

XỬ LÝ TIẾNG NÓI

Trịnh Văn Loan
Bộ môn Kỹ thuật Máy tính
Khoa CNTT, ĐHBK Hà Nội

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
- 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. 1999
- 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
- <http://dce.hut.edu.vn>

2

Nội dung

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

3

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...

4

Mục đích

- **Mã hoá** 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

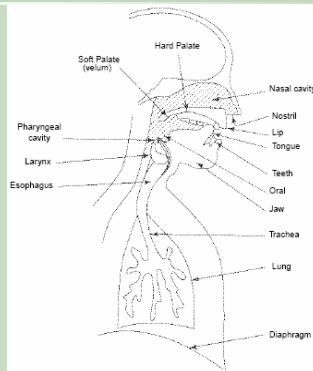
5

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 2 loại nguồn âm
 - tuần hoàn (dây thanh rung)
 - tạp âm (dây thanh không rung)

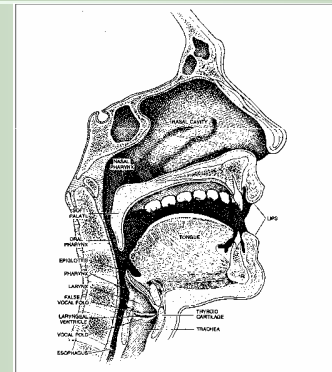
6

Bộ máy phát âm



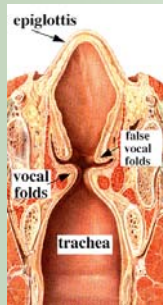
7

Bộ máy phát âm



8

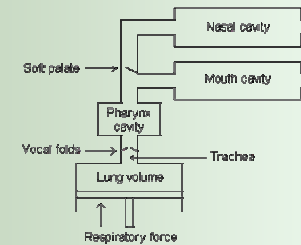
Bộ máy phát âm



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

9

Sơ đồ khối bộ máy phát âm



10

Thanh môn



11

Thanh môn

1. Một số khái niệm cơ bản

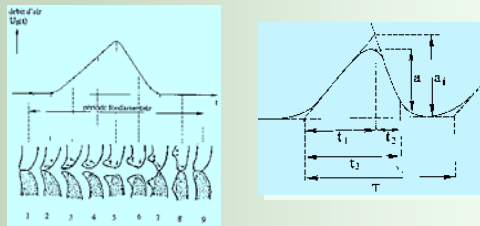
- Ở các vị trí hít, thở, phát âm, nói thì trào



A. Glotte pendant la respiration B. Glotte pour la phonation
 1. Glotte 2. Cordes vocales 3. Epiglottis 5. Cartilages aryénoïdes

12

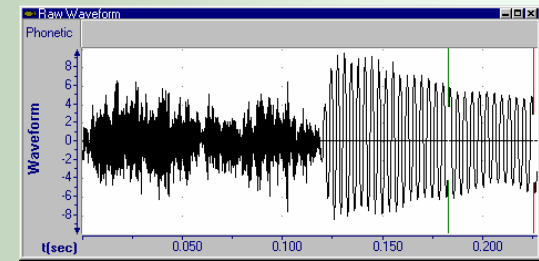
Dây thanh trong một chu kỳ dao động



13

Biểu diễn tín hiệu tiếng nói

■ Dạng sóng theo thời gian



14

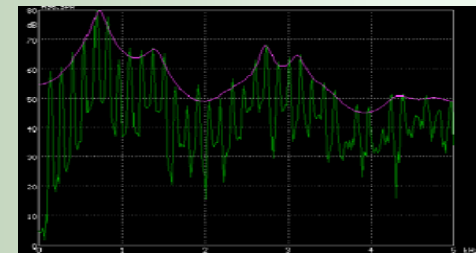
File WAV

- Tần số lấy mẫu: 8kHz, $F_1 = 11025$ Hz, 2F1, 4F1 (16kHz, 10kHz)
- Số bit/mẫu: 8, 16
- Mono, Stereo

15

Biểu diễn tín hiệu tiếng nói

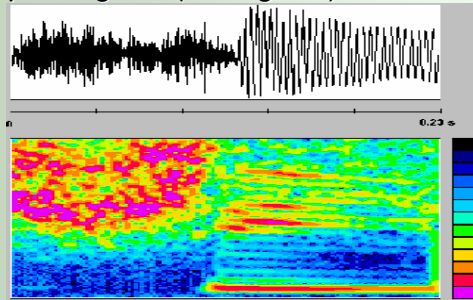
■ Phổ tín hiệu tiếng nói



16

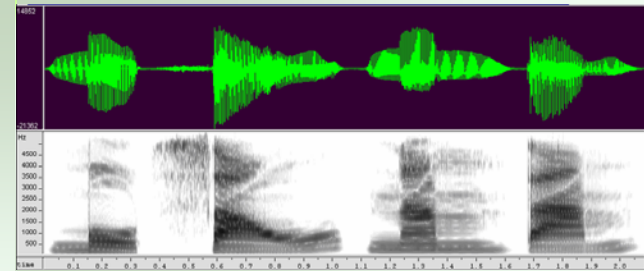
Biểu diễn tín hiệu tiếng nói

■ Spectrogram (Sonagram)



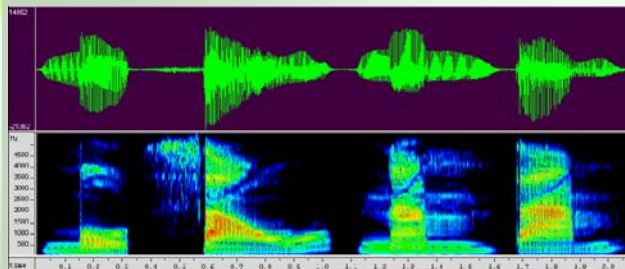
17

Biểu diễn tín hiệu tiếng nói



18

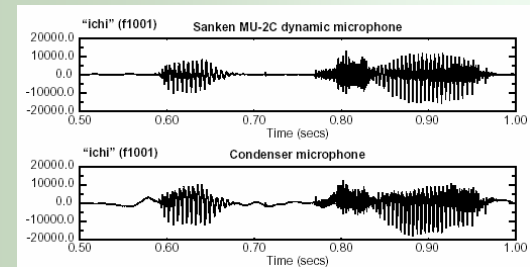
Biểu diễn tín hiệu tiếng nói



19

Biểu diễn tín hiệu tiếng nói

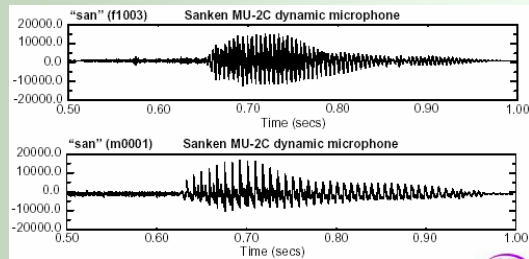
■ Thu bằng micro khác loại



20

Biểu diễn tín hiệu tiếng nói

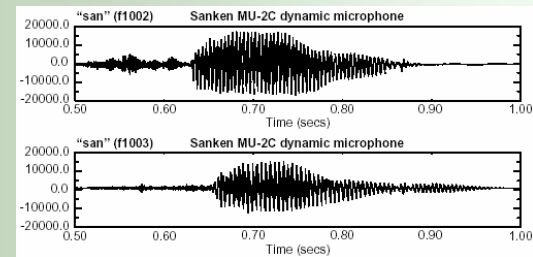
- Hai giọng khác nhau cho cùng một âm



21

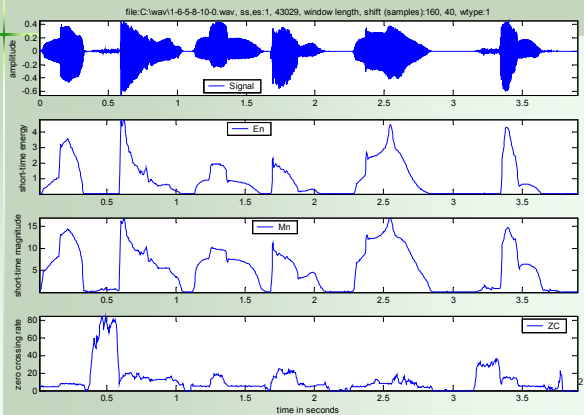
Biểu diễn tín hiệu tiếng nói

- Cùng người nói, cùng một âm



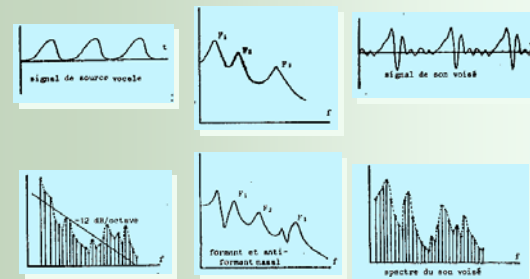
22

Năng lượng, tỷ lệ biến thiên qua giá trị không



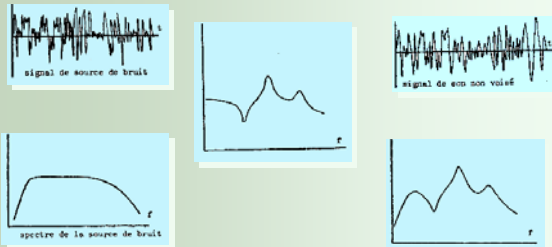
23

Tạo âm hữu thanh Formant và antiformant



24

Tạo âm vô thanh



25

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

26

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	ê chế
3	e	e dẻ
4	a	a ha
5	à	mất
6	ơ	bơ phở
7	â	ân cần
8	ư	từ từ
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 tiên
2	ua, uô (đọc ua)	tua rua, luôn
3	ưạ, ươ (đọc ưạ)	lừa thưa, lướt

27

Một số đặc điểm ngữ âm tiếng Việt

- Hệ thống âm vị: 22 phụ âm

1	b	bồng bênh
2	p	ốp ép
3	v	vấn vơ
4	ph	phôi pha
5	m	mơ màng
6	đ	đất đai
7	t	tin tưởng
8	th	thơ thần
9	d, gi	duyên, giữ
10	n	nóng
11	l	long lạnh

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

28

Một số đặc điểm ngữ âm tiếng Việt

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

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

29

Một số đặc điểm ngữ âm tiếng Việt

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

Hàng	hàng trước	hàng sau không tròn môi	hàng sau tròn môi
Độ mở			
hẹp	ĩ	ia, ye, ya, iê	ư ưa
hơi hẹp	ê	ơ â	ô
hơi rộng	e		o
rộng		a ă	

30

Một số đặc điểm ngữ âm tiếng Việt

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

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

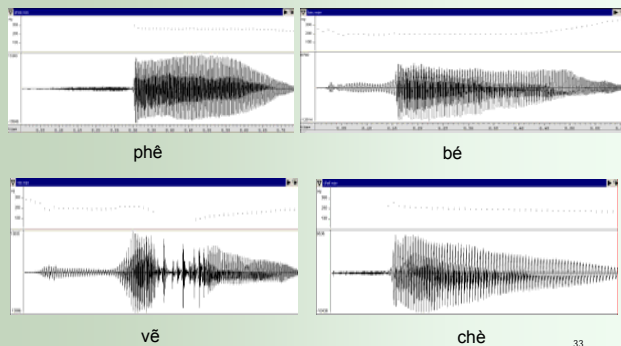
31

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.

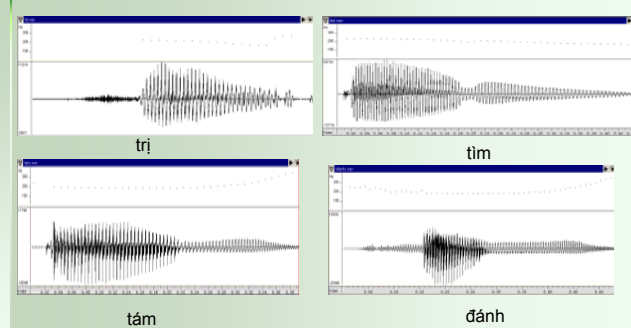
32

Dạng sóng một số từ tiếng Việt



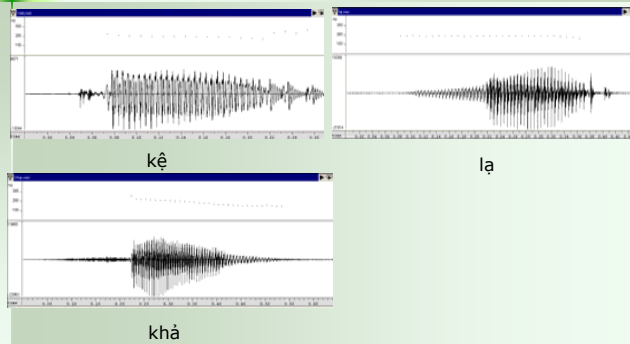
33

Dạng sóng một số từ tiếng Việt



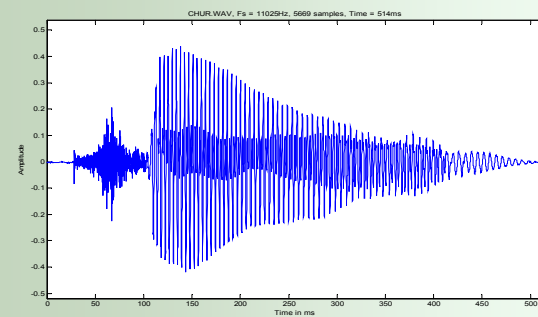
34

Dạng sóng một số từ tiếng Việt



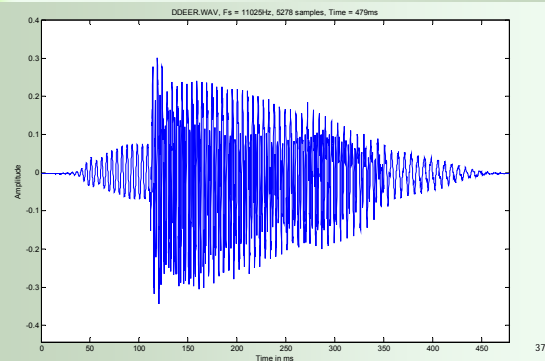
35

Dạng sóng một số từ tiếng Việt



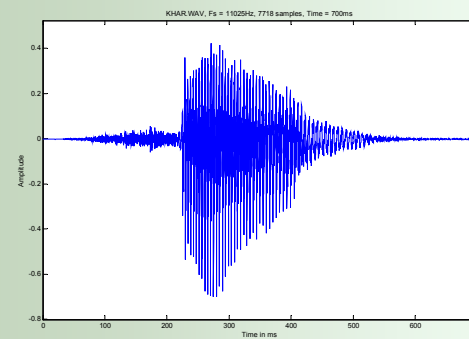
36

Dạng sóng một số từ tiếng Việt



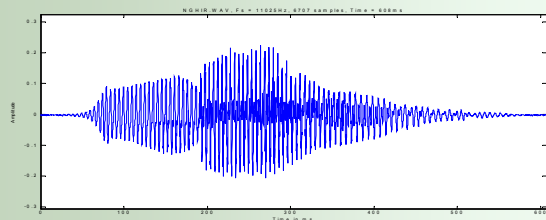
37

Dạng sóng một số từ tiếng Việt



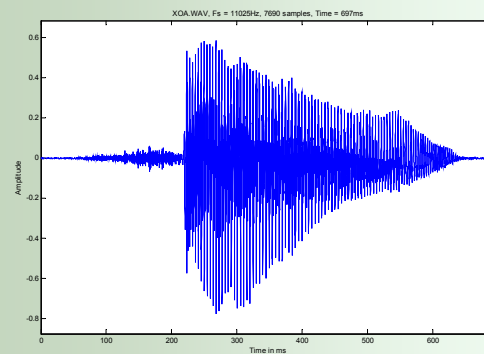
38

Dạng sóng một số từ tiếng Việt



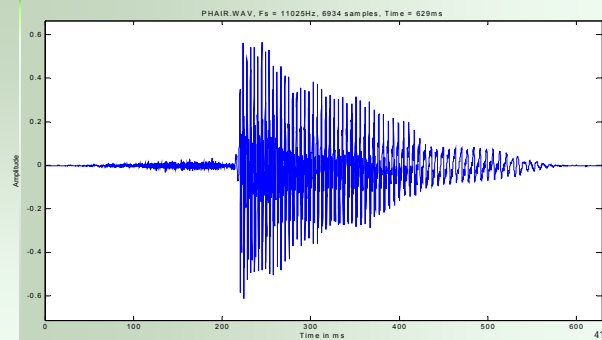
39

Dạng sóng một số từ tiếng Việt

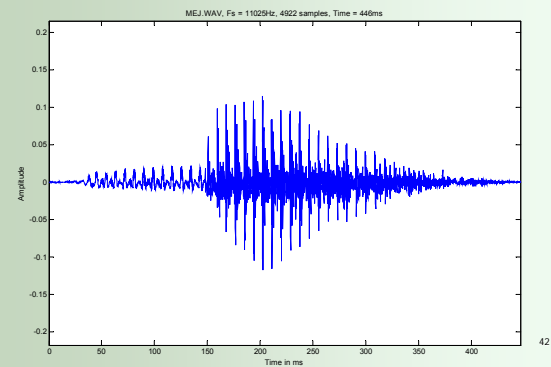


40

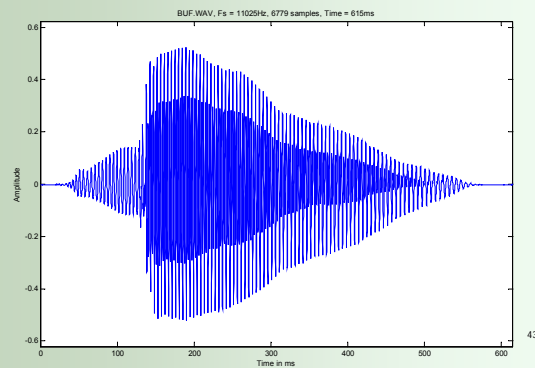
Dạng sóng một số từ tiếng Việt



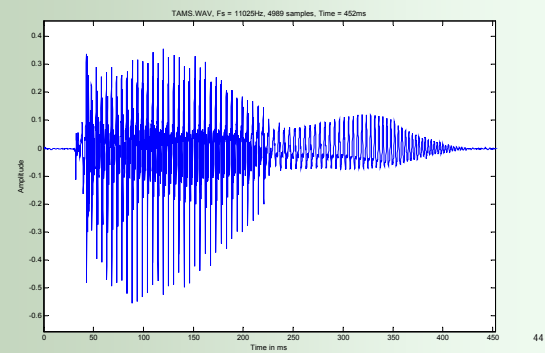
Dạng sóng một số từ tiếng Việt



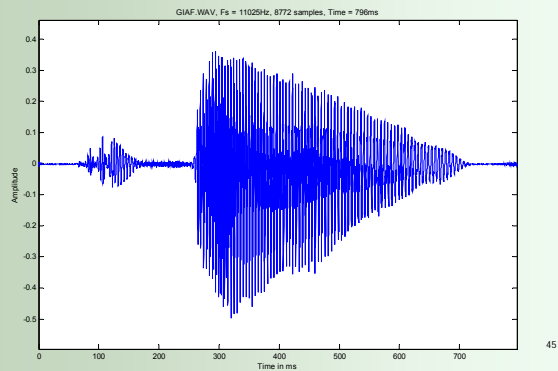
Dạng sóng một số từ tiếng Việt



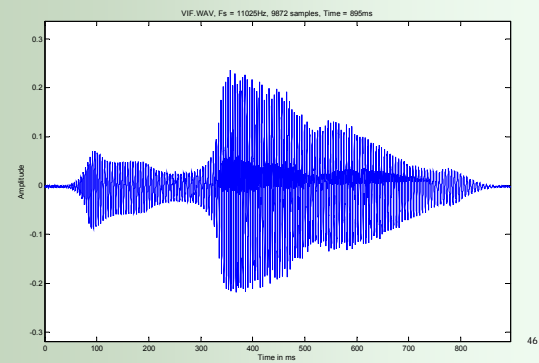
Dạng sóng một số từ tiếng Việt



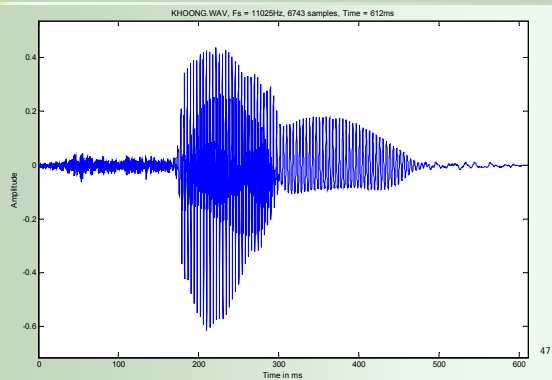
Dạng sóng một số từ tiếng Việt



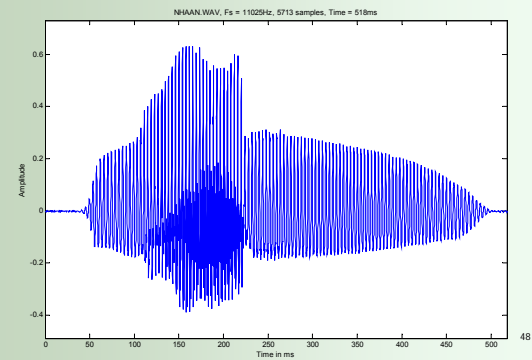
Dạng sóng một số từ tiếng Việt



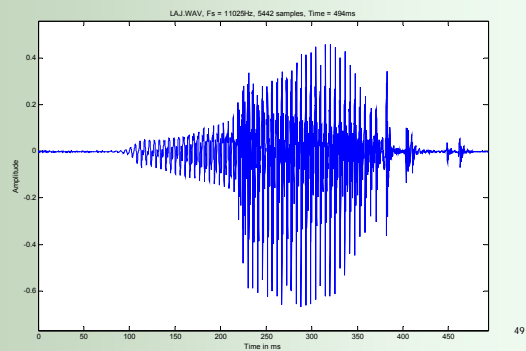
Dạng sóng một số từ tiếng Việt



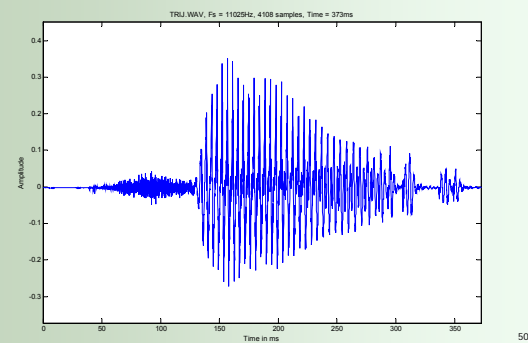
Dạng sóng một số từ tiếng Việt



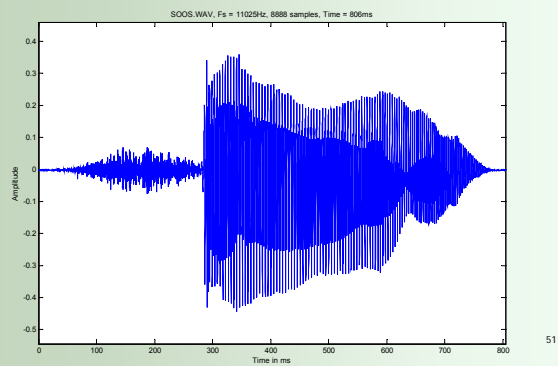
Dạng sóng một số từ tiếng Việt



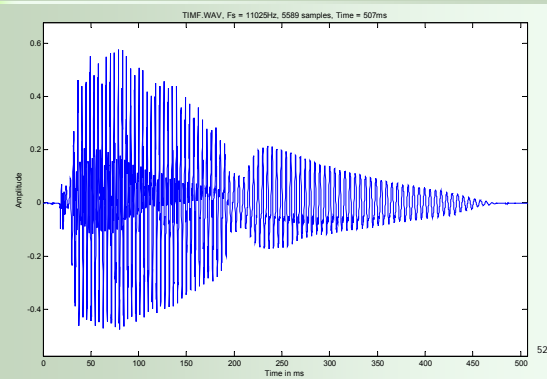
Dạng sóng một số từ tiếng Việt



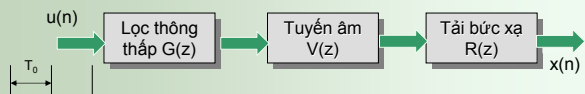
Dạng sóng một số từ tiếng Việt



Dạng sóng một số từ tiếng Việt



Mô hình tạo tiếng nói (Fant-1960)



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

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

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

53

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

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

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

$$x(n) + \sum_{i=1}^p a_i x(n-i) = \sigma u(n) \quad P = 2K+1$$

54

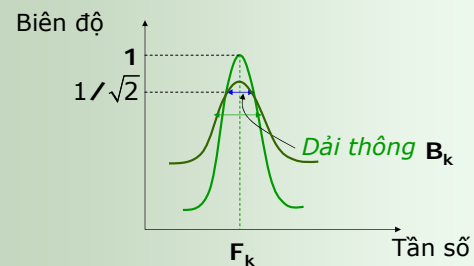
Mô hình ARMA

$$T(z) = \frac{\sigma_1}{A_1(z)} + \frac{\sigma_2}{A_2(z)} = \sigma \frac{C(z)}{A(z)} \quad C(z) = \sum_{i=0}^q c_i z^{-i} \quad c_0 = 1$$

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

55

Dải thông



56

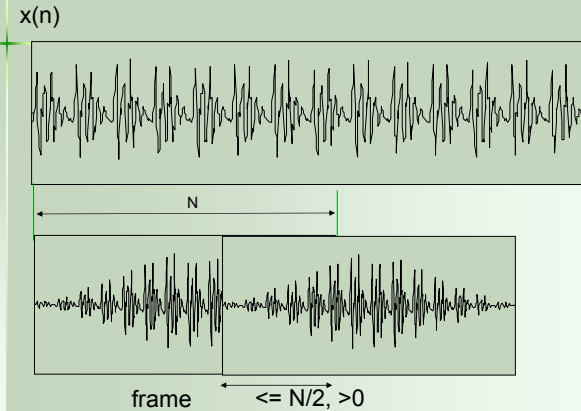
2. Xử lý tín hiệu tiếng nói

■ Phân tích phổ



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

57



58

Xử lý đồng hình (homomorphic)

■ $s(n) = h(n) * e(n)$ $S(\omega) = H(\omega) \cdot E(\omega)$

■ $\log[S(\omega)] = \log[H(\omega)] + \log[E(\omega)]$

$F^{-1}\{\log[S(\omega)]\} = F^{-1}\{\log[H(\omega)]\} + F^{-1}\{\log[E(\omega)]\}$

$F^{-1}\{\log[S(\omega)]\} = \hat{s}(n)$

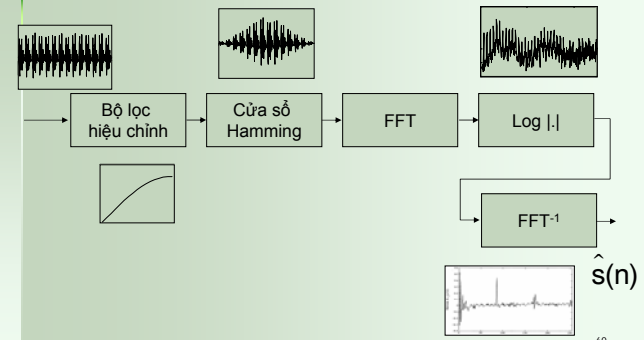
$F^{-1}\{\log[H(\omega)]\} = \hat{h}(n)$

$F^{-1}\{\log[E(\omega)]\} = \hat{e}(n)$

$\hat{s}(n) = \hat{h}(n) + \hat{e}(n)$

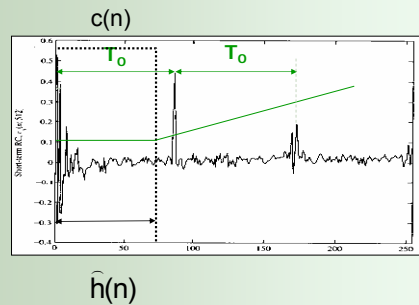
59

Sơ đồ khối xử lý đồng hình



60

Ví dụ



61

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

Sai số tiên đoán

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

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

Sai số bình phương toàn phần

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

Tối thiểu hóa sai số

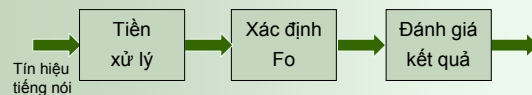
$$\frac{\partial E}{\partial \hat{a}_i} = 0, \quad i = 1, 2, \dots, p$$

62

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



63

Một số phương pháp xác định F_0

- 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

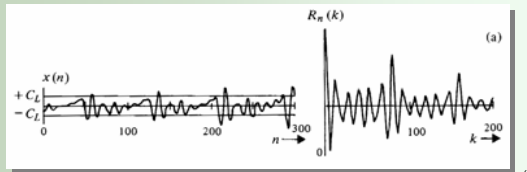
64

Dựa vào hàm tự tương quan

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

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

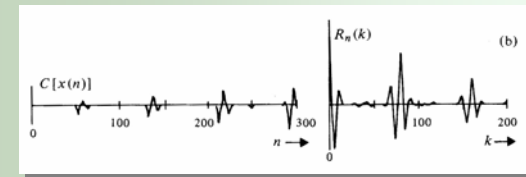
Fs = 10 kHz, N = 300, K = 150. Tìm cực đại trong khoảng (0, K)



65

Phương pháp tự tương quan có cải tiến

- Hạn chế, loại bỏ $|x| < C_L$



66

Dựa vào hàm vi sai biên độ trung bình (Average Magnitude Difference Function)

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

$$D(iP) = 0, \quad i = 0, 1, \dots, \quad \frac{1}{N} \sum_{n=0}^{N-1} |u(n)| \leq \left[\frac{1}{N} \sum_{n=0}^{N-1} u^2(n) \right]^{1/2}$$

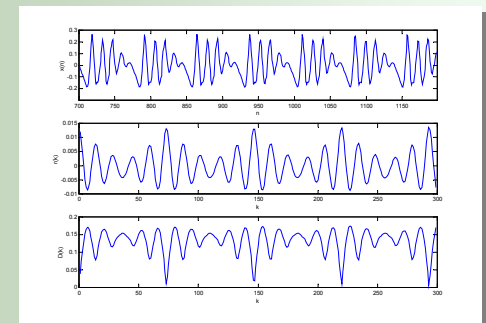
$$D(k) = \lambda \left\{ \frac{1}{N} \sum_{m=0}^{N-1} [x(n+m) - x(n+m-k)]^2 \right\}^{1/2}$$

$$= \lambda \left\{ \frac{1}{N} [2r(0) - 2r(k)] \right\}^{1/2} \quad k = 0, 1, \dots, K$$

với $\lambda < 1$

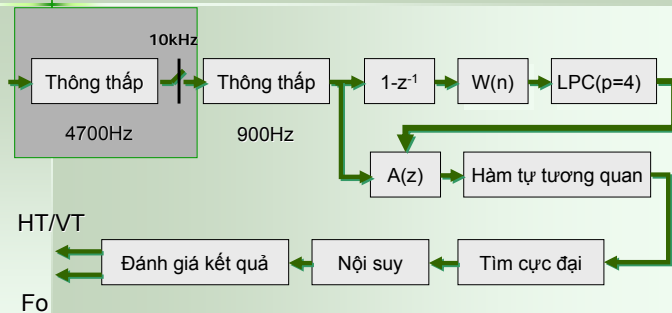
67

Ví dụ



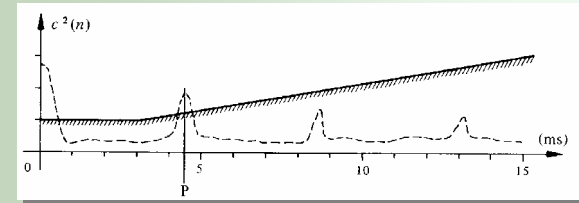
68

Dùng bộ lọc đảo (Simplified Inverse Filter Tracking)



69

Xử lý đồng hình



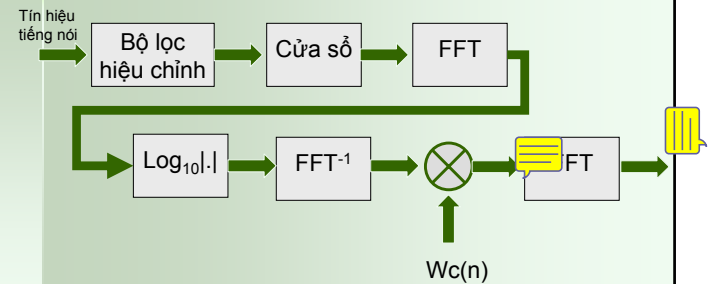
70

Xác định formant

- Tham số cần xác định
 - Formant F_k
 - Dải thông B_k
- Phương pháp
 - Xử lý đồng hình
 - LPC

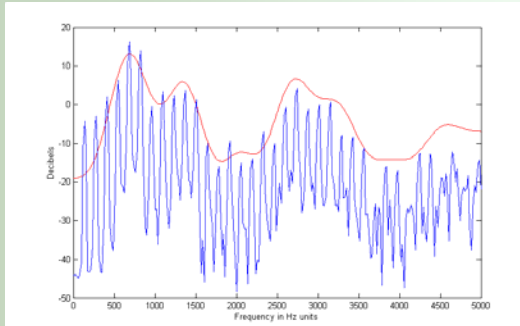
71

Xử lý đồng hình



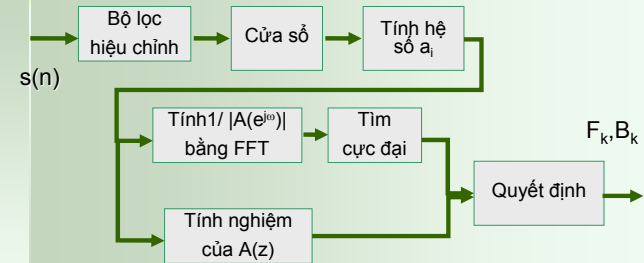
72

Xử lý đồng hình



73

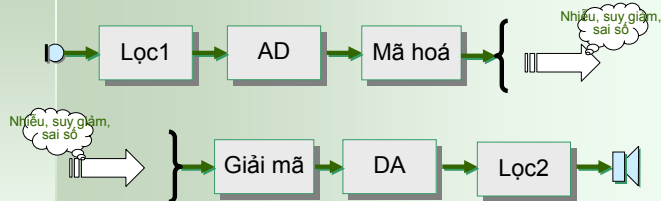
Phương pháp LPC



74

3. Mã hóa tiếng nói

■ Dãy thao tác mã hoá và giải mã

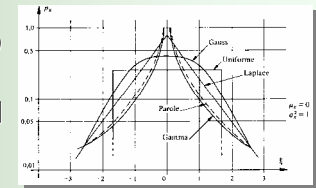


75

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



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

76

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

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

77

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 (tạp âm lượng tử) $e = Q(x) - x$
- Bước lượng tử : hiệu 2 mức tín hiệu kề nhau
 $\delta(i) = x(i) - x(i-1)$
- Thông lượng $I = bFs$ (bit/s). F_s : tần số lấy mẫu

78

Thông lượng

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

79

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

80

Lượng tử đều

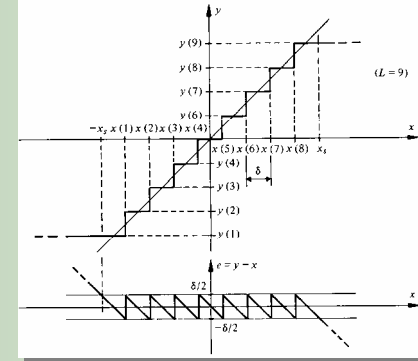
- Tổng quát, bước lượng tử là hàm của biên độ tín hiệu x (lượng tử không đều) \rightarrow đơn giản nhất là lượng tử đều.
- Mức lượng tử được chọn giữa 2 mức tín hiệu

$$y(i) = (1/2)[x(i-1) + x(i)]$$
- Luật lượng tử đều và đối xứng đặc trưng bởi:
 - các mức bão hoà $\pm x_s$
 - mức lượng tử L hoặc $(L+1) = 2^b$.
- Bước lượng tử $\delta = 2x_s/L$

81

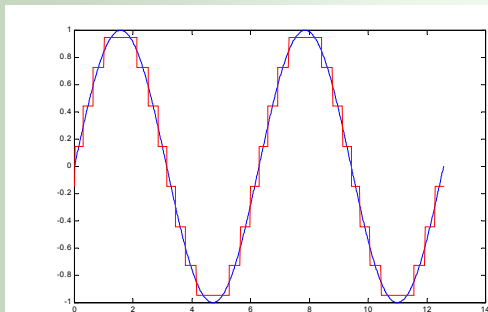
Lượng tử đều

- $L = 9$



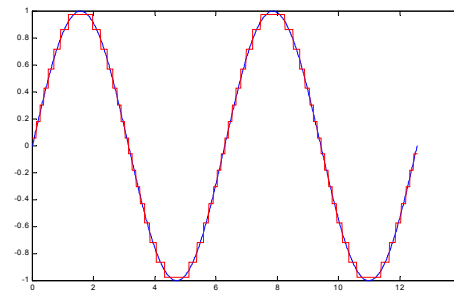
82

Lượng tử đều

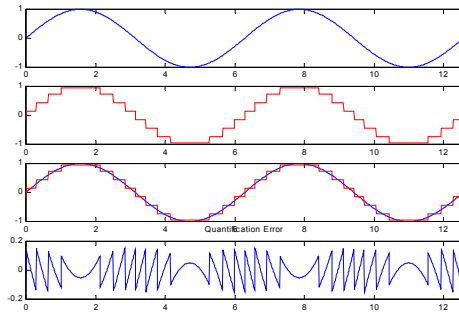


Lượng tử đều

- $L = 16$



Lượng tử đều



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 $+\delta/2$

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

- Trung bình tạp âm lượng tử = 0

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

86

Các tính chất lượng tử đều

- Tỷ số tín hiệu trên nhiễu

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

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

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

87

Tỷ số tín hiệu trên nhiễu

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

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

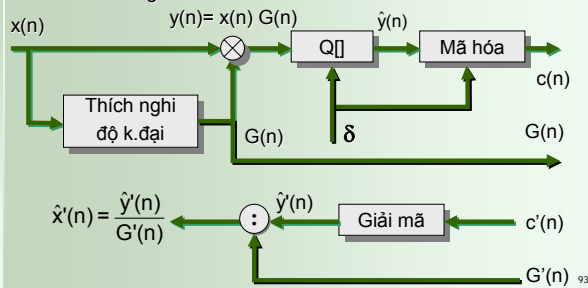
hoặc

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

88

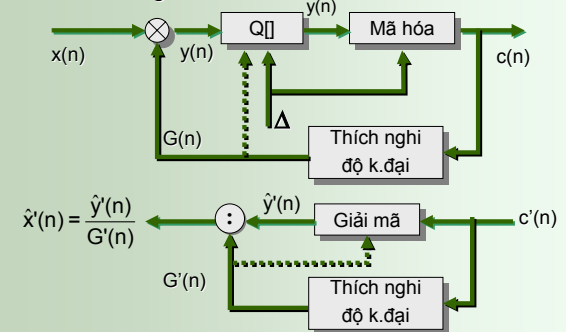
Lượng tử thích nghi

- Bước lượng tử tùy thuộc vào biên độ tín hiệu
 - Thích nghi trước



Lượng tử thích nghi

- Thích nghi sau



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 (6.4kbts/s), G728, G729 (8kbts/s)

4. Tổng hợp tiếng nói

- Tạo tiếng nói xuất phát từ biểu diễn ngữ âm 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

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)

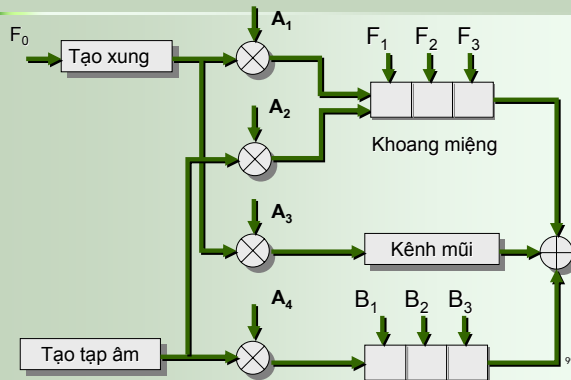
97

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
 - âm tiết (diphone)
 - từ
 - tổ hợp từ
 - câu

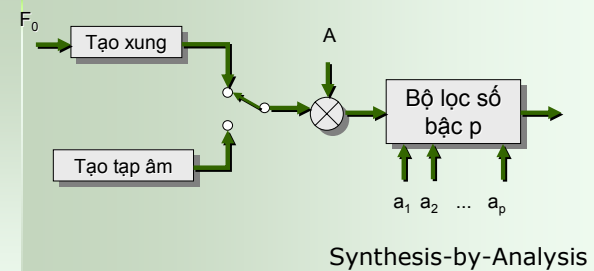
98

Tổng hợp formant



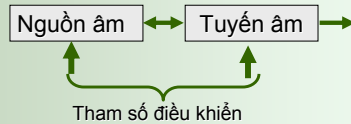
99

Tổng hợp LPC



100

Mô phỏng bộ máy phát âm

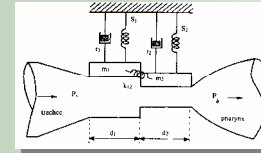


■ Mô phỏng nguồn âm (nguồn tuần hoàn)

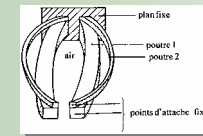
Mô phỏng dây thanh: Mô hình một khối, Mô hình hai khối, Mô hình nhiều khối, Mô hình hai dầm...

101

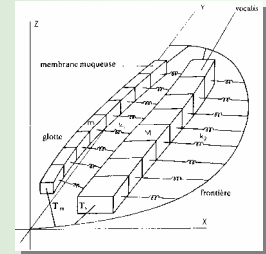
Mô hình nguồn âm



Mô hình 2 khối



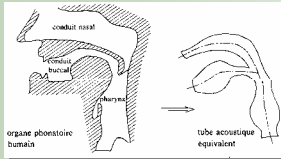
Mô hình 2 dầm



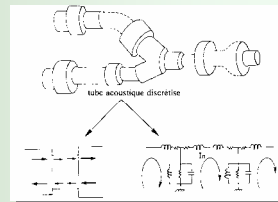
Mô hình nhiều khối

102

Mô phỏng tuyến âm



Rời rạc hóa



103

Mô hình phản xạ

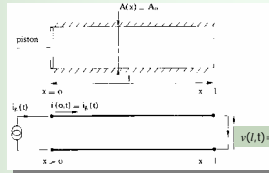
■ Giả thiết

- 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

104

Ổng tiết diện đều, không tổn hao

- Ổng tiết diện đều và đường dây tương đương



- Hệ phương trình Webster

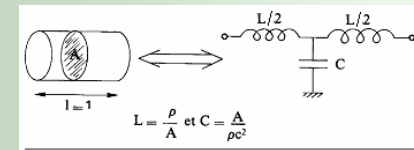
$$\begin{aligned} -\frac{\partial p}{\partial x} &= \frac{\rho_0}{A} \frac{\partial u}{\partial t} & u(x, t) &= u^+ \left(t - \frac{x}{c} \right) - u^- \left(t + \frac{x}{c} \right) \\ -\frac{\partial u}{\partial x} &= \frac{A}{\rho_0 c^2} \frac{\partial p}{\partial t} & p(x, t) &= \left[u^+ \left(t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

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

105

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



106

Xét trong miền tần số

- Sóng tới và sóng phản xạ có dạng

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

- Điều kiện biên tại thanh môn

$$u(0, t) = u_G(t) = U_G(\Omega) e^{j\Omega t}$$

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

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

$$Z_0(\Omega) = j\Omega \frac{\rho_0}{A}$$

107

Đáp ứng tần số

- Tại môi

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

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

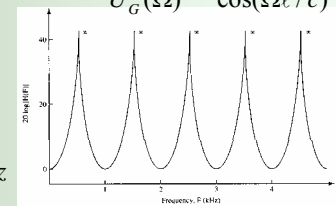
- Đáp ứng tần số

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

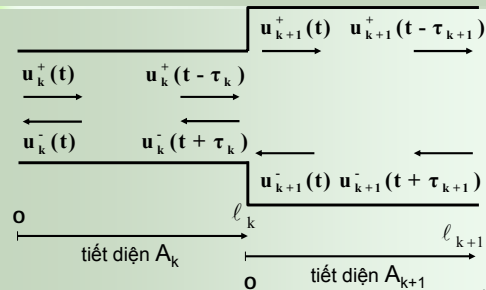
$$H(\Omega) \rightarrow \infty \text{ với } f = \frac{(2n+1)c}{4\ell}$$

$$\ell = 17,5 \text{ cm}, \quad c = 350 \text{ m/s}$$

$$f = 500, 1500, 2500 \dots \text{ Hz}$$



Mô hình phản xạ không tổn hao (Kelly-Lochbaum)



- Các ống cơ bản có cùng chiều dài $\tau_k = \tau_{k+1} = \frac{l}{c} = \tau$

109

Mô hình phản xạ không tổn hao (Kelly-Lochbaum)

- Tính liên tục của áp suất và thông lượng

$$p_k(\ell, t) = p_{k+1}(0, t)$$

$$u_k(\ell, t) = u_{k+1}(0, t)$$

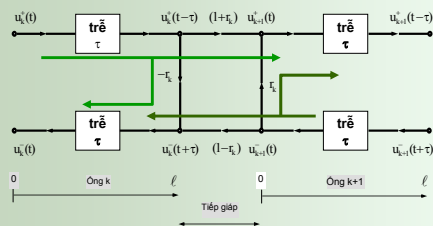
$$\begin{aligned} u_{k+1}^+(t) &= \frac{2A_{k+1}}{A_{k+1} + A_k} u_k^+(t - \tau) + \frac{A_{k+1} - A_k}{A_{k+1} + A_k} u_{k+1}^-(t) \\ u_k^-(t + \tau) &= -\frac{A_{k+1} - A_k}{A_{k+1} + A_k} u_k^+(t - \tau) + \frac{2A_k}{A_{k+1} + A_k} u_{k+1}^-(t) \end{aligned}$$

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

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

110

Phân bố sóng

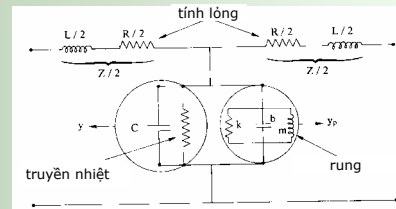


111

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

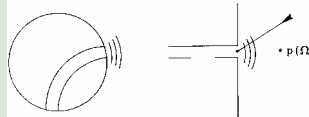


112

Hiệu ứng của các tổn hao

■ Tổn hao do bức xạ tại môi

- Mô hình quả bóng vô hạn

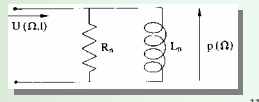


- Trở kháng bức xạ

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

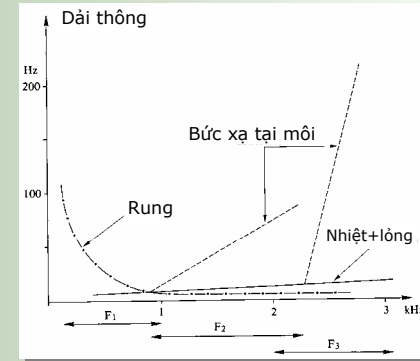
$$R_r = \frac{128}{9\pi^2}, L_r = \frac{8a}{3\pi c}$$

a : bán kính mở tại môi



113

Hiệu ứng chung của các tổn hao



114

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

115

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 (hàng 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

116

Nhận dạng người nói (Speaker Recognition)

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

117

Một số vấn đề đối với hệ thống nhận dạng tiếng nói

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

118