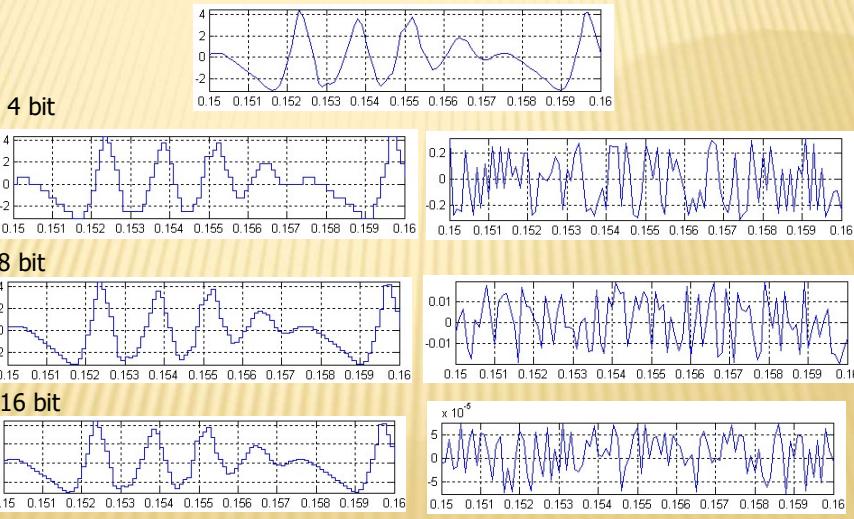
LƯỢNG TỬ ĐẦU



80

80

LƯỢNG TỬ ĐỀU



81

81

CÁC TÍNH CHẤT LƯỢNG TỬ ĐỀU

- ❖ Mật độ xác suất sai số lượng tử

$$p_e(\xi) = \sum_{i=-\ell}^{\ell} p_x(i\Delta + \xi), \quad \ell = (L-1)/2$$

phân bố đều giữa $-\Delta/2$ và $+\Delta/2$

$$p_e(\xi) = 1/\Delta, |\xi| \leq \Delta/2 \\ = 0, |\xi| > \Delta/2$$

- ❖ Trung bình tần âm lượng tử = 0

- ❖ Phương sai

$$\sigma_e^2 = \int_{-\delta/2}^{\delta/2} \xi^2 / \Delta d\xi = \Delta^2 / 12$$

82

82

CÁC TÍNH CHẤT LƯỢNG TỬ ĐỀU

- Tỷ số tín hiệu trên nhiễu

$$SN = \frac{\sigma_x^2}{\sigma_e^2}$$

$$SN(dB) = 10 \lg \left(\frac{\sigma_x^2}{\sigma_e^2} \right) = 6,02b + 4,77 - 20 \lg \left(\frac{x_{\max}}{\sigma_x} \right)$$

$$\text{Nếu } x_{\max} = 4\sigma_{\max} \rightarrow SN(dB) = 6b - 7,3$$

Với $b \geq 6$, tăng 6 dB mỗi khi tăng 1 bit lượng tử. Để có chất lượng thích hợp cần có $b \geq 11$

- Có thể tính SN như sau:

$$SN = \frac{\frac{1}{N} \sum_{n=0}^{N-1} x^2(n)}{\frac{1}{N} \sum_{n=0}^{N-1} e_q^2(n)} = \frac{\sum_{n=0}^{N-1} x^2(n)}{\sum_{n=0}^{N-1} e_q^2(n)}$$

83

83

TỶ SỐ TÍN HIỆU TRÊN NHIỄU

$$\times SN = \frac{\text{Năng lượng tín hiệu}}{\text{Năng lượng nhiễu}} = \frac{W_s}{W_n}$$

$$\times SN_{dB} = 10 \log_{10} SN$$

Hoặc

$$\times SN_{dB} = 20 \log_{10} \frac{\text{Biên độ tín hiệu}}{\text{Biên độ nhiễu}}$$

84

84

TỶ SỐ TÍN HIỆU TRÊN NHIỀU

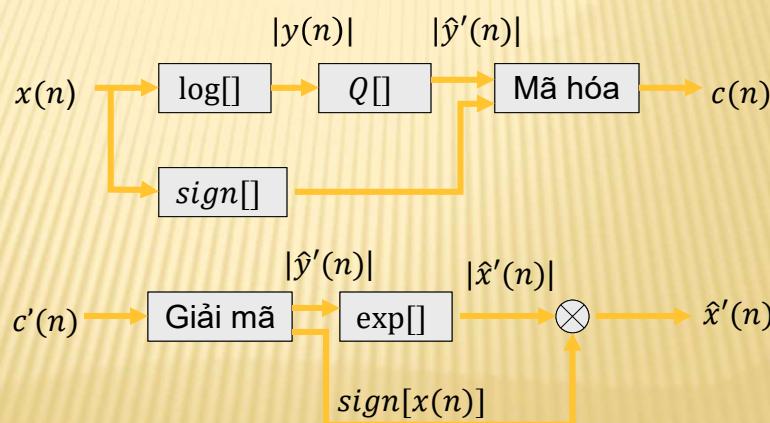
Năng lượng	SN (dB)
Tín hiệu = Nhiều	0
Tín hiệu = 2 Nhiều	2
Tín hiệu = 10 Nhiều	10
Tín hiệu = 100 Nhiều	20
Tín hiệu = 1000 Nhiều	30
Tín hiệu = 10^N Nhiều	$N \times 10$

85

85

LƯỢNG TỬ LOGARIT

- Sau khi lấy logarit biên độ tín hiệu sẽ mã hóa tuyến tính



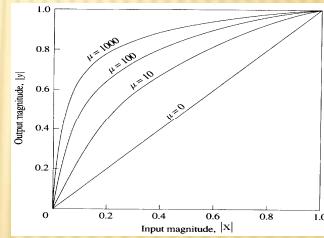
86

86

LƯỢNG TỬ LOGARIT

- ✖ Hai giải pháp dùng cho điện thoại
- + Luật μ (dùng ở Mỹ)

$$|y| = \frac{\log(1 + \mu|x|)}{\log(1 + \mu)}$$

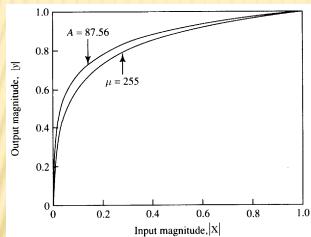


87

87

LƯỢNG TỬ LOGARIT

- ✖ Hai giải pháp dùng cho điện thoại
- + Luật A (dùng ở châu Âu)



$$|y| = \frac{1 + \log A |x|}{1 + \log A}$$

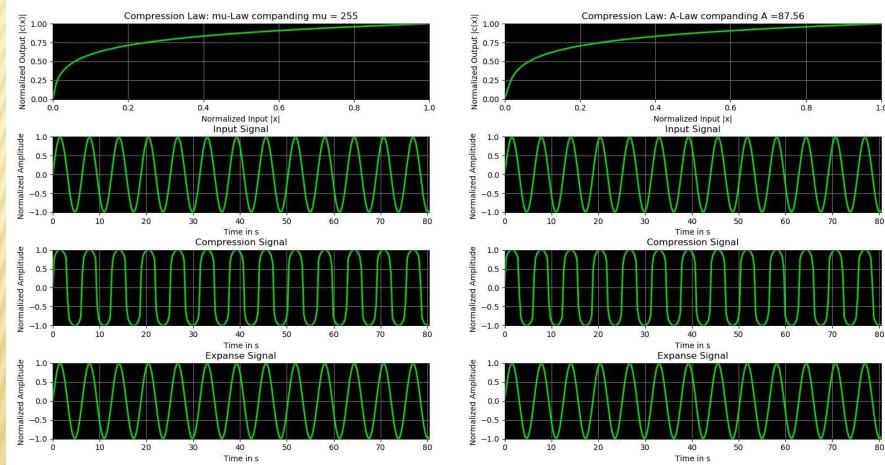
$$\mu = 255 \sim A = 87,56$$

8 bit logarit \sim 12 bit lượng tử đều

88

88

VÍ DỤ MU-LAW, A-LAW



89

89

Algorithms

For a given signal x , the output of the μ -law compressor is

$$y = \frac{V \log(1 + \mu|x|/V)}{\log(1 + \mu)} \operatorname{sgn}(x)$$

where V is the maximum value of the signal x , μ is the μ -law parameter of the compander, \log is the natural logarithm, and sgn is the signum function.

The output of the A-law compressor is

$$y = \begin{cases} \frac{A|x|}{1 + \log A} \operatorname{sgn}(x) & \text{for } 0 \leq |x| \leq \frac{V}{A} \\ \frac{V(1 + \log(A|x|/V))}{1 + \log A} \operatorname{sgn}(x) & \text{for } \frac{V}{A} < |x| \leq V \end{cases}$$

where A is the A-law parameter of the compander and the other elements are as in the μ -law case.

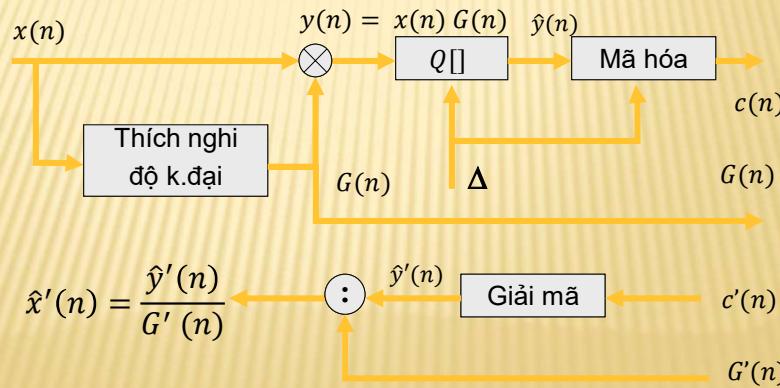
90

90

LƯỢNG TỬ THÍCH NGHI

- Bước lượng tử tuỳ thuộc vào biên độ tín hiệu

+ Thích nghi trước

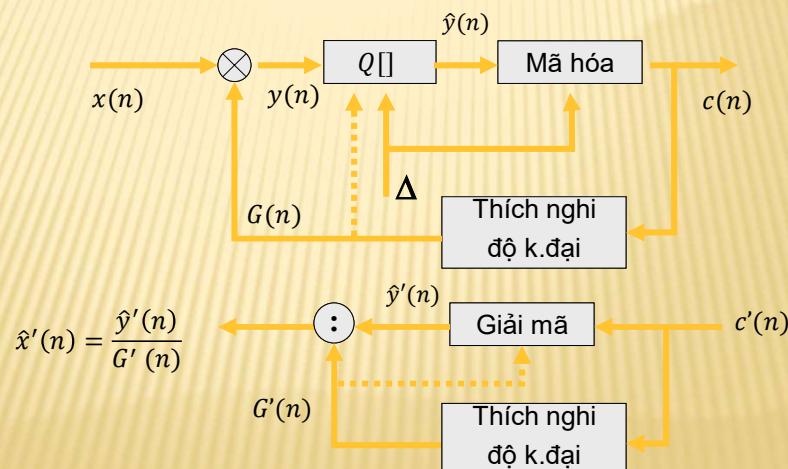


91

91

LƯỢNG TỬ THÍCH NGHI

+ Thích nghi sau



92

92

MỘT SỐ CHUẨN MÃ HOÁ ÂM THANH/TIẾNG NÓI

- ✖ G.721 : ADPCM, 32 kbps, 4bits, 8kHz
- ✖ G.722 : ~ADPCM, 48 đến 64 kbps,
- ✖ G.723 : ~ADPCM, 24 kbps, 3 bits, 8kHz
- ✖ G.728 : 16 Kbps
- ✖ GSM : điện thoại di động, 13 kbps
- ✖ Linear Predictive Encoding (Xerox), 5 kbps
- ✖ Code Excited Linear Prediction (CELP)
- ✖ Digital Video Interactive : ~ADPCM, 4 đến 8 bits
- ✖ VoIP: G723.1 (6.4kbits/s), G728, G729 (8kbits/s)

93

93

4. TỔNG HỢP TIẾNG NÓI

- ✖ Tạo tiếng nói xuất phát từ biểu diễn ngữ âm, ngữ nghĩa của lời nói
- ✖ Kỹ thuật tổng hợp tiếng nói:
 - + Tổng hợp trực tiếp
 - + Tổng hợp dựa trên mô hình
 - ✖ Bộ tổng hợp formant
 - ✖ Bộ tổng hợp dùng LPC
 - ✖ Bộ tổng hợp mô phỏng bộ máy phát âm
 - ✖ Bộ tổng hợp dùng HMM

94

94

PHÂN LOẠI

- ✖ Chất lượng bộ tổng hợp: Mức độ tự nhiên
 - + Mức độ rõ
 - + Thanh điệu
 - + Ngữ điệu
- ✖ Số lượng từ vựng:
 - + Hạn chế
 - + Không hạn chế
- ✖ Bộ tổng hợp tiếng nói từ văn bản
(Text-to-Speech)

95

95

ĐÁNH GIÁ CHẤT LƯỢNG TIẾNG NÓI TỔNG HỢP

- ✖ MOS: Mean Opinion Scores

96

96

TỔNG HỢP TRỰC TIẾP

✗ Ghi âm tiếng nói tự nhiên

- Đơn vị ghi âm
- Ghép các đơn vị ghi âm: từ, câu.

✗ Đơn vị ghi âm

+ âm vị : hiện tượng đồng cấu âm (coarticulation)

+ âm tiết (diphone - âm vị kép)

+ từ

+ tổ hợp từ

+ Câu

$$\text{nam} = n + a + m$$

$$= n + am$$

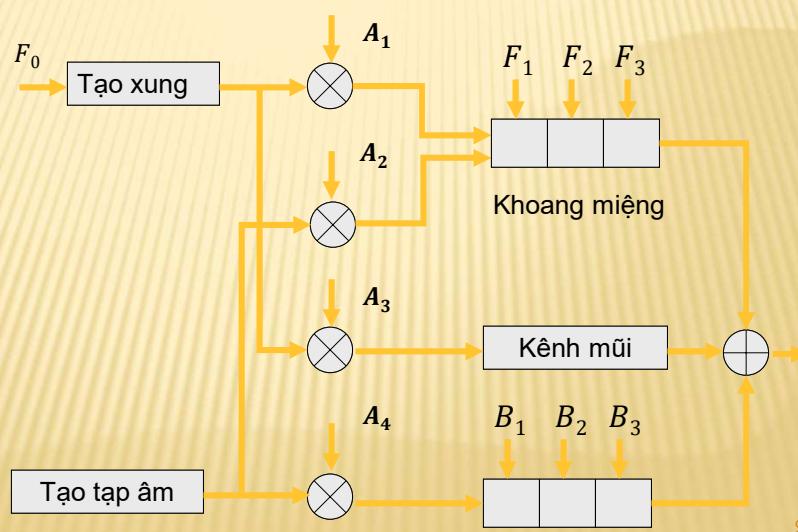
$$= na + m$$

$$= na + am$$

97

97

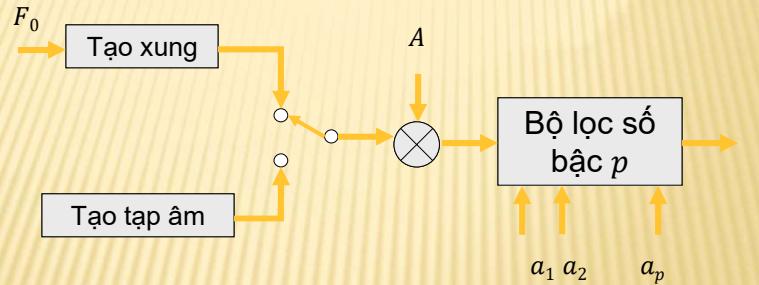
TỔNG HỢP FORMANT



98

98

TỔNG HỢP LPC

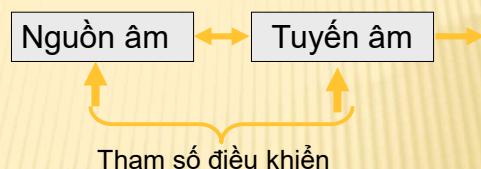


Synthesis-by-Analysis

99

99

MÔ PHỎNG BỘ MÁY PHÁT ÂM



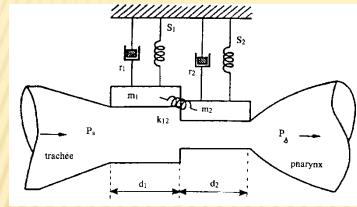
✖ Mô phỏng nguồn âm (nguồn tuần hoàn)

Mô phỏng dây thanh: Mô hình một khối, Mô hình hai khối, Mô hình nhiều khối, Mô hình hai dầm...

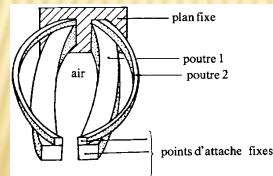
100

100

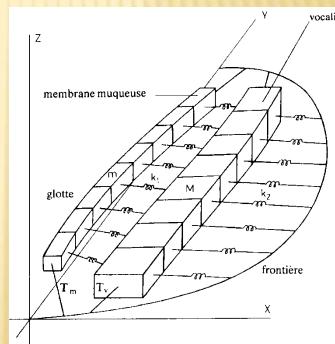
MÔ HÌNH NGUỒN ÂM



Mô hình 2 khối



Mô hình 2 đầm

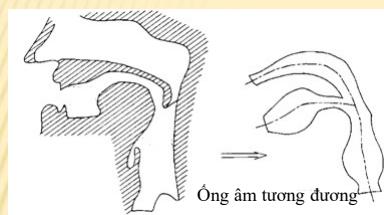


Mô hình nhiều khối

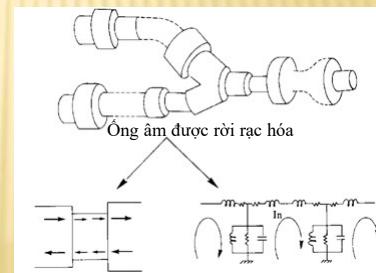
101

101

MÔ PHỎNG TUYẾN ÂM



Rời rạc hóa



102

102

MÔ HÌNH PHẢN XẠ

✖ Giả thiết

- ✚ Vách ngăn cứng
- ✚ Sóng truyền đơn hướng (dọc theo trục ống) chỉ xét các tần số < 5000 Hz, biến thiên diện tích không quá đột ngột
- ✚ Bỏ qua tổn hao: tính lỏng, truyền nhiệt

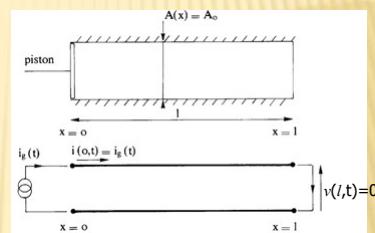
103

103

ỐNG TIẾT DIỆN ĐỀU, KHÔNG TỔN HAO

- ✖ Ống tiết điện đều và đường dây tương đương

- ✖ Hệ phương trình Webster



$$\begin{aligned} -\frac{\partial p}{\partial x} &= \frac{\rho_0}{A} \frac{\partial u}{\partial t} & u(x, t) &= u^+ \left(t - \frac{x}{c} \right) - u^- \left(t + \frac{x}{c} \right) \\ -\frac{\partial u}{\partial x} &= \frac{A}{\rho_0 c^2} \frac{\partial p}{\partial t} & p(x, t) &= \left[u^+ \left(t - \frac{x}{c} \right) + u^- \left(t + \frac{x}{c} \right) \right] \frac{\rho_0 c}{A} \end{aligned}$$

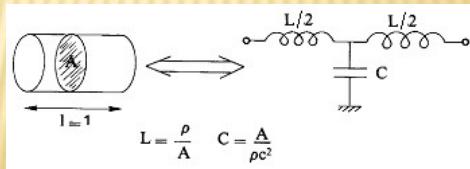
u : thông lượng, p : áp suất, ρ : mật độ không khí, c : vận tốc sóng âm

104

104

TƯƠNG TỰ ÂM HỌC – ĐIỆN HỌC

Âm học	Điện học
p : Áp suất	v : Điện áp
u : Thông lượng	i : Dòng điện
ρ_0/A : Điện cảm âm học	L : Điện cảm
$A/\rho_0 c^2$: Điện dung âm học	C : Điện dung



105

105

XÉT TRONG MIỀN TẦN SỐ

- Sóng tới và sóng phản xạ có dạng

$$u^+ \left(t - \frac{x}{c} \right) = K^+ e^{j\Omega(t-\frac{x}{c})}, \quad u^- \left(t + \frac{x}{c} \right) = K^- e^{j\Omega(t+\frac{x}{c})}$$

+ Điều kiện biên tại thanh mòn $u(0, t) = u_G(t) = U_G(\Omega) e^{j\Omega t}$

+ Điều kiện biên tại môi $p(\ell, t) = 0$

$$p(x, t) = jZ_0 \frac{\sin[\Omega(\ell-x)/c]}{\cos\Omega\ell/c} U_G(\Omega) e^{j\Omega t}, \quad u(x, t) = \frac{\cos[\Omega(\ell-x)/c]}{\cos\Omega\ell/c} U_G(\Omega) e^{j\Omega t}$$

$$Z_0 = \frac{\rho_0 c}{A}$$

106

106

ĐÁP ỨNG TẦN SỐ

❖ Tại môi $u(\ell, t) = U(\ell, \Omega)e^{j\Omega t}$

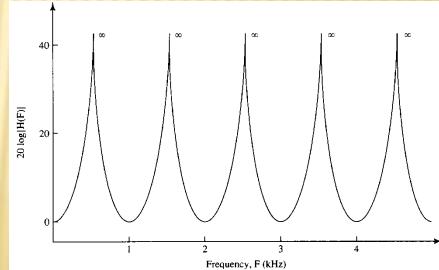
$$x = \ell \Rightarrow U(\ell, \Omega) = \frac{1}{\cos(\Omega\ell/c)} U_G(\Omega)$$

❖ Đáp ứng tần số

$$H(\Omega) = \frac{U(\ell, \Omega)}{U_G(\Omega)} = \frac{1}{\cos(\Omega\ell/c)}$$

$$H(\Omega) \rightarrow \infty \text{ với } f = \frac{(2n+1)c}{4\ell}$$

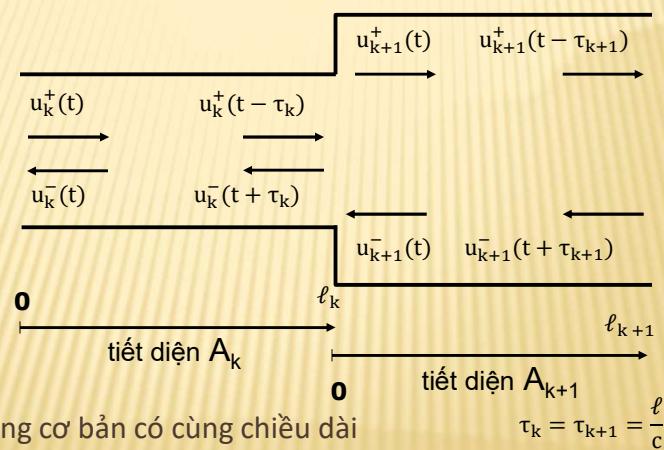
$$\begin{aligned} \ell &= 17,5 \text{ cm, } c=350 \text{ m/s} \\ f &= 500, 1500, 2500 \dots \text{ Hz} \end{aligned}$$



107

107

MÔ HÌNH PHẢN XẠ KHÔNG TỔN HAO (KELLY-LOCHBAUM)



❖ Các ống cơ bản có cùng chiều dài

$$\tau_k = \tau_{k+1} = \frac{\ell}{c}$$

108

108

MÔ HÌNH PHẢN XẠ KHÔNG TỔN HAO (KELLY-LOCHBAUM)

❖ Tính liên tục của áp suất và thông lượng

$$\begin{aligned} p_k(\ell, t) &= p_{k+1}(0, t) \\ u_k(\ell, t) &= u_{k+1}(0, t) \end{aligned}$$

$$\begin{aligned} u_{k+1}^+(t) &= \frac{2A_{k+1}}{A_{k+1} + A_k} u_k^+(t - \tau) + \frac{A_{k+1} - A_k}{A_{k+1} + A_k} u_{k+1}^-(t) \\ u_k^-(t + \tau) &= -\frac{A_{k+1} - A_k}{A_{k+1} + A_k} u_k^+(t - \tau) + \frac{2A_k}{A_{k+1} + A_k} u_{k+1}^-(t) \end{aligned}$$

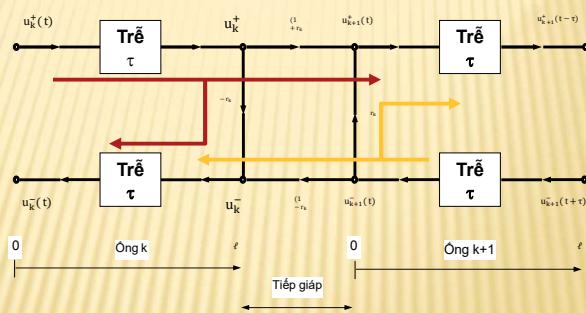
❖ Đặt hệ số phản xạ $r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k}$

$$\begin{aligned} u_{k+1}^+(t) &= (1 + r_k) u_k^+(t - \tau) + r_k u_{k+1}^-(t) \\ u_k^-(t + \tau) &= -r_k u_k^+(t - \tau) + (1 - r_k) u_{k+1}^-(t) \end{aligned}$$

109

109

PHÂN BỐ SÓNG

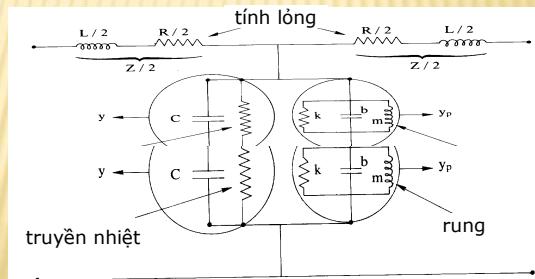


110

110

HIỆU ỨNG CỦA CÁC TỔN HAO

- ✖ Tổn hao do dịch chuyển không khí trong tuyến âm
 - + Do tính lỏng của không khí
 - + Do truyền nhiệt
 - + Do rung vách ngăn



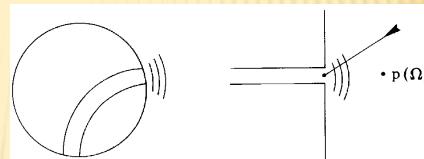
111

111

HIỆU ỨNG CỦA CÁC TỔN HAO

- ✖ Tổn hao do bức xạ tại môi
 - + Mô hình quả bóng vô hạn

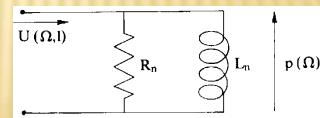
- + Trở kháng bức xạ



$$Z_r = \frac{p(\Omega)}{U(\Omega, \ell)} = \frac{j\Omega L_r R_r}{R_r + j\Omega L_r}$$

$$+ R_r = \frac{128}{9\pi^2}, L_r = \frac{8a}{3\pi}$$

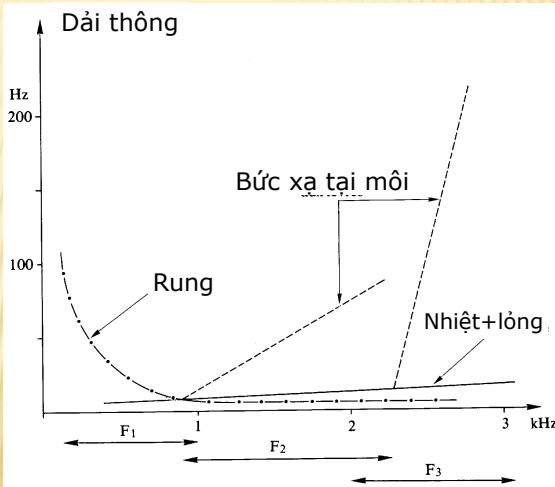
- + a : bán kính mở tại môi



112

112

HIỆU ỨNG CHUNG CỦA CÁC TỔN HAO



113

113

5. NHẬN DẠNG TIẾNG NÓI

- ✖ Hai giai đoạn: huấn luyện (học) – nhận dạng
- ✖ Phân loại theo
 - + Số lượng từ vựng
 - + Từ rời rạc – liên tục
 - + Một người nói – nhiều người nói
 - + Nhận dạng từ – câu

114

114

PHÂN LOẠI THEO ĐỘ PHỨC TẠP

- Nhận dạng từ riêng lẻ, từ vựng ít (<100), một người nói
- Từ vựng nhiều hơn (vài nghìn từ), một người nói
- Như trên nhưng cho hệ thống nhiều người nói
- Nhận dạng các từ đi với nhau, từ vựng ít (hang chục từ)
- Nhận dạng câu ngắn, từ vựng hạn chế, một người nói
- Như trên nhưng cho hệ thống nhiều người nói
- Nhận dạng lời nói liên tục, một hoặc nhiều người nói

115

115

NHẬN DẠNG NGƯỜI NÓI (SPEAKER RECOGNITION)

- Kiểm tra (xác thực) (verification) giọng nói
- Định danh (identification) giọng nói

116

116

MỘT SỐ VẤN ĐỀ ĐỐI VỚI HỆ THỐNG NHẬN DẠNG TIẾNG NÓI

- ✖ Phát hiện khoảng lặng, phát hiện tiếng nói
- ✖ Cải thiện chất lượng tín hiệu tiếng nói (giảm nhiễu)
- ✖ Tiếng nói được phát âm với thời hạn và nhịp điệu khác nhau
- ✖ Mô hình nhận dạng
 - + Mô hình Markov ẩn (Hidden Markov Model: HMM)
 - + Mạng nơ-ron

117