MỤC LỤC

Chương 7. Tăng cường chất lượng 5
7.1 Vị trí của chương trong cuốn sách 5
7.2 Ba thước đo mức 8
7.3 Chuẩn hoá theo độ to 13
7.4 Nhiễu và cách đo nó 17
7.5 Khử nhiễu bằng trừ phổ 20
7.6 Nhiễu nhạc — cái giá của trừ phổ 24
7.7 Dự án thực hành — giai đoạn 2 30
7.8 Thực hành, bài tập và tổng kết 32

DANH MỤC HÌNH VẼ

Hình 7.1 Năm mục kỹ thuật của chương 7 6
Hình 7.2 Ba tệp cùng đỉnh −1 dBFS, đo bằng ba thước khác nhau 6
Hình 7.3 Ba thước đo mức và việc của mỗi thước 8
Hình 7.4 Đáp ứng của thước K, và hệ quả khi đo 9
Hình 7.5 Ba đường độ to trên cùng một đoạn nhạc cổ điển 10
Hình 7.6 Cùng một giọng đọc, thêm 24 giây lặng, tính ba cách 11
Hình 7.7 Cùng một tín hiệu, mono và stereo, hai con số LUFS 12
Hình 7.8 Đỉnh mẫu và true peak trên sáu loại tư liệu 13
Hình 7.9 Mức đích và trần của sáu nơi giao nộp 14
Hình 7.10 Bốn bước chuẩn hoá theo độ to 15
Hình 7.11 Một bản nhạc cổ điển đưa về −14 LUFS, có và không có trần 15
Hình 7.12 Bốn tình huống và thước đo tương ứng 16
Hình 7.13 Cùng một giọng đọc ở hai mức nhiễu, nhìn hai cách 17
Hình 7.14 Hai cách đo SNR, và cách đọc con số ấy 18
Hình 7.15 Phổ nhiễu ước lượng từ đoạn lặng, so với phổ nhiễu thật 19
Hình 7.16 Bốn bước của phép trừ phổ 20
Hình 7.17 Một khung phổ, ba mức trừ, và chỗ nhiễu nhạc sinh ra 21
Hình 7.18 Toàn bộ thuật toán, và cách dùng nó 22
Hình 7.19 Bản sạch, bản có nhiễu, và bản đã trừ phổ 23
Hình 7.20 Vai trò của alpha và của beta 23
Hình 7.21 Nửa giây lặng, nhìn bằng phổ đồ, trước và sau khi khử 25
Hình 7.22 Mức nền, chất lượng tiếng và nhiễu nhạc theo alpha 26
Hình 7.23 Ba cách giảm nhiễu nhạc, và cách thứ tư 27
Hình 7.24 Ba cách khử nhiễu, đo bằng ba đại lượng 28
Hình 7.25 Bốn vùng SNR và quyết định tương ứng 29
Hình 7.26 Thứ tự đúng, và cái giá của làm sai thứ tự 30
Hình 7.27 Năm bước của giai đoạn 2 30
Hình 7.28 Sáu câu tổng kết chương 7 34

DANH MỤC BẢNG BIỂU

Bảng 7.1 Ba thứ gần giống mà thuộc chương khác....7
Bảng 7.2 Ba thang thời gian của phép đo độ to....10
Bảng 7.3 Đọc con số LRA....10
Bảng 7.4 Sáu chuẩn thường gặp....14
Bảng 7.5 Ngưỡng SNR và ý nghĩa thực tế....19
Bảng 7.6 Chọn hai tham số theo tình huống....24
Bảng 7.7 Ba đại lượng theo hệ số trừ vượt alpha....26
Bảng 7.8 Ba cách khử trên cùng một bản thu SNR 10 dB....28
Bảng 7.9 Kết quả cần có....31
Bảng 7.10 Bốn tiêu chí tự kiểm giai đoạn 2....31
Bảng 7.11 Bài tập của chương 7....33

CHƯƠNG 7
TĂNG CƯỜNG CHẤT LƯỢNG

Mục tiêu của chương

Sau chương này, người đọc phải làm được những việc sau.

- Giải thích được vì sao ba tệp cùng mức đỉnh lại không to bằng nhau, và chọn được thước đo phù hợp cho từng tình huống.
- Đo được độ to theo LUFS ở cả ba thang thời gian, đọc được LRA, và giải thích được vai trò của hai cái cổng trong BS.1770.
- Đo được true peak và nêu được vì sao trần luôn nằm dưới 0 dBFS.
- Chuẩn hoá được một tệp về mức đích của một nền tảng cụ thể, đủ bốn bước, có kiểm trần.
- Đo được SNR của một bản thu khi chỉ có bản đã trộn, và quyết định được có nên khử nhiễu hay không.
- Cài đặt được phép trừ phổ, chọn được hai tham số có căn cứ, và giải thích được vì sao pha được giữ nguyên.
- Nhận ra và *đo* được nhiễu nhạc, nêu được ba cách giảm nó cùng cái giá của mỗi cách.
- Sắp đúng thứ tự các bước, và nêu được cái giá bằng số của việc chuẩn hoá trước khi khử nhiễu.

> **Hai chỗ không có bữa trưa miễn phí"**
>
> Chương 5 và 6 dạy cách *làm ra* một bản dựng. Chương này hỏi bản dựng ấy đã nghe được chưa, và trả lời bằng hai việc: chỉnh mức cho đúng đích, và hạ nhiễu nền xuống.
>
> Cả hai việc đều đụng một bức tường. Chuẩn hoá độ to đụng trần true peak; khử nhiễu đụng nhiễu nhạc. Không việc nào có lời giải đúng tuyệt đối.
>
> Vì thế chương này không dạy "làm thế nào cho đúng" mà dạy "đo cái gì để chọn có ý thức". Mọi con số trong bài đều kiểm được trên máy của người đọc.

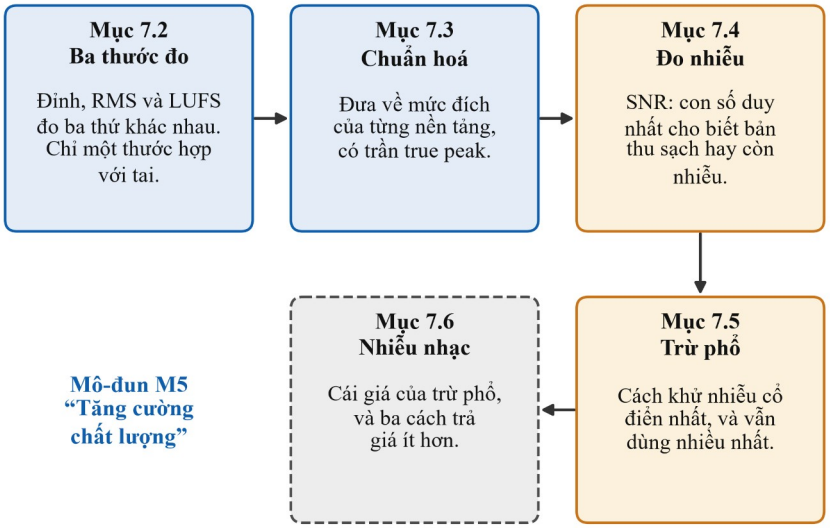
7.1 Vị trí của chương trong cuốn sách

7.1.1 Chương này bắt đầu từ đâu

Hai chương vừa qua dạy cách *làm ra* một bản dựng: đọc tệp, cắt, ghép, chỉnh gain, chuẩn hoá đỉnh. Kết quả là một tệp WAV chạy được và không có tiếng tách.

Chương này mang tên "Tăng cường chất lượng" và hỏi câu tiếp theo: bản dựng ấy đã *nghe được* chưa. Hai việc phải làm, và cả hai đều là chỗ mà một con số đo được thay thế cho một cuộc tranh cãi.

– **Chỉnh mức cho đúng đích.** Chương 5 chuẩn hoá theo đỉnh. Chương này cho thấy vì sao đỉnh không đủ, và thước nào mới đúng.
– **Hạ nhiễu nền xuống.** Đo được nhiễu bao nhiêu, khử bằng cách nào, và cái giá phải trả là gì.

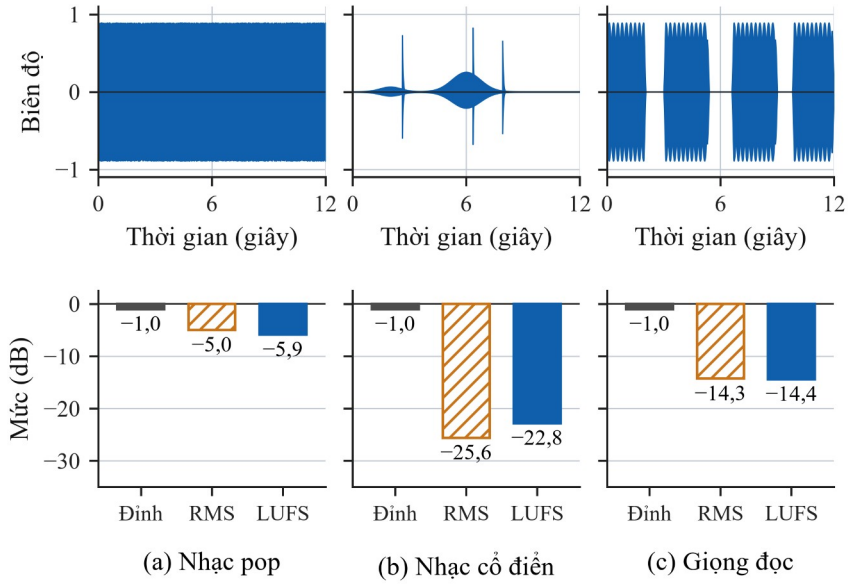


Hai việc: chỉnh mức cho đúng đích, và hạ nhiễu nền xuống. Cả hai đều có một con số đo được thay cho một cuộc tranh cãi, và đều có một cái giá phải trả.

Hình 7.1 Năm mục kỹ thuật của chương 7

7.1.2 Câu hỏi mở đầu — ba tệp cùng đỉnh, ba độ to

Cuối chương 6 đã dặn chuẩn bị ba tệp: một đoạn nhạc pop, một đoạn nhạc cổ điển và một đoạn giọng đọc. Hãy chuẩn hoá cả ba về đúng đỉnh −1 dBFS bằng hàm đã viết ở chương 5, rồi nghe lần lượt ở cùng một mức âm lượng loa.



Hình 7.2 Ba tệp cùng đỉnh −1 dBFS, đo bằng ba thước khác nhau

Cột xám của cả ba bằng nhau chính xác — đó là điều hàm chuẩn hoá đỉnh bảo đảm. Nhưng RMS chênh nhau 20,6 decibel và LUFS chênh 17,0 decibel. Và tai xác nhận: ba tệp *không* to bằng nhau, chênh nhau rất xa.

> **Một cách nói gọn về cả mục 7.2**
> Đỉnh trả lời câu hỏi "có vượt thang không". Đó là một câu hỏi kỹ thuật, và nó không liên quan gì tới việc tiếng nghe to hay nhỏ.
> Muốn biết to hay nhỏ thì phải hỏi một câu khác, và phải dùng một thước khác.
> Chương 5 đã nêu vấn đề này khi so đỉnh với RMS. Chương này đi tới cùng và giới thiệu thước mà cả ngành đang dùng.

7.1.3 Cách học chương này

- **Mỗi con số đều kiểm được trên máy của bạn. Mọi giá trị trong bài đều sinh ra từ mô-đun sigtools và kiểm được bằng vài dòng. Đừng tin con số vì nó in trong sách** — hãy chạy lại.
- **Chương này có hai chỗ "không có bữa trưa miễn phí".** Chuẩn hoá độ to đụng trần true peak; khử nhiễu đụng nhiễu nhạc. Cả hai chỗ đều không có lời giải đúng tuyệt đối, chỉ có lựa chọn có ý thức.
- **Nghe là bắt buộc, không phải tuỳ chọn.** Nhiễu nhạc ở mục 7.6 nhìn phổ đồ thì thấy, nhưng chỉ nghe mới hiểu vì sao nó khó chịu tới mức người dùng phải chấp nhận để lại nền ồn hơn.

7.1.4 Những gì chương này không làm

Bảng 7.1 Ba thứ gần giống mà thuộc chương khác

Thứ dễ gộp nhầm	Vì sao không thuộc chương 7	Thuộc chương nào
Bộ nén động, bộ giới hạn	Chúng *đổi* dải động chứ không chỉ dịch mức. Cần khái niệm ngưỡng, tỉ số, thời gian tấn công	Chương 8
Cân bằng tần số (EQ)	Cần thiết kế bộ lọc — nội dung chính của chương 8	Chương 8
Khử nhiễu bằng học máy	Không dựa trên giả định nhiễu dừng, và cần khối lượng dữ liệu huấn luyện ngoài phạm vi cuốn sách	Ngoài phạm vi sách

Riêng dòng cuối đáng nói thêm một câu. Các công cụ khử nhiễu bằng học máy hiện nay cho kết quả tốt hơn hẳn trừ phổ trên tiếng nói. Nhưng chúng là hộp đen: khi chúng cho kết quả sai, người dùng không biết vì sao và không chỉnh được gì. Học trừ phổ là học cái *khung suy nghĩ* — đo trước, chọn tham số có ý thức, đo lại cả hai mặt — và khung ấy dùng được với mọi công cụ, kể cả hộp đen.

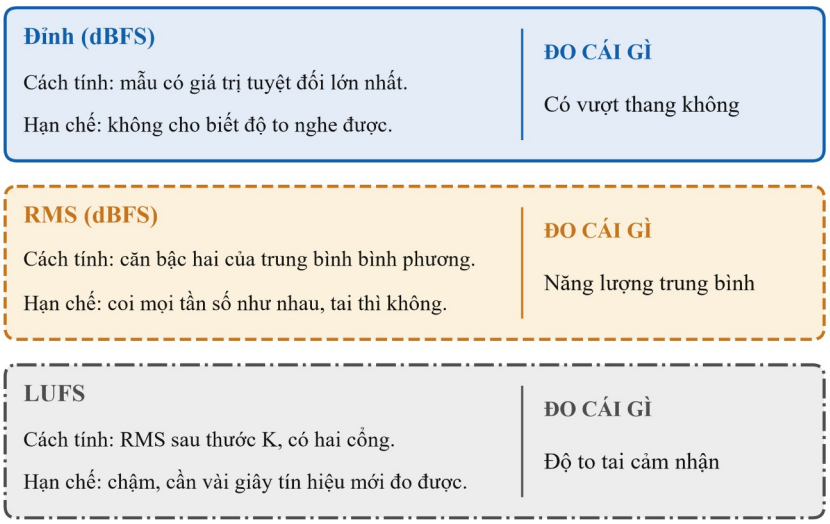
7.2 Ba thước đo mức

7.2.1 Đỉnh và RMS — nhắc lại từ chương 5

$$L_{\text{đỉnh}} = 20\log_{10}\max_n |x[n]| \; ; \qquad L_{\text{rms}} = 20\log_{10}\sqrt{\frac{1}{N}\sum_n x^2[n]} \tag{7.1}$$

$L_{\text{đỉnh}}$ là mức đỉnh, lấy mẫu có giá trị tuyệt đối lớn nhất. L_{rms} là mức hiệu dụng, căn bậc hai của trung bình bình phương trên N mẫu. Cả hai đều tính bằng dBFS — decibel so với toàn thang.

Hai thước này đo hai thứ khác nhau và không thay thế nhau. Đỉnh trả lời "có vượt thang không"; RMS trả lời "năng lượng trung bình là bao nhiêu". Khoảng cách giữa chúng là hệ số đỉnh, đại lượng đã gặp ở mục 5.7.



Hình 7.3 Ba thước đo mức và việc của mỗi thước

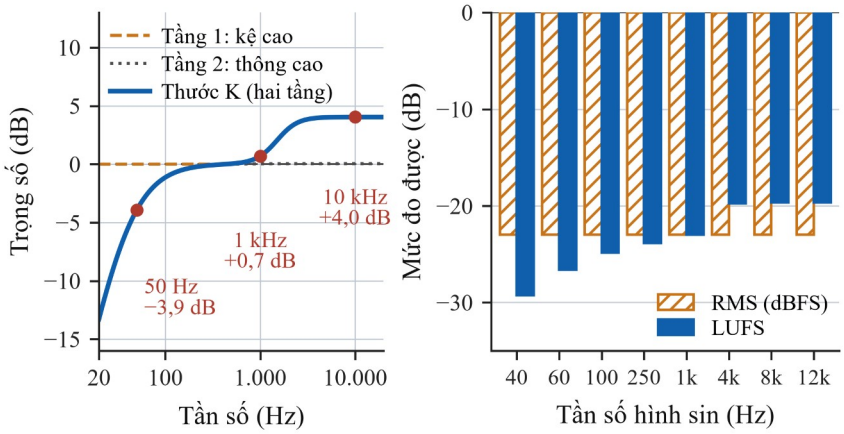
7.2.2 Vì sao RMS vẫn chưa đủ

RMS coi mọi tần số như nhau. Tai người thì không: ở mức nghe thông thường, tai kém nhạy hẳn với dải trầm và hơi nhạy hơn với dải 2–5 kHz. Đó chính là nội dung đường đẳng âm đã học ở mục 1.5.

Hậu quả rất cụ thể. Một bản nhạc nhiều tiếng trầm và một bản nhạc nhiều tiếng trung có thể có RMS bằng nhau chính xác mà nghe khác nhau vài decibel. Vì thế cần một thước có *trọng số* theo tần số.

7.2.3 Thước K và đơn vị LUFS

Khuyến nghị ITU-R BS.1770 định nghĩa một thước như thế. Nó gồm hai bộ lọc mắc nối tiếp, gọi chung là thước K.



(a) Đáp ứng của thước K (b) Cùng biên độ đỉnh −20 dBFS

Hình 7.4 Đáp ứng của thước K, và hệ quả khi đo

– **Tầng một — kệ cao.** Nâng dải trên 2 kHz lên gần 4 dB. Nó mô phỏng ảnh hưởng của chính cái đầu người lên âm tới từ phía trước.
– **Tầng hai — thông cao.** Cắt dải trầm: −3,9 dB ở 50 Hz, −13,3 dB ở 20 Hz. Nó mô phỏng việc tai kém nhạy với tần số thấp.

Sau khi qua hai tầng ấy, lấy bình phương trung bình như RMS, rồi cộng các kênh lại có trọng số.

$$z_i = \frac{1}{T}\sum_n y_i^2[n]\,; \qquad L_K = -0{,}691 + 10\log_{10}\sum_i G_i z_i \tag{7.2}$$

z_i là bình phương trung bình của kênh thứ i sau khi qua thước K, tính trên khoảng thời gian T. G_i là trọng số kênh: kênh trái, phải và giữa lấy 1; hai kênh vòng lấy 1,41. Hằng số −0,691 là để hiệu chỉnh cho một hình sin 997 Hz đọc đúng như quy ước của chuẩn.

Kết quả gọi là LUFS — Loudness Units relative to Full Scale. Một đơn vị *lu* bằng đúng một decibel; chữ *lu* chỉ dùng khi nói về hiệu hai mức độ to, còn LUFS dùng cho mức tuyệt đối.

Điều hình B02 cho thấy

Tám hình sin ở cột phải có biên độ *hệt nhau*. RMS của chúng bằng nhau chính xác — cột vàng phẳng ở −23,0 dBFS.

LUFS thì không: hình sin 40 Hz đọc −29,3 còn hình sin 8 kHz đọc −19,7. Chênh 9,6 decibel, đúng theo chiều mà tai cảm nhận.

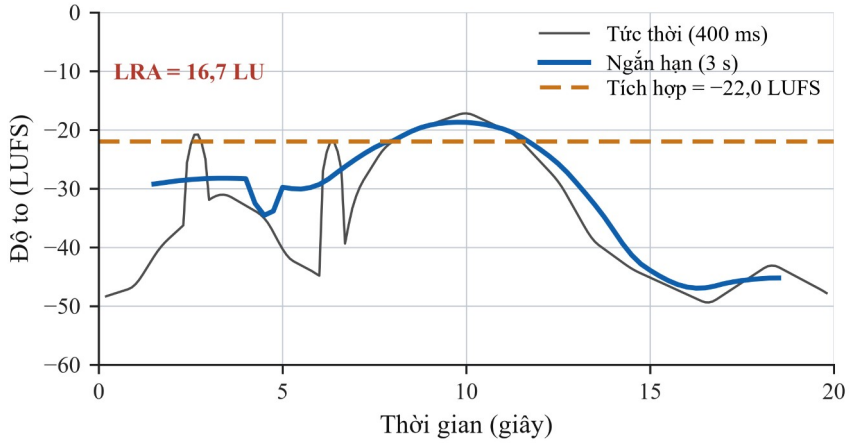
Đó là toàn bộ ý nghĩa của thước K, và cũng là toàn bộ khác biệt giữa RMS và LUFS.

7.2.4 Ba thang thời gian, ba con số

Một tệp không có một con số độ to mà có ba, tuỳ khoảng thời gian lấy trung bình.

Bảng 7.2 Ba thang thời gian của phép đo độ to

Tên	Cửa sổ	Dùng để làm gì
Tức thời (momentary)	400 ms	Dò chỗ đột ngột quá to; nhảy theo từng nốt nên khó đọc
Ngắn hạn (short-term)	3 s	Mượt hơn; đây là con số người dựng nhìn khi cân các đoạn với nhau
Tích hợp (integrated)	Cả tệp, có cổng	*Một* con số duy nhất cho cả tệp — và là con số mà nền tảng phát nhìn vào



Hình 7.5 Ba đường độ to trên cùng một đoạn nhạc cổ điển

Còn một đại lượng thứ tư đọc ra từ đường ngắn hạn: *Dải độ to*, ký hiệu LRA, tính bằng *lu*. Nó là khoảng giữa phân vị 10 % và phân vị 95 % của độ to ngắn hạn, sau khi đã bỏ các khối quá nhỏ. LRA nói lên điều mà một con số LUFS duy nhất không nói được: bản nhạc có chỗ rất nhỏ và chỗ rất to, hay phẳng lì từ đầu tới cuối.

Bảng 7.3 Đọc con số LRA

LRA	Nghĩa là	Thường gặp ở
Dưới 3 *lu*	Gần như phẳng — đã bị nén rất mạnh	Nhạc pop hiện đại, quảng cáo phát thanh
4 – 8 *lu*	Có biến thiên vừa phải	Podcast, phim truyền hình, nhạc nhẹ
9 – 20 *lu*	Dải động rộng	Nhạc cổ điển, phim rạp
Trên 20 *lu*	Rất rộng — cần kiểm xem có phải lỗi dựng không	Bản thu hoà nhạc chưa qua xử lý

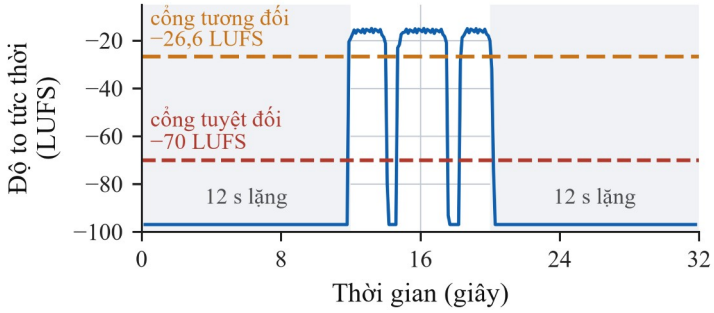
7.2.5 Hai cái cổng, và vì sao chúng cần thiết

Nếu chỉ lấy trung bình mọi khối 400 mili giây của cả tệp thì một đoạn im lặng dài sẽ kéo con số tụt xuống. Với podcast — thể loại nhiều khoảng lặng nhất — sai lệch ấy đủ để làm sai cả bản giao nộp.

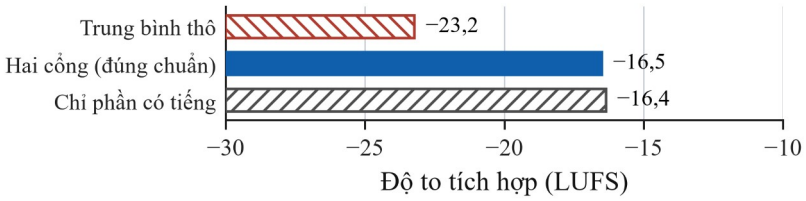
BS.1770 giải quyết bằng hai cái cổng, áp dụng lần lượt.

$$\Gamma_a = -70 \text{ LUFS}; \qquad \Gamma_r = L_K(\{j : l_j > \Gamma_a\}) - 10 \tag{7.3}$$

– **Cổng tuyệt đối.** Bỏ mọi khối có độ to dưới −70 LUFS. Đó là ngưỡng dưới đó gần như chắc chắn là im lặng.
– **Cổng tương đối.** Tính độ to trung bình của các khối còn lại, rồi bỏ tiếp mọi khối thấp hơn mức ấy 10 *lu*. Cổng này loại phần quá nhỏ so với phần chính của tệp.



(a) Giọng đọc thêm 24 s lặng



(b) Ba cách tính trên cùng tệp

Hình 7.6 Cùng một giọng đọc, thêm 24 giây lặng, tính ba cách

Kết quả trên hình: trung bình thô cho −23,2 LUFS; có hai cổng cho −16,5 LUFS; và chỉ đo phần có tiếng cho −16,4 LUFS. Nghĩa là hai cái cổng làm đúng việc của chúng — kết quả chỉ lệch 0,1 decibel so với khi tự tay cắt bỏ phần lặng.

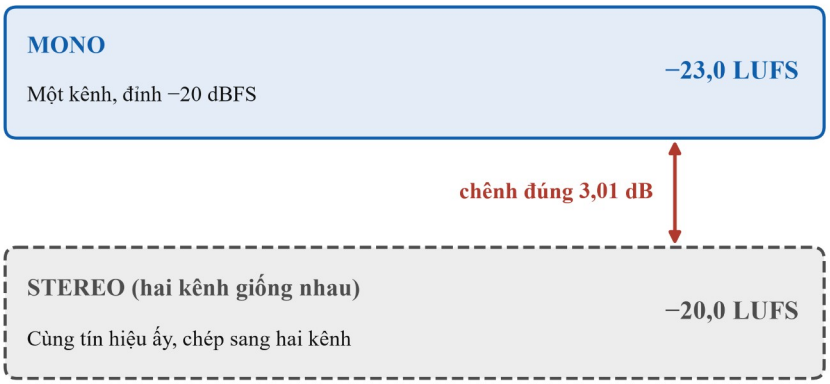
Hai cổng là chỗ hay bị cài đặt sai

Nhiều phần mềm có tuỳ chọn tắt cổng, hoặc mặc định đo theo cách khác. Kết quả là hai chương trình cho hai con số trên cùng một tệp, và người dùng không biết tin cái nào.

Cách kiểm: đo một tệp có 20 giây im lặng ở đầu, rồi đo lại chính tệp ấy sau khi cắt bỏ phần lặng. Nếu hai kết quả lệch nhau quá 0,3 dB thì phần mềm đang không dùng cổng đúng chuẩn.

Trong sigtools, hàm do_to_tich_hop thực hiện đúng hai cổng của BS.1770-4 và đã đối chiếu với bộ đo tham chiếu.

7.2.6 Một điểm dễ sai: LUFS phụ thuộc số kênh



Cùng một tiếng, cùng biên độ, cùng mức đỉnh, mà con số LUFS khác nhau 3 dB chỉ vì số kênh.

Khi giao nộp phải ghi rõ tệp mono hay stereo, và không so LUFS của tệp mono với LUFS của tệp stereo.

Hình 7.7 Cùng một tín hiệu, mono và stereo, hai con số LUFS

Nhìn lại công thức 7.2: các kênh được *cộng* vào nhau trước khi lấy lô-ga-rít. Chép một tín hiệu mono sang hai kênh nghĩa là nhân đôi tổng ấy, tức là cộng thêm đúng 3,01 decibel.

Điều này gây hiểu lầm thường xuyên khi giao nộp. Hai quy tắc thực hành để tránh.

– Luôn ghi rõ tệp là mono hay stereo khi báo con số LUFS.
– Không bao giờ so LUFS của một tệp mono với LUFS của một tệp stereo. Nếu phải so, chuyển cả hai về cùng số kênh trước.

7.2.7 Mã

Liệt kê 7.1 Đo ba thước trên một tệp

```
import numpy as np, soundfile as sf
import am

x, sr = sf.read("ban_dung.wav")

print("Đỉnh      :", am.peak_dbfs(x),        "dBFS")
print("True peak :", am.true_peak_dbtp(x),   "dBTP")
print("RMS       :", am.rms_dbfs(x),         "dBFS")
print("LUFS      :", am.do_to_tich_hop(x, sr))
print("LRA       :", am.dai_do_to(x, sr),    "LU")

# bốn dòng này nên có trong MỌI kịch bản giao nộp
```

Năm con số ấy là hồ sơ đầy đủ về mức của một tệp. Thói quen nên tập: in chúng ra trước khi làm bất cứ việc gì, và in lại sau khi làm xong. Mục 7.3 dùng đúng năm con số này.

7.3 Chuẩn hoá theo độ to

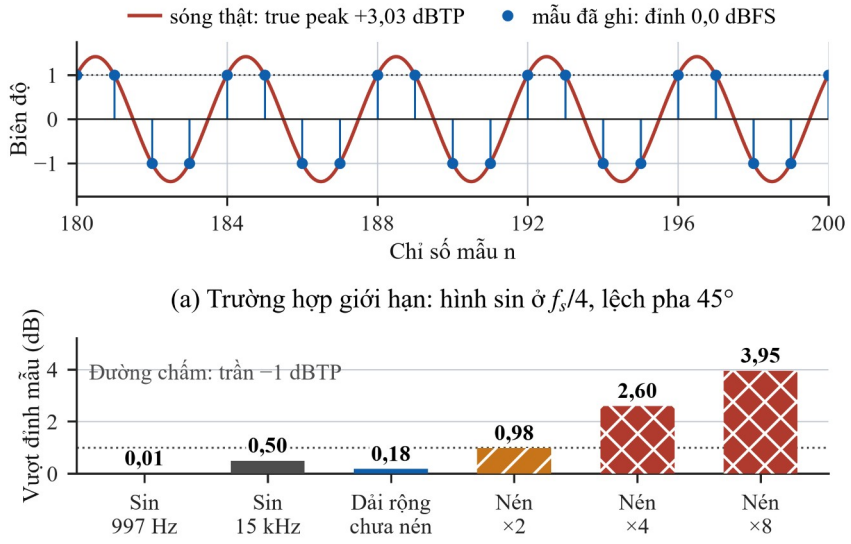
7.3.1 True peak — vì sao trần lại nằm dưới 0

Trước khi nói về chuẩn hoá, cần một khái niệm nữa. Chương 5 đã nêu nó ở mục 5.4.4 mà chưa dùng tới; chương này nó thành một ràng buộc bắt buộc.

Đỉnh mẫu chỉ nhìn các mẫu *đã ghi*. Nhưng bộ chuyển số–tương tự không phát ra các mẫu ấy — nó dựng lại một sóng liên tục đi qua chúng. Sóng ấy có thể vượt cao hơn mẫu cao nhất, và khi nó vượt quá toàn thang thì mạch khuếch đại phía sau bị cắt ngọn.

$$\mathrm{TP} = 20\log_{10}\max_{m}\left|x_{\uparrow M}[m]\right|, \qquad M \geq 4 \tag{7.4}$$

Cách đo theo BS.1770: lấy mẫu quá lên ít nhất bốn lần rồi mới tìm giá trị lớn nhất. Kết quả ghi bằng dBTP — decibel true peak.



(a) Trường hợp giới hạn: hình sin ở $f_s/4$, lệch pha 45°

(b) Sáu loại tư liệu cùng chuẩn hoá về 0,0 dBFS

Hình 7.8 Đỉnh mẫu và true peak trên sáu loại tư liệu

Hình bên trái là trường hợp giới hạn dựng sẵn để nhìn cho rõ: một hình sin ở đúng một phần tư tần số lấy mẫu, lệch pha 45 độ. Mọi mẫu của nó đều bằng ±0,707 lần biên độ thật, nên sau khi chuẩn hoá đỉnh mẫu về 0 dBFS, sóng thật tăng lên +3,03 dBTP.

Hình bên phải là điều đáng nhớ hơn: mức tăng thêm ấy phụ thuộc vào tư liệu đã bị nén biên độ tới mức nào. Nguồn dải rộng chưa nén chỉ tăng thêm 0,18 dB; nén biên độ gấp bốn lần thì tăng thêm 2,60 dB; nén gấp tám lần thì 3,95 dB.

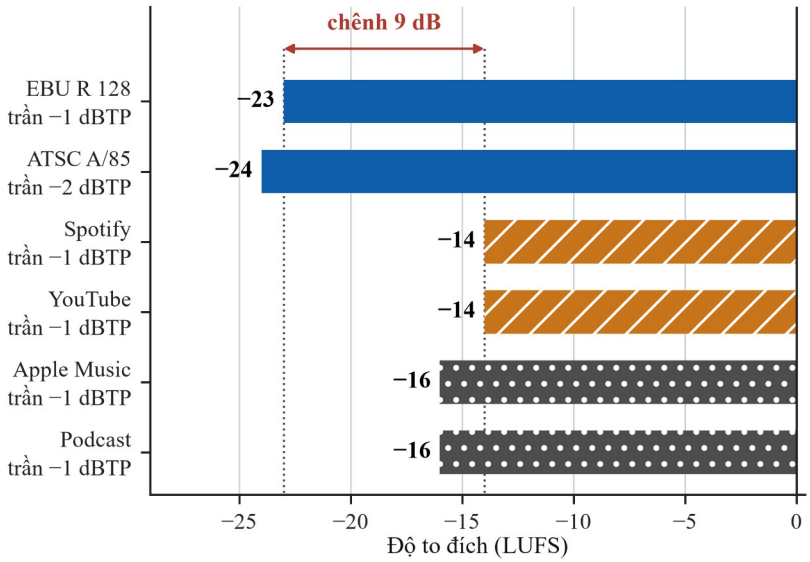
Vì sao trần là −1 chứ không phải 0

Mọi chuẩn phát sóng và mọi nền tảng phát trực tuyến đều đặt trần true peak dưới 0 dBFS — thường là −1 dBTP, phát thanh Bắc Mỹ là −2 dBTP.

Lý do thứ nhất: chừa chỗ cho phần tăng thêm giữa hai mẫu, đúng như hình bên trái.

Lý do thứ hai, thực tế hơn: mọi bộ mã hoá mất mát — MP3, AAC, Opus — đều làm đỉnh dịch đi một chút. Một tệp đúng 0,0 dBFS sau khi nén thành AAC có thể vượt lên trên 0 và méo.

7.3.2 Mức đích của từng nền tảng



Hình 7.9 Mức đích và trần của sáu nơi giao nộp

Không có một mức đích chung. Con số phụ thuộc vào nơi bản thu sẽ được nghe.

Bảng 7.4 Sáu chuẩn thường gặp

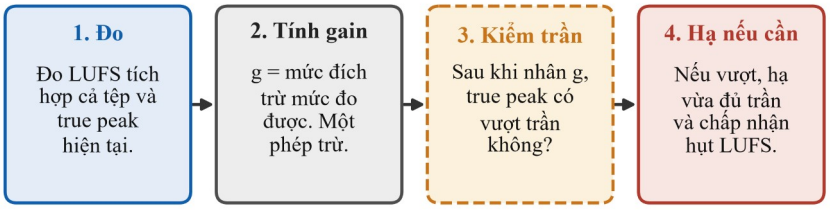
Nơi giao nộp	LUFS đích	Trần	Vì sao là con số ấy
EBU R 128 (phát thanh, TH châu Âu)	−23	−1 dBTP	Còn phải chừa dải động cho tiếng động lớn trong phim và thời sự
ATSC A/85 (truyền hình Bắc Mỹ)	−24	−2 dBTP	Cùng lý do; trần thấp hơn vì đường truyền cũ kém chịu đỉnh
Spotify, YouTube	−14	−1 dBTP	Người nghe thường ở nơi ồn, tai nghe điện thoại, xe cộ
Apple Music	−16	−1 dBTP	Chọn cân bằng hơn giữa độ to và dải động
Podcast (khuyến nghị chung)	−16	−1 dBTP	Chủ yếu là lời nói, nghe bằng tai nghe khi di chuyển

Chênh lệch giữa −23 và −14 là chín decibel — gần gấp ba lần về áp suất âm. Nộp một bản dựng theo chuẩn phát thanh lên nền tảng phát trực tuyến thì nó bị nghe là "nhỏ hẳn so với các bài khác", và ngược lại.

7.3.3 Bốn bước chuẩn hoá

$$g_{\text{dB}} = L_{\text{đích}} - L_{\text{đo}}\,; \qquad g_{\text{cuối}} = \min\left(g_{\text{dB}},\ \text{TP}_{\text{trần}} - \text{TP}_{\text{đo}}\right) \tag{7.5}$$

g_{dB} là lượng gain phải cộng để đạt mức đích: đơn giản là hiệu giữa mức đích và mức đo được. Nhưng nếu cộng chừng ấy mà true peak vượt trần thì phải lấy giá trị nhỏ hơn trong hai lựa chọn — và chấp nhận độ to hụt so với đích.



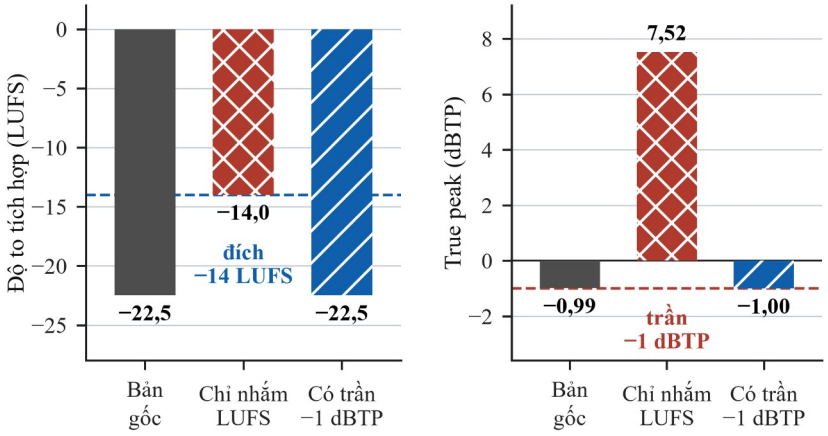
Bước 4 hay bị bỏ qua; bỏ qua thì tệp giao nộp bị nền tảng từ chối hoặc bị hạ mức tự động ở khâu phát.

Nếu phải hạ nhiều hơn 1–2 dB ở bước 4 thì vấn đề nằm ở chỗ khác: bản dựng đã bị nén quá tay từ trước. Luôn chuẩn hoá ở bước cuối.

Hình 7.10 Bốn bước chuẩn hoá theo độ to

Bước bốn là bước hay bị bỏ qua, và bỏ qua nó thì tệp giao nộp bị nền tảng từ chối hoặc bị hạ mức tự động ở khâu phát.

7.3.4 Khi trần chặn lại — trường hợp phải chọn



(a) Độ to đạt được (b) True peak tương ứng

Hình 7.11 Một bản nhạc cổ điển đưa về −14 LUFS, có và không có trần

Bản nhạc cổ điển trong hình có đỉnh cao hơn độ to trung bình 21,8 decibel. Đại lượng ấy gọi là PLR — tỉ số đỉnh trên độ to — và nó quyết định mọi chuyện ở đây.

Nếu chỉ nhắm đúng −14 LUFS mà không kiểm trần, true peak tăng lên +7,52 dBTP: méo mạnh ở khâu phát. Đặt trần −1 dBTP thì true peak về đúng chỗ, nhưng độ to dừng ở −22,5 LUFS, hụt 8,5 decibel so với đích.

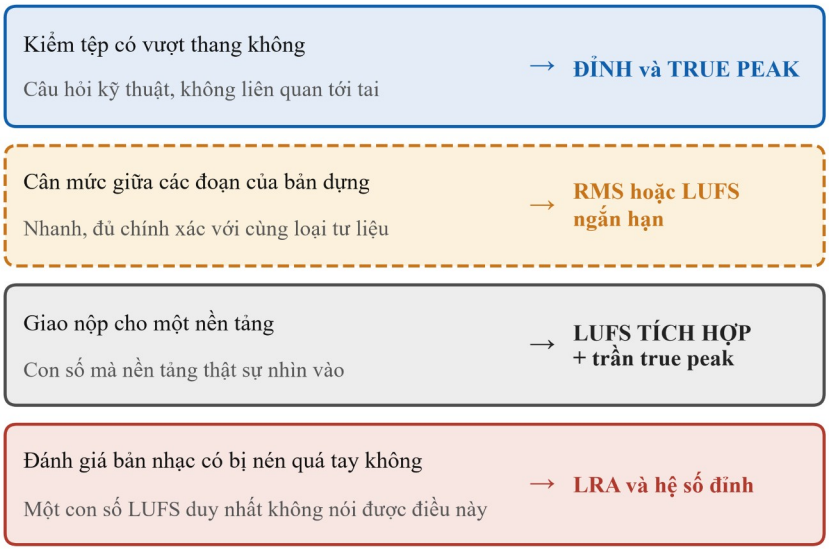
Phải chọn — và mọi chuẩn đều ưu tiên trần

Không có cách nào có cả hai. Điều duy nhất có thể làm là chọn có ý thức, và chuẩn nào cũng chọn giữ trần.

Nếu chỉ hụt 1–2 dB thì chấp nhận được: các nền tảng đều tự chỉnh lại ở khâu phát và người nghe không nhận ra.

Nếu hụt tới 8 dB như ví dụ trên thì vấn đề không nằm ở phép chuẩn hoá. Hoặc bản dựng có PLR quá cao và cần nén nhẹ trước — nội dung chương 8; hoặc đơn giản là tư liệu ấy không hợp với một nền tảng nhắm −14 LUFS.

7.3.5 Thước nào cho việc nào



Hình 7.12 Bốn tình huống và thước đo tương ứng

Dùng nhầm thước là nguyên nhân của phần lớn các bản giao nộp bị trả lại. Bảng trên gói lại thành bốn câu hỏi thực tế.

7.3.6 Mã

Liệt kê 7.2 Chuẩn hoá đủ bốn bước

```
def chuan_hoa_lufs(x, dich_lufs=-23.0, tran_tp=-1.0, sr=44100):
    """Chuẩn hoá theo độ to, CÓ trần true peak."""
    # 1. đo
    hien = do_to_tich_hop(x, sr)
    # 2. tính gain
    g = dich_lufs - hien
    y = x * db_to_lin(g)
    # 3. kiểm trần
    tp = true_peak_dbtp(y)
    bi_chan = tp > tran_tp
    # 4. hạ thêm nếu cần
    if bi_chan:
        g += (tran_tp - tp)
        y = x * db_to_lin(g)
    return y, g, bi_chan
```

```
y, g, chan = chuan_hoa_lufs(x, -16.0, -1.0, sr)
if chan:
    print(f"Tran chan lai. Do to that: "
          f"{do_to_tich_hop(y, sr):.1f} LUFS")
```

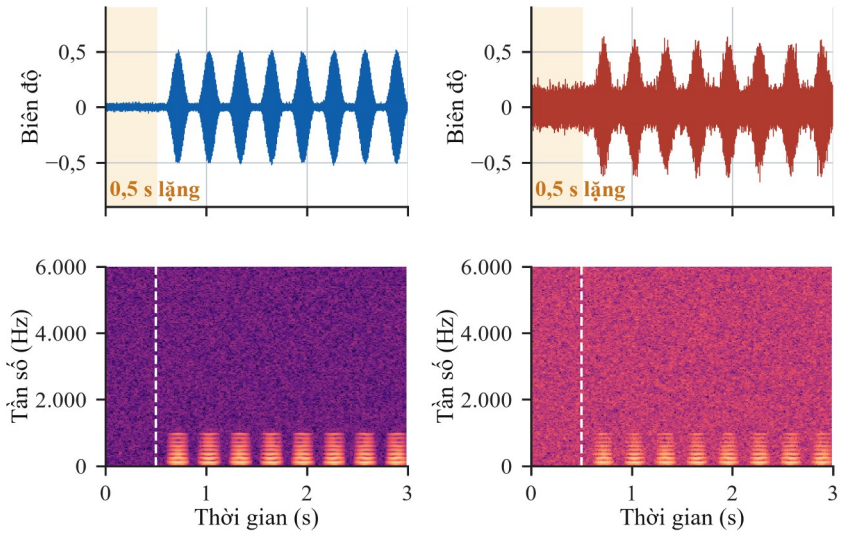
Giá trị trả về thứ ba đáng chú ý: hàm nói rõ cho người gọi biết trần có chặn hay không. Không có nó thì một bản hụt tám decibel so với đích vẫn lặng lẽ được ghi ra đĩa và không ai biết.

7.4 Nhiễu và cách đo nó

7.4.1 Nhiễu là gì trong phạm vi chương này

Chữ "nhiễu" trong xử lý âm thanh chỉ nhiều thứ. Chương này chỉ bàn một loại: nhiễu *nền dừng* — tiếng quạt, điều hoà, ù nguồn điện, tiếng rít của bộ tiền khuếch đại. Đặc điểm chung của chúng là phổ gần như không đổi theo thời gian.

Giả định "không đổi theo thời gian" chính là thứ làm cho toàn bộ mục 7.5 hoạt động được. Những gì *không* thoả giả định ấy — tiếng nói chuyện bên cạnh, tiếng xe chạy qua, tiếng cửa đóng — nằm ngoài phạm vi và phải xử lý bằng cách khác, thường là cắt bỏ bằng tay.



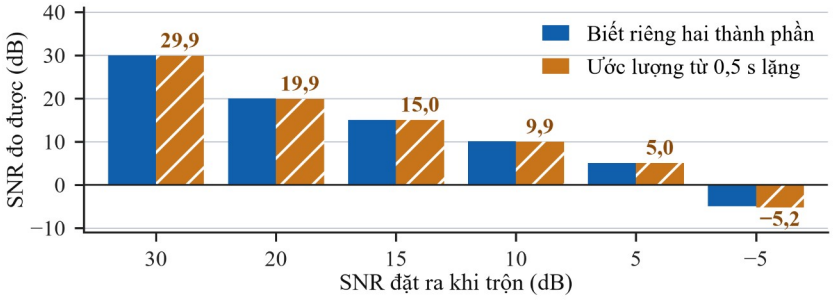
(a) SNR = 20 dB (b) SNR = 5 dB

Hình 7.13 Cùng một giọng đọc ở hai mức nhiễu, nhìn hai cách

7.4.2 SNR — một con số cho cả bản thu

$$\mathrm{SNR}_{\mathrm{dB}} = 10\log_{10}\frac{P_{\text{tín}}}{P_{\text{nhiễu}}} = 10\log_{10}\frac{P_{\text{tổng}} - P_{\text{nhiễu}}}{P_{\text{nhiễu}}} \tag{7.6}$$

Vế đầu là định nghĩa: tỉ số công suất tín hiệu trên công suất nhiễu, tính bằng decibel. Vế sau là cách tính khi chỉ có bản *đã trộn* — tình huống thật của mọi bản thu: đo công suất nhiễu trên một đoạn *lặng*, rồi trừ khỏi công suất tổng để ra công suất tín hiệu.



(a) Hai cách đo cho cùng một kết quả

Trên 40 dB	Phòng thu đã xử lý âm
25 – 40 dB	Bản thu tốt trong phòng yên
15 – 25 dB	Thu tại nhà, có quạt hoặc điều hoà
5 – 15 dB	Thu ngoài trời, quán, hội trường
Dưới 5 dB	Lời nói đã khó nghe rõ

(b) Đọc con số SNR: đo trước khi khử nhiễu

Hình 7.14 Hai cách đo SNR, và cách đọc con số ấy

Cột trái của hình là điều đáng chú ý về mặt thực hành: ước lượng từ nửa giây im lặng cho kết quả chính xác tới một phần mười decibel, so với khi biết riêng hai thành phần. Nói cách khác, chỉ cần nửa giây yên tĩnh là đo được chất lượng của cả bản thu.

Một thói quen đáng tập từ hôm nay

Mỗi lần bấm máy, thu dư ba tới năm giây yên tĩnh trước khi bắt đầu nói hoặc chơi nhạc.

Ba giây ấy dùng cho hai việc: đo SNR để biết bản thu có đủ sạch không, và ước lượng phổ nhiễu để khử nếu cần.

Không có ba giây ấy thì cả mục 7.5 không dùng được, và cách duy nhất còn lại là ước lượng nhiễu từ những khoảng lặng giữa câu — ngắn hơn nhiều và kém chính xác hơn nhiều.

7.4.3 Đọc con số SNR

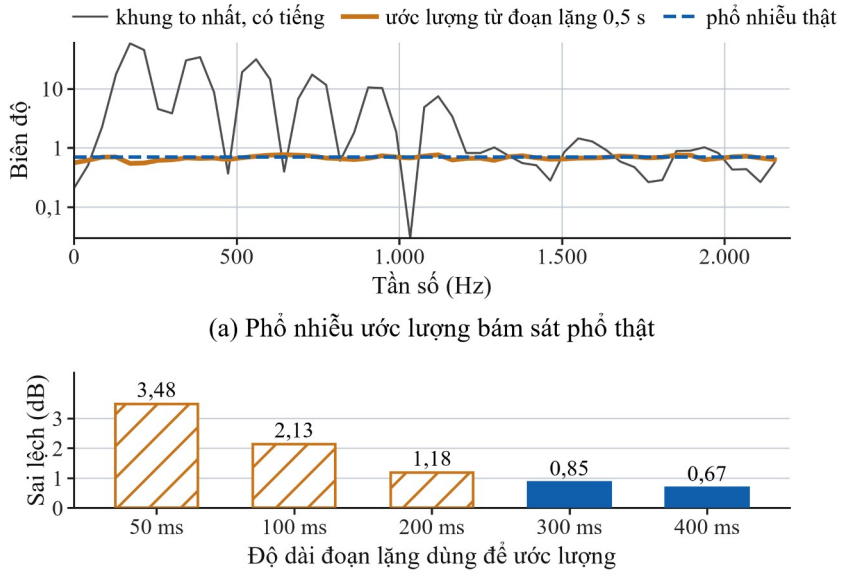
Bảng 7.5 Ngưỡng SNR và ý nghĩa thực tế

SNR	Tình huống điển hình	Nên làm gì
Trên 40 dB	Phòng thu đã xử lý âm học	Không cần khử
25 – 40 dB	Bản thu tốt trong phòng yên tĩnh	Không cần khử; khử chỉ tổ thêm nhiễu nhạc
15 – 25 dB	Thu tại nhà, có quạt hoặc điều hoà	Khử nhẹ, hệ số trừ vượt khoảng 1 – 1,5
5 – 15 dB	Thu ngoài trời, quán, hội trường	Khử vừa, hệ số khoảng 2, kèm san hệ số theo khung
Dưới 5 dB	Lời nói đã khó nghe rõ	Cân nhắc thu lại; khử đủ mạnh để nghe được thì tiếng đã méo mạnh

Bảng này trả lời câu hỏi *đầu tiên* của mọi việc khử nhiễu, và là câu hỏi mà người mới hay bỏ qua: có nên khử hay không. Trên 25 decibel thì câu trả lời là không, và lý do sẽ rõ ở mục 7.6.

7.4.4 Ước lượng phổ nhiễu

SNR là một con số cho cả bản thu. Muốn *khử* nhiễu thì cần biết nó phân bố thế nào theo tần số — tức là cần cả một phổ, chứ không phải một con số.



(a) Phổ nhiễu ước lượng bám sát phổ thật

(b) Đoạn lặng càng dài, ước lượng càng chuẩn

Hình 7.15 Phổ nhiễu ước lượng từ đoạn lặng, so với phổ nhiễu thật

Hình bên trái cho thấy phép ước lượng bám rất sát phổ thật, dù nó chỉ nhìn nửa giây đầu tệp. Hình bên phải cho thấy độ dài đoạn lặng quyết định độ chính xác: 50 mili giây cho sai lệch trung bình 3,48 dB, còn 400 mili giây chỉ 0,67 dB.

Liệt kê 7.3 Ước lượng phổ nhiễu và đo SNR

```
import am

x, sr = am.doc_wav("thu_tho.wav")

# doan lang: 0,0 den 0,45 giay
snr = am.snr_tu_doan_lang(x, 0.0, 0.45, sr)
print(f"SNR = {snr:.1f} dB")

if snr < 25:
    N = am.pho_nhieu_uoc_luong(x, 0.0, 0.45, sr)
    # N la mot mang bien do, dai n_fft // 2 + 1
else:
    print("Da du sach — khong can khu")
```

Giả định duy nhất, và chỗ nó sụp

Toàn bộ phép khử nhiễu của chương này đứng trên một giả định: phổ nhiễu *không đổi* theo thời gian. Nhờ thế, phổ đo trên nửa giây đầu dùng được cho cả tệp dài mười phút.

Giả định ấy đúng với quạt, điều hoà, ù nguồn, tiếng rít mạch.

Nó *sai* với tiếng nói chuyện bên cạnh, tiếng xe chạy qua, tiếng mưa lúc to lúc nhỏ. Với những thứ ấy, trừ phổ sẽ khử sai chỗ và làm sai lệch tiếng mà không hạ được nhiễu.

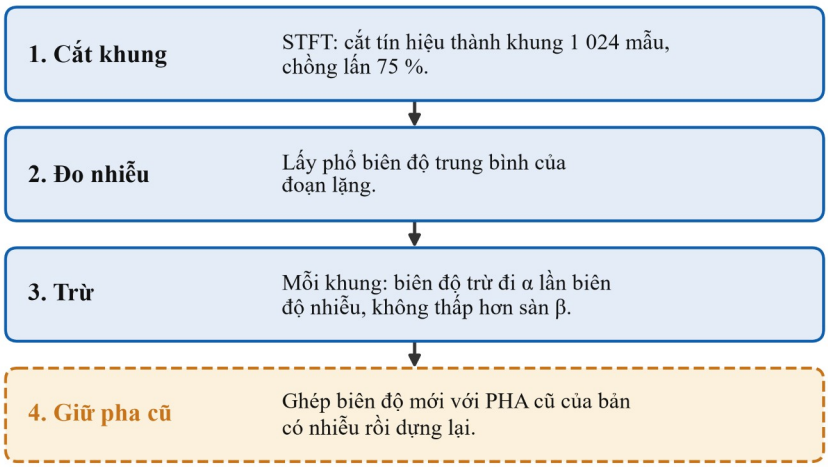
Cách kiểm nhanh: ước lượng phổ nhiễu ở *đầu* tệp và ở *cuối* tệp, rồi so hai phổ. Lệch quá 3 dB thì nhiễu không dừng, và đừng dùng phương pháp của mục sau.

7.5 Khử nhiễu bằng trừ phổ

7.5.1 Ý tưởng, gói trong một câu

Nếu biết nhiễu có biên độ bao nhiêu ở mỗi tần số, thì ở mỗi khung thời gian, cứ lấy biên độ đo được *trừ đi* biên độ nhiễu. Phần còn lại là tín hiệu.

Câu ấy đơn giản tới mức đáng ngờ, và nó đúng là đơn giản. Phương pháp có từ những năm 1970, thuật toán gói trong khoảng mười dòng, và đến nay vẫn là thứ nằm dưới phần lớn các nút "khử nhiễu" trong phần mềm biên tập.



Bước 4 là chỗ đáng ngạc nhiên nhất: pha không được xử lý gì cả. Tai ít nhạy với sai pha trong dải bị nhiễu lấn (xem mục 7.5.3).

Hình 7.16 Bốn bước của phép trừ phổ

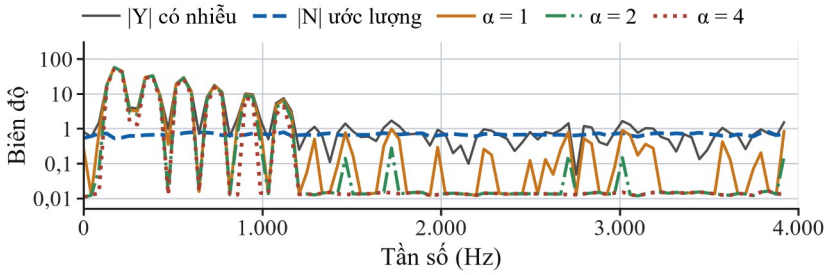
7.5.2 Công thức, và hai tham số

$$\left|\hat{S}[k]\right| = \max\left(\left|Y[k]\right| - \alpha\left|\hat{N}[k]\right|,\ \beta\left|\hat{N}[k]\right|\right) \tag{7.7}$$

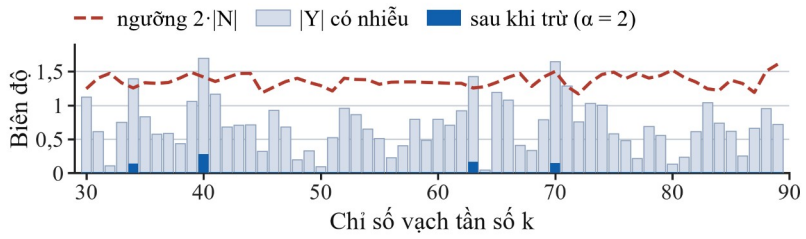
Y[k] là phổ của khung đang xét, N̂[k] là phổ nhiễu ước lượng, Ŝ[k] là phổ tín hiệu ước lượng. Hai tham số phải chọn.

– **Alpha — hệ số trừ vượt.** Trừ nhiều hơn mức ước lượng, để bù cho sai số của phép ước lượng. Alpha bằng 1 là trừ đúng bằng ước lượng; alpha bằng 2 là trừ gấp đôi.

– **Beta — sàn phổ.** Không cho biên độ tụt xuống dưới beta lần mức nhiễu. Nếu không có sàn này, những vạch bị trừ hết sẽ về đúng không, và chỗ trống ấy nghe còn khó chịu hơn chính nhiễu.



(a) Một khung phổ trừ ở ba mức α



(b) Chỉ 4 trong 60 vạch sống sót; 56 vạch bị nén xuống sàn

Hình 7.17 Một khung phổ, ba mức trừ, và chỗ nhiễu nhạc sinh ra

Cột phải của hình là chỗ đáng dừng lại lâu nhất trong cả chương. Trong một khung, chỉ những vạch vượt ngưỡng alpha·|N̂| mới sống sót — ở ví dụ này là bốn trên sáu mươi vạch. Năm mươi sáu vạch còn lại bị đưa xuống sàn.

Khung tiếp theo, vì nhiễu là ngẫu nhiên, lại là một tập bốn vạch *khác* sống sót. Cứ thế, mỗi khung một tập vạch khác nhau nhấp nháy. Mục 7.6 nói về hậu quả của điều đó.

7.5.3 Pha được giữ nguyên

$$\hat{s}[n] = \mathrm{ISTFT}\left(\left|\hat{S}[k]\right| \cdot e^{j\varphi_Y[k]}\right) \tag{7.8}$$

Bước dựng lại tín hiệu ghép biên độ *mới* với pha *cũ* — pha của bản có nhiễu, không xử lý gì cả. Đây là chỗ đáng ngạc nhiên nhất của phương pháp, và cũng là chỗ đáng giải thích.

– **Lý do thứ nhất: ước lượng pha rất khó.** Phổ biên độ của nhiễu thì đo được bằng cách lấy trung bình; pha của nhiễu thì không — nó ngẫu nhiên hoàn toàn và không có gì để lấy trung bình.

– **Lý do thứ hai: tai ít nhạy với sai pha ở đúng chỗ ấy.** Ở những vạch mà nhiễu lấn át, tai không phân biệt được pha đúng hay sai. Ở những vạch mà tín hiệu mạnh, pha đo được vốn đã gần đúng.

Đó là một đánh đổi rẻ, và là lý do phương pháp này sống được gần năm mươi năm. Điều đáng nhớ về mặt khái niệm: nó nối thẳng vào điều đã học ở mục 4.5 và 6.5 — pha quan trọng, nhưng *không* quan trọng như nhau ở mọi chỗ.

7.5.4 Mã

```
def tru_pho(x, pho_nhieu, alpha=2.0, beta=0.02,
            n_fft=1024, hop=256):
    X, i, w, tt = _stft(x, n_fft, hop)
    bien = np.abs(X)
    pha  = np.angle(X)        # giữ nguyên
    N = pho_nhieu[None, :]
    con = np.maximum(bien - alpha * N, beta * N)
    return _istft(con * np.exp(1j * pha),
                  i, w, tt, n_fft, hop)
```

(a) Toàn bộ thuật toán (chín dòng)

```
# 1. Đo SNR trước; trên 25 dB thì dừng
snr = snr_tu_doan_lang(x, 0.0, 0.45, sr)
if snr > 25: raise SystemExit("đủ sạch")
# 2. Ước lượng phổ nhiễu từ đoạn lặng
N = pho_nhieu_uoc_luong(x, 0.0, 0.45, sr)
# 3. Trừ phổ, rồi đo lại cả hai mặt
y = tru_pho(x, N, alpha=2.0)
print(san_nhieu_nen(y, sr))      # nền im đi
print(do_nhieu_nhac(y, 0.0, 0.45))  # cái giá
```

(b) Cách dùng và hai phép đo bắt buộc

Hình 7.18 Toàn bộ thuật toán, và cách dùng nó

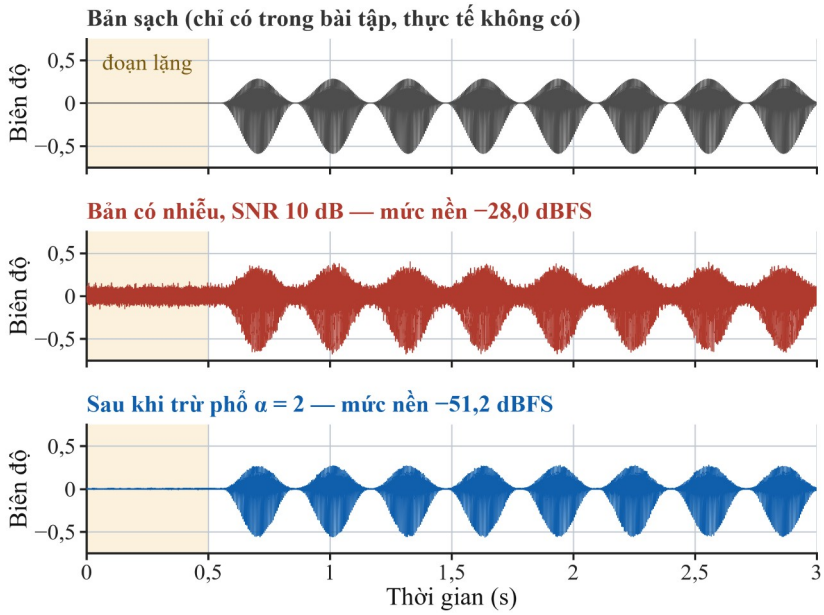
Liệt kê 7.4 Trừ phổ — bản đầy đủ

```
def tru_pho(x, pho_nhieu, alpha=2.0, beta=0.02,
            n_fft=1024, hop=256):
    """Tru pho: bien do tru di alpha lan bien do nhieu."""
    X, i, w, tt = _stft(x, n_fft, hop)
    bien = np.abs(X)
    pha  = np.angle(X)              # GIU NGUYEN
    N = np.asarray(pho_nhieu, dtype=float)[None, :]
    con = np.maximum(bien - alpha * N, beta * N)
    return _istft(con * np.exp(1j * pha), i, w, tt, n_fft, hop)


# dung: bao gio cung do TRUOC va do LAI SAU
print("SNR truoc :", snr_tu_doan_lang(on, 0.0, 0.45, sr))
N = pho_nhieu_uoc_luong(on, 0.0, 0.45, sr)
y = tru_pho(on, N, alpha=2.0)
print("Nen sau   :", san_nhieu_nen(y, sr), "dBFS")
print("Nhieu nhac:", do_nhieu_nhac(y, 0.0, 0.45, sr))
```

Ba dòng cuối là phần quan trọng hơn cả thuật toán: mỗi lần khử nhiễu phải đo *cả hai* mặt — nền im đi bao nhiêu, và trả giá bao nhiêu. Chỉ đo một mặt thì luôn thấy kết quả có vẻ tốt.

7.5.5 Kết quả, nhìn bằng dạng sóng

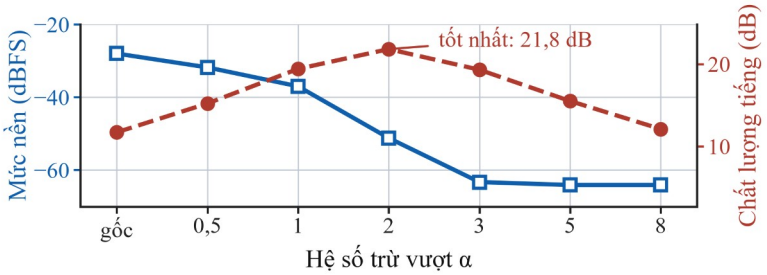


Hình 7.19 Bản sạch, bản có nhiễu, và bản đã trừ phổ

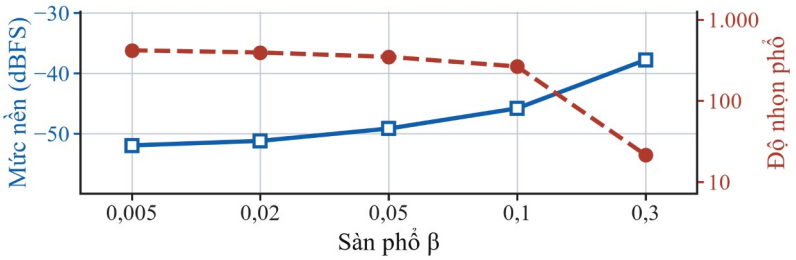
Nhìn dạng sóng thì kết quả có vẻ hoàn hảo: nửa giây đầu đã phẳng lì, mức nền hạ từ −28,0 xuống −51,2 dBFS, tức là 23 decibel.

Nhưng dạng sóng không nói được điều quan trọng nhất — cái *còn lại* trong nửa giây ấy nghe như thế nào. Đó là nội dung của mục sau, và là lý do mục này không kết thúc ở đây.

7.5.6 Chỉnh hai tham số



(a) α: có một giá trị tốt nhất



(b) β: đánh đổi giữa nền và nhiễu nhạc (α = 2)

Hình 7.20 Vai trò của alpha và của beta

Hai tham số có hai vai trò khác hẳn nhau, và hình trên tách chúng ra.

- **Alpha có một giá trị tốt nhất.** Trừ ít thì còn nhiễu; trừ quá tay thì gọt cả tiếng. Trên ví dụ của bài, chất lượng tiếng đạt cực đại 21,8 dB ở alpha bằng 2, và tụt xuống 12,1 dB ở alpha bằng 8.
- **Beta là một sự đánh đổi thuần tuý.** Beta nhỏ thì nền im hơn mà nhiễu nhạc nhiều hơn; beta lớn thì ngược lại. Không có giá trị tốt nhất, chỉ có lựa chọn.

Bảng 7.6 Chọn hai tham số theo tình huống

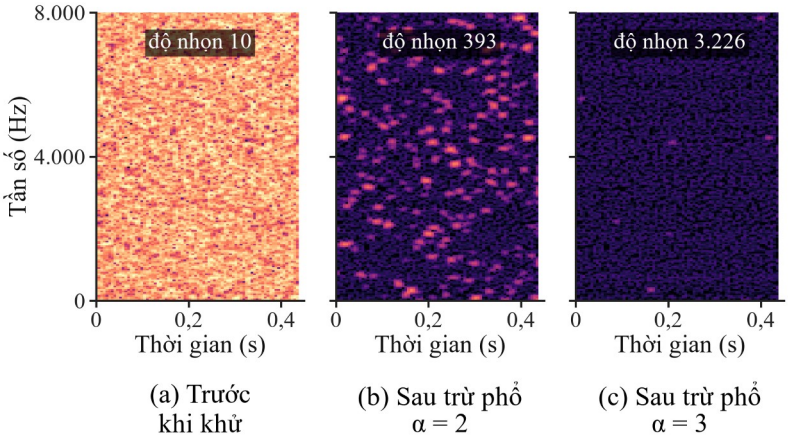
Tình huống	alpha	beta	Vì sao
SNR trên 25 dB	—	—	Đừng khử
Lời nói, SNR 15 – 25 dB	1,0 – 1,5	0,05	Mục tiêu là bớt chú ý tới nền, không phải xoá nền
Lời nói, SNR 5 – 15 dB	2,0	0,02	Vùng mà khử nhiễu thật sự có ích; nên kèm san khung
Nhạc, mọi mức SNR	1,0 – 1,5	0,1	Nhạc có nhiều thành phần yếu mà tai vẫn nghe; trừ mạnh làm mất đuôi vang và tiếng nhạc cụ nhỏ
Chỉ cần nghe rõ lời, không cần đẹp	3,0	0,01	Chấp nhận nhiễu nhạc để đổi lấy độ rõ — dùng cho tư liệu ghi chép, không dùng cho bản phát

7.6 Nhiễu nhạc — cái giá của trừ phổ

7.6.1 Nghe thấy gì, và vì sao

Nghe lại nửa giây lặng ở đầu tệp sau khi trừ phổ. Nền đã im hơn rất nhiều, nhưng cái còn lại không phải im lặng: nó là một chuỗi tiếng lích rích ngắn, cao, rời rạc, mà mọi tài liệu đều mô tả giống nhau — nghe như tiếng chuông gió.

Tên của nó trong tài liệu là *nhiễu nhạc* (musical noise), và nguyên nhân đã nằm ở cột phải của hình E02: mỗi khung một tập vạch khác nhau sống sót qua ngưỡng, rồi biến mất ở khung sau. Mỗi vạch nhấp nháy như thế là một tiếng lích rích ngắn ở đúng tần số của nó.



Hình 7.21 Nửa giây lặng, nhìn bằng phổ đồ, trước và sau khi khử

Ba phổ đồ trên vẽ trên cùng một thang màu để so được. Bên trái là nhiễu nền tự nhiên: một tấm màn đều. Sau khi trừ phổ, tấm màn ấy tách thành những chấm sáng rời rạc — và mỗi chấm là một tiếng lích rích.

7.6.2 Đo nhiễu nhạc bằng số

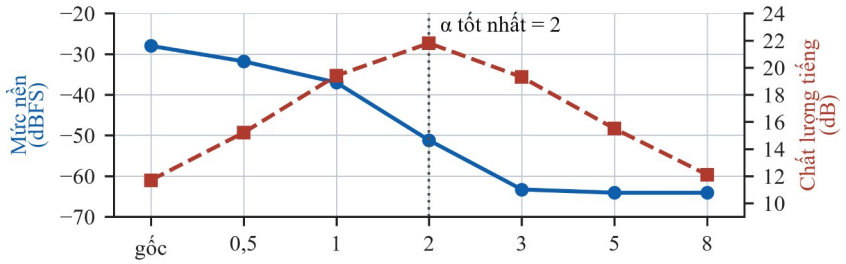
Mô tả "nghe như chuông gió" thì ai cũng nói được, nhưng không so sánh được hai cách khử với nhau. Cần một con số.

$$\kappa = \frac{1}{K\,\sigma_P^4} \sum_k (P[k] - \mu_P)^4 \tag{7.9}$$

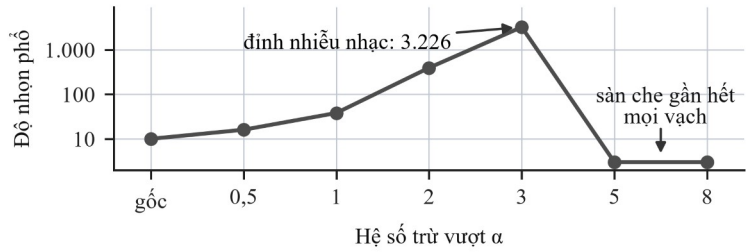
Đại lượng này là *độ nhọn* (kurtosis) của phổ công suất trong đoạn lặng. P[k] là công suất ở vạch k, còn mu và sigma là trung bình và độ lệch chuẩn của chúng.

Ý tưởng đơn giản: nhiễu nền tự nhiên trải đều nên độ nhọn thấp; nhiễu nhạc là những đỉnh cô lập nhô hẳn lên nên độ nhọn cao. Trên ví dụ của bài, nhiễu tự nhiên có độ nhọn 10, còn sau khi trừ phổ với alpha bằng 3 thì lên 3 226 — gấp hơn ba trăm lần.

7.6.3 Ba đại lượng, ba hình dạng



(a) Mức nền (đường liền) và chất lượng tiếng (đường đứt)



(b) Độ nhọn phổ ở đoạn lặng

Hình 7.22 Mức nền, chất lượng tiếng và nhiễu nhạc theo alpha

Đây là hình đáng học thuộc của cả chương. Ba đại lượng thay đổi theo alpha theo ba kiểu hoàn toàn khác nhau.

Bảng 7.7 Ba đại lượng theo hệ số trừ vượt alpha

alpha	Mức nền (dBFS)	Chất lượng tiếng (dB)	Độ nhọn phổ
gốc (chưa khử)	−28,0	11,7	10
0,5	−31,8	15,2	16
1	−37,0	19,4	38
2	−51,2	21,8 ← tốt nhất	393
3	−63,3	19,3	3 226 ← đỉnh
5	−64,1	15,5	3
8	−64,1	12,1	3

- **Mức nền giảm đều rồi bão hoà.** Từ alpha bằng 3 trở đi, gần như không giảm thêm được nữa — sàn beta đã chặn lại.
- **Chất lượng tiếng có một cực đại.** Ở alpha bằng 2. Trừ ít hơn thì nhiễu còn lấn; trừ nhiều hơn thì chính tiếng bị gọt. Ở alpha bằng 8, chất lượng còn kém hơn cả khi chưa khử gì.
- **Nhiễu nhạc tăng lên rồi sụp.** Đỉnh ở alpha bằng 3. Từ alpha bằng 5 trở đi độ nhọn về gần 3 — không phải vì hết nhiễu nhạc mà vì sàn đã che gần hết mọi vạch, kể cả vạch của tiếng. Đó là lý do phải nhìn cả cột chất lượng tiếng bên cạnh.

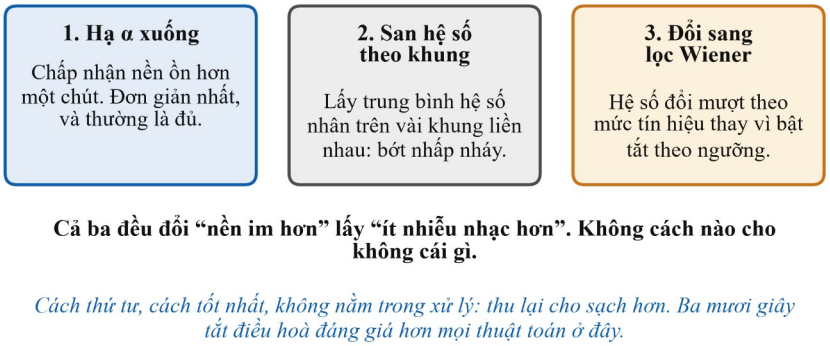
Bài học chung, không chỉ cho khử nhiễu

Nhìn *một* đại lượng thì luôn tưởng kết quả đang tốt. Chỉ nhìn mức nền thì alpha bằng 8 là lựa chọn tuyệt vời.

Mỗi phép xử lý âm thanh đều có ít nhất hai mặt, và mặt thứ hai thường khó đo hơn mặt thứ nhất. Việc của người kỹ thuật là tìm ra cách đo mặt thứ hai.

Điều này lặp lại ở chương 8 với bộ nén, ở chương 10 với hiệu ứng, và ở chương 12 với nén dữ liệu.

7.6.4 Ba cách trả giá ít hơn



Hình 7.23 Ba cách giảm nhiễu nhạc, và cách thứ tư

Cách thứ hai đáng nói kỹ vì nó rẻ nhất. Nhiễu nhạc sinh ra vì hệ số nhân của mỗi vạch *nhảy loạn* giữa các khung liền nhau. San hệ số ấy trên vài khung làm chúng bớt nhảy — đổi lại, tiếng bị nhoè nhẹ theo thời gian.

Liệt kê 7.5 Trừ phổ có san hệ số theo khung

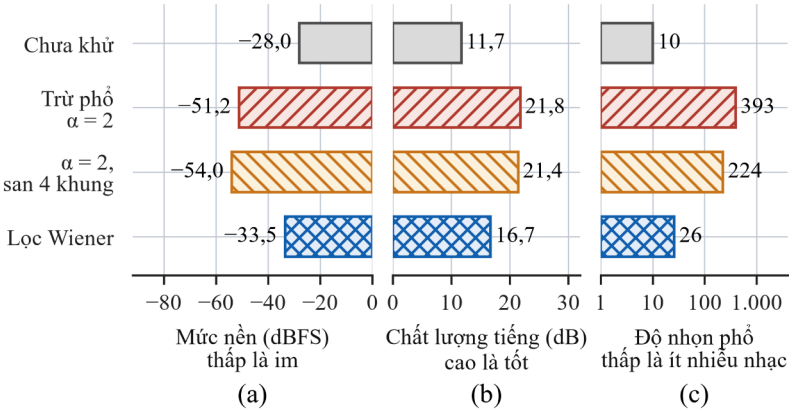
```
def tru_pho_san_khung(x, pho_nhieu, alpha=2.0, beta=0.02,
                      san=3, n_fft=1024, hop=256):
    X, i, w, tt = _stft(x, n_fft, hop)
    bien = np.abs(X)
    pha  = np.angle(X)
    N = np.asarray(pho_nhieu, dtype=float)[None, :]
    # he so nhan cua tung vach, tung khung
    he_so = (np.maximum(bien - alpha * N, beta * N)
             / np.maximum(bien, 1e-12))
    # SAN he so tren "san" khung lien nhau
    nhan = np.ones(int(san)) / float(san)
    he_so = sg.lfilter(nhan, 1.0, he_so, axis=0)
    return _istft(bien * he_so * np.exp(1j * pha),
                  i, w, tt, n_fft, hop)
```

Cách thứ ba là đổi hẳn công thức tính hệ số nhân. Thay vì bật tắt theo ngưỡng, cho hệ số đổi liên tục theo tỉ số tín hiệu trên nhiễu ở từng vạch.

$$H_w[k] = \frac{|\hat{S}[k]|^2}{|\hat{S}[k]|^2 + |\hat{N}[k]|^2} \tag{7.10}$$

Đây là bộ lọc Wiener ở dạng đơn giản nhất. Khi tín hiệu mạnh hơn nhiễu nhiều thì hệ số gần 1; khi nhiễu lấn át thì hệ số gần 0; và ở giữa nó chuyển dần chứ không nhảy. Chính chỗ "chuyển dần" ấy làm nó ít sinh nhiễu nhạc.

7.6.5 So ba cách trên cùng một bản thu



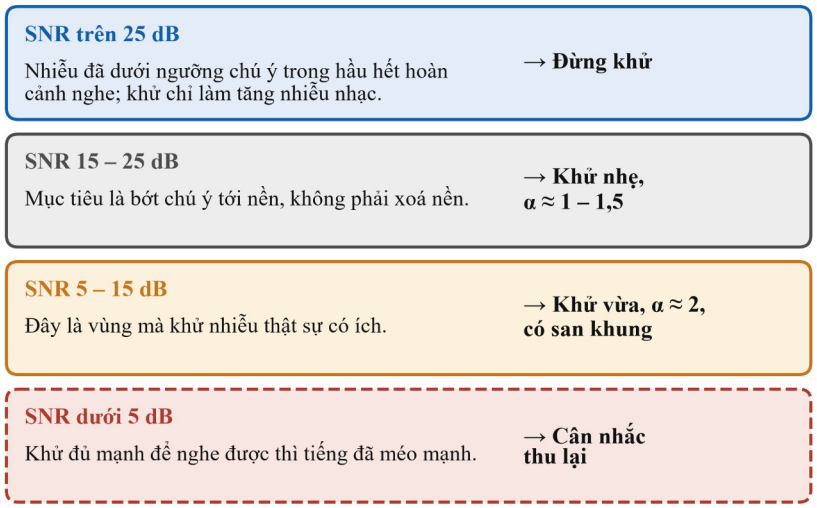
Hình 7.24 Ba cách khử nhiễu, đo bằng ba đại lượng

Bảng 7.8 Ba cách khử trên cùng một bản thu SNR 10 dB

Cách	Mức nền (dBFS)	Chất lượng tiếng (dB)	Độ nhọn phổ
Chưa khử	−28,0	11,7	10
Trừ phổ alpha = 2	−51,2	21,8	393
alpha = 2, san 4 khung	−54,0	21,4	224
Lọc Wiener	−33,5	16,7	26

Đọc ba cột cùng lúc thì thấy rõ không cách nào thắng ở cả ba. Trừ phổ hạ nền mạnh nhất nhưng để lại nhiễu nhạc nhiều nhất. San bốn khung giữ gần nguyên hai cột đầu mà hạ nhiễu nhạc gần một nửa — rẻ nhất trong ba cách, và là lựa chọn mặc định nên dùng. Wiener gần như không sinh nhiễu nhạc nhưng chỉ hạ nền được 5,5 decibel.

7.6.6 Khi nào đừng khử

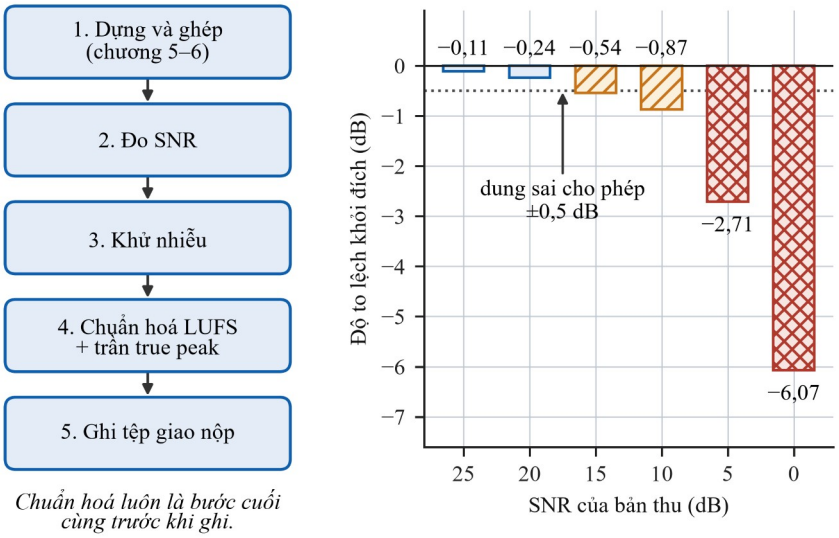


Hình 7.25 Bốn vùng SNR và quyết định tương ứng

Quyết định đầu tiên của mọi việc khử nhiễu là quyết định có khử hay không, và nó dựa trên một con số đo được. Trên 25 decibel thì nhiễu đã dưới ngưỡng chú ý trong hầu hết hoàn cảnh nghe, và khử chỉ tổ đổi một thứ không ai để ý lấy một thứ ai cũng để ý.

> **Cách thứ tư, và là cách tốt nhất**
>
> Không có thuật toán nào trong chương này bằng được việc *thu lại* cho sạch hơn.
>
> Ba mươi giây tắt điều hoà, dời micro gần miệng thêm hai mươi xăng-ti-mét, hoặc treo một tấm chăn sau lưng người nói — mỗi việc ấy đáng giá hơn mọi tham số alpha và beta cộng lại.
>
> Nói ra điều này *không* làm mất giá trị của mục 7.5. Nó đặt mục ấy đúng chỗ: khử nhiễu là thứ dùng khi không thu lại được, chứ không phải thứ thay cho việc thu cẩn thận.

7.6.7 Thứ tự các bước



(a) Thứ tự đúng (b) Độ hụt độ to khi chuẩn hoá trước

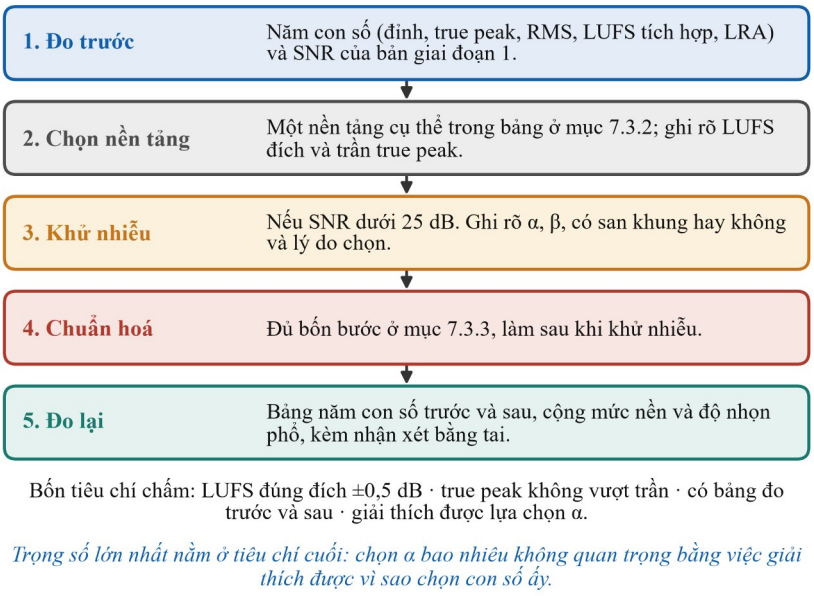
Hình 7.26 Thứ tự đúng, và cái giá của làm sai thứ tự

Chuẩn hoá *luôn* là bước cuối cùng trước khi ghi tệp. Lý do có thể đo được: nếu chuẩn hoá trước rồi mới khử nhiễu, phần nhiễu bị lấy đi cũng mang theo một phần độ to, và bản giao nộp hụt so với đích.

Hụt bao nhiêu thì tuỳ bản thu bẩn tới đâu: 0,11 decibel ở SNR 25 dB — không đáng kể; nhưng 6,07 decibel ở SNR 0 dB — sai hẳn. Cùng một đoạn mã, chỉ đổi thứ tự hai dòng.

7.7 Dự án thực hành — giai đoạn 2

7.7.1 Yêu cầu



Hình 7.27 Năm bước của giai đoạn 2

Giai đoạn 2 nên làm xong trước khi sang chương 10, làm tiếp trên chính bản dựng đã có ở giai đoạn 1. Không cần tư liệu mới; nếu bản giai đoạn 1 quá sạch (SNR trên 25 dB) thì thu thêm một đoạn ba mươi giây ở nơi có nhiễu nền rõ để có cái mà xử lý.

Bảng 7.9 Kết quả cần có

Hạng mục	Yêu cầu cụ thể
Bảng đo trước	Năm con số của bản giai đoạn 1: đỉnh, true peak, RMS, LUFS tích hợp, LRA. Cộng SNR đo từ đoạn lặng
Chọn nền tảng	Một nền tảng cụ thể trong bảng ở mục 7.3.2, ghi rõ LUFS đích và trần true peak
Khử nhiễu	Nếu SNR dưới 25 dB thì khử, ghi rõ alpha, beta, có san khung hay không, và *lý do* chọn từng con số
Chuẩn hoá	Đủ bốn bước ở mục 7.3.3, làm *sau* khi khử nhiễu. Ghi rõ trần có chặn lại hay không
Bảng đo sau	Cùng năm con số ấy, cộng mức nền và độ nhọn phổ nếu có khử nhiễu
Nhận xét	Nửa trang: nghe thấy khác gì, và có chỗ nào đánh đổi mà bạn thấy chưa thoả đáng

7.7.2 Cách tự kiểm

Bảng 7.10 Bốn tiêu chí tự kiểm giai đoạn 2

Tiêu chí	Trọng số	Cách kiểm
LUFS đúng đích ±0,5 dB	2	Đo tệp kết quả bằng bộ đo độc lập. Nếu trần chặn lại thì phải nói rõ trong bảng đo sau, và khi ấy vẫn coi là đạt
True peak không vượt trần	2	Đo bằng lấy mẫu quá 4 lần. Vượt trần là không đạt, không có ngoại lệ
Có đủ bảng đo trước và sau	2	Đủ năm con số ở cả hai bảng; các con số phải khớp khi chạy lại phép đo trên tệp kết quả
Giải thích được lựa chọn alpha	4	Đây là tiêu chí có trọng số lớn nhất. Chọn alpha bao nhiêu không quan trọng bằng việc nêu được đã thử những giá trị nào và vì sao dừng ở giá trị đã chọn

> **Vì sao tiêu chí cuối có trọng số lớn nhất**
>
> Ba tiêu chí đầu đều kiểm được bằng máy và đều có đáp án đúng duy nhất. Chúng kiểm xem quy trình có được làm đúng không.
>
> Tiêu chí cuối không có đáp án đúng duy nhất. Nó kiểm xem người làm có *hiểu rằng đang đánh đổi* cái gì lấy cái gì hay không.
>
> Một bản chọn alpha bằng 1,5 và giải thích được vì sao có giá trị hơn một bản chọn alpha bằng 2 mà không nói được lý do — dù bản thứ hai nghe có thể hay hơn.

7.7.3 Ba lỗi hay gặp nhất ở giai đoạn này

- **Chuẩn hoá trước khi khử nhiễu.** Đúng cái lỗi ở mục 7.6.7. Bản kết quả hụt so với đích mà người làm không biết, vì họ đo LUFS ngay sau bước chuẩn hoá chứ không đo lại trên tệp cuối cùng.
- **Khử nhiễu trên tệp đã sạch.** Không đo SNR trước, cứ khử vì thấy có nút khử. Kết quả là một bản có nhiễu nhạc rõ ràng hơn hẳn bản gốc — và trong bảng đo sau, độ nhọn phổ tố cáo điều đó.
- **Báo LUFS của tệp mono rồi xuất tệp stereo. Hoặc ngược lại. Chênh đúng 3,01 decibel, và lộ ra** ngay khi chạy lại phép đo.

7.8 Thực hành, bài tập và tổng kết

7.8.1 Ba bài thực hành

Ba bài dưới đây làm ngay sau khi đọc xong chương trên tư liệu của chính mình, mỗi bài khoảng mười phút.

- **Bài 1 — ba tệp, ba độ to.** Chuẩn hoá ba tệp đã chuẩn bị về đúng đỉnh −1 dBFS. Nghe lần lượt ở cùng mức loa và *xếp hạng* chúng theo độ to cảm nhận. Rồi đo LUFS của cả ba và so với thứ hạng vừa xếp. Nếu thứ tự khớp thì tai và thước đang nói cùng một chuyện.
- **Bài 2 — trần chặn lại.** Lấy tệp có PLR lớn nhất trong ba tệp và chuẩn hoá về −14 LUFS. Kiểm true peak. Nếu vượt trần thì đặt trần và ghi lại độ to thật đạt được. Trả lời: hụt bao nhiêu, và có chấp nhận được không.
- **Bài 3 — nghe cho được nhiễu nhạc.** Trên tệp có nhiễu, thử alpha bằng 1, 2, 3 và 6. Với mỗi giá trị, nghe riêng nửa giây *lặng* ở đầu và mô tả bằng lời cái nghe thấy. Rồi in độ nhọn phổ và xem con số có khớp với mô tả không.

7.8.2 Bảy bài tập

Bảng 7.11 Bài tập của chương 7

Bài	Nội dung	Đáp số hoặc tiêu chí
1	Một tệp đo được −31 LUFS. Cần cộng bao nhiêu dB để đạt chuẩn EBU R 128? Và để đạt chuẩn Spotify?	+8 dB; +17 dB
2	Một tệp mono đo −18 LUFS. Chép sang stereo (hai kênh giống hệt) thì đo được bao nhiêu?	−14,99 LUFS (tăng 3,01)
3	Một bản có LUFS −20 và true peak −0,4 dBTP. Chuẩn hoá về −16 LUFS với trần −1 dBTP thì đạt được độ to thật là bao nhiêu?	−20,6 LUFS — trần chặn lại, hụt 4,6 dB so với đích
4	Viết hàm bao_cao_muc(x, sr) in ra đủ năm con số của mục 7.2.7, kèm một dòng cảnh báo nếu true peak vượt −1 dBTP.	Hàm chạy và bắt được tệp vượt trần
5	Trộn nhiễu trắng vào một tệp sạch để đạt SNR đúng 12 dB. Kiểm lại bằng hàm ước lượng từ đoạn lặng.	Hai con số lệch nhau dưới 0,2 dB
6	Trên tệp ở bài 5, quét alpha từ 0,5 tới 6 theo bước 0,5. Vẽ ba đường: mức nền, chất lượng tiếng, độ nhọn phổ. Chỉ ra alpha tốt nhất.	Đồ thị có đủ ba đường; alpha tốt nhất nằm trong khoảng 1,5 – 2,5
7	Chứng minh bằng số rằng chuẩn hoá *trước* khi khử nhiễu cho kết quả sai. Làm trên hai mức SNR khác nhau và so hai sai lệch.	Sai lệch lớn hơn ở mức SNR thấp hơn

7.8.3 Sáu câu phải trả lời được

Ba tệp cùng đỉnh −1 dBFS có to bằng nhau không?	Không. Trên ba tư liệu của bài, LUFS chênh nhau 17,0 dB; đỉnh không nói gì về độ to.
Thước K làm gì?	Hạ dải trầm (−3,9 dB ở 50 Hz) và nâng dải cao (+4,1 dB ở 10 kHz) cho khớp với tai người.
Vì sao LUFS cần hai cổng?	Không có cổng, 24 s im lặng kéo con số tụt 6,7 dB; tệp nhiều khoảng lặng bị chuẩn hoá quá mức.
True peak khác đỉnh mẫu thế nào?	Đo sau khi lấy mẫu quá 4 lần. Tư liệu bị cắt đỉnh có thể vượt tới +3,95 dB so với đỉnh mẫu.
Đo SNR bằng cách nào khi chỉ có bản đã trộn?	Lấy công suất nhiễu từ một đoạn lặng, rồi trừ khỏi công suất tổng. Sai lệch dưới 0,1 dB.
Nhiễu nhạc từ đâu ra, và trả giá thế nào?	Từ những vạch nhấp nháy qua ngưỡng. Đỉnh ở $\alpha = 3$ với độ nhọn gấp hơn ba trăm lần nhiễu tự nhiên.

Chương 8 mở mô-đun M6 (hiệu ứng) từ chính vấn đề đã gặp ở mục 7.3.4: bộ nén động.

Hình 7.28 Sáu câu tổng kết chương 7

Cả sáu câu đều có đáp số bằng số và tự kiểm được trên máy. Nếu còn câu nào chưa trả lời được, hãy quay lại đúng mục tương ứng trước khi sang chương 8.

7.8.4 Ba điều đọng lại từ chương này

- **Đỉnh, RMS và LUFS đo ba thứ khác nhau.** Ba tệp cùng đỉnh −1 dBFS có LUFS chênh nhau 17 decibel. Dùng nhầm thước là nguyên nhân của phần lớn các bản giao nộp bị trả lại.
- **Mọi phép xử lý đều có ít nhất hai mặt.** Chuẩn hoá độ to đụng trần true peak. Khử nhiễu đụng nhiễu nhạc. Việc của người kỹ thuật là tìm ra cách *đo* mặt thứ hai, vì mặt thứ nhất bao giờ cũng dễ đo hơn.
- **Quyết định đầu tiên luôn là "có nên làm không".** Trên 25 decibel SNR thì đừng khử nhiễu. Nếu trần chặn lại quá 2 decibel thì vấn đề nằm ở bản dựng chứ không ở phép chuẩn hoá. Cả hai quyết định ấy đều dựa trên một con số đo được, không phải cảm tính.

7.8.5 Chuẩn bị cho chương 8

Chương 8 mở phần Hiệu ứng và bắt đầu từ chính thứ đã làm sai bản nhạc cổ điển ở mục 7.3.4: PLR quá cao thì không chuẩn hoá về mức to được. Công cụ chữa nó là bộ nén động.

- Giữ lại bản nhạc cổ điển đã dùng ở chương này. Chương 8 sẽ nén nó và đo lại PLR để xem chuẩn hoá về −14 LUFS có được không.
- Chuẩn bị thêm một đoạn thu có mức lên xuống thất thường — chẳng hạn một cuộc phỏng vấn hai người ngồi cách micro không đều nhau. Đó là tư liệu chuẩn cho phần nén động.

– Đọc lại mục 4.6 về phương trình sai phân và mục 6.4 về phép trễ. Chương 8 xây thẳng trên hai mục ấy: bộ lọc là phương trình sai phân, còn hiệu ứng trễ và hồi tiếp là phép trễ đặt trong vòng phản hồi.

7.8.6 Mô-đun am sau chương 7

Liệt kê 7.6 Tệp mới thêm vào mô-đun

```
am/
    __init__.py
    io.py
    bien_do.py
    thoi_gian.py
    khoi.py
    do_to.py        <- MOI: thuoc K, LUFS, LRA, true peak
    khu_nhieu.py    <- MOI: SNR, tru pho, Wiener
    test_am.py

# am/do_to.py
bo_loc_k, ap_thuoc_k
do_to_tuc_thoi, do_to_ngan_han, do_to_tich_hop, dai_do_to
true_peak_dbtp, chenh_true_peak_db
gain_de_dat_lufs, chuan_hoa_lufs, do_to_theo_nen_tang

# am/khu_nhieu.py
snr_db, snr_tu_doan_lang, san_nhieu_nen
pho_nhieu_uoc_luong, tru_pho, tru_pho_san_khung, loc_wiener
do_nhieu_nhac, meo_sau_khu_nhieu
```

Mười tám hàm của chương 6 nay thành ba mươi bốn. Cấu trúc không đổi, chỉ mọc thêm hai tệp — đúng như đã nói ở mục 6.7.4. Nhớ quy ước: mỗi hàm mới phải kèm ít nhất một phép kiểm trong test_am.py trước khi được dùng trong dự án thực hành.

Quy ước từ chương 6 vẫn giữ nguyên: mỗi hàm mới phải kèm ít nhất một phép kiểm trước khi được dùng trong dự án thực hành. Bốn loại phép kiểm ở mục 6.7.3 áp dụng thẳng vào các hàm mới của chương này.

Liệt kê 7.7 Bốn phép kiểm cho các hàm mới

```
# 1. GIA TRI DA BIET — tinh duoc bang tay
#    sin 1 kHz -20 dBFS mono phai doc -23,0 LUFS
x = sine(1000, 4.0, amp=0.1)
assert abs(do_to_tich_hop(x, sr) + 23.0) < 0.05

# 2. TINH CHAT BAT BIEN — dung voi moi dau vao
#    chuyen mono sang stereo thi LUFS tang dung 3,01
st = np.stack([x, x], axis=1)
assert abs(do_to_tich_hop(st, sr)
           - do_to_tich_hop(x, sr) - 3.01) < 0.05

# 3. TRUONG HOP BIEN — dau vao o ria mien xac dinh
#    doan ngan hon mot khoi 400 ms thi khong do duoc
assert not np.isfinite(do_to_tich_hop(sine(1000, 0.1), sr))

# 4. SO VOI CACH CHAM — hai cach, mot ket qua
#    true peak phai >= dinh mau, luon luon
assert true_peak_dbtp(x) >= peak_dbfs(x) - 1e-9
```

Một phép kiểm đáng thêm cho riêng chương này

Các hàm khử nhiễu khó kiểm bằng một dấu bằng, vì kết quả phụ thuộc tư liệu. Nhưng có một tính chất luôn đúng và kiểm được: trừ phổ với alpha bằng 0 phải trả về gần đúng tín hiệu vào.

Viết assert cho tính chất ấy thì bắt được mọi lỗi ở khâu cắt khung và dựng lại — vốn là chỗ hay sai nhất khi tự cài đặt STFT.

Bộ tự kiểm của sigtools dùng chính cách này, và nhờ nó phát hiện ra một lỗi thật ở phần đệm hai biên khi dựng sách này.

Tài liệu tham khảo của chương

Chuẩn và khuyến nghị — nguồn gốc của mọi con số

- ITU-R BS.1770-4 — Algorithms to measure audio programme loudness and true-peak audio level. Đây là nguồn *trực tiếp* của mục 7.2 và 7.3: thước K, công thức độ to, hai cái cổng, và phép đo true peak bằng lấy mẫu quá bốn lần.
- EBU R 128 — Loudness normalisation and permitted maximum level of audio signals. Nguồn của mức đích −23 LUFS và của khái niệm dải độ to.
- EBU Tech 3341 và 3342 — hai tài liệu kỹ thuật đi kèm R 128, mô tả chi tiết ba thang thời gian và cách tính LRA.
- ATSC A/85 — khuyến nghị tương ứng của Bắc Mỹ, mức đích −24 LKFS.
- AES17 — quy ước đo lường cho thiết bị âm thanh số.

Sách và bài báo

- Boll, S. F. (1979) — Suppression of Acoustic Noise in Speech Using Spectral Subtraction. Bài báo gốc của phương pháp ở mục 7.5. Chỉ chín trang và đọc được với kiến thức của chương này.
- Berouti, M., Schwartz, R., Makhoul, J. (1979) — Enhancement of Speech Corrupted by Acoustic Noise. Nguồn của hai tham số alpha và beta ở mục 7.5.2.
- Loizou, P. C. — Speech Enhancement: Theory and Practice. Sách chuyên khảo đầy đủ nhất về đề tài của mục 7.5 và 7.6.
- Katz, B. — Mastering Audio: The Art and the Science. Chương về độ to và dải động, viết từ góc nhìn người làm nghề chứ không phải người làm chuẩn.
- Uemura, Y. và cộng sự (2009) — Musical noise generation analysis for noise reduction methods based on spectral subtraction. Nguồn của cách dùng độ nhọn để đo nhiễu nhạc ở mục 7.6.2.

Công cụ nên biết

- pyloudnorm — thư viện Python đo LUFS theo BS.1770. Dùng để đối chiếu kết quả của mô-đun tự viết; các con số trong sách này đã được đối chiếu với nó và lệch dưới 0,05 dB.

- ffmpeg với bộ lọc ebur128 — đo LUFS, LRA và true peak từ dòng lệnh, tiện cho việc kiểm hàng loạt tệp.
- Youlean Loudness Meter hoặc bất kỳ bộ đo nào tuân thủ EBU R 128 — để nhìn ba thang thời gian chạy theo thời gian thực.

Đọc thêm theo hướng

- **Muốn đi sâu về độ to:** tìm đọc về mô hình độ to của Moore và Glasberg. Chúng chính xác hơn thước K nhiều nhưng tốn tính toán hơn nhiều, và không dùng được cho đo thời gian thực — lý do BS.1770 chọn một mô hình đơn giản hơn.
- **Muốn đi sâu về khử nhiễu:** tìm đọc về ước lượng nhiễu *thích nghi* — cập nhật phổ nhiễu liên tục trong lúc xử lý thay vì đo một lần ở đầu tệp. Đó là cách các công cụ hiện nay xử lý được cả nhiễu không dừng.
- **Muốn đi sâu về đánh giá chất lượng:** tìm đọc PESQ và POLQA — hai phép đo tự động cho điểm chất lượng tiếng nói tương quan với đánh giá của người nghe. Chúng làm đúng việc mà hàm meo_sau_khu_nhieu của bài chỉ làm được ở mức thô sơ.

Ghi chú về phiên bản thư viện

Mã trong bài viết theo numpy 1.26 trở lên và scipy 1.11 trở lên. Hai chỗ đáng lưu ý.

- scipy.signal.resample_poly dùng cho phép đo true peak: bộ lọc mặc định của nó có gợn nhỏ, nên kết quả có thể lệch khoảng 0,02 dB so với phép nội suy lý tưởng. Không đáng kể cho mục đích của bài.
- Nếu đo true peak trên một đoạn cắt cụt ở hai đầu, phần dựng lại ở biên có thể tăng lên giả tạo. Trong bài, mọi tín hiệu dùng để đo true peak đều đã fade hai đầu.

Một lưu ý về các con số trong bài

> **Cách đọc các con số**
>
> Các con số về đáp ứng thước K, về mức đích của các nền tảng, và về quan hệ mono–stereo là kết quả tính chính xác từ chuẩn.
>
> Các con số về ba tư liệu (nhạc pop, cổ điển, giọng đọc) là kết quả *đo* trên tư liệu dựng sẵn mô tả trong chú thích từng hình. Trên tư liệu thật, con số sẽ khác; điều không đổi là thứ tự và bậc độ lớn của chênh lệch.
>
> Các con số về khử nhiễu — mức nền, chất lượng tiếng, độ nhọn phổ — phụ thuộc vào cả tư liệu lẫn mức nhiễu. Điều đáng nhớ không phải con số mà là *hình dạng* của ba đường ở hình 7.24.
>
> Phần độ to đã được đối chiếu với thư viện tham chiếu pyloudnorm và lệch dưới 0,05 dB. Mọi phép tính trong bài sinh ra từ mô-đun sigtools.py, đã qua 268 phép tự kiểm.