TRƯỜNG ĐẠI HỌC XÂY DỰNG HÀ NỘI
KHOA CÔNG NGHỆ THÔNG TIN

LẬP TRÌNH ÂM THANH

CHƯƠNG 15

Bản quyền, đạo đức, phát hiện vi phạm

Sau chương này người học phải làm được

- Đọc dấu vết codec trên tệp lạ; chứng minh bằng hai con số một tệp bitrate cao là bản dựng lại.
- Giải thích stereo ghép đôi làm ảnh stereo hẹp lại và đo được mức hẹp ấy.
- Cài ba sơ đồ thuỷ vân; đo độ bền bằng tỉ lệ bit sai qua bộ phép tấn công.
- Dựng vân tay bằng chòm sao đỉnh phổ và khoá băm từng cặp; đo khả năng nhận dạng dưới nhiễu.
- Chọn đặc trưng phát hiện giọng tổng hợp, đặt ngưỡng bằng EER (Equal Error Rate – tỉ lệ lỗi cân bằng), kiểm trên người nói chưa gặp.
- Chỉ ra chỗ máy dò mù: nó dò dấu vết vật lý, không dò "AI (Artificial Intelligence – trí tuệ nhân tạo)".
- Phân biệt ba lớp quyền; trả lời năm câu hỏi trước khi phát hành bản có giọng do AI sinh.
- Ghép ba công cụ bảo vệ vào đường ống chương 14 mà không làm sai lệch bản giao.

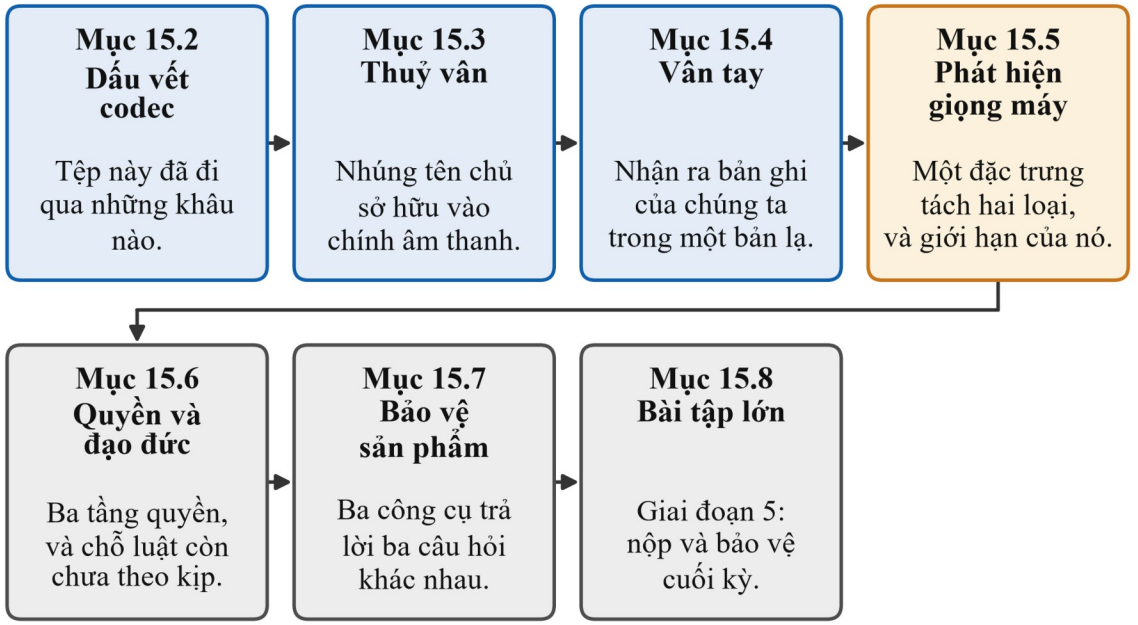
Chương cuối, và có một mục không đo được

Bốn mục kỹ thuật (dấu vết codec, thuỷ vân, vân tay, giọng tổng hợp) đều đo được; mọi phép nén chạy bằng ffmpeg. Mục 15.6 không có con số nào, và đó là chủ ý: kỹ thuật nói làm được gì, không nói nên làm gì.

MỤC 15.1

Vị trí chương học trong học phần

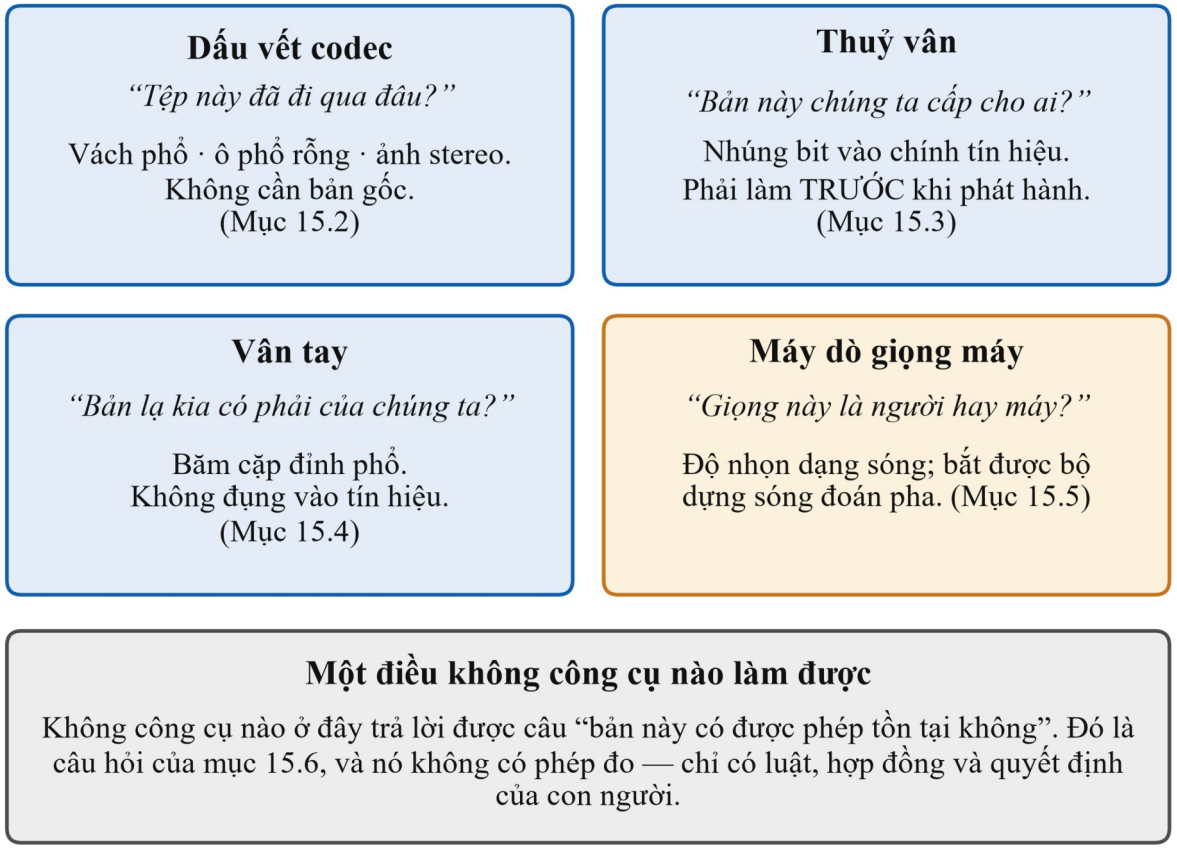
- Mười bốn chương trước trả lời một câu hỏi: làm ra âm thanh như thế nào.
- Chương cuối đổi loại câu hỏi: bản này ở đâu ra, và được phép làm gì với nó.
- Câu thứ nhất đo được bằng số; câu thứ hai cần luật, hợp đồng và quyết định của con người.



Mười bốn chương trước lo việc LÀM RA âm thanh. Chương cuối lo hai câu hỏi khác: bản này ở đâu ra, và chúng ta được phép làm gì với nó.

Bốn câu hỏi, bốn công cụ

- Dấu vết codec: tệp này đã đi qua đâu.
- Thuỷ vân: bản này đã cấp cho ai.
- Vân tay: bản lạ kia có phải của mình không.
- Máy dò giọng máy: giọng này là người hay máy.
- Dùng nhầm công cụ thì không ra câu trả lời nào.



Hình 15.2 — Bốn câu hỏi và bốn công cụ tương ứng

Điều kiện dùng được của từng công cụ

Công cụ	Trả lời câu	Cần gì trước đó	Đụng vào tín hiệu
Dấu vết codec	Tệp đã đi qua đâu	Không cần gì	Không
Thuỷ vân	Bản cấp cho ai	Nhúng trước khi phát hành	Có, dưới ngưỡng nghe
Vân tay	Bản lạ có phải của mình	Dựng kho từ bản gốc	Không
Máy dò giọng máy	Người hay máy	Hiệu chỉnh ngưỡng trên dữ liệu có nhãn	Không

Điều không công cụ nào làm được

Không công cụ nào trả lời được câu "bản này có được phép tồn tại không". Đó là ranh giới thật, và mục 15.6 chỉ ra năm chỗ cụ thể mà ranh giới ấy đi qua.

Mọi phép nén chạy bằng ffmpeg

- Dùng chính bộ mã hoá của người dùng thật: libmp3lame, aac, libopus, libvorbis; không qua mô hình giả lập.
- Bản nhạc gốc 48 kHz dùng cho các mục 15.2, 15.3, 15.4.
- Bản trộn stereo có tương quan hai kênh 0,9615 dùng cho mục 15.2.4.
- Giọng mô phỏng: nguồn xung thanh môn có jitter, shimmer, lọc qua ba formant.
- Giọng máy: chính giọng ấy đi qua phổ mel rồi Griffin-Lim.

Vì sao dùng Griffin-Lim làm giọng máy

Mục 15.5 đo dấu vết của khâu dựng lại sóng, và Griffin-Lim để lại đúng loại dấu vết ấy vì cũng phải đoán pha. Bộ dựng sóng học được hiện đại dựng pha tốt hơn nhiều; mục 15.5.5 đo thẳng điều đó.

MỤC 15.2

Dấu vết codec: tệp này đã đi qua đâu

- Bộ mã hoá có mất mát luôn để lại dấu vết trên phổ.
- Hai dấu vết chính: vách phổ và tỉ lệ ô phổ rỗng.
- Hai dấu vết phụ: stereo ghép đôi và tiền vọng.
- Chuyển mã là trường hợp gặp nhiều nhất trong thực tế.

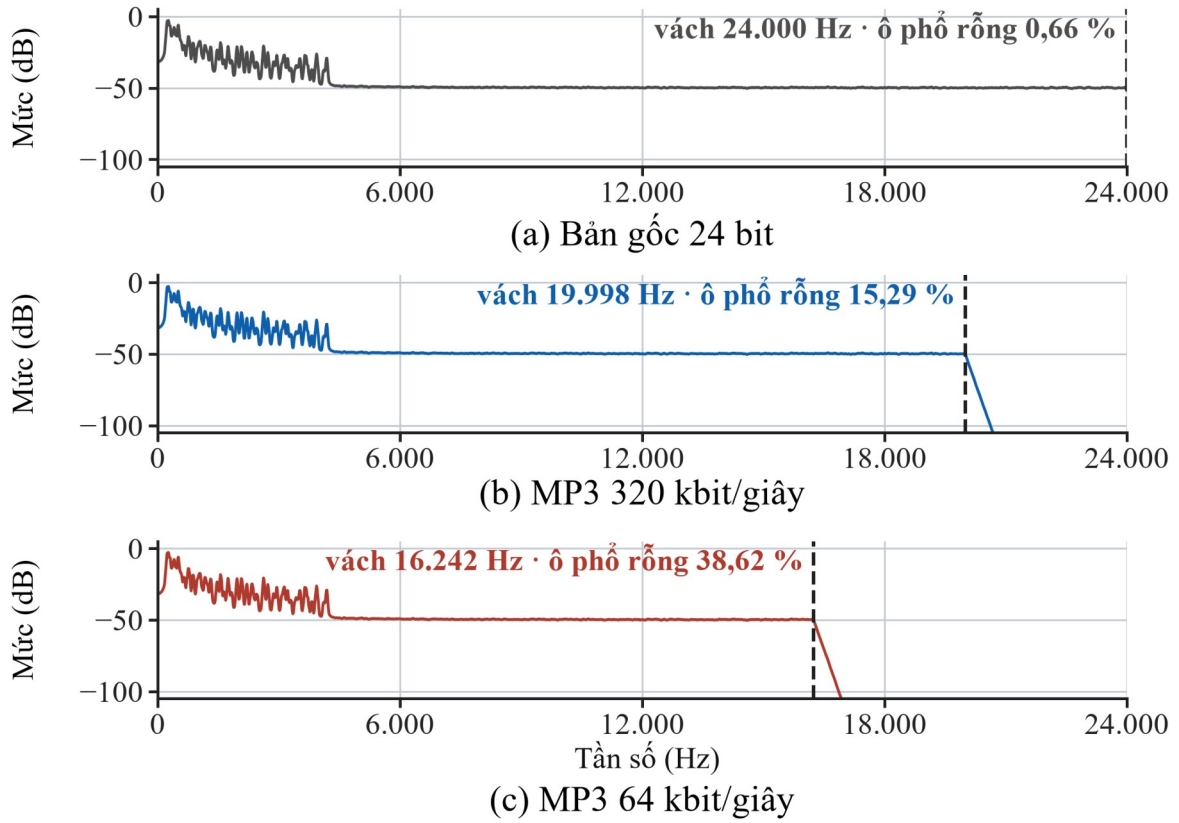
Vì sao bộ mã hoá có mất mát để lại dấu vết

- Bộ mã hoá chia phổ thành dải, tính tai nghe được bao nhiêu ở mỗi dải rồi cấp bit theo đó.
- Dải tai không nghe thì cấp không bit, nghĩa là loại bỏ hẳn.
- Âm thanh tự nhiên không có ô phổ rỗng tuyệt đối, cũng không tắt hẳn từ một tần số trở lên.
- Tệp đã qua nén có cả hai dấu hiệu ấy.

Ba loại dấu vết và điều kiện đọc được

Dấu vết	Đọc được khi nào	Bền tới đâu
Vách phổ	Chỉ cần chính tệp ấy	Sống qua mọi phép xử lý sau, kể cả nén lại
Tỉ lệ ô phổ rỗng	Luôn luôn	Bền nhất; không phụ thuộc tần số lấy mẫu
Ảnh stereo hẹp	Khi tệp có hai kênh	Bền, nhưng cần biết bản gốc rộng bao nhiêu
Tiền vọng	Chỉ khi có bản gốc để so	Yếu: mọi phép xử lý sau đều làm nhoè

Phổ trung bình của một bản nhạc qua ba đường



Hình 15.3 — Phổ trung bình của cùng một bản nhạc qua ba đường

Tỉ lệ ô phổ bị đưa về 0

$$L_{\text{rỗng}} = \frac{100}{TK} \left| \{ (t,k) : \, 20 \log_{10} |X(t,k)| < X_{\max} - \Delta \} \right|$$

(15.1)

- T là số khung, K là số vạch tần số của mỗi khung.
- delta là khoảng cách tính từ ô lớn nhất của cả phổ đồ.
- Bài dùng delta bằng 90 decibel, đủ sâu để không đếm nhầm chỗ vốn yên.
- Vách phổ dò bằng hàm đã dựng ở mục 14.2.5.

Đếm tỉ lệ ô phổ bị đưa về 0

```
def lo_pho_pt(x, nguong_db=-90.0, n_fft=2048, hop=512):
    X, _, _, _ = _stft(sang_mono(x), n_fft, hop)
    db = 20.0 * np.log10(np.maximum(np.abs(X), 1e-15))
    return float(100.0 * np.mean(db - db.max() < nguong_db))
```

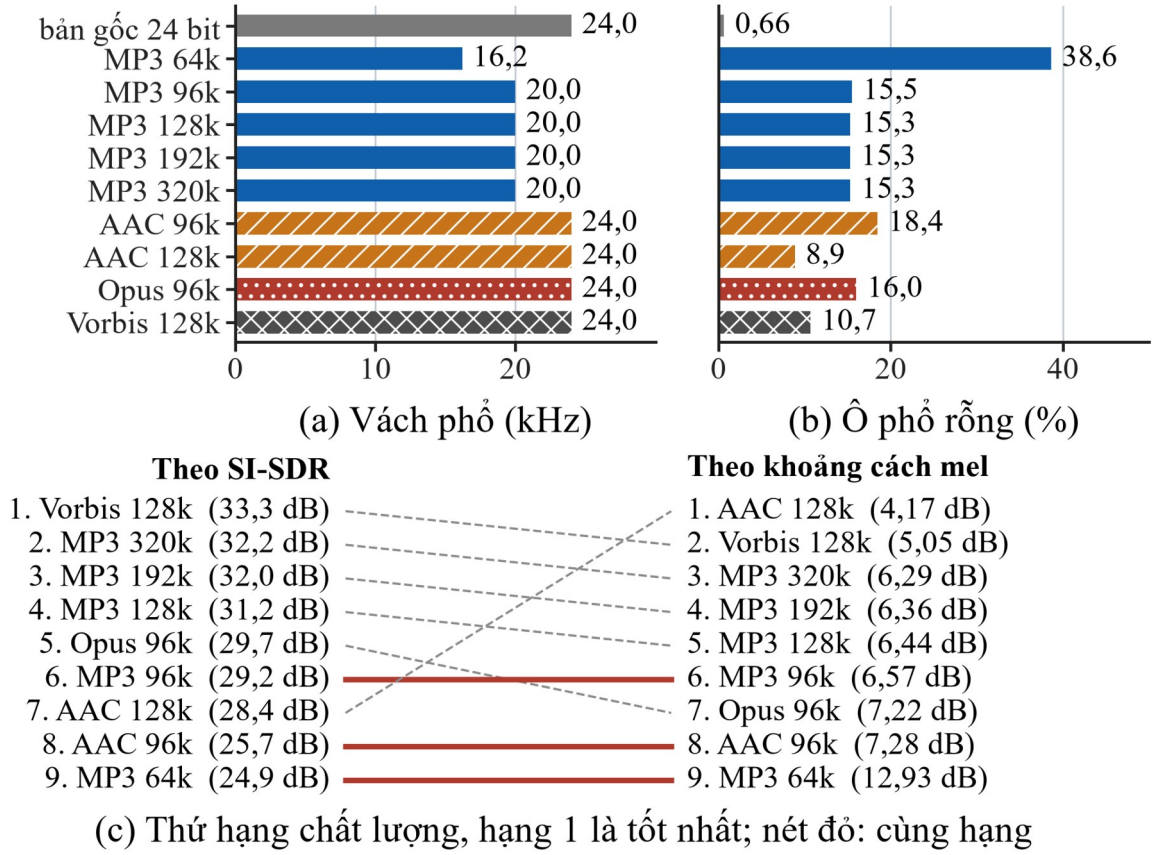
Chín đường nén trên cùng một bản nhạc tám giây

Bản	Vách phổ	Ô phổ rỗng	SI-SDR	Tiền vọng
bản gốc 24 bit	24.000 Hz	0,66 %	—	—
MP3 64k	16.242 Hz	38,62 %	24,9 dB	7,3 dB
MP3 128k	19.992 Hz	15,30 %	31,2 dB	5,7 dB
MP3 320k	19.998 Hz	15,29 %	32,2 dB	0,6 dB
AAC 128k	24.000 Hz	8,86 %	28,4 dB	14,8 dB
Opus 96k	24.000 Hz	16,00 %	29,7 dB	2,2 dB
Vorbis 128k	24.000 Hz	10,68 %	33,3 dB	2,4 dB

Rút gọn từ Bảng 15.4 (7 trong 10 dòng; lược MP3 96k, MP3 192k, AAC 96k và cột khoảng cách mel).

Ba điều đọc được từ phép quét

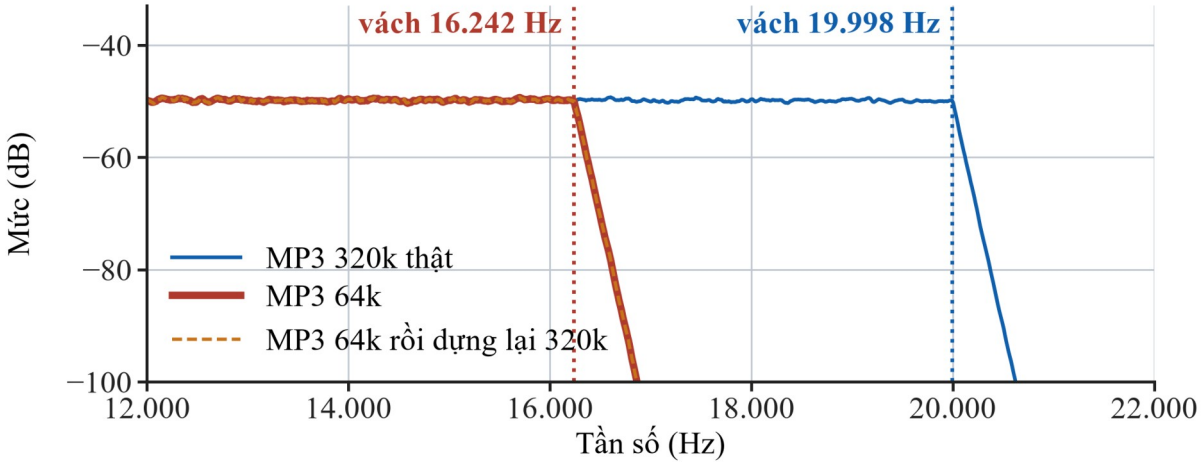
- MP3 cắt, các codec khác thì không: Lame cắt ở khoảng 20 kHz từ 96 kbit/giây, và ở 16,2 kHz tại 64 kbit/giây.
- Ô rỗng tách bản gốc khỏi mọi bản nén: 0,66 % so với 8,86 % của bản nén sạch nhất.
- SI-SDR (Scale-Invariant Signal-to-Distortion Ratio – tỉ số tín hiệu trên méo bất biến tỉ lệ) và khoảng cách mel cho hai thứ hạng khác nhau: chỉ 2 trong 9 hạng trùng nhau.



Hình 15.4 — Ba cách nhìn cùng một bảng: vách, lỗ phổ, và thứ hạng chất lượng

Bitrate ghi trong tệp không đúng thực tế

- Trường hợp gặp nhiều nhất: lấy bản 64 kbit/giây, giải nén, nén lại ở 320 kbit/giây.
- Hai tệp cùng 322,6 kbit/giây nhưng hai con số phổ hoàn toàn khác nhau.



(a) Vách phổ không đổi khi dựng lại ở bitrate cao

Bản	Vách phổ	Ô phổ rỗng	Bitrate thật
bản gốc	24.000 Hz	0,66 %	1.152 kbit/giây
MP3 64k	16.242 Hz	38,62 %	64,6 kbit/giây
MP3 64k rồi dựng lại 320k	16.242 Hz	38,61 %	322,6 kbit/giây
MP3 320k thật	19.998 Hz	15,29 %	322,6 kbit/giây

(b) Hai tệp cùng 322,6 kbit/giây, hai con số phổ khác nhau

Hình 15.5 — Hai tệp cùng 322,6 kbit/giây, hai con số phổ hoàn toàn khác

Bản gốc, bản 64k, bản chuyển mã và bản 320k thật

Bản	Vách phổ	Ô phổ rỗng	Bitrate thật
bản gốc	24.000 Hz	0,66 %	1.152 kbit/giây
MP3 64k	16.242 Hz	38,62 %	64,6 kbit/giây
MP3 64k rồi dựng lại 320k	16.242 Hz	38,61 %	322,6 kbit/giây
MP3 320k thật	19.998 Hz	15,29 %	322,6 kbit/giây

Hai dòng cuối là toàn bộ mục này

Cùng 322,6 kbit/giây, nhưng một tệp có vách ở 16.242 Hz và 38,61 % ô rỗng, tệp kia ở 19.998 Hz và 15,29 %. Nén lại không thêm được thông tin: đừng đánh giá chất lượng bằng bitrate ghi trong tệp.

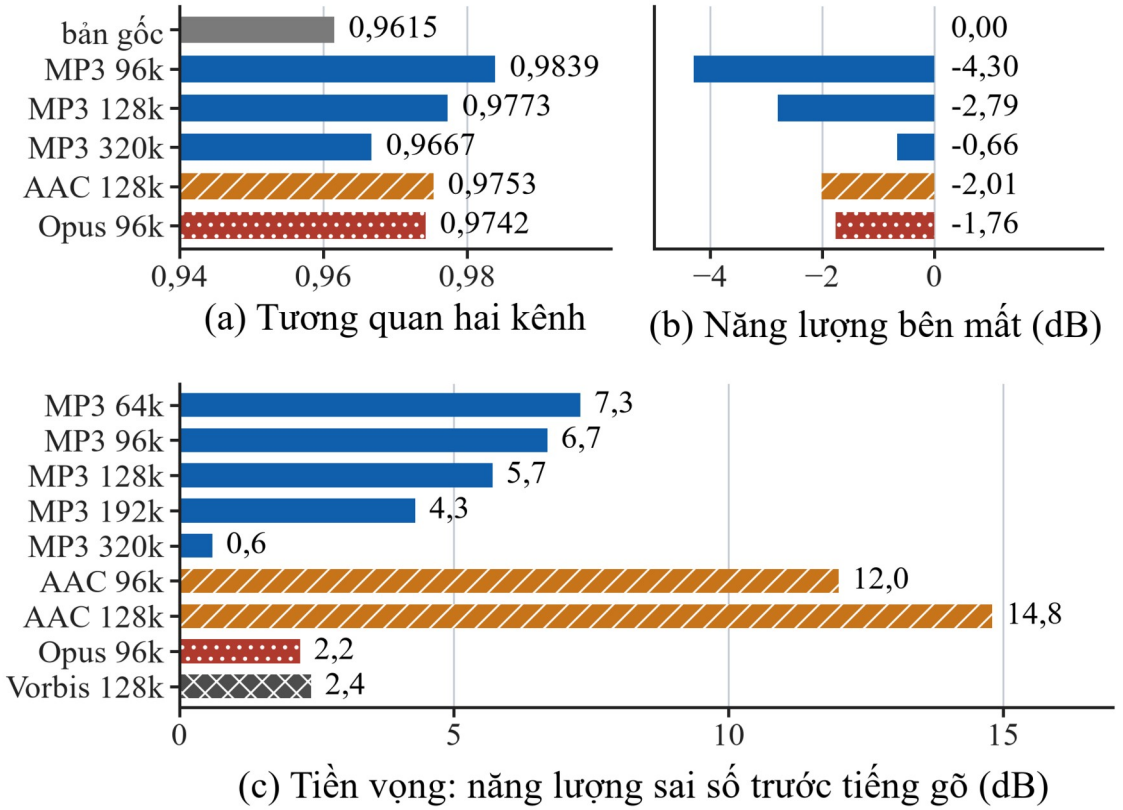
Kênh giữa và kênh bên

$$M = \frac{L+R}{2}, \qquad S = \frac{L-R}{2}, \qquad \eta = 10\log_{10}\frac{\langle S^2\rangle}{\langle M^2\rangle}$$

(15.2)

- Bộ mã hoá đổi từ trái/phải sang kênh giữa và kênh bên.
- Bản trộn thường: kênh giữa mang gần hết năng lượng, kênh bên rất nhỏ (−17,08 dB ở vật liệu của bài).
- Kênh bên là thứ bị loại bỏ trước tiên khi thiếu bit.

Hai dấu vết stereo và một dấu vết thời gian



Hình 15.6 — Hai dấu vết stereo và một dấu vết thời gian

Stereo ghép đôi ăn vào kênh bên như thế nào

Bản	Tương quan hai kênh	Bên trên giữa	Năng lượng bên mất	SI-SDR bên
bản gốc	0,9615	−17,08 dB	—	—
MP3 96k	0,9839	−20,90 dB	−4,30 dB	−5,3 dB
MP3 128k	0,9773	−19,39 dB	−2,79 dB	−2,9 dB
MP3 320k	0,9667	−17,72 dB	−0,66 dB	6,3 dB
AAC 128k	0,9753	−19,03 dB	−2,01 dB	−3,1 dB
Opus 96k	0,9742	−18,84 dB	−1,76 dB	−1,2 dB

Rút gọn: lược cột SI-SDR kênh giữa.

Ảnh stereo hẹp lại, và ý nghĩa pháp lý

- Ở 96 kbit/giây, kênh giữa đạt SI-SDR 18,2 dB nhưng kênh bên chỉ −5,3 dB: phần sai lớn hơn phần đúng.
- Nghe được: tương quan hai kênh đi từ 0,9615 lên 0,9839, bản nhạc nghe gần mono hơn.
- Pháp lý: tệp tự nhận là bản thu stereo gốc mà tương quan cao bất thường là lý do để đo tiếp.
- Đó chưa phải bằng chứng: một bản trộn vốn đã hẹp cũng cho con số ấy.

Tiền vọng: năng lượng sai số ngay trước tiếng gõ

Bản	Tiền vọng	Nhận xét
MP3 64k	7,3 dB	Nghe rõ nhất, như tiếng rít nhẹ trước nhịp
MP3 128k	5,7 dB	Giảm theo bitrate
MP3 320k	0,6 dB	Gần như hết
AAC 128k	14,8 dB	Cao hơn MP3 vì cửa sổ dài hơn
Opus 96k	2,2 dB	Thấp, vì Opus đổi cửa sổ linh hoạt
Vorbis 128k	2,4 dB	Thấp, cùng lý do

Tiền vọng xuất hiện vì sai số lượng tử trải đều trong cả khối và rò ra phần trước tiếng gõ.

Một phép đo đã thử và không dùng được

Dò lưới khung (MP3 khối 576 mẫu, AAC 1.024) không dùng được trên vật liệu này: bản gốc đã có thành phần mạnh đúng tại tần số ứng với 576 mẫu, nên đo nội dung nhạc chứ không đo lưới khung. Phép đo không dùng được thì nói rõ.

MỤC 15.3

Thuỷ vân âm thanh

- Thuỷ vân giấu một chuỗi bit vào tín hiệu: tai không nghe, máy đọc lại được.
- Trả lời câu hỏi: bản này đã cấp cho ai.
- Ba sơ đồ: bit thấp nhất, trải phổ, tiếng vọng.
- Thước đo độ bền: tỉ lệ bit đọc sai qua các phép tấn công.

Ba chiều của một sơ đồ thuỷ vân

Chiều	Nghĩa là gì	Đo bằng gì
Dung lượng	Giấu được bao nhiêu bit mỗi giây	Số bit trên tổng số giây
Độ bền	Còn đọc được sau những phép xử lý nào	Tỉ lệ bit đọc sai sau mỗi phép tấn công
Độ nghe thấy	Nghe ra được hay không	SI-SDR và khoảng cách mel (hay bất đồng)

Ba sơ đồ và tam giác đánh đổi của chúng

Bit thấp nhất

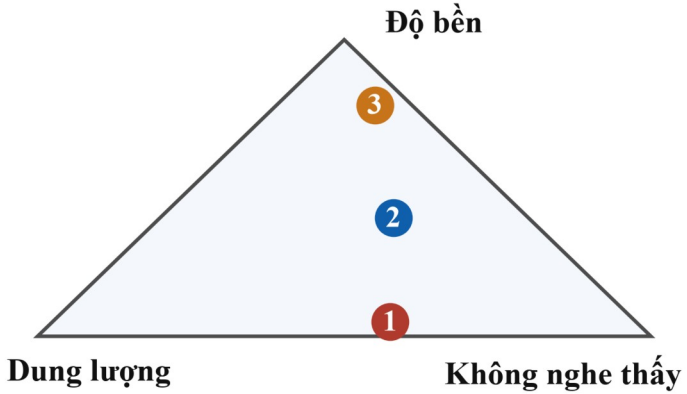
Ghi bit vào bit cuối của mẫu 16 bit.
48.000 bit/giây
SI-SDR 87,6 dB
mất sạch sau mọi phép xử lý

Trải phổ

Cộng ±α·(chuỗi giả ngẫu nhiên) vào từng khối.
5,86 bit/giây
SI-SDR 26 dB
sống qua MP3 64k

Tiếng vọng

Cộng chính tín hiệu trễ 0,8 ms (bit 0) hoặc 1,6 ms (bit 1).
2,93 bit/giây
mel chỉ 1,50 dB
đọc bằng cepstrum



1. Bit thấp nhất: dung lượng lớn, không bền
2. Trải phổ: cân bằng, yếu khi mất đồng bộ
3. Tiếng vọng: bền nhất, dung lượng nhỏ nhất

Vị trí trong tam giác chỉ minh hoạ định tính; không sơ đồ nào mạnh cả ba chiều.

Hình 15.7 — Ba sơ đồ và tam giác đánh đổi của chúng

Bit thấp nhất: dung lượng lớn nhất, không nghe thấy

- Ghi bit vào bit thấp nhất của mỗi mẫu đã lượng tử hoá.
- Dung lượng bằng tần số lấy mẫu: 48.000 bit mỗi giây.
- Không nghe thấy: SI-SDR 87,6 dB, khoảng cách mel 0,01 dB.
- Về hai chiều đầu, không sơ đồ nào bằng.

Nhúng thuỷ vân bit thấp nhất

```
def thuy_van_lsb_nhung(x, bit, so_bit_mau=16):
    q = np.round(np.clip(x, -1, 1) * (2**(so_bit_mau-1) - 1))
    q = q.astype(np.int64)
    q[:len(bit)] = (q[:len(bit)] & ~1) | np.asarray(bit)
    return q.astype(float) / (2**(so_bit_mau-1) - 1)
```

Bit thấp nhất chết vì mọi thứ

Giảm mức 6 dB: tỉ lệ bit sai 0,485. Đổi tần số lấy mẫu: 0,482. Nén MP3: khoảng 0,49. Giá trị 0,5 là đoán ngẫu nhiên, tức thuỷ vân biến mất hoàn toàn. Chỉ dùng được để kiểm toàn vẹn trong hệ thống khép kín.

Trải phổ: nhúng

$$y[n] = x[n] + (2b - 1)\,\alpha\,\sigma_{\text{khối}}\,p[n], \qquad \alpha = 10^{G/20}$$

(15.3)

- Mỗi khối N mẫu mang một bit.
- Cộng vào khối một chuỗi giả ngẫu nhiên p, dấu tuỳ theo bit.
- Biên độ tỉ lệ với mức hiệu dụng của chính khối: nhạc to thì thuỷ vân mạnh, chỗ yên thì yếu.
- Đúng theo đường che lấp của mục 1.7.

Trải phổ: đọc bằng tương quan

$$\hat{b} = \frac{1}{2}\left[1 + \operatorname{sgn}\sum_{n=0}^{N-1} y[n]p[n]\right]$$

(15.4)

- Đọc lại là một phép nhân vô hướng với chuỗi p.
- Bit là 1 thì tổng dương, bit là 0 thì tổng âm.
- Chuỗi p đã biết trước nên không cần bản gốc.

Nhúng thuỷ vân trải phổ

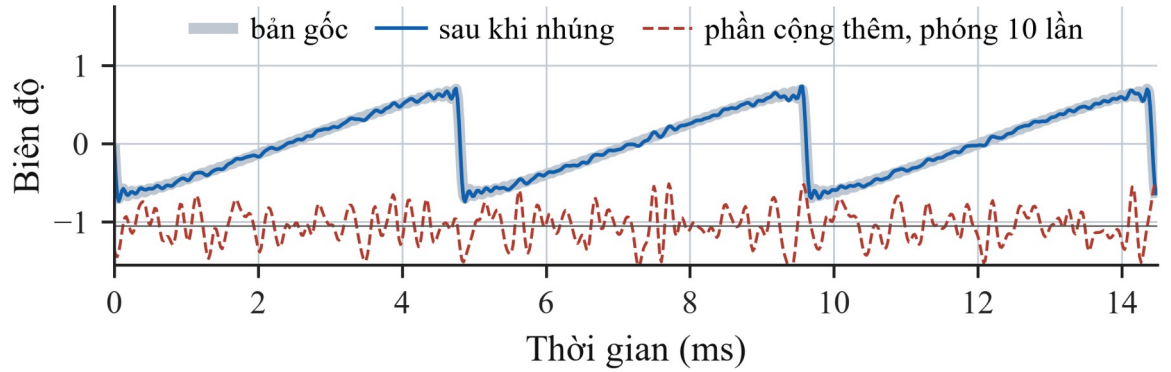
```
def thuy_van_trai_pho_nhung(x, bit, sr, chip=8192, muc_db=-26.0,
                            hat=0):
    y = x.copy()
    p = chuoi_pn(chip, hat, sr)          # da loc ve 500 - 5000 Hz
    g = db_to_lin(muc_db)
    for k, b in enumerate(bit):
        i = k * chip
        seg = y[i:i+chip]
        a = g * np.sqrt(np.mean(seg**2))
        y[i:i+chip] = seg + (1.0 if b else -1.0) * a * p
    return y
```

Giấu vào chỗ codec không dám đụng

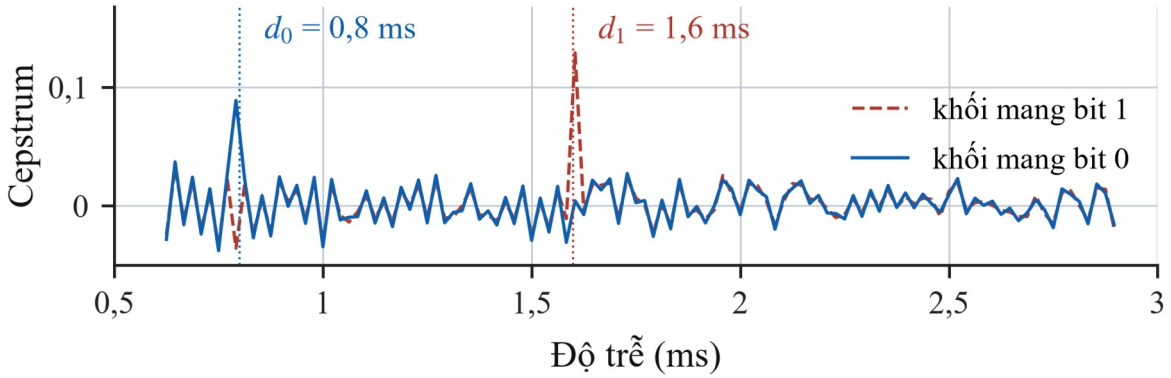
Chuỗi giả ngẫu nhiên được lọc về dải 500 đến 5.000 Hz. Codec cắt phần trên vách phổ và loại ô nhỏ ở vùng cao, nên thuỷ vân đặt ở đó sẽ mất ngay. Đặt vào dải tai nghe rõ nhất thì thuỷ vân sống.

Thuỷ vân trải phổ và thuỷ vân tiếng vọng

- Trải phổ nhìn trên dạng sóng; tiếng vọng nhìn trên cepstrum.
- Độ bền đo bằng tỉ lệ bit đọc sai: 0 là đọc đúng hoàn toàn.
- Giá trị 0,5 là đoán ngẫu nhiên, thuỷ vân đã mất sạch.



(a) Trải phổ: SI-SDR 26,0 dB, khoảng cách mel 8,35 dB



(b) Tiếng vọng: SI-SDR 11,8 dB, khoảng cách mel 1,50 dB

Hình 15.8 — Thuỷ vân trải phổ trên dạng sóng, và thuỷ vân tiếng vọng trên cepstrum

Tỉ lệ bit đọc sai

$$\mathrm{BER} = \frac{1}{N_{\mathrm{bit}}} \sum_{i=1}^{N_{\mathrm{bit}}} [\hat{b}_i \neq b_i]$$

(15.5)

- Giá trị 0: đọc đúng hoàn toàn.
- Giá trị 0,5: đoán ngẫu nhiên.
- Dùng chung cho cả ba sơ đồ và mọi phép tấn công.

Ẩn bằng tiếng vọng

$$y[n] = x[n] + a\,x[n - d_b], \qquad c[m] = \mathsf{F}^{-1}\{\log|Y(f)|\}$$

(15.6)

- Cộng chính tín hiệu bị trễ một khoảng rất ngắn; bit nằm ở khoảng trễ được chọn.
- Khoảng trễ d: 0,8 ms cho bit 0 và 1,6 ms cho bit 1.
- Tiếng vọng ở độ trễ d làm phổ loga gợn với chu kỳ 1/d.
- Cepstrum có một đỉnh đúng tại d.

Đọc thuỷ vân tiếng vọng bằng cepstrum

```
def thuy_van_vong_doc(y, so_bit, sr, d0=0.0008, d1=0.0016,
                      khoi=16384):
    n0, n1 = int(d0*sr), int(d1*sr)
    out = []
    for k in range(so_bit):
        seg = y[k*khoi:(k+1)*khoi] * np.hanning(khoi)
        X = np.fft.rfft(seg)
        c = np.fft.irfft(np.log(np.maximum(np.abs(X), 1e-12)))
        out.append(1 if c[n1] > c[n0] else 0)
    return np.array(out)
```

SI-SDR 11,8 dB nhưng khoảng cách mel chỉ 1,50 dB

Thuỷ vân tiếng vọng có SI-SDR 11,8 dB, đọc theo thói quen là sai nghiêm trọng; nhưng khoảng cách mel 1,50 dB, nhỏ hơn cả bản trải phổ (SI-SDR 26 dB). Thứ cộng thêm là tín hiệu trễ, tai nghe như tiếng phòng. Đây là lần thứ năm một thước đo khách quan lệch khỏi tai.

Tỉ lệ bit đọc sai qua chín phép tấn công

- Bit thấp nhất mất hoàn toàn sau mọi phép xử lý.
- Trải phổ và tiếng vọng sống qua gần hết, kể cả MP3 64 kbit/giây.
- Mỗi sơ đồ có một điểm yếu riêng.

Phép tấn công	Bit thấp nhất	Trải phổ	Tiếng vọng
không tác động	0,000	0,000	0,000
giảm mức 6 dB	0,485	0,000	0,000
cộng nhiễu −30 dB	0,499	0,000	0,000
đổi 48 → 44,1 → 48 kHz	0,482	0,000	0,000
thông thấp 8 kHz	0,502	0,087	0,000
MP3 128 kbit/giây	0,496	0,000	0,000
MP3 64 kbit/giây	0,501	0,000	0,043
AAC 128 kbit/giây	0,493	0,000	0,000
cắt 1.000 mẫu ở đầu	0,499	0,696	0,000

đọc gần như không sai (không quá 0,01)
sai một phần (dưới 0,1)
từ 0,48 trở lên: thuỷ vân không còn gì (0,5 là đoán ngẫu nhiên)

Hình 15.9 — Tỉ lệ bit đọc sai của ba sơ đồ qua chín phép tấn công

Tỉ lệ bit đọc sai: 0 là đúng, 0,5 là ngẫu nhiên

Phép tấn công	Bit thấp nhất	Trải phổ	Tiếng vọng
không đụng gì	0,000	0,000	0,000
giảm mức 6 dB	0,485	0,000	0,000
đổi 48 → 44,1 → 48 kHz	0,482	0,000	0,000
thông thấp 8 kHz	0,502	0,087	0,000
MP3 64 kbit/giây	0,501	0,000	0,043
cắt 1.000 mẫu ở đầu	0,499	0,696	0,000
Dung lượng (bit/giây)	48.000	5,86	2,93

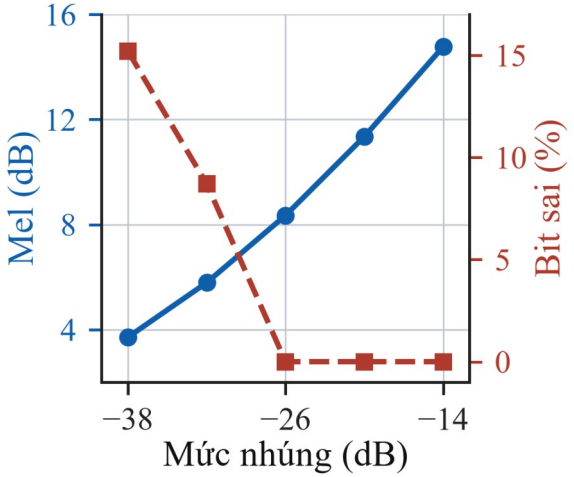
Rút gọn: lược ba phép tấn công (nhiễu −30 dB, MP3 128k, AAC 128k).

Sơ đồ bền nhất có dung lượng nhỏ nhất

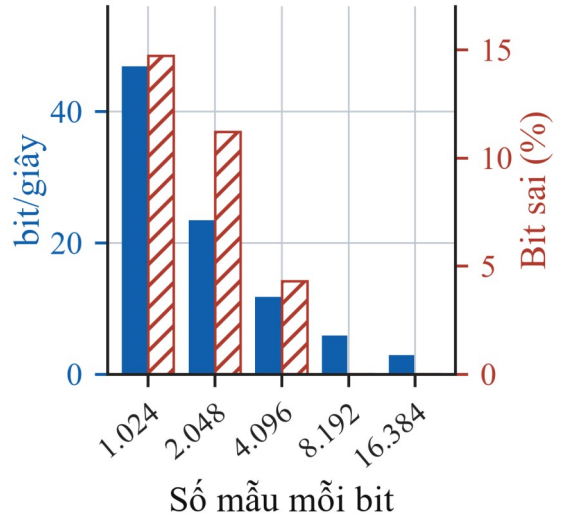
- Sơ đồ bền nhất có dung lượng nhỏ hơn sơ đồ yếu nhất mười sáu nghìn lần.
- Mã định danh vài chục bit trong bản nhạc vài phút chỉ cần rất ít dung lượng.
- Với việc đó, 2,93 bit mỗi giây là quá đủ.

Kẻ thù thật của trải phổ

- Trải phổ sống qua MP3 64 kbit/giây nhưng chết vì cắt 1.000 mẫu (21 ms).
- Tín hiệu không lỗi; mọi khối bị lệch chỗ.
- Mỗi khối đọc ra là một phần của hai khối nhúng.



(a) Quét mức nhúng
(bit sai: sau MP3 64 kbit/giây)



(b) Quét độ dài khối
(bit sai: sau MP3 64 kbit/giây)

Cắt 1.000 mẫu (21 ms) ở đầu

đọc thẳng: bit sai 69,6 %
(mọi khối bị lệch chỗ)

Tìm lại đồng bộ rồi đọc

độ lệch tìm được 7.192 mẫu
(đúng 7.192); bit sai 0,0 %

(c) Mất đồng bộ và cách chữa

Hình 15.10 — Hai phép đánh đổi, và cách chữa mất đồng bộ

Tìm lại đồng bộ bằng chính chuỗi giả ngẫu nhiên

```
def tim_dong_bo_trai_pho(y, sr, chip=8192, hat=0):
    p = chuoi_pn(chip, hat, sr)
    tot, best = -1.0, 0
    for d in range(chip):
        v = sum(abs(np.dot(y[d+k*chip : d+(k+1)*chip], p))
                for k in range(4))
        if v > tot:
            tot, best = v, d
    return best
```

Cắt đầu và cách chữa

Đại lượng	Giá trị
Số mẫu bị cắt	1.000
Tỉ lệ bit sai nếu đọc thẳng	0,696
Độ lệch tìm được	7.192 mẫu
Độ lệch đúng	7.192 mẫu
Tỉ lệ bit sai sau khi đồng bộ lại	0,000

Hệ thống thật nhúng thêm mẫu đồng bộ rồi tìm bằng tương quan; nhanh hơn và chịu được cắt giữa tệp (bài tập 3).

Quét mức nhúng: độ nghe thấy đổi lấy độ bền

Mức nhúng	SI-SDR	Khoảng cách mel	Bit sai (sạch)	Bit sai (MP3 64k)
−14 dB	14,0 dB	14,77 dB	0,000	0,000
−20 dB	20,0 dB	11,35 dB	0,000	0,000
−26 dB	26,0 dB	8,35 dB	0,000	0,000
−32 dB	32,0 dB	5,79 dB	0,087	0,087
−38 dB	38,0 dB	3,71 dB	0,130	0,152

Quét độ dài khối: dung lượng đổi lấy độ bền

Số mẫu mỗi bit	Dung lượng	Bit sai (sạch)	Bit sai (MP3 128k)	Bit sai (MP3 64k)
1.024	46,88 bit/giây	0,144	0,144	0,147
2.048	23,44 bit/giây	0,107	0,107	0,112
4.096	11,72 bit/giây	0,022	0,032	0,043
8.192	5,86 bit/giây	0,000	0,000	0,000
16.384	2,93 bit/giây	0,000	0,000	0,000

Có một ngưỡng; vượt ngưỡng thì sai rất nhanh. Cấu hình dùng ở mục 15.7: −26 dB, khối 8.192 mẫu.

MỤC 15.4

Vân tay âm thanh: nhận ra một bản ghi

- Vân tay trả lời: bản lạ kia có phải của mình hay không.
- Không đụng vào tín hiệu, không cần chuẩn bị trước khi phát hành.
- Dựng trên đỉnh phổ, băm từng cặp đỉnh, khớp theo độ lệch thời gian.

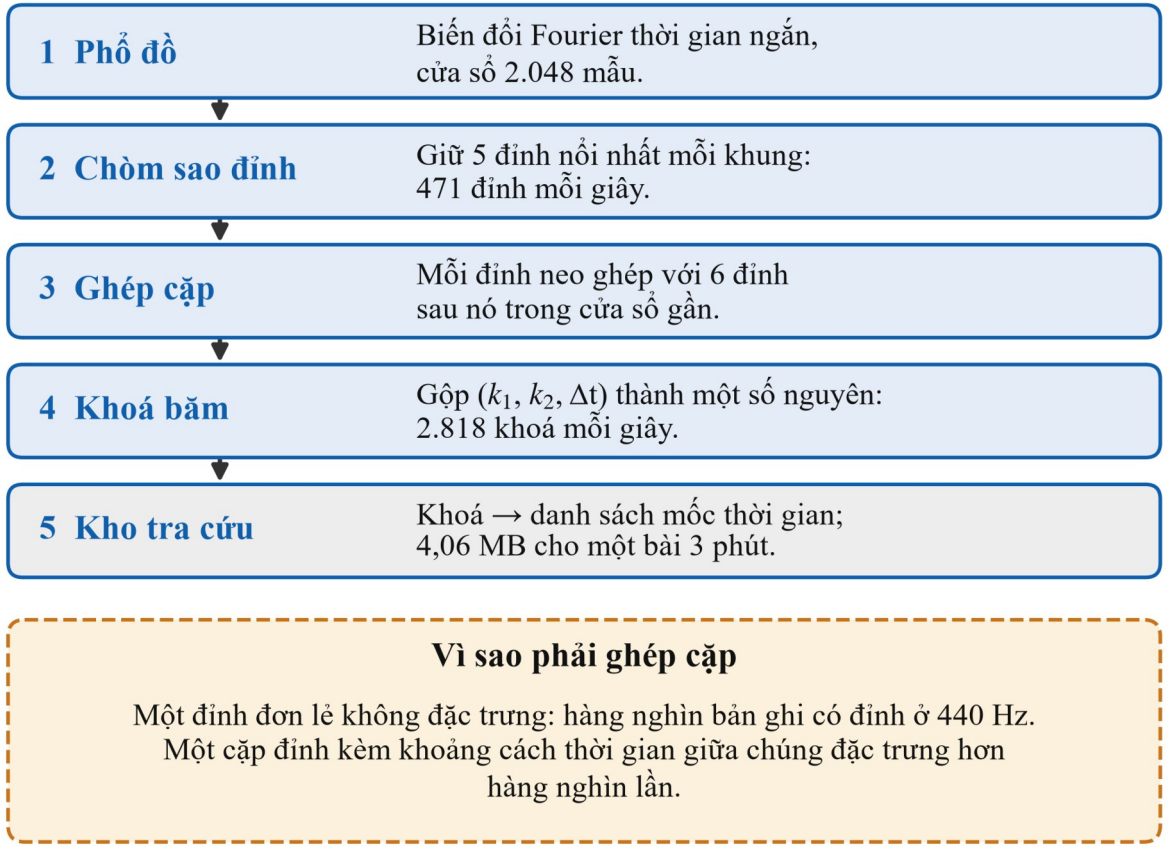
Ba điều kiện một vân tay phải thoả

Điều kiện	Nghĩa là gì
Bền	Không đổi sau nén, sau nhiễu, sau khi qua loa và micrô
Phân biệt được	Hai bản ghi khác nhau cho vân tay khác nhau
Tra cứu được	Tìm được trong kho hàng triệu bài trong thời gian chấp nhận được

Thuỷ vân phải nhúng trước khi phát hành; vân tay giải đúng bài toán còn lại.

Năm bước từ phổ đồ tới khoá tra cứu

- Điều kiện thứ nhất quyết định dựng vân tay trên đỉnh phổ.
- Bộ mã hoá loại ô nhỏ mà tai không nghe, nhưng không dám đụng vào đỉnh.



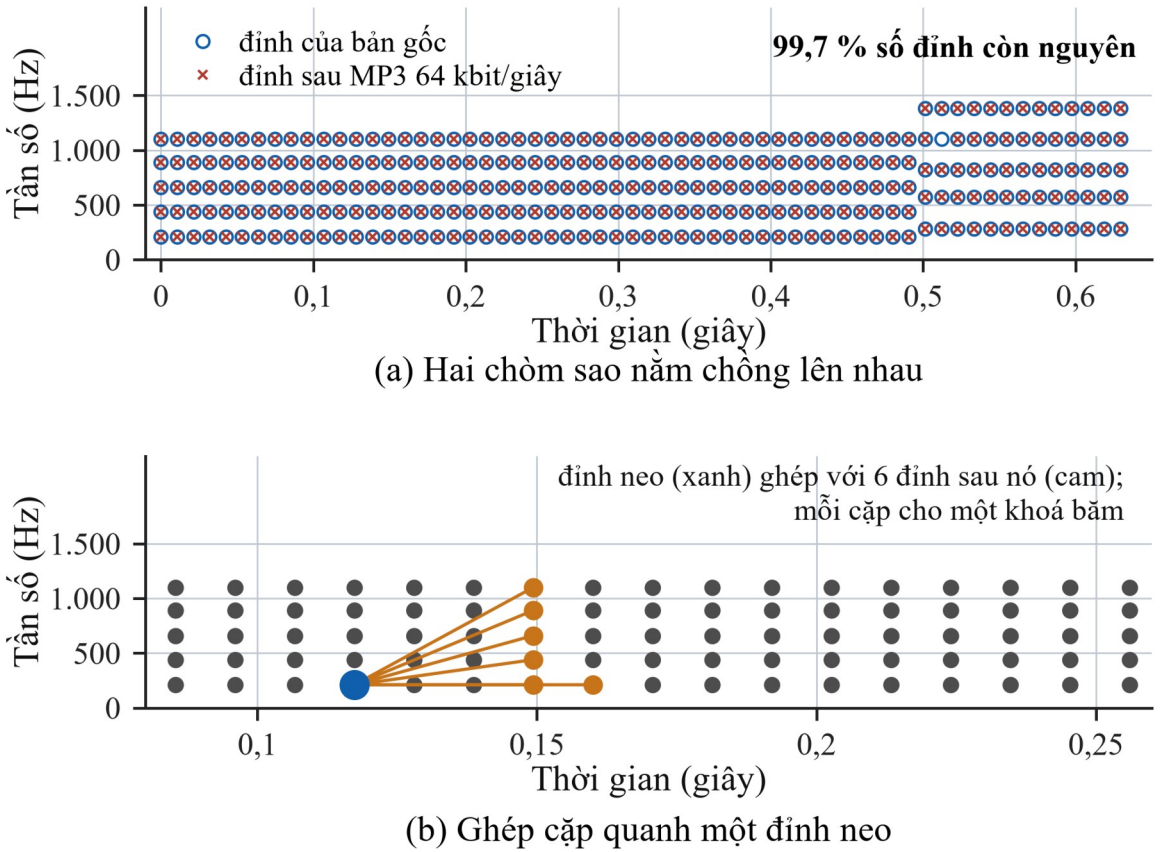
Hình 15.11 — Năm bước từ phổ đồ tới một khoá tra cứu

Vì sao đỉnh phổ sống sót

Bản	SI-SDR	Lệch độ lớn phổ (trung vị)	Đỉnh còn nguyên
MP3 128 kbit/giây	31,2 dB	0,77 dB	99,7 %
MP3 64 kbit/giây	24,9 dB	1,64 dB	99,7 %

Đỉnh là thứ tai nghe rõ nhất nên bộ mã hoá giữ lại; đó là toàn bộ lý do phương pháp chạy được.

Hai chòm sao chồng lên nhau, ghép cặp quanh đỉnh neo



Hình 15.12 — Hai chòm sao chồng lên nhau, và cách ghép cặp quanh một đỉnh neo

Khoá băm từng cặp đỉnh

$$h = (k_1 \ll 20) \text{ OR } (k_2 \ll 8) \text{ OR } \Delta t \tag{15.7}$$

- Một đỉnh đơn lẻ không đặc trưng: hàng nghìn bản có đỉnh ở 440 Hz.
- Một cặp đỉnh kèm khoảng cách thời gian đặc trưng hơn hàng nghìn lần.
- Khoá gộp ba số: vạch tần số của đỉnh neo, của đỉnh sau, và khoảng cách thời gian (số khung).

Băm từng cặp đỉnh quanh một đỉnh neo

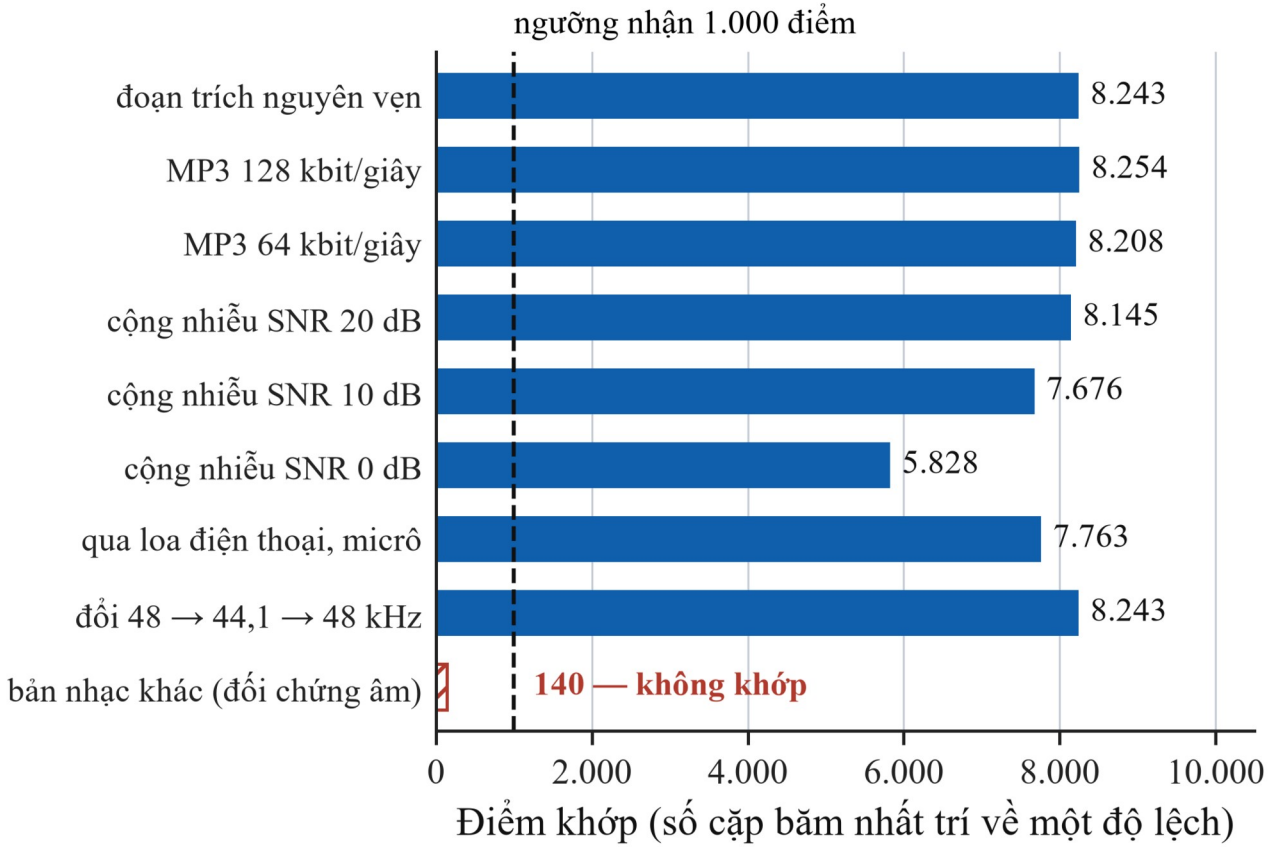
```
def bam_cap_dinh(dinh, quat_t=24, quat_f=48, toi_da=6):
    dinh = sorted(dinh); ra = []
    for i, (t1, f1) in enumerate(dinh):
        dem = 0
        for t2, f2 in dinh[i+1:]:
            dt = t2 - t1
            if dt <= 0: continue
            if dt > quat_t: break
            if abs(f2 - f1) > quat_f: continue
            ra.append((((f1 << 20) | (f2 << 8) | dt), t1))
            dem += 1
            if dem >= toi_da: break
    return ra
```

Điểm khớp theo độ lệch thời gian

$$D = \max_{\delta} \left| \{ (h, t) : \; t_{\text{kho}}(h) - t = \delta \} \right| \tag{15.8}$$

- Điểm khớp là số cặp khoá cùng trỏ về một độ lệch thời gian, không phải số khoá trùng.
- Đoạn trích ba giây vẫn khớp vì mọi cặp nhất trí về cùng một độ lệch.
- Khoá trùng ngẫu nhiên thì rải đều.

Nhận dạng đoạn trích ba giây dưới tám phép thử



Hình 15.13 — Nhận dạng một đoạn trích ba giây dưới tám phép thử

Truy vấn bằng đoạn trích ba giây từ giây thứ 2,5

Phép thử	Điểm khớp	Độ lệch trả về	Đúng chỗ
đoạn trích nguyên vẹn	8.243	234	có
MP3 64 kbit/giây	8.208	234	có
cộng nhiễu SNR 10 dB	7.676	234	có
cộng nhiễu SNR 0 dB	5.828	234	có
qua loa điện thoại rồi micrô	7.763	234	có
bản nhạc khác (đối chứng âm)	140	469	không

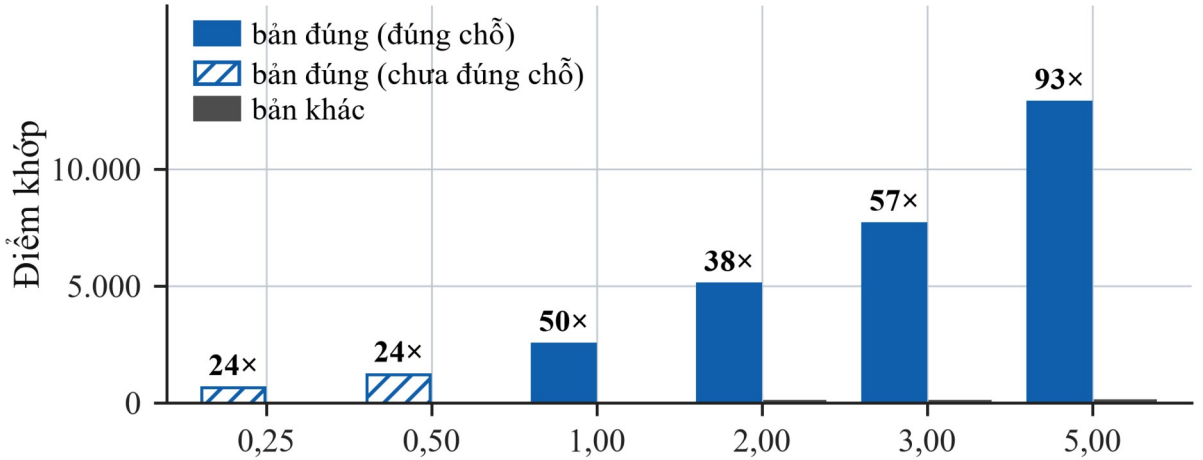
Rút gọn: lược MP3 128k, nhiễu SNR (Signal-to-Noise Ratio – tỉ số tín hiệu trên nhiễu) 20 dB và đổi tần số lấy mẫu.

Hai con số đáng nhớ

SNR 0 dB nghĩa là nhiễu mạnh bằng nhạc, vậy mà phép khớp vẫn cho 5.828 điểm, đúng chỗ. Bản nhạc khác chỉ cho 140 điểm: khoảng cách 42 lần, đủ rộng để đặt ngưỡng 1.000 điểm.

Đoạn trích ngắn tới đâu, kho lớn tới đâu

- Đoạn trích và kích thước kho là hai đại lượng cần cân.
- Bảng bên dưới cho số liệu theo độ dài và giá của kho.



(a) Từ 1 giây trở lên là khớp đúng chỗ

Đỉnh giữ lại mỗi giây	**471**
Khoá băm mỗi giây	**2.818**
Thời gian dựng mỗi giây nhạc	**8,0 ms**
Dung lượng một bài ba phút	**4,06 MB**
Kho cho một triệu bài	**khoảng 4,1 TB**

(b) Giá của một kho vân tay

Hình 15.14 — Độ dài đoạn trích cần thiết, và giá của một kho vân tay

Đoạn trích lẫn nhiễu SNR 10 dB, theo độ dài

Độ dài	Điểm bản đúng	Điểm bản khác	Tỉ số	Đúng chỗ
0,25 giây	646	27	24×	không
0,50 giây	1.214	51	24×	không
1,00 giây	2.517	50	50×	có
2,00 giây	5.089	133	38×	có
3,00 giây	7.676	134	57×	có
5,00 giây	12.870	139	93×	có

Dưới một giây: biết "bài nào" nhưng chưa biết "chỗ nào". Đủ để nhận dạng, chưa đủ để đối chiếu vi phạm.

Giá của một kho vân tay

Đại lượng	Giá trị
Đỉnh giữ lại mỗi giây	471
Khoá băm mỗi giây	2.818
Thời gian dựng mỗi giây nhạc	8,0 mili giây
Dung lượng một bài ba phút	4,06 MB
Kho cho một triệu bài	khoảng 4,1 TB

Bốn terabyte là một máy chủ, không phải một trung tâm dữ liệu; nên dựng kho cho sản phẩm của chính mình.

MỤC 15.5

Phát hiện giọng tổng hợp

- Không có đặc trưng của "AI" nói chung, chỉ có đặc trưng của một cách dựng sóng cụ thể.
- Dấu vết khai thác: độ nhọn của dạng sóng giọng người.
- Ngưỡng đặt bằng EER, kiểm trên người nói chưa gặp và sau nén.
- Kết thúc bằng chỗ máy dò mù hẳn.

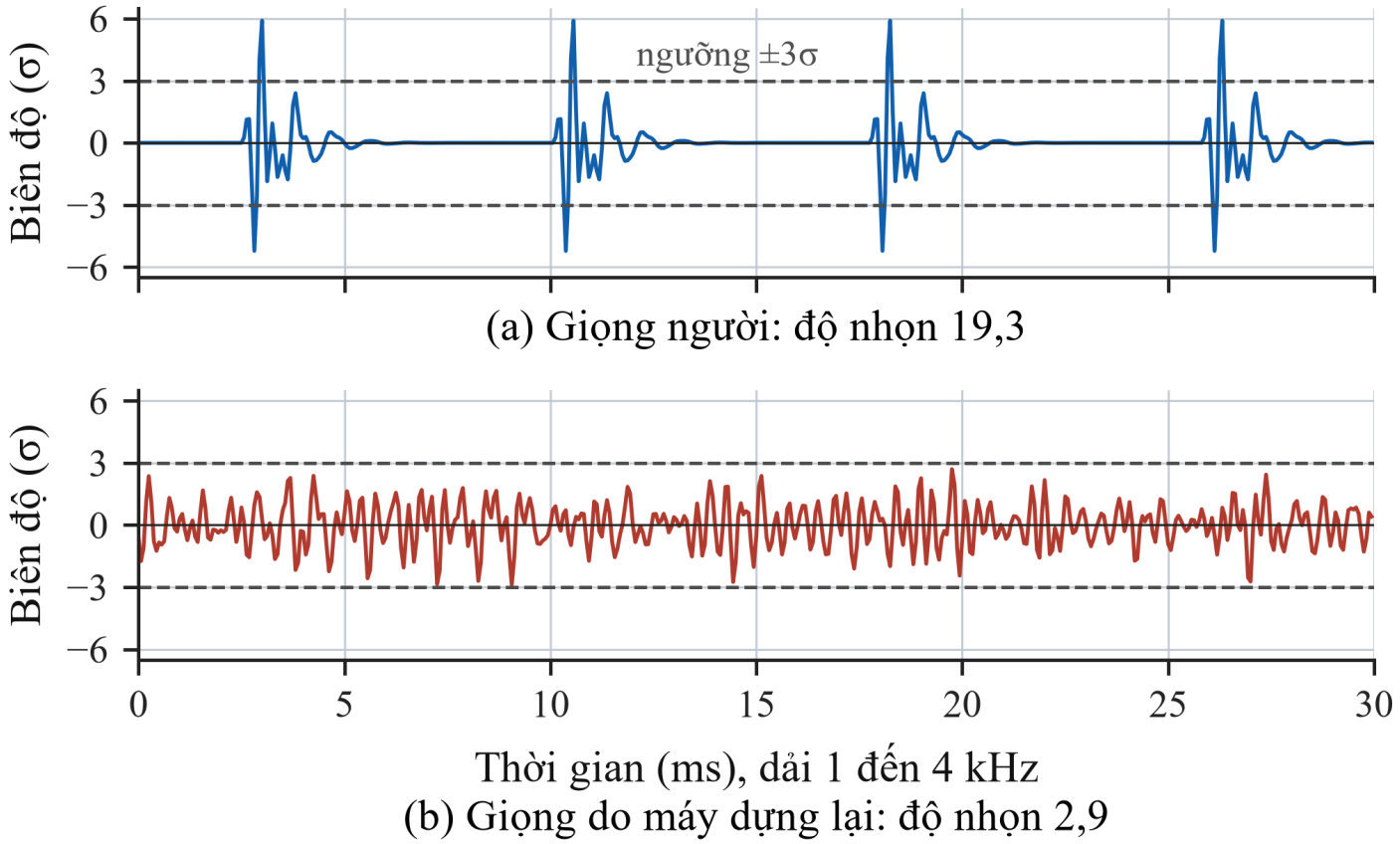
Ba cách đặt bài toán, hai cách sai

Sai: tìm một đặc trưng phân biệt AI với người, vì AI là rất nhiều kiến trúc. Sai: huấn luyện trên dữ liệu có nhãn, vì chỉ học được hệ thống có trong dữ liệu. Đúng: tìm dấu vết vật lý mà một lớp phương pháp buộc phải để lại, rồi đo nó.

Dạng sóng của giọng người rất nhọn

- Nguồn giọng người là chuỗi xung thanh môn; mọi hài khoá pha với các xung ấy.
- Dạng sóng trong dải trung vì thế có những gai rõ rệt.
- Bộ dựng sóng đoán pha làm sai sự khoá pha: hài vẫn đủ nhưng không cộng thành gai.
- Dạng sóng giống nhiễu Gauss.

Cùng một câu nói, cùng phổ mel, khác nhau ở pha



Hình 15.15 — Cùng một câu nói, cùng một phổ mel — khác nhau ở PHA

Độ nhọn: mô-men bậc bốn chuẩn hoá

$$\kappa = \frac{\mathbb{E}[(v-\mu)^4]}{\sigma^4}$$

(15.9)

- Phân phối Gauss có độ nhọn đúng bằng 3.
- Chuỗi xung có độ nhọn lớn hơn nhiều vì mô-men bậc bốn phạt rất mạnh giá trị xa trung bình.
- Đo theo từng khung, trong dải 1 đến 4 kHz.

Đo độ nhọn theo từng khung, dải 1 đến 4 kHz

```
def do_nhon_theo_khung(x, sr, f1=1000.0, f2=4000.0, n=2048,
                       hop=1024, nguong_rms=0.02):
    xb = loc_dai(x, f1, f2, sr)
    ra = []
    for i in range(0, x.size - n, hop):
        if np.sqrt(np.mean(x[i:i+n]**2)) < nguong_rms:
            continue                        # bo qua doan lang
        v = xb[i:i+n] - xb[i:i+n].mean()
        ra.append(float(np.mean((v / np.std(v))**4)))
    return np.array(ra)
```

Bốn đặc trưng đã thử trên cùng một cặp giọng

Đặc trưng	Giọng người	Giọng máy	AUC
Độ nhọn dải 1–4 kHz	18,79 ± 5,74	3,44 ± 1,56	1,000
Đỉnh trên hiệu dụng (dB)	16,19 ± 1,05	11,03 ± 1,45	0,981
Tỉ số hài trên nhiễu (dB)	6,07 ± 3,04	1,60 ± 2,55	0,876
Độ phẳng phổ 4–8 kHz	0,490 ± 0,130	0,484 ± 0,077	0,460

Độ phẳng phổ vùng cao vô dụng vì Griffin-Lim giữ đúng độ lớn phổ, chỉ làm sai pha.

AUC: xác suất xếp hạng đúng

$$\mathrm{HNR} = 10 \log_{10} \frac{r(T_0)}{1 - r(T_0)}$$

(15.10)

- AUC (Area Under the Curve – diện tích dưới đường cong ROC) là xác suất một khung giọng người có đặc trưng lớn hơn một khung giọng máy.
- Giá trị 0,5: đoán ngẫu nhiên.
- Giá trị 1,0: hai đống tách hẳn nhau.

Cách tính AUC

$$\mathrm{AUC} = \Pr\left(\kappa_{\text{thật}} > \kappa_{\text{máy}}\right)$$

(15.11)

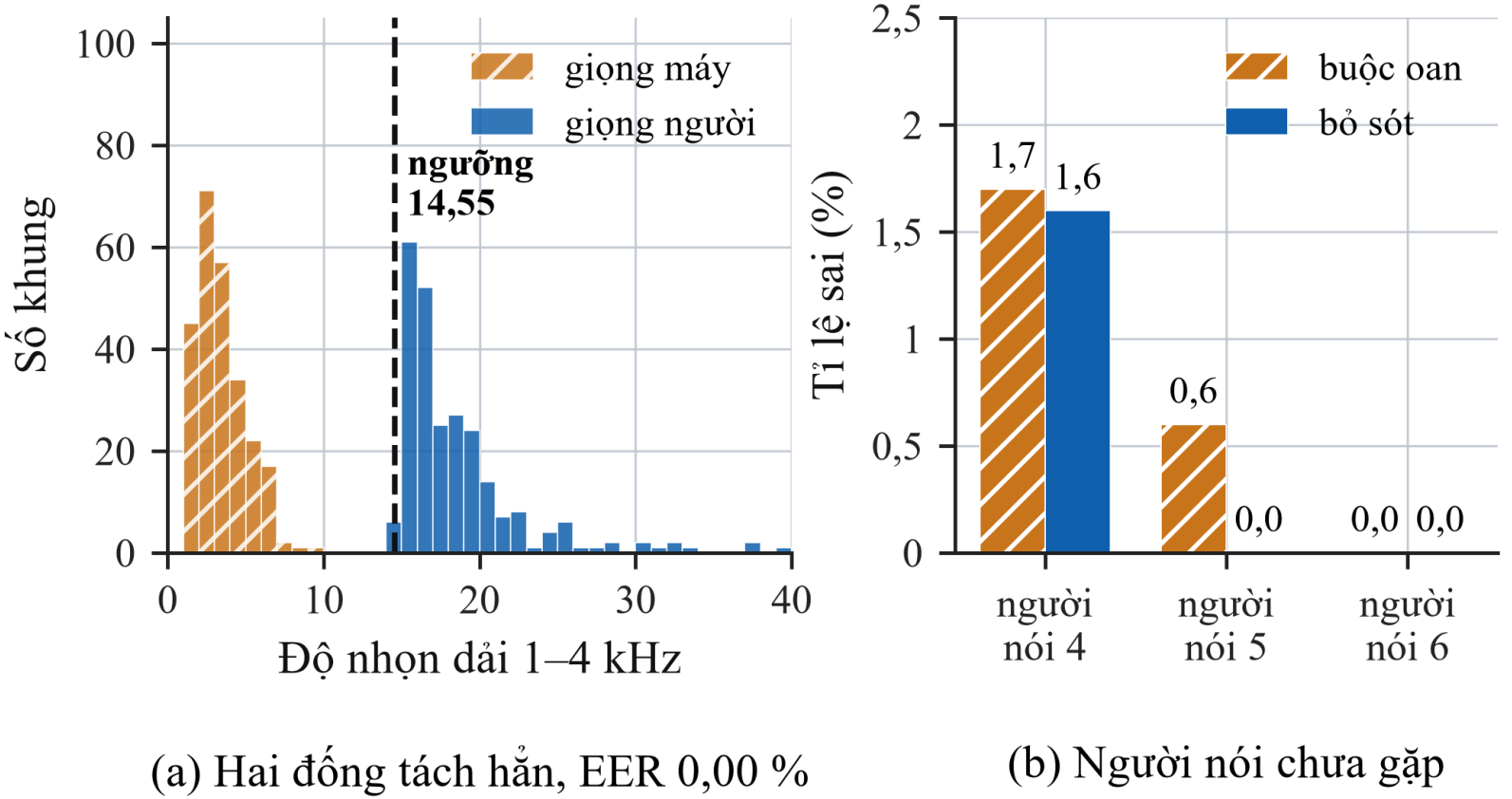
- Công thức đếm số cặp (người, máy) được xếp đúng thứ tự.
- Không phụ thuộc vào ngưỡng đã chọn.
- Dùng để so các đặc trưng với nhau trước khi đặt ngưỡng.

EER: tỉ lệ bỏ sót bằng tỉ lệ buộc oan

$$\text{EER}: \quad P_{\text{bỏsót}}(\theta) = P_{\text{buộcoan}}(\theta) \tag{15.12}$$

- Chưa biết giá của hai loại lỗi thì báo kết quả chuẩn bằng EER.
- Chọn ngưỡng sao cho tỉ lệ bỏ sót bằng tỉ lệ buộc oan.
- Báo chính tỉ lệ ấy; bài dùng ngưỡng 14,55.

Phân bố độ nhọn, ngưỡng và phép thử người nói mới



(a) Hai đống tách hẳn, EER 0,00 %

(b) Người nói chưa gặp

Hình 15.16 — Phân bố độ nhọn, ngưỡng, và phép thử trên người nói chưa gặp

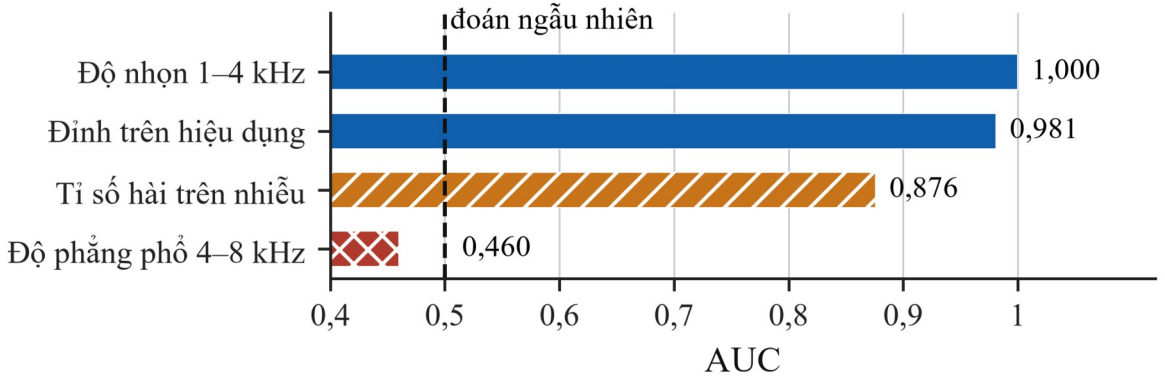
Ngưỡng đặt trên một người nói, dùng cho ba người khác

Người nói	Người (TB)	Máy (TB)	Buộc oan	Bỏ sót
hiệu chỉnh	18,79	3,44	0,0 %	0,0 %
người nói 4	18,65	3,56	1,7 %	1,6 %
người nói 5	18,69	3,50	0,6 %	0,0 %
người nói 6	18,67	3,49	0,0 %	0,0 %

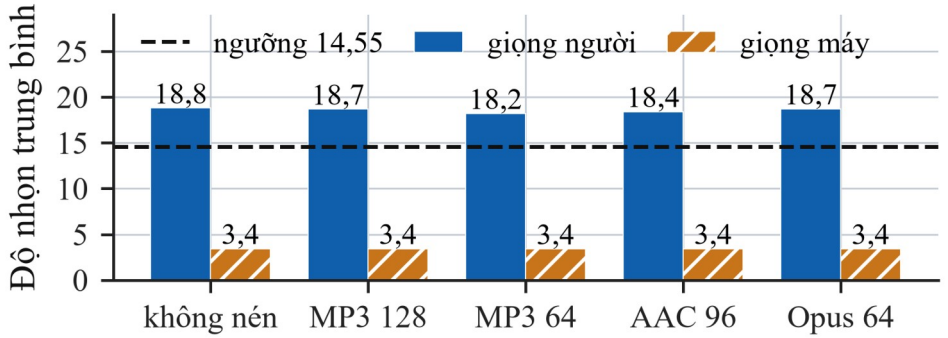
Giới hạn: ba người nói do cùng một mô-đun sinh ra; thử trên giọng thu thật là bài tập 5.

Bốn đặc trưng, và máy dò sau khi tệp đã qua nén

- Gần như mọi tệp đến tay người dùng đã qua nén ít nhất một lần.
- Nén không xoá được dấu vết: bộ mã hoá giữ nguyên các đỉnh mà tai nghe rõ.
- Chính những đỉnh ấy tạo nên các gai của dạng sóng.



(a) Bốn đặc trưng, một đặc trưng vượt trội



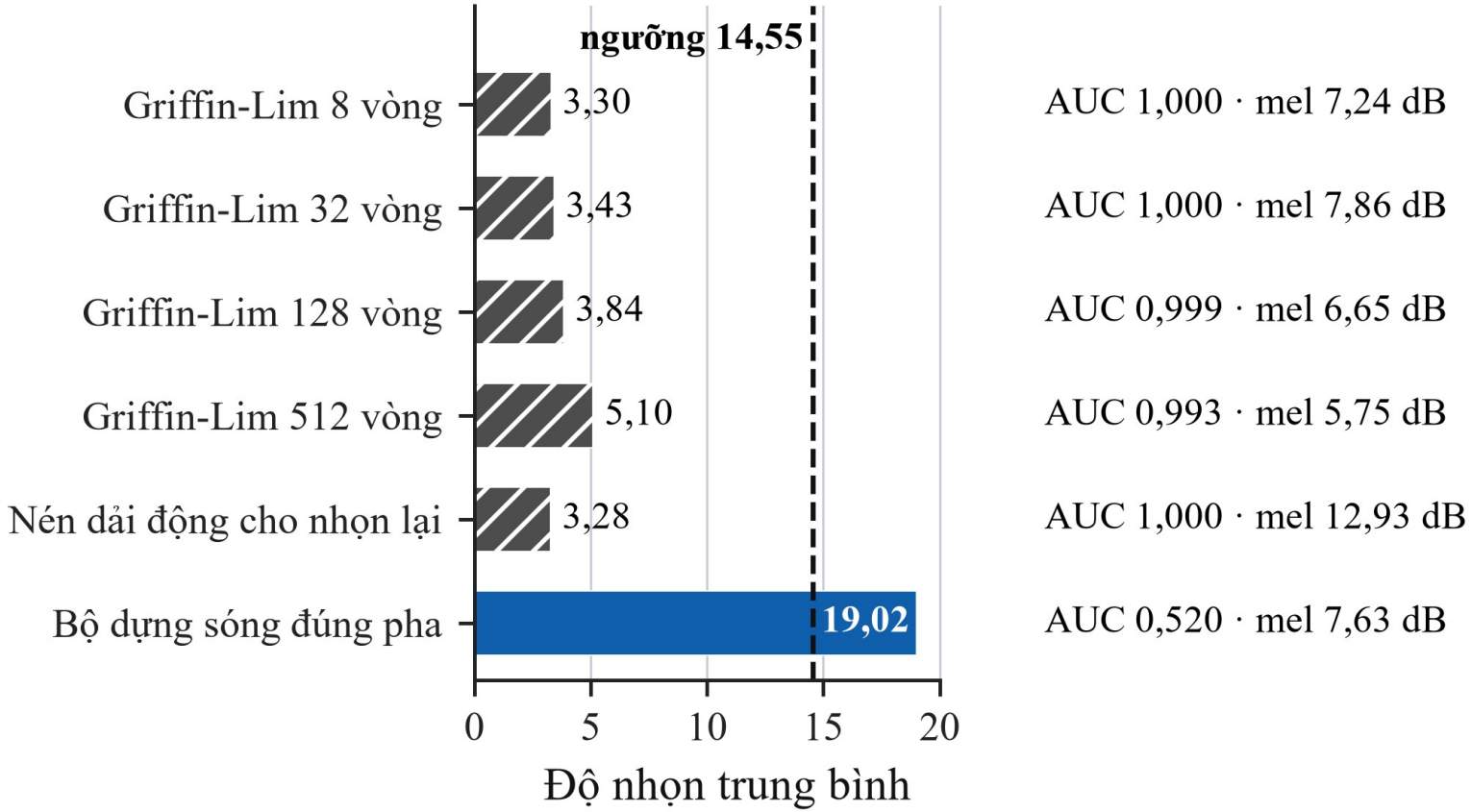
(b) Sau khi tệp qua nén: AUC vẫn bằng 1,000

Hình 15.17 — Bốn đặc trưng, và máy dò sau khi tệp đã qua nén

Độ nhọn sau các đường nén

Bản	Giọng người	Giọng máy	AUC
không nén	18,79	3,44	1,000
MP3 128 kbit/giây	18,66	3,43	1,000
MP3 64 kbit/giây	18,19	3,40	1,000
AAC 96 kbit/giây	18,40	3,43	1,000
Opus 64 kbit/giây	18,69	3,42	1,000

Sáu cách chữa: năm cách không qua, một cách qua ngay



Hình 15.18 — Sáu cách chữa, năm cách không qua được và một cách qua ngay

Sáu cách chữa và kết quả

Cách chữa	Độ nhọn	AUC	Khoảng cách mel	Qua ngưỡng
Griffin-Lim 8 vòng	3,30	1,000	7,24 dB	không
Griffin-Lim 32 vòng	3,43	1,000	7,86 dB	không
Griffin-Lim 128 vòng	3,84	0,999	6,65 dB	không
Griffin-Lim 512 vòng	5,10	0,993	5,75 dB	không
nén dải động	3,28	1,000	12,93 dB	không
bộ dựng sóng đúng pha	19,02	0,520	7,63 dB	Qua

Máy dò đo pha, không đo "AI"

Năm cách đầu chữa triệu chứng và đều không qua; 512 vòng chỉ đạt độ nhọn 5,10, còn xa ngưỡng 14,55. Bộ dựng sóng đúng pha có độ nhọn 19,02 và AUC 0,520: máy dò đoán ngẫu nhiên. Đây là cuộc chạy đua, không phải bài toán đã giải xong.

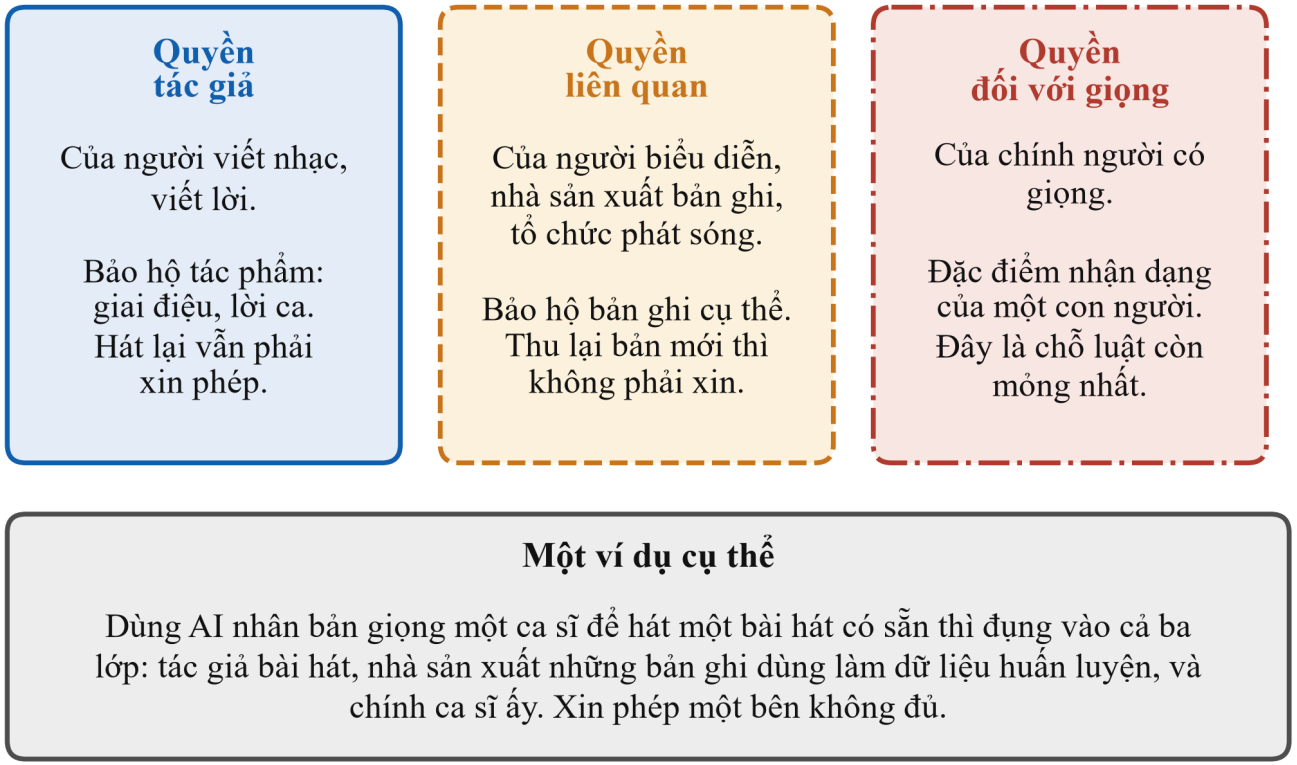
MỤC 15.6

Quyền, đạo đức, và chỗ kỹ thuật dừng lại

- Một bản ghi mang ít nhất ba lớp quyền của ba người khác nhau.
- Năm câu hỏi phải trả lời trước khi phát hành bản có giọng do AI sinh.
- Mục duy nhất trong học phần không có con số nào.

Ba loại quyền trên cùng một bản ghi

- Quyền tác giả: người viết nhạc, viết lời.
- Quyền liên quan: người biểu diễn, nhà sản xuất bản ghi, tổ chức phát sóng.
- Quyền đối với giọng: chính người có giọng.
- Xin phép một bên không đủ.



Hình 15.19 — Ba loại quyền trên cùng một bản ghi

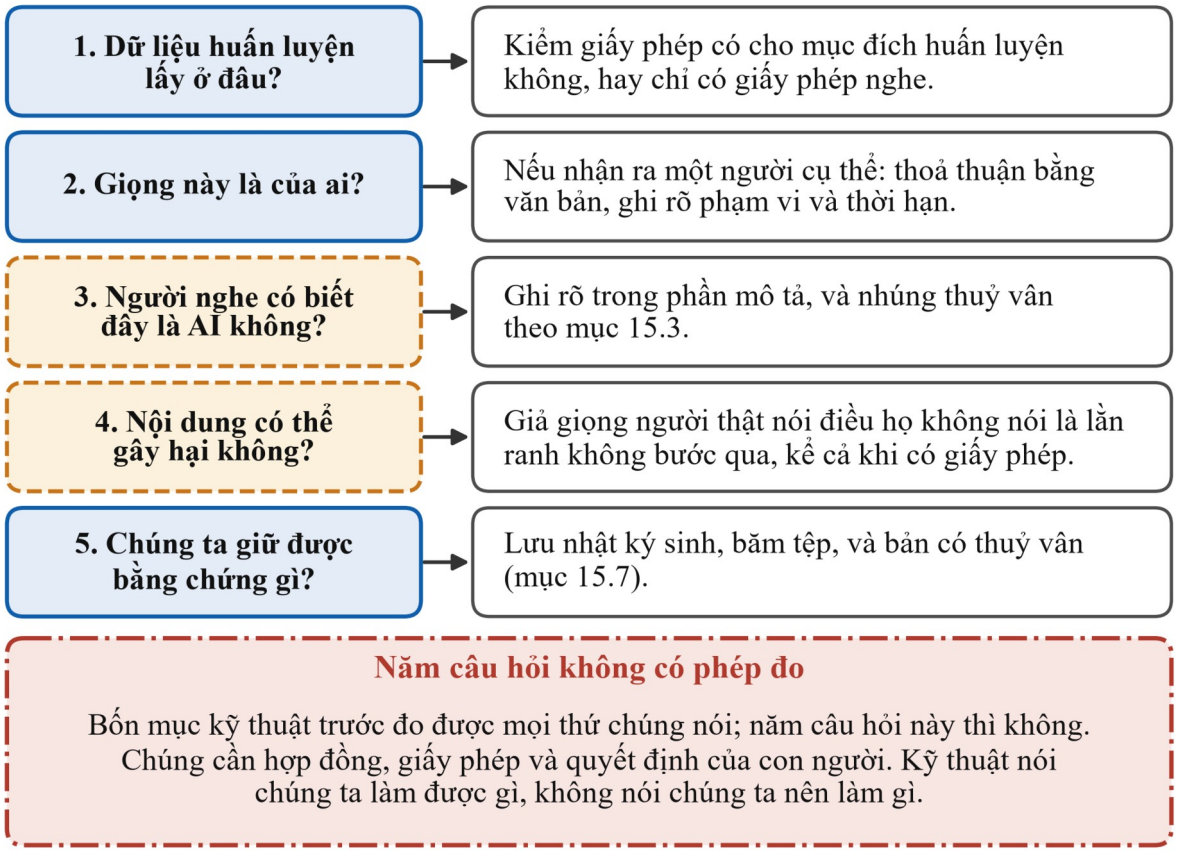
Ba lớp quyền và điều mỗi lớp bảo hộ

Lớp	Của ai	Bảo hộ cái gì	Hát hoặc thu lại
Quyền tác giả	Người viết nhạc, lời	Giai điệu, lời ca	Vẫn phải xin phép
Quyền liên quan	Người biểu diễn, nhà sản xuất	Bản ghi cụ thể	Thu bản mới thì không phải xin
Quyền đối với giọng	Chính người có giọng	Đặc điểm nhận dạng một người	Luật còn mỏng nhất

Ở Việt Nam, hai lớp đầu nằm trong Luật Sở hữu trí tuệ; cần tra văn bản hiện hành vì phần này thay đổi nhanh.

Năm câu hỏi trước khi phát hành

- Dữ liệu huấn luyện lấy ở đâu.
- Giọng này là của ai.
- Người nghe có biết đây là AI không.
- Nội dung có thể gây hại không.
- Giữ được bằng chứng gì.



Hình 15.20 — Năm câu hỏi phải trả lời trước khi phát hành một bản có giọng do AI sinh

Năm câu hỏi và điều phải làm

Câu hỏi	Phải làm gì
Dữ liệu huấn luyện lấy ở đâu	Kiểm giấy phép có cho mục đích huấn luyện không
Giọng này là của ai	Có thoả thuận bằng văn bản, ghi rõ phạm vi và thời hạn
Người nghe có biết là AI không	Ghi rõ trong mô tả; nhúng thuỷ vân theo mục 15.3
Nội dung có gây hại không	Không giả giọng người thật nói điều họ không nói
Giữ được bằng chứng gì	Lưu nhật ký sinh, mã băm tệp và bản có thuỷ vân

Chỗ kỹ thuật dừng lại

Không có phép đo nào trả lời được năm câu hỏi này. Kỹ thuật nói làm được gì, không nói nên làm gì. Ranh giới ấy không do luật vẽ ra, vì luật luôn đi sau, mà do chính người làm nghề tự giữ.

MỤC 15.7

Bảo vệ sản phẩm của chính người học

- Ghép ba công cụ: thuỷ vân, vân tay, mã băm SHA-256.
- Đặt thuỷ vân đúng chỗ trong đường ống chương 14.
- Bảo vệ không được làm sai lệch bản giao.

Ba công cụ dưới bảy phép thử

Phép thử	Thuỷ vân đọc đúng	Sau khi đồng bộ	Điểm vân tay	SHA-256
không đụng gì	100,0 %	100,0 %	22.030	khớp
đổi 1 bit cuối của 1 mẫu	100,0 %	100,0 %	22.030	sai
giảm mức 6 dB	100,0 %	100,0 %	22.030	sai
MP3 128 kbit/giây	100,0 %	100,0 %	22.075	sai
MP3 64 kbit/giây	100,0 %	100,0 %	22.010	sai
đổi 48 → 44,1 → 48 kHz	100,0 %	100,0 %	22.030	sai
trích 3 giây ở giữa	60,9 %	100,0 %	8.091	sai

SHA-256 chỉ trả lời được câu "còn nguyên vẹn từng bit không"; với âm thanh, câu ấy gần như luôn là không.

Ba công cụ không thay được nhau: chúng trả lời ba câu hỏi khác nhau, và một bản giao nghiêm túc cần cả ba.

Hình 15.21 — Ba công cụ dưới bảy phép thử

Ba công cụ trên một bản nhạc đã nhúng thuỷ vân

Phép thử	Thuỷ vân đúng	Sau đồng bộ lại	Điểm vân tay	SHA-256
không đụng gì	100,0 %	100,0 %	22.030	khớp
đổi 1 bit cuối của 1 mẫu	100,0 %	100,0 %	22.030	Hỏng
MP3 64 kbit/giây	100,0 %	100,0 %	22.010	Hỏng
đổi 48 → 44,1 → 48 kHz	100,0 %	100,0 %	22.030	Hỏng
trích 3 giây ở giữa	60,9 %	100,0 %	8.091	Hỏng

Rút gọn: lược giảm mức 6 dB và MP3 128k (kết quả như dòng liền kề).

Không công cụ nào thay được công cụ nào

- SHA-256 chỉ trả lời "còn nguyên từng bit không"; là công cụ duy nhất chứng minh tệp đúng là tệp đã nộp.
- Thuỷ vân trả lời "bản này cấp cho ai", đọc đúng 100 % qua mọi phép thử sau khi đồng bộ lại.
- Vân tay trả lời "bản lạ kia có phải của mình"; còn dùng được khi bản đã phát hành mà bỏ sót thuỷ vân.

Xác suất trùng mã băm

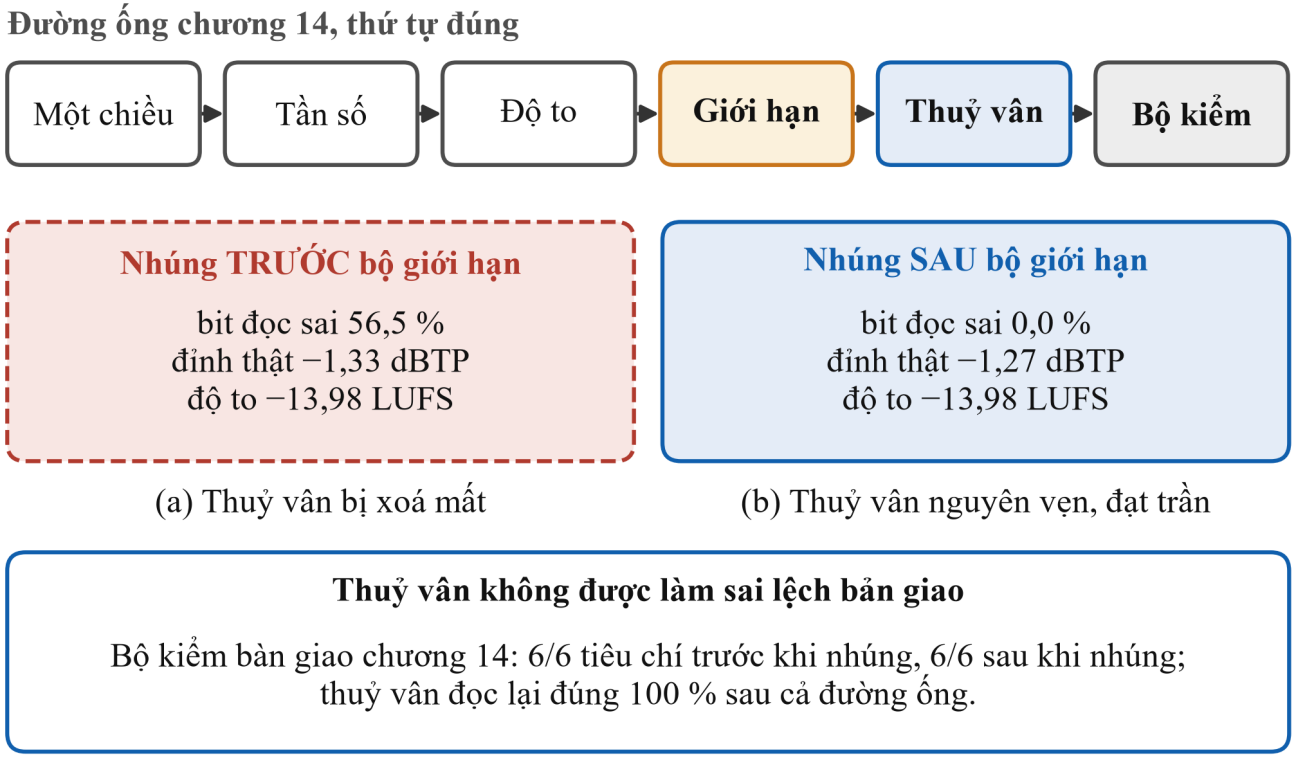
$$N_{\text{vachạm}} \approx 1{,}177\sqrt{2^{b}}$$

(15.13)

- Hai tệp khác nhau chỉ có xác suất trùng mã băm đáng kể sau khoảng 2 mũ 128 tệp.
- Con số ấy lớn hơn mọi thứ đếm được.
- Trên thực tế mã băm là định danh tuyệt đối.

Thuỷ vân trước hay sau bộ giới hạn

- Chương 14 chứng minh thứ tự các khối quyết định kết quả.
- Thuỷ vân được thêm vào đúng đường ống ấy.
- Bộ kiểm bàn giao chạy sau cùng.



Hình 15.22 — Thuỷ vân trước hay sau bộ giới hạn, và bộ kiểm bàn giao

Hai vị trí, hai kết quả

Vị trí	Bit đọc sai	Đỉnh thật	Độ to
nhúng trước bộ giới hạn	56,5 %	−1,33 dBTP	−13,98 LUFS
nhúng sau bộ giới hạn	0,0 %	−1,27 dBTP	−13,98 LUFS

Bộ giới hạn hạ gain không đều theo thời gian nên phá tương quan. Nếu đỉnh thật vượt trần, hạ trần bộ giới hạn xuống −1,5 dBTP (decibel đỉnh thật, True Peak) trước, không hạ mức thuỷ vân.

Bộ kiểm bàn giao chương 14, trước và sau khi nhúng

Tiêu chí	Trước khi nhúng	Sau khi nhúng
Số tiêu chí đạt	6 / 6	6 / 6
Thuỷ vân đọc lại	—	100,0 %

Điều kiện bắt buộc: bảo vệ không được làm sai lệch bản giao.

Tổng thời gian máy của chuỗi bảo vệ

$$C_{\text{bảovệ}} = \sum_{i} c_i \quad [\text{mschomỗigiây}]$$

(15.14)

- Tổng thời gian là tổng ba công cụ, tính trên mỗi giây âm thanh.
- Thuỷ vân 1,7 ms; kho vân tay 9,3 ms; SHA-256 0,5 ms; tổng 11,5 ms.
- Bảo vệ một bản ba phút tốn khoảng hai giây máy.

Đưa cả ba vào đường ống

So với 85 ms mỗi giây của cả đường ống hậu kỳ ở mục 14.6.3, bảo vệ là phần rẻ nhất. Đặt ngay sau bộ giới hạn và trước bộ kiểm để chạy tự động, không phụ thuộc trí nhớ của ai; bỏ sót thuỷ vân thì không sửa được về sau.

MỤC 15.8

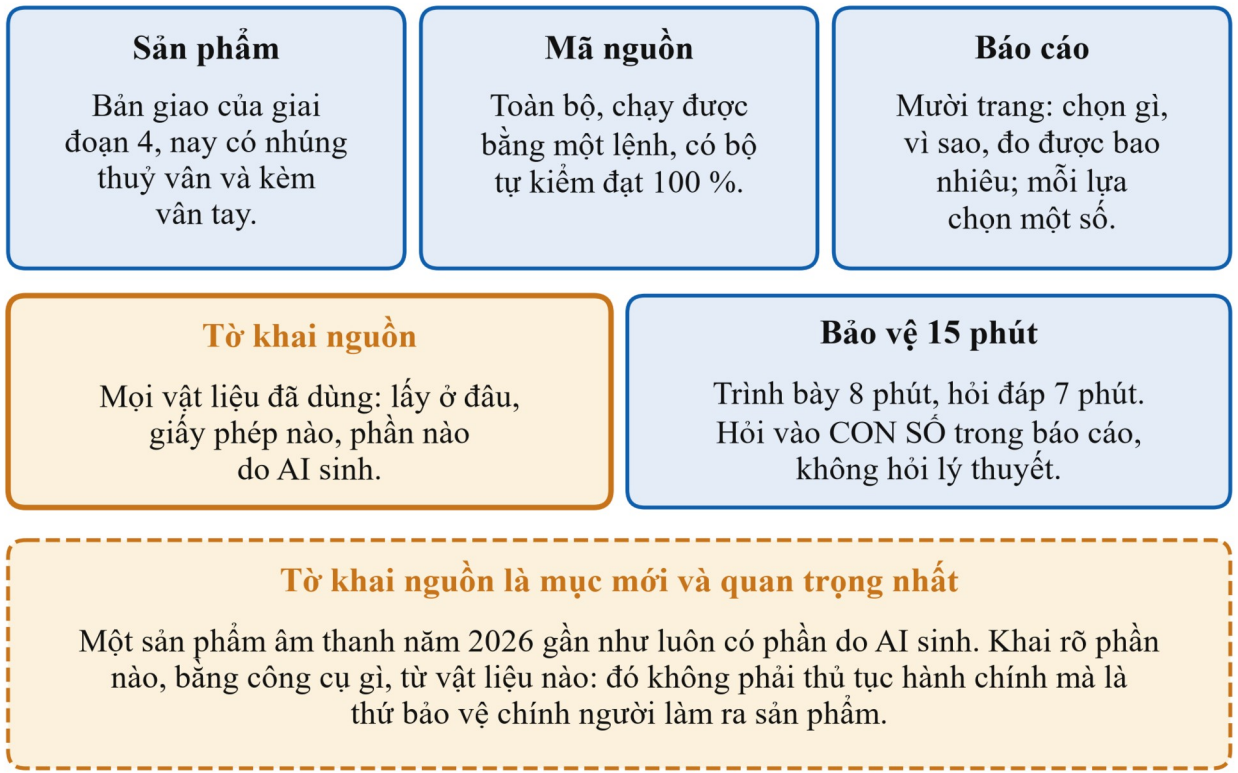
Bài tập lớn: giai đoạn 5, nộp và bảo vệ

- Giai đoạn 5 gộp lại bốn giai đoạn trước, không yêu cầu làm lại.
- Hai việc: bọc sản phẩm bằng lớp bảo vệ của mục 15.7, và trình bày mọi lựa chọn bằng số đo.

Năm giai đoạn của bài tập lớn

Giai đoạn	Mốc	Chương	Trạng thái
1	Sinh và đo tín hiệu	Chương 4	Đã nộp
2	Chuỗi hiệu ứng có đo	Chương 9	Đã nộp
3	Bản trộn chạy trên trình duyệt	Chương 12	Đã nộp
4	Bản giao đạt bộ kiểm	Chương 14	Đã nộp
5	Bảo vệ và tài liệu hoàn chỉnh	Chương 15	Nộp cuối kỳ

Năm thứ phải nộp ở giai đoạn 5



Hình 15.23 — Năm thứ phải nộp ở giai đoạn 5

Năm thứ phải nộp và tiêu chuẩn

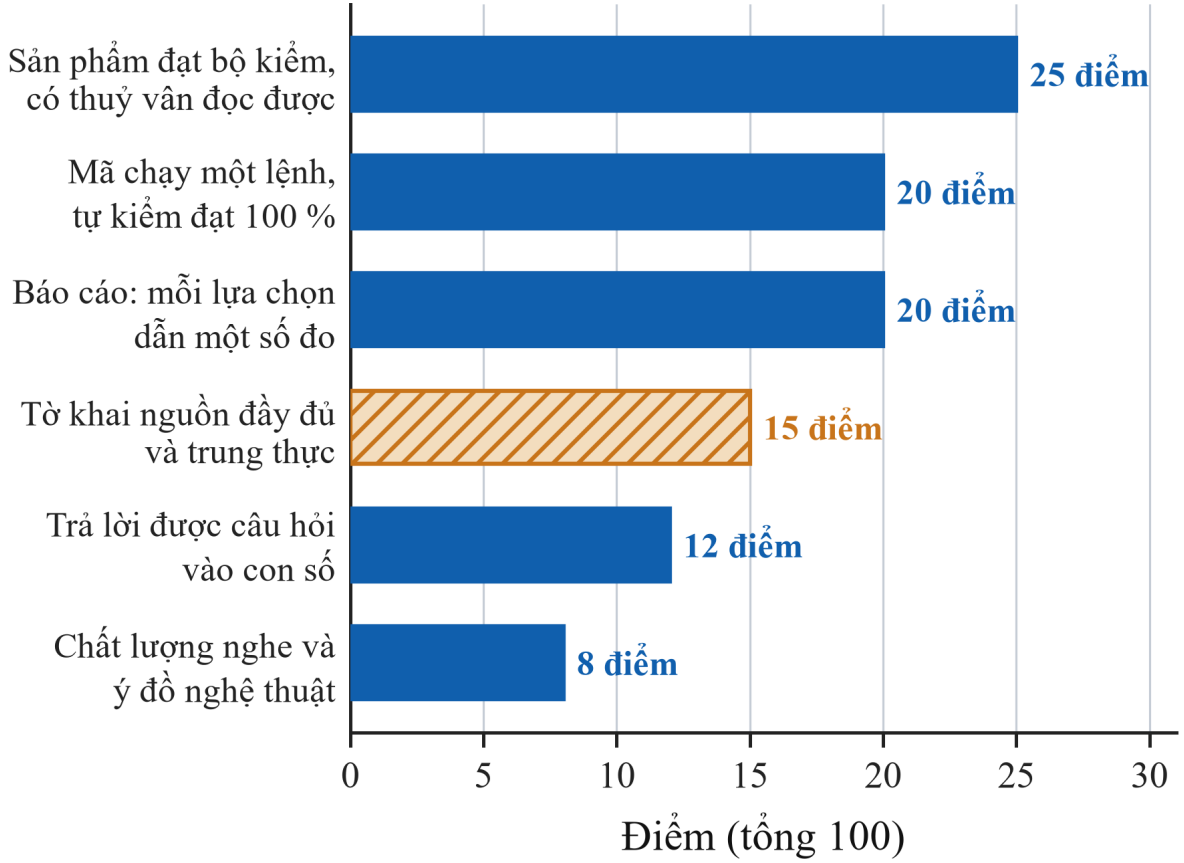
Nộp gì	Tiêu chuẩn
Sản phẩm	Bản giao giai đoạn 4 có thuỷ vân, kho vân tay và mã băm SHA-256
Mã nguồn	Chạy một lệnh từ vật liệu nguồn tới tệp giao; tự kiểm đạt 100 %
Báo cáo	Mười trang; mỗi lựa chọn tham số dẫn ít nhất một số đo
Tờ khai nguồn	Mọi vật liệu: nguồn, giấy phép, phần do AI sinh và công cụ
Bảo vệ 15 phút	Trình bày 8 phút, hỏi đáp 7 phút, hỏi vào con số trong báo cáo

Mục mới và quan trọng nhất

Sản phẩm năm 2026 gần như luôn có phần do AI sinh. Khai rõ phần nào, công cụ nào, vật liệu nào là thứ bảo vệ chính người làm ra sản phẩm khi có tranh chấp. Đây là chỗ sự trung thực không kiểm được bằng máy: khai thiếu một vật liệu bị trừ hết mười lăm điểm.

Thang chấm giai đoạn 5, tổng 100 điểm

- Tám mươi tám điểm đầu không dựa vào tai người chấm.
- Phần nghe chỉ có ý nghĩa khi phần kỹ thuật đã sạch.



Hình 15.24 — Thang chấm giai đoạn 5, tổng 100 điểm

Thang chấm chi tiết

Mục	Điểm	Chấm bằng gì
Sản phẩm đạt bộ kiểm, thuỷ vân đọc được	25	Bộ kiểm chương 14 và hàm đọc thuỷ vân
Mã chạy một lệnh, tự kiểm đạt 100 %	20	Chạy hai lần từ thư mục sạch, so hai kết quả
Báo cáo: mỗi lựa chọn dẫn một số đo	20	Đếm lựa chọn không có số đi kèm
Tờ khai nguồn đầy đủ, trung thực	15	Đối chiếu với mã nguồn và vật liệu nộp kèm
Trả lời được câu hỏi vào con số	12	Hỏi thẳng vào một số trong báo cáo
Chất lượng nghe, ý đồ nghệ thuật	8	Nghe

Bảy câu hỏi được công khai từ trước

- Tần số lấy mẫu của mô hình nguồn là bao nhiêu, và nhóm biết bằng cách nào?
- Số cạnh của bộ lọc đổi tần số là bao nhiêu, và vì sao lại là con số ấy?
- Hệ số tương quan tại mối nối dài nhất là bao nhiêu, và luật mờ chéo nào được chọn theo đó?
- Đỉnh thật của bản giao là bao nhiêu, đo ở bước nào của đường ống?
- Mức nhúng thuỷ vân là bao nhiêu dB, và đã thử tấn công nào?
- Chạy lại mã hai lần có cho tệp giống hệt không, kiểm bằng lệnh gì?
- Phần nào do AI sinh, và dựa vào giấy phép nào?

Khác nhau ở chỗ có dẫn số đo hay không

Sai: "chúng em chọn c bằng 16 vì đó là giá trị thường dùng". Đúng: "chọn c bằng 16 vì với c bằng 8 dải giữ nguyên vẹn chỉ còn 76,7 phần trăm Nyquist, còn với c bằng 32 độ trễ lên 99,5 ms mà độ triệt chỉ hơn 1 dB", kèm vị trí bảng đo trong báo cáo.

Ba lỗi hay gặp nhất ở buổi bảo vệ

Lỗi	Vì sao xảy ra	Cách tránh
Không nhớ số của nhóm	Một người viết, người khác trình bày	Cả nhóm cùng chạy lại mã trước buổi bảo vệ
Trả lời bằng lý thuyết chung	Đọc tài liệu nhưng không tự đo	Mỗi khẳng định có một dòng lệnh sinh ra nó
Không giải thích được chỗ không đo	Ngại nói giới hạn của nhóm	Nói rõ giới hạn là điểm cộng

MỤC 15.9

Tổng kết chương và tổng kết học phần

- Ba bài thực hành và bảy bài tập củng cố bốn mục kỹ thuật.
- Sáu con số gói lại chương; năm thói quen mang ra khỏi học phần.
- Năm hướng đi tiếp sau học phần.

Ba bài thực hành

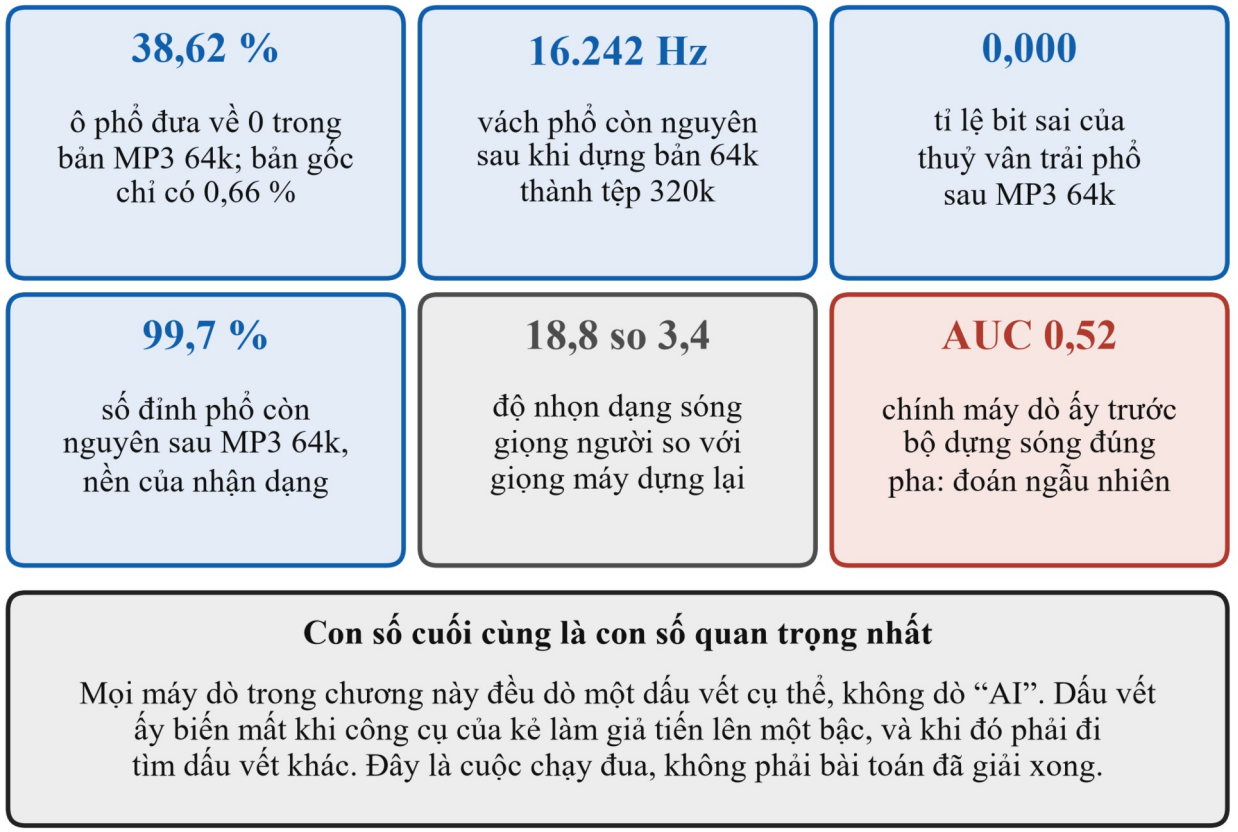
- Thực hành 1, dấu vết codec: nén ở bốn bitrate, đo vách phổ và ô rỗng, rồi đoán bitrate của tệp lạ.
- Thực hành 2, thuỷ vân: nhúng trải phổ ở −26 dB, tự nghĩ ba phép tấn công mới và đo tỉ lệ bit sai.
- Thực hành 3, vân tay: dựng kho cho một đoạn nhạc, truy vấn bằng đoạn phát qua loa và thu bằng micrô.

Bảy bài tập

- Bài 1: điều tra ba tệp lạ bằng vách phổ, ô rỗng và tương quan hai kênh.
- Bài 2: dựng lại phép thử chuyển mã 64k rồi 320k, chứng minh bằng hai con số.
- Bài 3: thêm mẫu đồng bộ cho trải phổ, đo lại hàng cắt 1.000 mẫu, nêu cái giá về dung lượng.
- Bài 4: dựng kho vân tay hai mươi bài, đo tỉ lệ nhận đúng.
- Bài 5: dựng lại bảng mục 15.5.3 trên giọng thu thật; ngưỡng 14,55 còn dùng được không.
- Bài 6: nghĩ ra đặc trưng thứ năm, đo AUC, nêu dấu vết vật lý nó nhìn vào.
- Bài 7: viết tờ khai nguồn theo năm câu hỏi ở mục 15.6.2; bài không có phép đo.

Chương 15 gói lại trong sáu con số

- Sáu con số là sáu bằng chứng, mỗi con số gắn với một mục kỹ thuật.
- Bảng tiếp theo nêu ý nghĩa của từng con số.



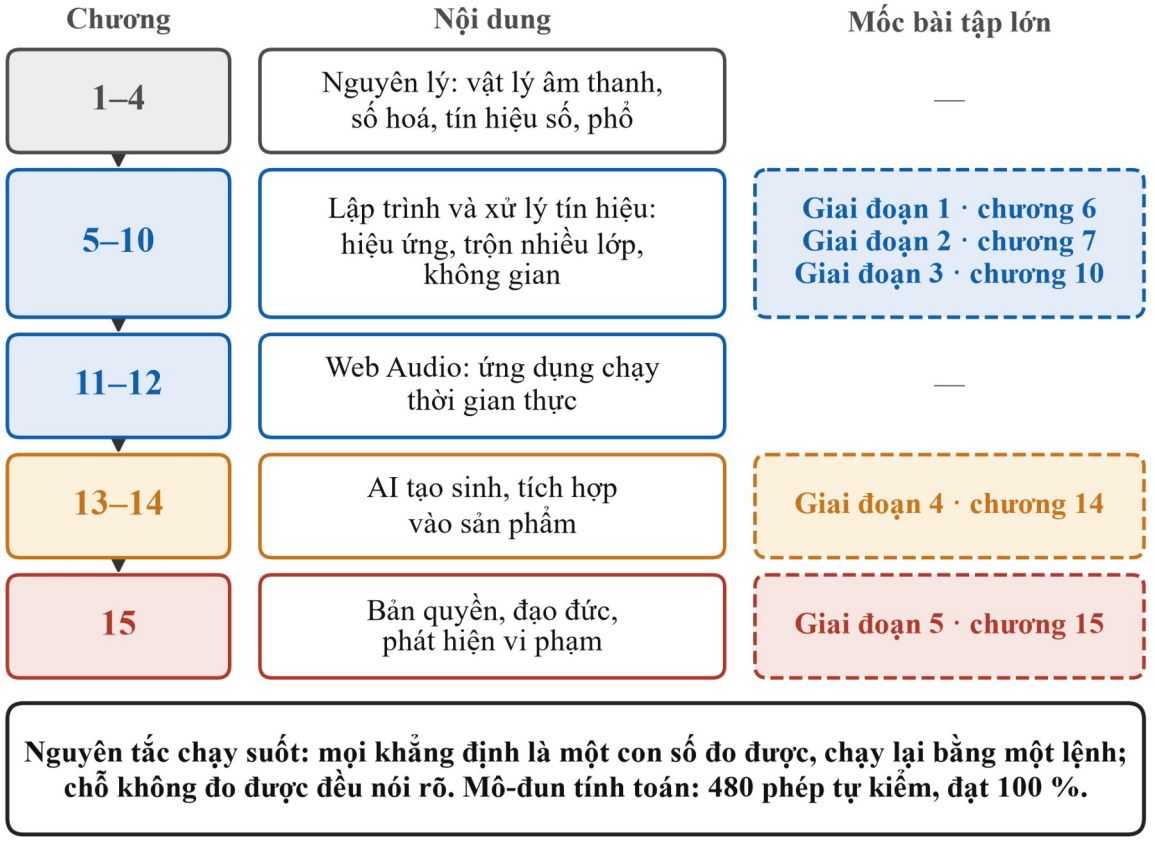
Hình 15.25 — Chương 15 gói lại trong sáu con số

Sáu con số và ý nghĩa

Con số	Của cái gì	Nói lên điều gì
38,62 %	ô phổ rỗng, MP3 64k	Bản gốc chỉ 0,66 %: dấu vết nén không giấu được
16.242 Hz	vách phổ của bản 64k dựng thành 320k	Bitrate ghi trong tệp không nói gì về chất lượng
0,000	bit sai của trải phổ sau MP3 64k	Giấu đúng chỗ codec không dám đụng thì sống
99,7 %	đỉnh phổ còn nguyên sau MP3 64k	Nền của mọi phép nhận dạng bản ghi
18,8 so 3,4	độ nhọn giọng người so giọng máy	AUC 1,000, sống qua mọi phép nén
AUC 0,52	máy dò trước bộ dựng sóng đúng pha	Dò dấu vết, không dò "AI"

Mười lăm chương trên một đường đi

- Đi từ vật lý của sóng âm tới câu hỏi một bản ghi có được phép tồn tại không.
- Một nguyên tắc chạy suốt: mọi khẳng định là một con số đo được, chạy lại được bằng một lệnh.
- Chỗ không đo được thì nói rõ là không đo được.
- Mô-đun tính toán khép lại với 480 phép tự kiểm, đạt 100 %.



Hình 15.26 — Mười lăm chương trên một đường đi

Năm thói quen đáng mang ra khỏi học phần

Thói quen	Học ở chương nào
Đo trước rồi mới kết luận, và đo ở cuối đường xử lý	B7 · 14.4.5 · 14.6
Tính trần trên của bài toán trước khi giải	13.7 · 15.4.2
Hỏi thước đo đang phạt gì trước khi tin một con số kém	13.8 · 14.6.2 · 15.3.4
Dựng phép đo có kiểm soát, lặp lại, lấy trung vị	14.5.3
Nói rõ chỗ không đo được thay vì lặng lẽ bỏ qua	13.5.3 · 15.2.4 · 15.5.5

Năm hướng đi tiếp

Hướng	Bắt đầu từ	Việc làm được sau đó
Xử lý tín hiệu thời gian thực	B11, B12	Viết plugin, thiết bị nhúng, hệ thống biểu diễn
Học máy cho âm thanh	B13	Huấn luyện mô hình sinh hoặc nhận dạng tiếng Việt
Hậu kỳ và phát hành	B10, B14	Kỹ thuật viên hậu kỳ, đường ống phát hành tự động
Pháp y âm thanh	B15	Giám sát bản quyền, hỗ trợ giám định
Âm học và không gian	B2, B10	Thiết kế phòng thu, âm thanh cho phim

Ba việc trong tháng đầu sau học phần

Giữ mô-đun tính toán và bộ tự kiểm, bổ sung mỗi khi gặp phép đo mới. Chạy bộ kiểm bàn giao chương 14 trên ba sản phẩm cũ của mình. Dựng kho vân tay cho mọi thứ đã phát hành: mất vài giờ máy, làm một lần.

Những điều cần nhớ

1. Dấu vết codec (vách phổ, ô rỗng, stereo, tiền vọng) cho biết tệp đã đi qua đâu mà không cần bản gốc.
2. Bitrate ghi trong tệp không nói về chất lượng thật; hai con số phổ nói sự thật trong vài giây.
3. Thuỷ vân trải phổ ở −26 dB, khối 8.192 mẫu bền qua nén; kẻ thù thật của nó là mất đồng bộ.
4. Vân tay dựng trên đỉnh phổ và khoá băm từng cặp, nhận ra đoạn ba giây cả ở SNR 0 dB.
5. Máy dò giọng tổng hợp đo dấu vết pha, không đo "AI", và mù trước bộ dựng sóng đúng pha.
6. Kỹ thuật nói làm được gì, không nói nên làm gì: ba lớp quyền và năm câu hỏi thuộc về người làm nghề.