DAY 27 · Part 3 · What It Can Do

How does your phone write down what you say? — Voice

Five-Minute AI Lessons · Oct 3, 2026

A voice is vibration chopped into tiny pieces, and text is the answer you get by matching each piece with its neighbors.

In a nutshell

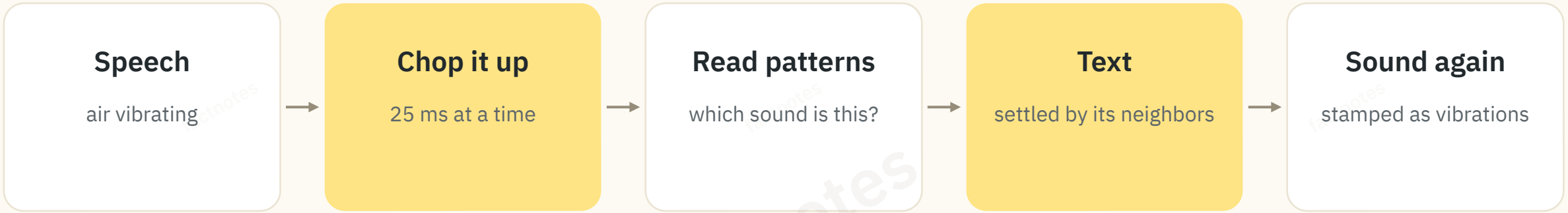
1. AI doesn't hear your words in one go — it **chops them into tiny pieces**. It writes the sound down as numbers 16,000 times a second, then sweeps over them with a 25-millisecond window.

2. Each piece on its own is just a "pitch pattern." What you actually said only gets decided **by looking at the pieces around it** — like the video you saw yesterday, one slot alone doesn't tell you.

3. Speaking is the same road in reverse. It turns text into a sound pattern, then follows that pattern to **stamp out air vibrations more than 16,000 times a second** and send them out of the speaker.

In one picture

Speech → text → speech, one full loop



The two middle boxes are the key: chop it small, and look at the neighbors together.

Key idea: a voice is vibration chopped into tiny pieces, and text is what you pick by matching each piece with its neighbors

How "Can you write it down" becomes text — and then sound again

1. Chop up the sound

Cut 1 second into tiny slots

0–25ms | 25–50ms | 50–75ms | 75–100ms | ... | 1 second

One piece lasts just a sliver of a single blink. It has to be sliced this thin so that close sounds like "b" an...

2. Pieces into words

The neighbors decide the word

Can **you** **write** **it** **down** right

By sound alone, "write" and "right" are exactly the same. So AI picks by checking which one fits with the words...

3. Words back into so...

Pattern first, then vibrations

Text | **Pitch pattern** | **Vibration values** | **Speaker** | Done

Speech is really just a record of air vibrating. Stamp those values out in order and you get a voice — it can e...

Tap the buttons to change steps. You'll see speech get chopped into pieces, the pieces join into words, and the...

Try it yourself in the web lesson · next-token.net/en/learn/day-27/

Quick check

Just three questions

Q1

What's the first thing AI does when it turns speech into text?

① Chops it into very short pieces
② Understands the whole sentence at once
③ Looks the words up in a dictionary
④ Looks at the speaker's face

Q2

When two words sound the same, like "write" and "right," how does AI choose?

① It picks whichever was louder
② It checks which one fits the words around it
③ It always picks the shorter word
④ It asks a person every time

Q3

When text is turned into a voice, what comes out at the end?

① A photo
② A list of words
③ Air vibration values
④ A search term

Know the why, too

Q1

① It writes the sound down as numbers 16,000 times a second and sweeps over it in 25-millisecond pieces. Chopping it up is what lets similar sounds be told apart.

Q2

② One piece is only a pitch pattern. The word gets settled by checking what fits with its neighbors. That's why names it's never heard often come out wrong.

Q3

③ It makes a pitch pattern from the text, then stamps out vibration values thousands of times a second and sends them to the speaker. It's writing-it-down run backwards.

A voice is vibration chopped into tiny pieces.
and text is the answer you get by matching each piece with its neighbors.



Tomorrow: How is it different from search? — search hands you a list, AI answers in one chunk. What's really different?