

**TẬP ĐOÀN BUÙ CHÍNH VIỄN THÔNG VIỆT NAM
HỌC VIỆN CÔNG NGHỆ BUÙ CHÍNH VIỄN THÔNG**
Đại học

**BÀI GIẢNG
XỬ LÝ ÂM THANH VÀ HÌNH ẢNH**

*Chuyên ngành Điện tử Viễn thông
(Lưu hành nội bộ)*

Biên soạn: TS. Lê Nhật Thăng

MỤC LỤC

MỤC LỤC	i
LỜI NÓI ĐẦU	iii
CÁC THUẬT NGỮ VIẾT TẮT	iv
Chương 1: Giới thiệu chung	1
1.1. Các khái niệm và lý thuyết cơ sở	1
1.2. Vai trò của xử lý âm thanh và hình ảnh trong truyền thông đa phương tiện	8
1.3 Kết luận chương 1	11
Hướng dẫn ôn tập chương 1	11
Chương 2: Kỹ thuật xử lý âm thanh	12
2.1. Các đặc trưng cơ bản của âm thanh	12
2.1.1. Khái niệm về âm thanh và các tham số đánh giá	12
2.1.2. Các đặc điểm của hệ thống thính giác con người	14
2.2. Mã hóa tín hiệu thoại	20
2.2.1. Quá trình tạo ra tiếng nói	20
2.2.2. Tổng quan về mã hóa tín hiệu thoại	25
2.2.3. Các phương pháp mã hóa	27
2.2.3.1. Mã hóa dạng sóng	29
2.2.3.2. Mã hóa tham số	32
2.2.3.3. Mã hóa lai	35
2.3. Mã hóa âm thanh	37
2.4. Kết luận chương 2	41
Hướng dẫn ôn tập chương 2	42
Chương 3: Kỹ thuật xử lý ảnh	43
3.1. Giới thiệu chung	43
3.1.1. Khái niệm cơ bản về ảnh và xử lý ảnh, video	43
3.1.2. Các ứng dụng phổ biến của xử lý ảnh	44
3.1.3. Các bước xử lý ảnh số	49
3.1.4. Các thành phần của hệ thống xử lý ảnh số	50
3.1.5. Đồ họa và các kiểu dữ liệu ảnh	51
3.1.6. Màu sắc trong ảnh và video	56

3.1.7. Cơ bản về video	60
3.2. Cơ sở kỹ thuật xử lý ảnh	65
3.2.1. Cơ sở của cảm nhận thị giác	65
3.2.2. Quá trình thu tín hiệu hình ảnh	68
3.2.3. Lấy mẫu và lượng tử hóa	71
3.2.4. Xử lý ảnh trong miền không gian	78
3.2.5. Xử lý ảnh trong miền tần số	80
3.3. Kỹ thuật nén ảnh	84
3.3.1. Tổng quan về nén ảnh	84
3.3.2. Hiệu quả của quá trình nén và chất lượng ảnh	88
3.3.3 Phân loại các phương pháp nén ảnh	89
3.3.4. Các phương pháp mã hóa dùng trong kỹ thuật nén không tổn thất	91
3.3.5. Các phương pháp mã hóa dùng trong kỹ thuật nén có tổn thất	100
3.4. Nén trong ảnh	105
3.5. Nén liên ảnh	114
3.6. Kết luận chương 3	118
Hướng dẫn ôn tập chương 3	119
 Chương 4: Các chuẩn mã hóa âm thanh và nén ảnh trong truyền thông đa phương tiện	121
4.1. Các chuẩn mã hóa tín hiệu thoại	121
4.2. Các chuẩn mã hóa âm thanh	123
4.3. Các chuẩn nén ảnh JPEG	125
4.4. Các chuẩn nén Video MPEG-1, 2, 4, 7 và MPEG-21	130
4.5. Các chuẩn nén Video H26x của ITU	143
4.6. Kết luận chương 4	143
Hướng dẫn ôn tập chương 4	144
 TÀI LIỆU THAM KHẢO	145

LỜI NÓI ĐẦU

Trong những năm gần đây, cùng với sự tiến bộ của khoa học kỹ thuật, sự phát triển mạnh mẽ của máy tính, thông tin di động và của Internet thì nhu cầu trao đổi các dịch vụ truyền thông đa phương tiện trên mạng thông tin là rất lớn đặc biệt là các ứng dụng truyền âm thanh và video thời gian thực qua mạng IP như âm nhạc theo yêu cầu, video phone, video-conferencing, tele-medical hay video theo yêu cầu... Cho nên, vấn đề xử lý âm thanh và hình ảnh sao cho có hiệu quả cao, đảm bảo tiết kiệm băng thông truyền dẫn, giảm bớt không gian lưu trữ để truyền thông tin trên mạng một cách dễ dàng và nhanh chóng trở nên cấp thiết hơn bao giờ hết.

Bài giảng “Xử lý âm thanh và hình ảnh” sẽ giới thiệu những kiến thức cơ bản về các kỹ thuật xử lý âm thanh, hình ảnh, video và đặc biệt chú trọng đến các phương pháp nén, lưu trữ, các tiêu chuẩn nén âm thanh, hình ảnh, video đã và đang được ứng dụng trong truyền thông đa phương tiện nhằm đảm bảo chất lượng âm thanh, hình ảnh và tăng hiệu suất truyền dẫn thông tin.

Nội dung của bài giảng bao gồm:

- Chương 1: Giới thiệu chung: Giới thiệu tổng quan các khái niệm và lý thuyết cơ sở phục vụ cho môn học và vai trò của xử lý âm thanh và hình ảnh ứng dụng trong truyền thông đa phương tiện.
- Chương 2: Kỹ thuật xử lý âm thanh: Giới thiệu các đặc trưng cơ bản của âm thanh, phân tích các đặc điểm của cơ quan phát âm và tạo ra tiếng nói của con người, các phương pháp mã hóa thoại, âm thanh.
- Chương 3: Kỹ thuật xử lý ảnh: Tập trung trình bày các khái niệm cơ bản về ảnh và video; giới thiệu về kỹ thuật xử lý ảnh, nén ảnh tĩnh, nén video.
- Chương 4: Các chuẩn nén âm thanh và hình ảnh: Giới thiệu các chuẩn nén thoại, âm thanh, các chuẩn nén ảnh JPEG và nén video MPEG, H26x.

Mặc dù đã có nhiều cố gắng, song do đây là lần biên soạn đầu tiên và còn có nhiều hạn chế về thời gian nên bài giảng này không tránh khỏi thiếu sót. Rất mong nhận được sự đóng góp của các đồng nghiệp, các học viên, sinh viên và bạn đọc để bài giảng này được hoàn thiện hơn.

Những ý kiến đóng góp xin gửi về:

Bộ môn Kỹ thuật Chuyển mạch- Khoa Viễn thông 1

Học Viện Công nghệ Bưu chính Viễn thông

Địa chỉ: Km 10, đường Hà Nội – Hà Đông, Hà Đông, Hà Nội

Tel: 0433820860; 0438549352; 0904342557

Fax: 0433511405

E-mail: thangln@ptit.edu.vn

Xin trân trọng cảm ơn!

Hà Nội, tháng 7 năm 2010

TS. Lê Nhật Thăng

CÁC THUẬT NGỮ VIỆT TẮT

ADSL	Asymmetric Digital Subscriber Line	Đường dây thuê bao số bất đối xứng
ADPCM	Adaptive Differential Pulse Code Modulation	Điều chế xung mã vi sai thích ứng
BRR	Bit Rate Reduction	Sự giảm tốc độ bit
BSC	Binary Symmetric Channel	Kênh nhị phân đối xứng
CABAC	Context Adaptive Binary Arithmetic Coding	Mã hoá thuật toán nhị phân theo tình huống
CCD	Charge Coupled Device	Thiết bị cảm biến quang điện bán dẫn
CIF	Common Intermediate Format	Khuôn dạng trung gian chung
DCT	Discrete Cosine Transform	Biến đổi Cosin rời rạc
DDC	Double Delta Coding	Mã hoá delta kép
DFT	Discrete Fourier Transform	Biến đổi rời rạc Fourier
DPCM	Differential Pulse Code Modulation	Điều chế xung mã vi sai
DSL	Digital Subscriber Line	Đường dây thuê bao số
DSLAM	Digital Subscriber Line Access Multiplexer	Bộ tập trung đường dây thuê bao số
DSM	Digital Storage Media	Phương tiện lưu trữ số
DVB	Digital Video Broadcasting	Quảng bá truyền hình số
DVD	Digital Video Disc	Đĩa ảnh số (quang)
DWT	Descrete Wavelet Transform	Biến đổi Wavelet rời rạc
EBCOT	Embedded Block Coding with Optimal Truncation	Mã hóa khối nhúng với cắt giảm tối ưu
EDTV	Extended Definition TeleVision	Truyền hình mở rộng
EOB	End of Block	Kết thúc khối
FIR	Finite Impulse Response	Đáp ứng xung hữu hạn
GIF	Graphics Interchange Format	Định dạng trao đổi ảnh
GOP	Group of Picture	Nhóm các khung ảnh
GOV	Group of VOPs	Nhóm các GOV
HDTV	High-Definition TeleVision	Truyền hình độ phân giải cao
HVS	Human Vision System	Hệ thống thị giác của người
ICT	Irreversible color transform	Chuyển đổi màu không thuận nghịch
IGMP	Internet Group Management Protocol	Giao thức quản lý nhóm Internet
IIR	Infinte impulse response	Đáp ứng xung vô hạn
IPTV	Internet Protocol Television	Truyền hình dựa trên Internet
ISO	International Organization for Standardization	Tổ chức chuẩn quốc tế
JIF	JPEG Interchange Format	Định dạng trao đổi JPEG
JPEG/ JVT	Joint Photographic Experts Group/ Joint Video Team	Nhóm chuyên gia ghép nối đồ họa

KLT	Karhunen – Loeve Transform	Chuyển đổi Karhunen – Loeve
LZW	Lempel Ziv-Wench Transform	Chuyển đổi Lempel Ziv-Wench
MoD	Music on Demand	Âm nhạc theo yêu cầu
MPEG	Moving Picture Experts Group	Nhóm chuyên gia về ảnh động
NTSC	National Television System Committee	Ủy ban hệ thống truyền hình quốc gia
PAL	Phase Alternating Line	PAL
PCM	Pulse Code Modulation	Điều chế xung mã
PDF	Portable Document Format	Định dạng tài liệu linh động
PON	Passive Optical Networks	Mạng quang thụ động
QCIF	Quarter Common Intermediate Format	Định dạng có độ phân giải $\frac{1}{4}$ CIF
RAC	Relative Address Coding	Mã hóa địa chỉ tương đối
RCT	Reversible Color Transform	Chuyển đổi thuận nghịch
RLC/ RLE	Run Length Coding/ Encoding	Mã hóa độ dài chạy
RMS	Root Mean Square	Độ lệch trung bình bình phương
SECAM	Sequential Color with Memory	SECAM
SNR	Signal to Noise Ratio	Tỷ số tín hiệu trên nhiễu
STB	Set Top Box	Set Top Box
TIFF	Tagged Image File Format	Khuôn dạng file tiêu chuẩn
TVoD	TeleVision on Demand	Tivi theo yêu cầu
VDSL	Very High Speed Digital Subscriber Line	Đường thuê bao số tốc độ rất cao
VLC	Variable Length Code	Mã hóa độ dài thay đổi
VO	Video Object	Đối tượng hình ảnh
VoD	Video on Demand	Video theo yêu cầu
VOL	Video Object Layer	Lớp đối tượng hình ảnh chuyển động
VOP	Video Object Plane	Mặt phẳng đối tượng hình ảnh chuyển động

Chương 1: Giới thiệu chung

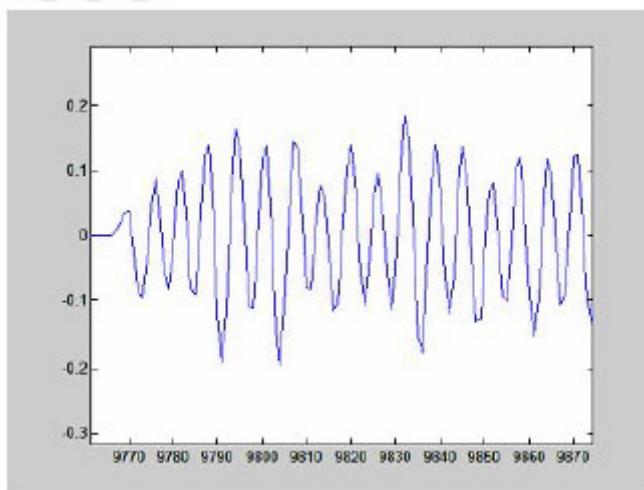
1.1. Các khái niệm và lý thuyết cơ sở

1.1.1. Tín hiệu

1.1.1.1. Định nghĩa

Tín hiệu là biểu diễn vật lý của thông tin. Trong thực tế, các tín hiệu nhìn thấy là các sóng ánh sáng mang thông tin tới mắt của con người và các tín hiệu nghe thấy là các sự biến đổi của áp suất không khí truyền thông tin tới tai chúng ta.

Về mặt toán học, tín hiệu được biểu diễn bởi hàm của một hoặc nhiều biến số độc lập. Ví dụ, tín hiệu âm thanh có biên độ âm biến đổi theo thời gian như ở hình vẽ dưới đây.



Hình 1.1: Tín hiệu âm thanh

Tổng quát hơn, tín hiệu có thể biến đổi theo hai chiều: không gian/thời gian. Ví dụ với ảnh, có màu biến đổi theo không gian hai chiều; với video, màu biến đổi theo cả không gian và thời gian.

1.1.1.2. Tín hiệu liên tục

Nếu biến độc lập của sự biểu diễn toán học của một tín hiệu là liên tục, thì tín hiệu đó được gọi là liên tục. Dựa theo biên độ, người ta có thể phân loại tín hiệu liên tục thành: tín hiệu tương tự và tín hiệu lượng tử hóa.

Nếu biên độ của tín hiệu liên tục là liên tục thì tín hiệu đó được gọi là tín hiệu tương tự. Còn nếu biên độ của tín hiệu liên tục là rời rạc thì tín hiệu đó được gọi là tín hiệu lượng tử hóa.

1.1.1.3. Tín hiệu rời rạc

Nếu tín hiệu được biểu diễn bởi hàm của các biến rời rạc thì tín hiệu đó được gọi là tín hiệu rời rạc. Dựa theo biên độ, người ta có thể phân loại tín hiệu rời rạc thành: tín hiệu lấy mẫu và tín hiệu số.

Nếu biên độ của tín hiệu rời rạc là liên tục (không được lượng tử hóa) thì tín hiệu đó được gọi là tín hiệu lấy mẫu. Còn nếu biên độ của tín hiệu rời rạc là rời rạc thì tín hiệu đó được gọi là tín hiệu số.

1.1.2. Số hóa tín hiệu tương tự

Nói chung tín hiệu tương tự thì liên tục theo thời gian và giá trị. Theo quan điểm lý thuyết thông tin, lượng thông tin chứa trong tín hiệu tương tự là vô hạn. Rõ ràng, điều này tạo ra quan hệ với các tín hiệu này một nhiệm vụ khó khăn trong điều kiện dung lượng bộ nhớ và năng lực xử lý của máy tính bị hạn chế. Mặt khác, các tín hiệu số chỉ xuất hiện trong những khoảng thời gian nhất định và chỉ được biểu diễn bằng các giá trị biên độ rời rạc. Sự suy giảm thông tin này là mục tiêu làm cho quá trình xử lý thêm hữu ích và trên thực tế là những bước nén đầu tiên.

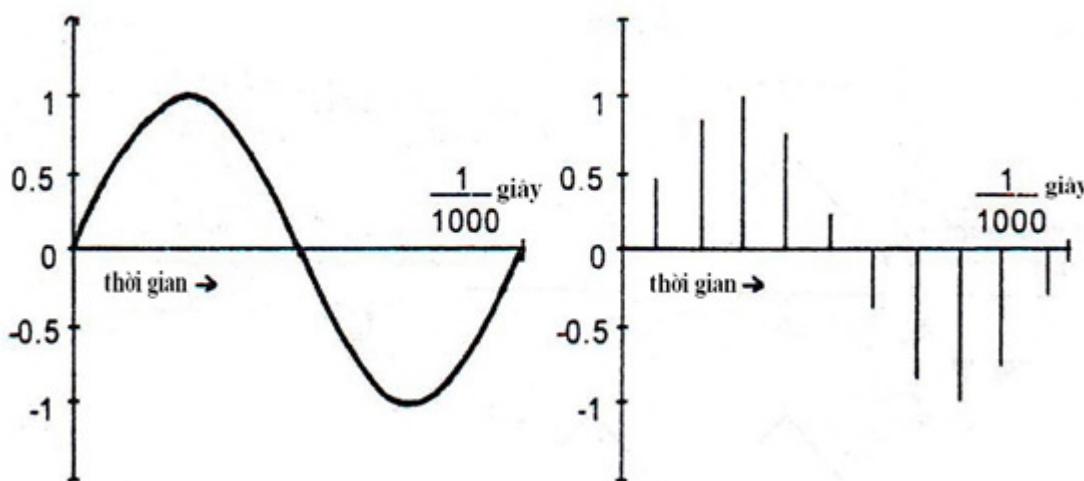
Số hóa là phương pháp giảm lượng thông tin đến mức hợp lý bằng cách lấy những giá trị đại diện có tính toán cân nhắc kỹ. Việc này làm thành hai phần. Phần lấy mẫu theo thời gian và lấy mẫu biên độ. Theo lý thuyết cả hai bước độc lập nhau, trong thực tế, chúng thường được thực hiện bởi cùng phần tử xử lý là *bộ chuyển đổi tương tự thành số (ADC)*. Đó là sự số hóa trong giới hạn để thu được thông tin mong muốn có ích chứa trong tín hiệu tương tự và loại bỏ thông tin dư thừa không cần thiết. Cho nên chúng ta phải biết các thuộc tính của các tín hiệu cần thiết được số hóa để thực hiện biến đổi tín hiệu tương tự sang tín hiệu số một cách thích hợp.

Thuộc tính chung của tín hiệu video và âm thanh bao gồm băng tần, tỉ số tín hiệu trên nhiễu, tỉ số tín hiệu trên méo, và dài động. Độ rộng băng tần miêu tả sự thay đổi tín hiệu tương tự khả dụng trong quãng thời gian cho trước, nó lần lượt xác định số lượng mẫu được lấy trong một đơn vị thời gian để bảo toàn được thông tin chứa trong tín hiệu. Thông tin về dài động và các nhân tố khác (ví dụ như nhiễu chồng lấn tín hiệu) xác định độ chính xác biên độ của tín hiệu phải được giữ để chống lại bất kỳ tạp âm chú ý hay không mong muốn.

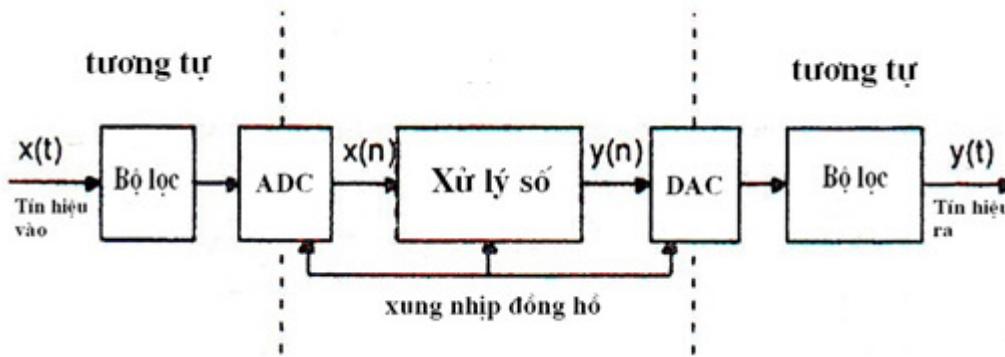
Để chuyển đổi tín hiệu tương tự sang tín hiệu số, tín hiệu tương tự thường được lấy mẫu tại những khoảng thời gian bằng nhau. Và biên độ của mỗi mẫu được lượng tử hóa

rồi được gán với một từ mã số. Vì thế tín hiệu số là một chuỗi với tốc độ bit không đổi hình thành từ quá trình xử lý lấy mẫu với mã số nhị phân độ dài bằng nhau.

Hình 1.2 mô tả việc lấy mẫu tín hiệu. Tín hiệu vào tương tự liên tục theo thời gian $x(t)$ được lọc thông qua bộ lọc ngoài. Sau đó đi qua bộ lấy mẫu, bộ này là một mạch điện lấy mẫu với tần số f_s lớn hơn hai lần tần số lớn nhất của tín hiệu. Bộ lấy mẫu biến đổi tín hiệu tương tự thành tín hiệu rời rạc theo thời gian, tín hiệu này sau đó, trong đoạn sau của bộ chuyển đổi ADC, được lượng tử hóa và gán bởi một từ mã nhị phân. Toàn bộ quá trình trên được minh họa trong hình 1.3.



Hình 1.2: Lấy mẫu tín hiệu tương tự



Hình 1.3: Nguyên lý cơ bản của xử lý số tín hiệu

Tín hiệu được lấy mẫu và được lượng tử hóa như trên được gọi là điều chế xung mã PCM (Pulse Code Modulation) vì mỗi một mẫu được mã hóa độc lập với các mẫu khác và các từ mã có chiều dài không đổi. Mỗi từ mã bao gồm nhiều bit: 8 đến 10 bit được sử dụng cho tín hiệu video; 8 bit cho tín hiệu âm thanh ở dải tần thấp và 16 đến 20 bit dùng cho tín hiệu âm thanh yêu cầu chất lượng cao.

1.1.3. Biến đổi Fourier

Biến đổi là công cụ khá mạnh cho việc mô hình hóa nội dung thông tin và áp dụng cho các nguyên lý nén. Trong lĩnh vực âm thanh, một biến đổi cho phép ta thấy nội dung thông qua phổ âm thanh. Trong lĩnh vực video (hình ảnh động), các phép biến đổi có thể giúp ta phân tích tần số không gian trong từng bức tranh đơn lẻ; và nó có thể cũng được sử dụng ở các chu kỳ theo chiều cao hoặc độ rộng của bức tranh.

Phân tích Fourier dựa trên việc bắt kỳ một dạng sóng tín hiệu tuần hoàn nào đều có thể được tái cấu trúc thành một số các tín hiệu hình sin có biên độ và pha thay đổi và có quan hệ điều hòa với nhau.

Biến đổi Fourier là một công cụ mạnh và được ứng dụng khá nhiều trong xử lý âm thanh và hình ảnh. Lưu ý rằng, các tín hiệu âm thanh và video hiếm khi là tổ hợp của các tín hiệu có tính chu kỳ nên chúng ta cần xác định rõ cửa sổ thời gian hoặc không gian mà chúng ta sẽ áp dụng khi biến đổi.

1.1.3.1. Biến đổi Fourier thuận

Nếu dãy $x(n)$ thoả mãn điều kiện:

$$\sum_{n=-\infty}^{\infty} |x(n)| < \infty \quad (1.1)$$

thì sẽ tồn tại phép biến đổi Fourier như sau:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega.n} \quad (1.2)$$

Biến đổi Fourier đã chuyển dãy số $x(n)$ thành hàm phức $X(e^{j\omega})$, (1.2) là biểu thức biến đổi Fourier thuận và được ký hiệu như sau:

$$FT[x(n)] = X(e^{j\omega}) \quad (1.3)$$

hay:

$$x(n) \xrightarrow{FT} X(e^{j\omega}) \quad (1.4)$$

(FT là chữ viết tắt của thuật ngữ tiếng Anh *Fourier Transform*).

Ký hiệu $X(e^{j\omega})$ để phân biệt phép biến đổi Fourier của dãy số $x(n)$ $FT[x(n)] = X(e^{j\omega})$ với phép biến đổi Fourier của hàm liên tục $x(t)$:

$$FT[x(t)] = \dot{X}(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \quad (1.5)$$

Biểu thức biến đổi *Fourier* của dãy số $x(n)$ (1.2) là suất phát từ biểu thức biến đổi *Fourier* của hàm liên tục $x(t)$, vì khi hàm dưới dấu tích phân là dãy rời rạc thì phải thay dấu tích phân bằng dấu tổng.

Do tính chất tuần hoàn của hàm mũ $e^{j\omega}$, nên $X(e^{j\omega})$ là hàm tuần hoàn của biến ω với chu kỳ 2π :

$$X(e^{j(\omega+k.2\pi)}) = \sum_{n=-\infty}^{\infty} x(n) e^{-j(\omega+k.2\pi).n} = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega.n} = X(e^{j\omega})$$

Điều đó có nghĩa là chỉ cần nghiên cứu hàm tần số $X(e^{j\omega})$ của các dãy rời rạc $x(n)$ với $\omega \in (-\pi, \pi)$ hoặc $\omega \in (0, 2\pi)$.

Sử dụng biến đổi *Fourier* cho phép nghiên cứu phổ của tín hiệu số và đặc tính tần số của hệ xử lý số. Nếu $x(n)$ là tín hiệu số thì $FT[x(n)] = X(e^{j\omega})$ là phổ của tín hiệu $x(n)$, còn với $h(n)$ là đặc tính xung của hệ xử lý số thì $FT[h(n)] = H(e^{j\omega})$ là đặc tính tần số của hệ xử lý số.

1.1.3.2. Biến đổi Fourier ngược

Biến đổi Fourier ngược cho phép tìm dãy $x(n)$ từ hàm ảnh $X(e^{j\omega})$. Để tìm biểu thức của phép biến đổi Fourier ngược, xuất phát từ biểu thức *Fourier* thuận (1.2):

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega.n} \quad (1.6)$$

Nhân cả hai vế của (1.6) với $e^{j\omega m}$ rồi lấy tích phân trong khoảng $(-\pi, \pi)$, nhận được:

$$\int_{-\pi}^{\pi} X(e^{j\omega}).e^{j\omega.m} d\omega = \int_{-\pi}^{\pi} \sum_{n=-\infty}^{\infty} x(n).e^{-j\omega.n}.e^{j\omega.m} d\omega = \sum_{n=-\infty}^{\infty} x(n) \int_{-\pi}^{\pi} e^{j\omega.(m-n)} d\omega$$

Vì : $\int_{-\pi}^{\pi} e^{j\omega.(m-n)} d\omega = \begin{cases} 2\pi & \text{khi } m=n \\ 0 & \text{khi } m \neq n \end{cases}$

Nên : $\int_{-\pi}^{\pi} X(e^{j\omega}).e^{j\omega.m} d\omega = 2\pi.x(n)$

Từ đó suy ra biểu thức của phép biến đổi *Fourier* ngược:

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}).e^{j\omega.n} d\omega \quad (1.7)$$

Phép biến đổi *Fourier* ngược được ký hiệu như sau:

$$IFT[X(e^{j\omega})] = x(n) \quad (1.8)$$

Hay :

$$X(e^{j\omega}) \xrightarrow{IFT} x(n) \quad (1.9)$$

(*IFT* là chữ viết tắt của thuật ngữ tiếng Anh *Inverse Fourier Transform*).

Biểu thức biến đổi *Fourier* thuận (1.6) và biểu thức biến đổi *Fourier* ngược (1.7) hợp thành cặp biến đổi *Fourier* của dãy số $x(n)$.

1.1.4. Biến đổi Cosin rời rạc

Phép biến đổi được xem là tốt nhất cho nén ảnh là phép biến đổi cosin rời rạc (DCT). DCT là một trường hợp đặc biệt của biến đổi Fourier.

Biến đổi DCT là một công đoạn chính trong các phương pháp nén sử dụng biến đổi. Hai công thức ở đây minh họa cho 2 phép biến đổi DCT thuận nghịch đối với mỗi khối ảnh có kích thước 8 x 8. Giá trị $x(n_1, n_2)$ biểu diễn các mức xám của ảnh trong miền không gian, $X(k_1, k_2)$ là các hệ số sau biến đổi DCT trong miền tần số.

$$X(k_1, k_2) = \frac{\varepsilon_{k1}\varepsilon_{k2}}{4} \sum_{n_1=0}^7 \sum_{n_2=0}^7 x(n_1, n_2) \cos \frac{(2n_1+1)k_1\pi}{16} \cos \frac{(2n_2+1)k_2\pi}{16} \quad (1.10)$$

$$x(n_1, n_2) = \frac{\varepsilon_{k1}\varepsilon_{k2}}{4} \sum_{k_1=0}^7 \sum_{k_2=0}^7 X(k_1, k_2) \cos \frac{(2n_1+1)k_1\pi}{16} \cos \frac{(2n_2+1)k_2\pi}{16} \quad (1.11)$$

với

$$\varepsilon_{kl} = \begin{cases} \sqrt{2} & \text{khi } k_1 = 0 \\ 0 & \text{khi } 1 < k_1 < 8 \end{cases}$$

và

$$\varepsilon_{kl} = \begin{cases} \sqrt{2} & \text{khi } k_2 = 0 \\ 0 & \text{khi } 1 < k_2 < 8 \end{cases}$$

1.1.5. Biến đổi Wavelet

Biến đổi Wavelet là phép biến đổi được sử dụng để phân tích các tín hiệu không ổn định (non-stationary) – là những tín hiệu có đáp ứng tần số thay đổi theo thời gian.

Biến đổi Wavelet được thực hiện theo cách: tín hiệu được nhân với hàm Wavelet rồi thực hiện biến đổi riêng rẽ cho các khoảng tín hiệu khác nhau trong miền thời gian tại các tần số khác nhau. Cách tiếp cận như vậy còn được gọi là: phân tích đa phân giải – MRA (Multi Resolution Analysis): phân tích tín hiệu ở các tần số khác nhau và cho các độ phân giải khác nhau. MRA khi phân tích tín hiệu cho phép: phân giải thời gian tốt và phân giải tần số kém ở các tần số cao; phân giải tần số tốt và phân giải thời gian kém ở các tần số

thấp. Như vậy kỹ thuật này rất thích hợp với những tín hiệu: có các thành phần tần số cao xuất hiện trong khoảng thời gian ngắn, các thành phần tần số thấp xuất hiện trong khoảng thời gian dài chẳng hạn như ảnh và khung ảnh video.

1.1.5.1. Biến đổi Wavelet liên tục

Bằng cách lấy thang tỉ lệ (scaling) và dịch chuyển một hàm thời gian $\psi(t)$ gọi wavelet mẹ hay wavelet cơ sở, ta được một họ wavelet:

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (1.12)$$

trong đó a là thông số thang tỉ lệ chỉ sự co giãn của wavelet, b là thông số dịch chuyển chỉ vị trí thời gian của wavelet. Dạng sóng tổng quát của các wavelet trong cùng họ được bảo toàn trong mọi co giãn và tịnh tiến.

Biến đổi wavelet liên tục (CWT) của một hàm thời gian (tín hiệu) $x(t)$ được định nghĩa như sau:

$$W_x(a, b) = \int_{-\infty}^{\infty} x(t) \psi_{ab}^*(t) dt = \langle x(t), \psi_{ab}(t) \rangle \quad (1.13)$$

trong đó $*$ chỉ liên hiệp phức, $\langle \cdot, \cdot \rangle$ chỉ tích nội. Biến đổi wavelet $W_x(a, b)$ diễn tả sự tương quan giữa tín hiệu $x(t)$ và wavelet $\psi_{a,b}(t)$. Biến đổi thuận ở trên là phân tích, ngược lại là tổng hợp để phục hồi tín hiệu thời gian.

1.1.5.2. Biến đổi Wavelet rời rạc

Biến đổi wavelet liên tục chứa nhiều trùng lặp và đòi hỏi tính toán công phu nên ít được dùng. Cả hai trở ngại trên được giải quyết đồng thời bằng cách rời rạc hóa thông số a, b :

$$a = a_0^m, \quad b = nb_0 a_0^m \quad (1.14)$$

trong đó m, n là số nguyên. Họ wavelet ở (1.12) trở thành:

$$\Psi_{m,n}(t) = a_0^{-m/2} \psi(a_0^{-m}t - nb_0) \quad (1.15)$$

Thông dụng nhất là rời rạc hóa dạng bát phân (octave) hay lũy thừa của 2 (dyadic) với $a_0=2, b_0 = 1$, kết quả:

$$\Psi_{m,n}(t) = 2^{-m/2} \psi(2^{-m}t - n) \quad (1.16)$$

Với sự chọn lựa thông số a, b như trên ta có biến đổi wavelet rời rạc (DWT) có các hệ số wavelet là:

$$c_{m,n} = \langle x(t), \Psi_{m,n}(t) \rangle = \int_{-\infty}^{\infty} x(t) \Psi_{m,n}^*(t) dt \quad (1.17)$$

Việc tổng hợp sẽ cho lại tín hiệu thời gian:

$$x(t) = \sum_m \sum_n c_{m,n} \Psi_{m,n}(t) \quad (1.18)$$

1.2. Vai trò của xử lý âm thanh và hình ảnh trong truyền thông đa phương tiện

Truy cập thông tin đa phương tiện khắp mọi nơi bây giờ là động lực chính cho việc thiết kế những mạng máy tính và mạng truyền thông thế hệ mới. Hơn nữa, các sản phẩm đang được phát triển để mở rộng khả năng tại tất cả các kết nối mạng hiện có để hỗ trợ lưu lượng truyền thông đa phương tiện. Đây là một sự chuyển dịch từ mạng điện thoại tương tự phát triển bởi Bell System đến mạng chuyển mạch gói dữ liệu – cơ sở của mạng Internet đến mạng truyền thông hợp nhất hỗ trợ người dùng ở khắp mọi nơi.

Đa phương tiện là sản phẩm của quá trình kết hợp dữ liệu, thoại, đồ họa, âm thanh, hình ảnh và video theo một cách thức nhất định để phục vụ nhu cầu truyền thông của con người. Truyền thông đa phương tiện liên quan đến việc truyền thông tin đa phương tiện qua mạng truyền thông.

Ngày nay, các công nghệ truyền thông hiện đại đã trở thành một phần không thể thiếu trong giao tiếp hằng ngày của chúng ta. Nó đã làm thay đổi nhanh chóng cách sống của chúng ta, cách tiếp nhận sự giáo dục, cách làm việc và là phần cơ bản tất yếu giúp thực hiện nhiệm vụ một cách suôn sẻ trong xã hội đương thời cũng như cuộc sống cá nhân của mỗi con người. Sự lớn mạnh một cách nhanh chóng trong kỹ thuật truyền thông của chúng ta là một cuộc cách mạng đã làm thay đổi xã hội chỉ trong một thời gian ngắn cuối thế kỷ 20, đặc biệt là trong hai thập kỷ cuối.

Trong những tiến bộ của cuộc cách mạng truyền thông gần đây, chúng ta thấy có bốn phát triển kỹ thuật đã làm thay đổi toàn cảnh trong lĩnh vực viễn thông. Đầu tiên là sự tăng nhanh của tốc độ trao đổi thông tin, với sự đột phá của công nghệ cáp quang, việc truyền thông tin đã được nâng lên từ khoảng 100Mbps cho một sợi quang bắt đầu từ năm 1980 cho đến bây giờ là 400Gbps. Dung lượng sợi quang đến 4000 lần chỉ trong vòng 20 năm qua.

Thứ 2 là, sự có mặt khắp nơi của các mạng chuyển mạch gói bởi sự phổ biến một cách nhanh chóng của Internet và các trang Web. Sự ra đời và phát triển của Internet và Web đã tạo ra một nền tảng chung cho chúng ta chia sẻ nhiều loại thông tin một cách nhanh chóng trong nhiều mối quan hệ theo nhiều cách khác nhau. So sánh với các mạng chuyển mạch khen truyền thông thì các mạng chuyển mạch gói có chi phí hiệu quả hơn và có năng lực xử lý và phục vụ tốt hơn. Hơn nữa, việc bổ sung thêm nhiều dịch vụ mới và những ứng dụng mới dễ dàng và linh hoạt hơn so với mạng chuyển mạch khen.

Thứ 3 là sự phát triển của truyền thông không dây. Hơn hai thập niên trước đây, tất cả mọi người đều biết rất ít về truyền thông cá nhân không dây, nhưng ngày nay nó đã được đón nhận nồng nhiệt bởi cả cộng đồng và công việc kinh doanh của nó đang lớn mạnh từng ngày ở khắp mọi nơi. Công nghệ truyền thông không dây đã phát triển từ những hệ thống tương tự đầu tiên (1G) cho đến thế hệ thứ hai (2G) là hệ thống kỹ thuật số, và nó hiện đang tiếp tục tiến triển đến thế hệ thứ 3 (3G), thứ 4 (4G) mà ở đó có sự tối ưu cho cả các dịch vụ truyền thông thoại, dữ liệu, ảnh và video hợp nhất.

Thứ 4 là nhu cầu phát triển truy nhập băng thông rộng qua các kết nối như đường dây thuê bao số DSL (Digital Subscriber Line) hoặc cáp tới mạng Internet là rất lớn... Chính sự phát triển này cho thấy trước được truyền thông trong tương lai gần sẽ tiến tới các mạng chuyển mạch gói dung lượng cao, tốc độ truyền tải lớn với truy nhập băng rộng không dây vào bất kỳ lúc nào và ở bất kỳ đâu.

Theo các nhà nghiên cứu, truyền thông đa phương tiện bao gồm rất nhiều chủ đề:

- **Xử lý đa phương tiện và mã hóa:** bao gồm phân tích nội dung đa phương tiện, tìm kiếm đa phương tiện dựa trên nội dung, an ninh đa phương tiện, xử lý âm thanh, hình ảnh và video, nén ...
- **Đa phương tiện hỗ trợ và hệ thống mạng:** bao gồm các giao thức mạng, Internet, các hệ điều hành, máy chủ và khách, chất lượng dịch vụ (QoS), và cơ sở dữ liệu.
- **Các công cụ đa phương tiện, hệ thống đầu cuối, và các ứng dụng.** Chúng bao gồm hệ thống siêu đa phương tiện (hypermedia), giao diện người dùng, hệ thống phân quyền, tương tác đa phương thức, và tích hợp: có mặt khắp nơi "ubiquity" - thiết bị duyệt Web ở khắp mọi nơi, giáo dục đa phương tiện, bao gồm cả máy tính hỗ trợ dạy học và thiết kế, và các ứng dụng của môi trường ảo.

Từ đây chúng ta có thể thấy rằng xử lý âm thanh, hình ảnh và video mà đặc biệt là các kỹ thuật nén âm thanh, thoại, hình ảnh và video là một trong những nội dung nghiên cứu của truyền thông đa phương tiện, hỗ trợ đắc lực cho việc truyền tải cũng như lưu trữ các nội dung đa phương tiện một cách có hiệu quả nhất.

Ví dụ như một tín hiệu âm thanh chất lượng cao cần xấp xỉ 1.5 Mbps hay một tín hiệu video màu độ phân giải thấp chất lượng TV chứa 30 khung hình/giây, với mỗi khung hình chứa 640x480 điểm ảnh (24 bit cho mỗi điểm ảnh màu) cần hơn 210 Megabit/giây cho lưu trữ. Do đó, một giờ phim màu số hóa cần xấp xỉ 95 Gigabyte để lưu trữ. Với tín hiệu video có độ phân giải cao – HDTV (High-Definition Television) có độ phân giải 1280x720 với 60 khung hình/giây thì đòi hỏi lưu trữ càng nhiều hơn nữa. Một giờ phim màu số hóa của tín hiệu video HDTV sẽ cần khoảng 560 Gigabyte lưu trữ. Hình chụp X-quang số hóa kích thước 14x17 inch² chiếm gần 45 Megabyte lưu trữ.

Hơn thế nữa, việc truyền những tín hiệu đa phương tiện có yêu cầu băng thông rất lớn qua những kênh truyền thông với băng thông giới hạn là một thách thức lớn và đôi khi không thể truyền được ở dạng thô của những tín hiệu đó. Mặc dù giá thành lưu trữ đã giảm đáng kể qua thập niên vừa rồi, nhưng nhu cầu của những ứng dụng lưu trữ và xử lý dữ liệu đang phát triển bùng nổ vượt qua tiến bộ này. Hầu hết những tín hiệu như ảnh, video, và âm thanh thường chứa nhiều thông tin dư thừa trong biểu diễn của chúng. Nén dữ liệu nói chung cũng như nén âm thanh, thoại, hình ảnh và video là giảm dư thừa trong biểu diễn dữ liệu để giảm đòi hỏi lưu trữ dữ liệu và do đó giảm chi phí truyền thông.

Những lợi ích và ứng dụng của nén dữ liệu được liệt kê dưới đây:

- Giảm không gian lưu trữ dữ liệu.
- Giảm chi phí khi truyền khói lượng lớn dữ liệu trên đường dài qua việc tối ưu băng thông đường truyền có sẵn.
- Tăng chất lượng hiển thị qua kênh truyền có băng thông giới hạn. Do đó, người dùng có thể thưởng thức những tín hiệu nghe nhìn chất lượng cao. Ví dụ: kênh TV 6 MHz có thể mang tín hiệu HDTV với chất lượng âm thanh, hình ảnh tốt hơn ở tốc độ cao hơn và độ phân giải cao hơn mà không cần thêm băng thông đường truyền.
- Vì việc giảm tốc độ dữ liệu bởi việc nén, mạng máy tính và Internet đang ngày càng trở nên thân thiện hơn về âm thanh và đồ họa, hơn là chỉ tập trung vào dữ liệu và văn bản như trước đây.
- Tăng cường bảo mật dữ liệu nhờ mã hóa và truyền dữ liệu phân tán từ những tập tin cơ sở dữ liệu đã nén nhằm ngăn việc truy xuất những thông tin đã được sở hữu.
- Tăng đáng kể tốc độ tính toán nhập-xuất trong thiết bị nhờ biểu diễn dữ liệu ngắn hơn.
- Giảm chi phí sao lưu và khôi phục dữ liệu nhờ lưu trữ bản sao của những tập tin cơ sở dữ liệu ở dạng nén.
- Những lợi ích này sẽ cho phép nhiều ứng dụng đa phương tiện hơn với giá thành giảm và do đó hướng tới nhiều người dùng hơn trong một tương lai gần.

Tóm lại, nén dữ liệu đã tạo ra nhiều cơ hội cho những ứng dụng sáng tạo như thư viện số, lưu trữ số, hội nghị truyền hình từ xa, giải trí số...

Hiện tại, có rất nhiều các chuẩn nén âm thanh, thoại, hình ảnh và video đang được sử dụng phổ biến trong truyền thông đa phương tiện như: G711, G729; JPEG; MPEG; H264...

1.3. Kết luận chương 1

Chương 1 đã giới thiệu những khái niệm cơ bản liên quan đến âm thanh, hình ảnh và video, trình bày về xu hướng phát triển của viễn thông: các nhu cầu về dịch vụ băng thông rộng, tốc độ cao và các hạn chế của công nghệ truyền dẫn, chuyển mạch liên quan để thấy được vai trò của xử lý âm thanh, thoại, hình ảnh và video trong truyền thông đa phương tiện.

Hướng dẫn ôn tập chương 1

1. Phân biệt các loại tín hiệu (liên tục, rời rạc).
2. Quá trình số hóa tín hiệu tương tự.
3. Ý nghĩa của các biến đổi Fourier, DCT và Wavelet.
4. Những lợi ích và ứng dụng của nén dữ liệu trong truyền thông đa phương tiện.
5. Một tín hiệu hình Sin có biên độ 5V cần được biến đổi thành dạng số sao cho nhận được tỷ số tín hiệu trên tạp âm lượng tử hóa không thấp hơn 25 dB. Yêu cầu cần bao nhiêu bước lượng tử hóa như nhau và cần có bao nhiêu bít để mã hóa mỗi thành phần rời rạc.
6. Giả sử một tín hiệu có phân bố đều (uniform), được lượng tử đều 256 mức, có tỷ số S/N là 18dB. Nếu muốn tăng tỷ số S/N của tín hiệu thành 30dB thì số mức lượng tử sẽ phải là bao nhiêu?

Chương 2: Kỹ thuật xử lý âm thanh

2.1. Các đặc trưng cơ bản của âm thanh

2.1.1. Khái niệm về âm thanh và các tham số đánh giá

Âm thanh trong thế giới tự nhiên về bản chất là những sóng âm được tạo ra từ dao động của vật thể và được truyền đi trong một môi trường truyền âm nhất định.

Âm thanh của tiếng nói, tương tự, là những sóng âm được tạo ra từ dao động của các bộ phận trong bộ máy phát âm và được truyền đi trong môi trường truyền âm (thường là không khí). Những sóng âm này truyền đến tai người nghe, dập vào màng nhĩ - một màng mỏng rất nhạy cảm của tai người – làm cho màng nhĩ cũng dao động, các dây thần kinh của màng nhĩ sẽ nhận được cảm giác âm khi tần số dao động của sóng đạt đến một độ lớn nhất định và người nghe nhận biết được lời nói. Liên lạc thông tin bằng tiếng nói là truyền thông tin từ não người nói sang não người nghe. Có thể xem như tiếng nói (thoại) là một trường hợp riêng nhưng phổ biến của âm thanh.

Âm thanh có các tham số đánh giá đặc trưng sau đây:

1. Tần số: Tần số của âm đơn là số lần dao động của không khí truyền dẫn âm trong một đơn vị thời gian là 1 giây. Tần số biểu thị độ cao (pitch) của âm thanh. Tần số càng lớn thì âm thanh càng cao và ngược lại. Đơn vị để đo tần số của âm thanh là Hertz (viết tắt là Hz).

Tai con người chỉ cảm thụ được những dao động có tần số từ khoảng 16Hz đến khoảng 20000Hz. Dải tần số từ 16Hz đến 20000Hz được gọi là dải tần số âm thanh hay âm tần hoặc sóng âm. Những âm có tần số nhỏ hơn 16Hz gọi là sóng hạ âm, còn những âm có tần số lớn hơn 20000 Hz gọi là sóng siêu âm và con người không cảm nhận được các sóng âm này nhưng có khá nhiều loài vật có thể cảm nhận được (ví dụ loài dơi có thể nghe được sóng siêu âm). Sóng âm, sóng siêu âm và hạ âm không chỉ truyền trong không khí mà còn có thể lan truyền tốt ở những môi trường rắn, lỏng, do đó sử dụng rất nhiều trong các thiết bị máy móc hiện nay.

Ứng với mỗi tần số dao động f , có chu kỳ dao động T là một bước sóng λ của âm thanh được xác định theo biểu thức $\lambda = c.T$ (c là tốc độ lan truyền của âm thanh trong không khí = 340m/s). Do đó, bước sóng của âm thanh trong dải âm tần là từ 21.25m đến 0.017m.

Trong thực tế, một âm phát ra thường không phải là một âm đơn mà là một âm phức bao gồm một âm đơn và một số âm hài có tần số gấp 2, 3 hoặc 4... lần âm đơn. Ngoài ra, trong dải âm tần người ta chia ra: tiếng trầm từ 16Hz đến 300Hz; tiếng vừa từ

300Hz đến 3000Hz và tiếng bỗng (tiếng thanh) 3000Hz đến 12000Hz. Tiếng nói con người thường có dải tần số từ 300Hz đến 3400Hz.

2. Áp suất âm thanh: Áp suất âm thanh hay còn gọi là thanh áp. Âm thanh truyền lan đến đâu thì làm thay đổi áp suất không khí ở đó. Áp suất do âm thanh tạo thêm ra ở một điểm gọi là thanh áp ở điểm đó. Đơn vị đo thanh áp là bar. Một bar là thanh áp tác động lên một diện tích 1cm² một lực là 1dyn. 1 bar = 1dyn/cm². Tuy nhiên, ngày nay, người ta thường dùng đơn vị Pascan (Pa) để đo thanh áp. 1 bar = 10 Kpa; 1 Pa = 1 N/m².

3. Công suất âm thanh: Là năng lượng âm thanh đi qua một diện tích S trong thời gian một giây. Công suất âm thanh P có thể tính bằng biểu thức:

$$P = p.S.v \quad (2.1)$$

Trong đó p là thanh áp, v là tốc độ dao động của một phần tử không khí tại đó và S là diện tích. Công suất âm thanh tính theo đơn vị oát (W).

Thông thường máy bay phản lực có công suất âm thanh là 10.000W; ô tô vận tải phóng nhanh: 0.12W; nói chuyện bình thường: 0.0003W.

4. Cường độ âm thanh: Cường độ âm thanh I là công suất âm thanh đi qua một đơn vị diện tích là 1cm².

$$I = P/S = p.v \quad (2.2)$$

Ba đại lượng: áp suất âm thanh, công suất âm thanh; cường độ âm thanh gắn liền với nhau. Cả ba đều biểu thị độ lớn nhỏ của âm thanh. Âm thanh có năng lượng càng lớn thì công suất, cường độ và áp suất của âm thanh càng lớn.

5. Độ mạnh (Intensity): Do biên độ dao động của vật thể quyết định. Biên độ dao động là trị số lớn nhất mà dao động đạt tới trong một nửa chu kỳ. Biên độ dao động càng lớn, âm thanh càng vang to và ngược lại. Đơn vị đo độ mạnh của âm thanh là décibel (viết tắt là dB). Trong lời nói của một người, độ mạnh của âm thanh là yếu tố cơ bản tạo nên âm lượng của âm và trọng âm của từ.

6. Độ dài (Length): Do thời gian dao động của vật thể quyết định. Độ dài của âm thanh tạo nên sự tương phản giữa các bộ phận của lời nói, là yếu tố tạo nên trọng âm, tạo nên các nguyên âm đối lập nhau về độ dài. Hai từ "tang" và "tăng" trong tiếng Việt có sự đối lập âm a dài (trong "tang") và âm a ngắn (trong "tăng").

7. Âm sắc (Timbre): Phụ thuộc vào độ cao, độ dài và độ mạnh tham gia bổ sung vào các thành phần kết cấu của âm. Đây là vẻ riêng biệt của một âm. Âm sắc được quyết định bởi: thể chất của vật thể dao động, tính chất phức hợp do hiện tượng cộng hưởng âm thanh và phương pháp làm cho vật thể dao động. Một âm có cùng độ cao, độ mạnh, độ dài được phát ra từ dây tờ sõ khác với từ một dây đồng; từ một ống sáo to dài, sõ khác với từ một ống sáo nhỏ, ngắn; từ việc gẩy sõ khác với từ việc gõ, búng, cọ xát hoặc thổi.

Âm sắc chính là cái sắc thái riêng của từng âm. Âm sắc còn được quyết định bởi vật thể dao động theo chu kì đều đặn hay không đều đặn; dao động theo chu kì đều đặn thì tạo ra âm vang (sonants), chu kì không đều đặn thì tạo ra âm ồn hay âm có nhiều tiếng động (non - sonants hoặc bruyants).

Xử lý âm thanh bao gồm nhiều lĩnh vực khác nhau, và tất cả đều liên quan đến sự hiện diện của âm thanh đối với người nghe. Chúng bao gồm: (1) Tái tạo lại âm nhạc với độ trung thực cao (high fidelity music) như là âm thanh ở các đĩa Compact (CD – audio compact discs); (2) Viễn thông thoại hoặc là các mạng điện thoại; (3) tổng hợp thoại (synthetic speech) mà ở đó các máy tính tạo và nhận dạng các mẫu thoại của con người.

Các lĩnh vực cụ thể liên quan đến xử lý âm thanh (trong đó có thoại) bao gồm: Nén âm thanh, thoại, nhận dạng tiếng nói; tổng hợp tiếng nói; nâng cao chất lượng (enhancement) và hồi phục tiếng nói (restoration).

Nén các tín hiệu thoại có rất nhiều ứng dụng, ví dụ như trong công nghệ thông tin di động số có nhiều người sử dụng dùng chung băng tần số. Nói cách khác, kỹ thuật nén cho phép nhiều người sử dụng chia sẻ tài nguyên của hệ thống hơn và làm cho việc sử dụng tài nguyên hệ thống có hiệu quả hơn. Một ví dụ khác nữa là lưu trữ thoại số (các máy trả lời điện thoại tự động), với một dung lượng bộ nhớ cho trước, nén cho phép lưu nhiều bản tin dài hơn.

Để có thể xử lý âm thanh/thoại cần hiểu rõ được các cơ sở của quá trình tạo ra tiếng nói và đặc trưng hệ thống thính giác con người.

2.1.2. Các đặc điểm của hệ thống thính giác con người

2.1.2.1. Cấu trúc tai người

Hệ thống thính giác của con người - Tai người là một cơ quan cực kỳ phức tạp. Hình 2.1 minh họa những kết cấu chính của tai người. Tai ngoài (outer ear) gồm có hai phần, vành tai – lớp da và sụn, gắn vào phía cạnh bên của đầu và ống tai có đường kính 0.5cm và cắm sâu vào trong đầu khoảng chừng 3cm. Cơ cấu tai ngoài này dẫn hướng âm thanh từ môi trường bên ngoài tới các cơ quan nhạy cảm là tai giữa và tai trong nằm an toàn ở trong hộp sọ của người. Cuối ống tai là một màng mỏng được gọi là màng nhĩ (*tympanic membrane*) hay còn gọi là trống tai (*ear drum*). Các sóng âm thanh đập vào màng nhĩ và làm cho nó rung động.

Tai giữa là một tập hợp các xương nhỏ có nhiệm vụ truyền những rung động của màng nhĩ tới ốc tai (*cochlea*) hay còn gọi là tai trong (*inner ear*), mà ở đó những rung động được biến đổi thành các xung thần kinh (neural impulses). Tai trong là một ống nhỏ có chứa dung dịch chất lỏng, có đường kính khoảng 2 mm và dài khoảng 3 cm. Mặc dầu được minh họa ở dạng duỗi thẳng ở hình 2.1, trên thực tế, tai trong được cuộn lại và trống