LÊ KHẮC TUẤN

SÁCH THAM KHẢO

LẬP TRÌNH ÂM THANH

Từ sóng âm đến Web Audio và AI tạo sinh — nền tảng tham khảo cho người yêu thích lập trình và âm thanh

Cơ sở âm thanh và tín hiệu số · Lập trình xử lý âm thanh · Hiệu ứng và không gian âm thanh · Ứng dụng âm thanh trên web · AI tạo sinh, tích hợp và đạo đức

www.lekhactuan.com

HÃY CỨ KHÁT KHAO, HÃY CỨ DẠI KHỜ

HÀ NỘI · 2026

MỤC LỤC

Mở đầu. Tổng quan cuốn sách 5
Vì sao cần cuốn sách Lập trình âm thanh 5
Mục đích, đối tượng và phạm vi 5
Ba ý xuyên suốt cuốn sách 6
Cấu trúc cuốn sách 6
Cấu trúc một chương 7
Cách sử dụng cuốn sách 8
Môi trường và công cụ 8
Quy ước trình bày 9
Đọc xong cuốn sách, người đọc làm được gì 9

DANH MỤC HÌNH VẼ

Hình MĐ.1 Mười lăm chương trong năm phần: từ sóng âm đến một sản phẩm âm thanh chạy được..... 6

DANH MỤC BẢNG BIỂU

Bảng MĐ.1 Năm phần của cuốn sách.....7
Bảng MĐ.2 Gợi ý đường đọc theo mục tiêu.....8
Bảng MĐ.3 Ý nghĩa của các khung màu.....9

MỞ ĐẦU
TỔNG QUAN CUỐN SÁCH

Mục tiêu của chương

Đọc xong phần này, người đọc nắm được:

- Cuốn sách Lập trình âm thanh viết về điều gì, viết cho ai, và vì sao nên học theo cách tự viết mã.
- Cấu trúc năm phần, mười lăm chương và mối liên hệ giữa chúng.
- Cách đọc, cách luyện tập và cách chọn đường đọc phù hợp với mục tiêu của mình.
- Môi trường làm việc cần chuẩn bị và các quy ước trình bày dùng trong sách.

Vì sao cần cuốn sách Lập trình âm thanh

Âm thanh có mặt trong hầu hết sản phẩm số: nhạc và phim, trò chơi, cuộc gọi video, trợ lý ảo, podcast, ứng dụng trên web. Nhưng với đa số người làm phần mềm, âm thanh vẫn là một chiếc hộp đen: gọi một hàm để phát, gọi một thư viện để lọc, rồi hy vọng kết quả nghe được. Khi kết quả sai thì không biết sai ở đâu. Tiếng bị "tách", bị méo, quá to hay quá nhỏ, lệch pha giữa hai kênh, tệp nặng bất thường, độ trễ làm người dùng khó chịu — tất cả đều có nguyên nhân cụ thể mà người chỉ biết gọi hàm không thấy được.

Cuốn sách này mở chiếc hộp đen ấy ra. Điều cần nhớ đầu tiên là một tệp âm thanh trên máy tính, về bản chất, chỉ là một dãy số. Lọc, nén, thêm vang, trộn nhiều lớp, sinh nhạc bằng AI đều là cách biến đổi dãy số đó. Hiểu dãy số ấy từ đâu ra, mang ý nghĩa vật lý gì và biến đổi nó bằng phép tính nào thì người đọc tự viết được các công cụ xử lý âm thanh, thay vì chỉ dùng công cụ của người khác.

Bối cảnh hiện nay làm điều đó càng quan trọng. Các mô hình AI sinh nhạc và giọng nói cho ra kết quả trong vài giây; các trợ lý lập trình viết được mã xử lý theo yêu cầu. Nhưng kết quả của AI vẫn là một dãy số, và người dùng vẫn phải kiểm tra nó có đúng mức, đúng tần số lấy mẫu, có méo hay có tiếng tách không. Người đã tự viết được các phép xử lý cơ bản thì đọc và kiểm tra được thứ mà AI trả về; người chưa từng tự viết thì chỉ có thể tin.

Mục đích, đối tượng và phạm vi

Mục đích. Sau khi đọc hết cuốn sách, người đọc tự viết được một bộ thư viện xử lý âm thanh của riêng mình, đưa được bộ thư viện ấy vào một sản phẩm chạy được (một ứng dụng web hoặc một trò chơi), và biết đo chất lượng bằng con số chứ không chỉ bằng tai.

Đối tượng. Cuốn sách viết cho người yêu thích lập trình và âm thanh: người tự học, người làm sản phẩm đa phương tiện, và bất kỳ ai làm web, trò chơi hay phần mềm có dùng âm thanh. Người đọc cần biết lập trình cơ bản (biến, hàm, vòng lặp, mảng) và toán phổ thông (luỹ thừa, logarit, lượng giác). Không cần biết trước nhạc lý hay điện tử. Những khái niệm khó hơn như số phức, tích chập, phương trình sai phân được giải thích ngay tại chỗ gặp đầu tiên.

Phạm vi. Cuốn sách tập trung vào kỹ thuật xử lý tín hiệu âm thanh bằng mã. Các nội dung sau chỉ được nhắc ở mức cần thiết chứ không đi sâu: âm học kiến trúc, nhạc lý và sáng tác, thiết kế âm thanh nghệ thuật, điện tử âm thanh, và lý thuyết học máy chi tiết. Người muốn đi sâu vào những hướng ấy sẽ tìm thấy sách chuyên khảo ở phần tài liệu tham khảo cuối mỗi chương.

Ba ý xuyên suốt cuốn sách

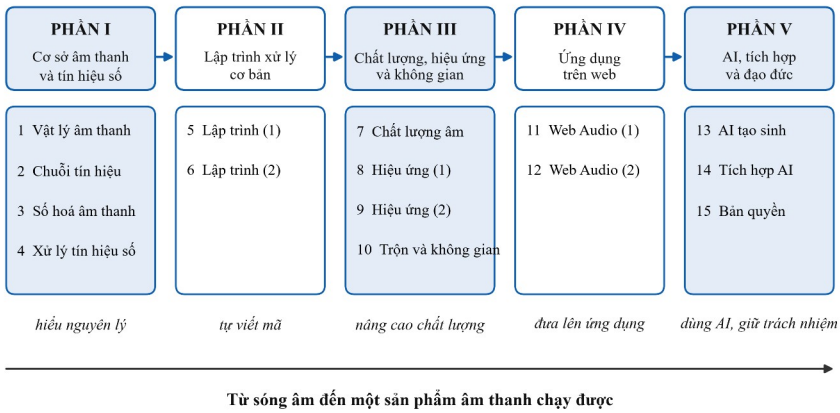
– **Mọi thứ là một dãy số.** Từ sóng áp suất không khí tới tệp WAV, từ phổ tới đường bao, từ tín hiệu của một nhạc cụ tới bản trộn của cả bài hát. Mỗi lần gặp một khái niệm mới, hãy hỏi: nó là phép biến đổi nào của dãy số?

– **Mọi điều cần được đo bằng số.** Mức tín hiệu theo dBFS, độ to theo LUFS, độ trễ theo mili giây, tỉ số tín hiệu trên nhiễu theo dB. Cuốn sách luôn đi từ cảm giác ("nghe to", "nghe đục") sang con số kiểm chứng được, và đoạn mã nào cũng kèm cách kiểm tra kết quả.

– **Nguyên lý trước, mã sau; tự viết trước, dùng thư viện sau.** Bốn chương đầu chưa yêu cầu viết dòng mã nào mà dựng nền nguyên lý. Từ chương 5, mọi dòng mã là cách diễn đạt lại một nguyên lý đã học bằng ngôn ngữ máy tính. Thư viện có sẵn chỉ được dùng để đối chiếu, sau khi người đọc đã tự viết được phiên bản của mình.

Cấu trúc cuốn sách

Cuốn sách gồm mười lăm chương, chia thành năm phần theo đường đi từ hiểu nguyên lý, tới tự viết mã, tới xử lý nâng cao, tới ứng dụng, rồi tới trí tuệ nhân tạo. Hình MĐ.1 vẽ bản đồ ấy, và Bảng MĐ.1 ghi nội dung chính của từng phần.



Hình MĐ.1 Mười lăm chương trong năm phần: từ sóng âm đến một sản phẩm âm thanh chạy được

Bảng MĐ.1 Năm phần của cuốn sách

Phần	Chương	Nội dung chính	Người đọc làm được
I. Cơ sở âm thanh và tín hiệu số	1 – 4	Bản chất vật lý của sóng âm; chuỗi tín hiệu analog; lấy mẫu, lượng tử hoá và dBFS; biến đổi Fourier, phổ, cửa sổ, bộ lọc số và tích chập.	Giải thích và tính toán các đại lượng của âm thanh; đọc đúng một biểu đồ phổ.
II. Lập trình xử lý cơ bản	5 – 6	Mảng tín hiệu, đọc ghi tệp WAV, chỉnh gain theo decibel, trộn, cắt ghép, chuẩn hoá; fade, crossfade, trễ, đảo, xử lý theo khối; đóng gói mô-đun.	Tự viết các hàm xử lý nền và đóng gói thành thư viện dùng lại được.
III. Chất lượng, hiệu ứng và không gian	7 – 10	Đo mức và độ to, khử nhiễu; bộ nén, giới hạn, cổng, lọc lược, hồi tiếp, điều biến; bộ lọc bậc hai, cân bằng tần số, vang; trộn nhiều lớp, panning, binaural.	Làm sạch bản ghi, dựng hiệu ứng và trộn nhiều lớp thành một bản cân bằng.
IV. Ứng dụng trên web	11 – 12	Mô hình đồ thị node của Web Audio, ngữ cảnh và đồng hồ âm thanh, AudioWorklet, lập lịch tham số, rack hiệu ứng, tối ưu tài nguyên, kiểm thử.	Dựng một ứng dụng âm thanh thời gian thực chạy trong trình duyệt.
V. AI tạo sinh, tích hợp và đạo đức	13 – 15	Biểu diễn âm thanh cho mô hình, các họ mô hình sinh nhạc và giọng, cách đánh giá; đường ống hậu kỳ cho kết quả AI; dấu vết codec, thuỷ vân, vân tay, phát hiện giọng tổng hợp, bản quyền.	Đưa kết quả của AI vào sản phẩm đúng kỹ thuật và đúng trách nhiệm.

Sự phụ thuộc giữa các chương. Phần I và phần II nên đọc theo thứ tự, vì chương sau dùng lại khái niệm của chương trước: decibel, dBFS, phổ, gain. Từ phần III trở đi, mỗi chương có thể đọc tương đối độc lập, miễn đã nắm phần I – II. Riêng phần IV cần chương 5 – 6 vì dựng lại bằng JavaScript những thuật toán đã viết bằng Python.

Cấu trúc một chương

Các chương được viết theo cùng một khuôn để người đọc quen nhịp đọc. Một chương gồm:

- **Mục tiêu của chương** và vị trí của chương so với các chương trước và sau.
- **Các mục nội dung chính,** mỗi mục mở bằng một câu hỏi hoặc một hiện tượng quen thuộc, rồi đi đến nguyên lý, ví dụ số và hình minh hoạ.

- **Hướng dẫn thực hành** với đoạn mã chạy được và kết quả mẫu để đối chiếu.
- **Bài tập tính toán có đáp số,** kèm cách trình bày bài làm; câu hỏi ôn tập và bài thảo luận bằng tình huống thực tế.
- **Tóm tắt,** bản đồ khái niệm, bảng thuật ngữ Anh – Việt và bảng tự kiểm trước khi sang chương sau.
- **Tài liệu tham khảo có chú giải:** mỗi nguồn ghi rõ nó là nguồn của phần nào trong chương, kèm gợi ý đọc thêm theo hướng quan tâm.

Cách sử dụng cuốn sách

Cách đọc. Đọc để hiểu, rồi gõ lại và chạy từng đoạn mã. Với mỗi hình có đường cong hay phổ, hãy thử đoán hình trước khi đọc chú thích. Làm bài tập tính toán trên giấy, đối chiếu đáp số, rồi mới xem lại phần giải thích khi lệch. Phần nghe là bắt buộc: các đoạn mã đều nên được nghe lại bằng tai nghe, vì có những lỗi con số không cho thấy ngay. Không nhất thiết phải đọc từ đầu đến cuối: Bảng MĐ.2 gợi ý đường đọc theo mục tiêu.

Dùng AI khi học. Trợ lý AI là công cụ tốt để hỏi về cú pháp hay giải thích một dòng mã. Quy tắc nên giữ: chỉ nhờ AI viết khi đã biết mình cần nó làm gì và đã tự dự đoán kết quả; sau khi có mã, giải thích lại được từng dòng quan trọng bằng lời của mình, và kiểm tra kết quả bằng một phép đo cụ thể.

Bảng MĐ.2 Gợi ý đường đọc theo mục tiêu

Muốn	Nên đọc	Cần đã đọc
Hiểu nền tảng của âm thanh số	Chương 1 – 4	Không
Tự viết các hàm xử lý cơ bản	Chương 5 – 6	Chương 1, 3 (decibel, dBFS, lấy mẫu)
Làm sạch bản ghi, chuẩn hoá độ to	Chương 7	Chương 4 (phổ), 5 – 6
Làm hiệu ứng: nén, lọc, vang	Chương 8 – 9	Chương 4 (bộ lọc), 5 – 6
Trộn nhiều lớp, âm thanh không gian	Chương 10	Chương 7 – 9
Dựng ứng dụng âm thanh trên web	Chương 11 – 12	Chương 5 – 6
Dùng kết quả của AI trong sản phẩm	Chương 13 – 14	Chương 3, 7
Hiểu bản quyền và phát hiện vi phạm	Chương 15	Chương 3 (nén), 4 (phổ)

Môi trường và công cụ

Phần lập trình của cuốn sách dùng Python 3 (từ phiên bản 3.10 trở lên) với các thư viện: numpy để giữ và tính trên dãy mẫu, scipy để lọc và biến đổi, soundfile để đọc ghi tệp âm thanh, matplotlib để vẽ dạng sóng và phổ (thêm sounddevice để

phát âm thanh). Cài đặt và đoạn mã kiểm tra môi trường nằm ở mục thực hành của chương 1. Phần ứng dụng trên web (chương 11 – 12) cần một trình duyệt hiện đại như Chrome hoặc Edge và một trình soạn mã bất kỳ.

Nên cài thêm Audacity hoặc một phần mềm tương tự để nhìn và nghe lại kết quả: nó là cách đối chiếu độc lập với mã tự viết. Các chương về AI (13 – 14) không đòi hỏi huấn luyện mô hình: phần thực hành làm việc trên kết quả mà mô hình đã sinh, tập trung vào đo, kiểm tra và xử lý kết quả ấy.

Quy ước trình bày

- **Đánh số.** Mục đánh số tối đa ba cấp (ví dụ 4.2.1). Hình, bảng, công thức và đoạn mã đánh số theo chương: Hình 4.5, Bảng 4.3, công thức (4.1), Liệt kê 4.1. Chú thích hình đặt dưới hình, chú thích bảng đặt trên bảng.
- **Số và đơn vị.** Dấu phẩy là dấu thập phân, dấu chấm tách hàng nghìn (ví dụ 44.100 Hz, 0,5 dB). Đơn vị viết theo hệ SI, tên viết tắt tiếng Anh như dBFS, LUFS, FFT giữ nguyên và được giải thích ở lần xuất hiện đầu.
- **Thuật ngữ.** Thuật ngữ tiếng Anh đặt trong ngoặc ở lần đầu, kèm bảng Anh – Việt ở cuối mỗi chương. Chữ in nghiêng nhấn mạnh khái niệm quan trọng.
- **Mã nguồn.** Đoạn mã đặt trong khung xám, phông đơn cách, đánh số "Liệt kê". Tên hàm và biến đặt bằng tiếng Việt không dấu cho dễ đọc và thống nhất giữa các chương.

Các khung màu trong bài có ý nghĩa quy ước, tóm tắt ở Bảng MĐ.3.

Bảng MĐ.3 Ý nghĩa của các khung màu

Khung	Dùng để
Khung xanh lá	Gợi ý cách học, việc nên thử ngay, mẹo thực hành.
Khung vàng	Một quy ước hoặc một điều cần lưu ý khi áp dụng.
Khung đỏ nhạt	Lỗi hay gặp, điều dễ hiểu sai.
Khung xanh dương	Giải thích bổ sung, liên hệ với chương khác.

Đọc xong cuốn sách, người đọc làm được gì

- Giải thích được bản chất vật lý của âm thanh và dùng thành thạo các thang đo: tần số, biên độ, decibel, dBFS, độ to.
- Mô tả được đường đi từ sóng âm thành dãy số trong máy tính, và chọn đúng tần số lấy mẫu, độ sâu bit cho từng mục đích.
- Đọc đúng một biểu đồ phổ và phổ đồ; hiểu ảnh hưởng của cửa sổ và độ dài khung đến độ phân giải.
- Tự viết được các hàm xử lý nền (gain, trộn, cắt ghép, fade, trễ) và đóng gói thành mô-đun dùng lại được.

- Làm sạch bản ghi và chuẩn hoá độ to; dựng được hiệu ứng nén, lọc, cân bằng, vang; trộn nhiều lớp và đặt chúng vào không gian âm thanh.
- Dựng được ứng dụng âm thanh thời gian thực trên Web Audio và kiểm thử nó bằng phép đo.
- Đưa kết quả của mô hình AI vào sản phẩm đúng kỹ thuật: đổi tần số lấy mẫu, chuẩn hoá mức, ghép đoạn, hậu kỳ, đo kiểm tự động.
- Hiểu các vấn đề bản quyền, đạo đức và các kỹ thuật phát hiện vi phạm: dấu vết codec, thuỷ vân, vân tay âm thanh, phát hiện giọng tổng hợp.

> **Lời khuyên để đọc cuốn sách này hiệu quả**
> Gõ lại mọi đoạn mã, đừng chép. Nghe lại mọi kết quả bằng tai nghe. Tự dự đoán đáp số trước khi chạy. Khi hai thứ không khớp — tai nghe khác con số, mã khác đáp số — đó chính là chỗ học được nhiều nhất, đừng bỏ qua.