آموزش پروژه‌محور پردازش صوت و گفتار با پایتون

فصل اول: سیگنال، ریاضیات و ابزارها

درس 1:

کالبدشکافی صدا در پایتون

طراح و مدرس: مصطفی کرمانی‌نیا

تحت نظارت علمی: دکتر مهدی سیفی‌پور

Made with GAMMA

صدا؛ دیتای قدرتمندی که نمی‌بینیم!

از تشخیص آهنگ با Shazam تا دستیارهای صوتی و حتی تشخیص بیماری از روی صدا، همه این سیستم‌ها از یک منطق مشترک استفاده می‌کنند.

با هم یاد می‌گیریم چطور امواج نامرئی صدا را به داده تبدیل کنیم و با استفاده از پایتون روی آن‌ها پردازش انجام دهیم.

Made with GAMMA

نقشه راه: اولین دیدار با صدا

ماهیت فیزیکی صدا
صدا دقیقا چیست و چطور به زبان ریاضی ترجمه می‌شود؟

دیجیتالی‌سازی (ADC)
کامپیوتر که فقط صفر و یک می‌فهمد، چطور صدای پیوسته را درک می‌کند؟

نرخ نمونه‌برداری و عمق بیت
مفاهیم نرخ نمونه‌برداری و عمق بیت به زبان ساده.

قضیه نایکوئیست
قانون طلایی برای ضبط درست صدا.

کدتایم!
پیاده‌سازی عملی با `librosa` و `matplotlib` در پایتون.

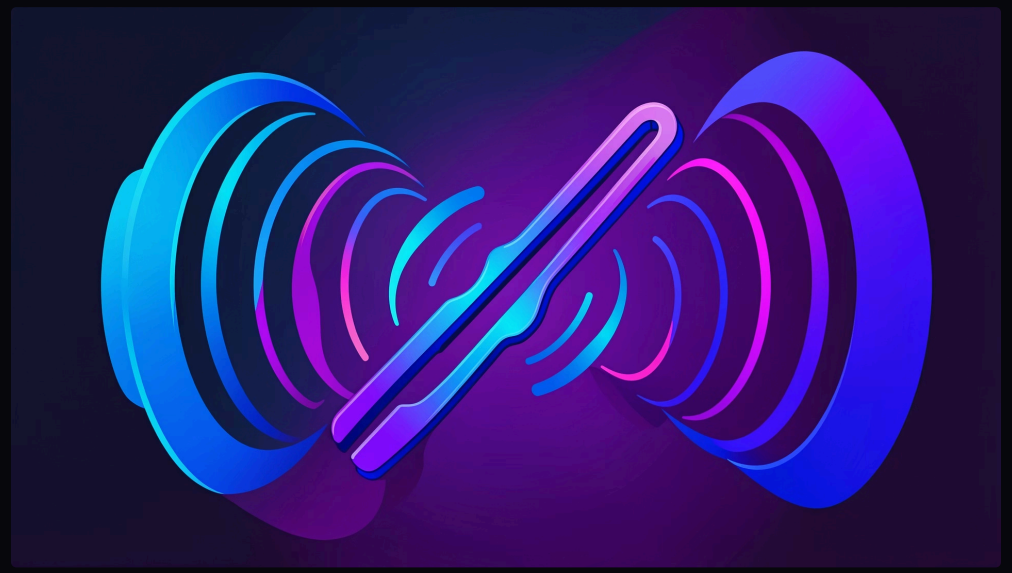
Made with GAMMA

صدا چیست؟

صدا یک موج مکانیکی است که از ارتعاشات (Vibrations) به وجود می‌آید و از طریق تغییرات فشار هوا منتقل می‌شود.

اگر بخواهیم ساده‌ترین شکل صدا (یک بوق خالص) را مدل‌سازی کنیم، به یک موج سینوسی می‌رسیم:

$$y(t) = A\sin(2\pi f t)$$



A — دامنه

Amplitude → بلندی صدا (Loudness)

f — فرکانس

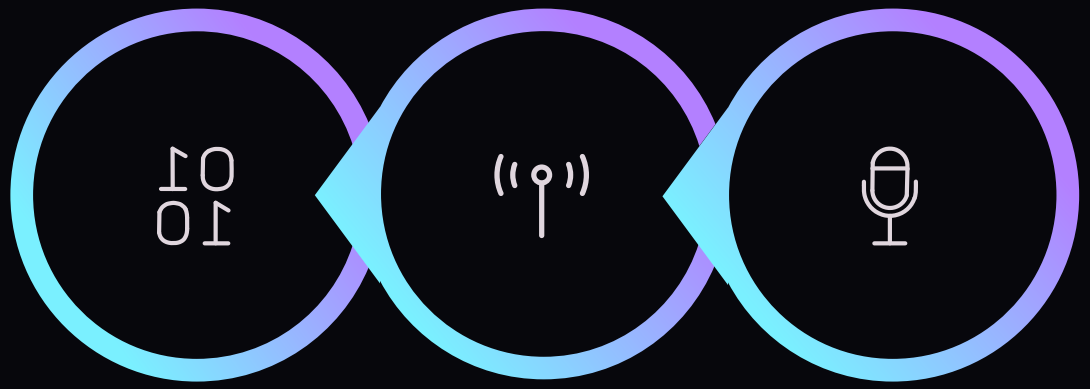
Frequency → زیر و بمی صدا (Pitch)

t — زمان

Time → متغیر اصلی حرکت موج

Made with GAMMA

پل ارتباطی دنیای واقعی و کامپیوتر



ضبط صوت

نمونه‌برداری ADC

داده‌های دیجیتال

کامپیوترها امواج پیوسته فیزیکی رو درک نمی‌کنن – فقط دیتا می‌فهمن! قطعه **ADC** در کارت صدا مثل یک مترجم عمل می‌کنه و با گرفتن «عکس‌های لحظه‌ای» از موج، اون رو به اعداد دیجیتال تبدیل می‌کنه.

Made with GAMMA

چطور کیفیت صدا را اندازه بگیریم؟

نرخ نمونه‌برداری

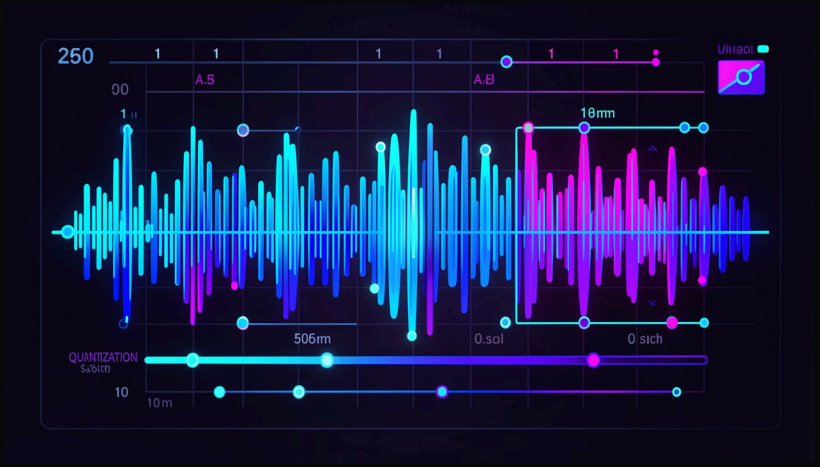
در هر ثانیه چند بار از موج «عکس» می‌گیریم؟ واحد: هرتز (Hz)

مثال: کیفیت استاندارد = CD 44,100 Hz

عمق بیت (Bit Depth)

هر بار که عکس می‌گیریم، با چه دقتی بلندی صدا (دامنه) را ثبت می‌کنیم؟

مثال: صدای bit-16 → تا 65,536 سطح مختلف برای دامنه!



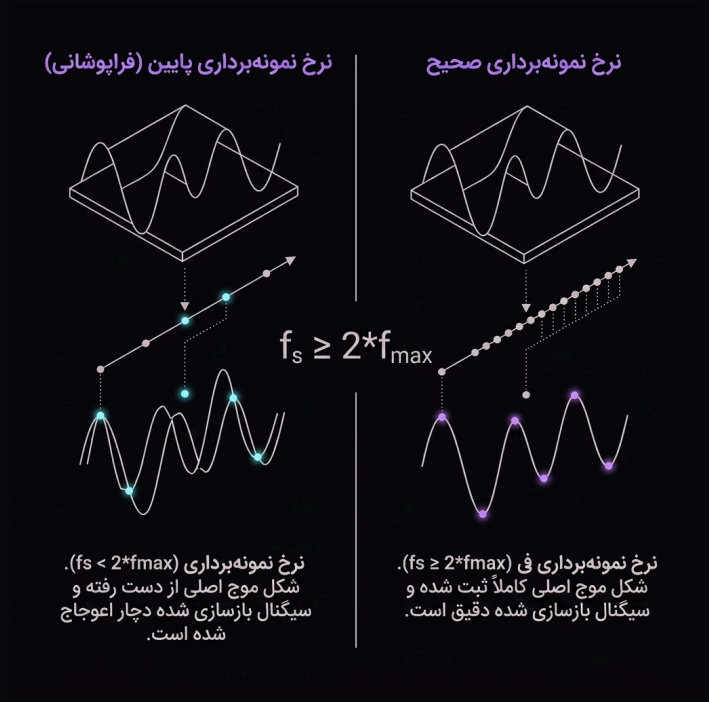
Made with GAMMA

قانون طلایی نمونه‌برداری: قضیه نایکوئیست

فرمول اصلی

$$f_s \geq 2f_{max}$$

نرخ نمونه‌برداری باید حداقل دو برابر بالاترین فرکانس موجود در سیگنال باشه



مثال جالب:

- گوش انسان حداکثر تا فرکانس ۲۰,۰۰۰ هرتز را می‌شنود.
- پس حداقل به نرخ ۴۰,۰۰۰ هرتز نیاز داریم تا تمام صداهای قابل شنیدن به درستی ثبت شوند.
- پس استاندارد CD روی ۴۴,۱۰۰ هرتز تنظیم شده است.

Made with GAMMA

کدتایم!

کالبدشکافی صدا در عمل

نصب کتابخانه

لود کردن فایل صوتی

رسم شکل موج

Made with GAMMA