EdSutra

Teaching guide

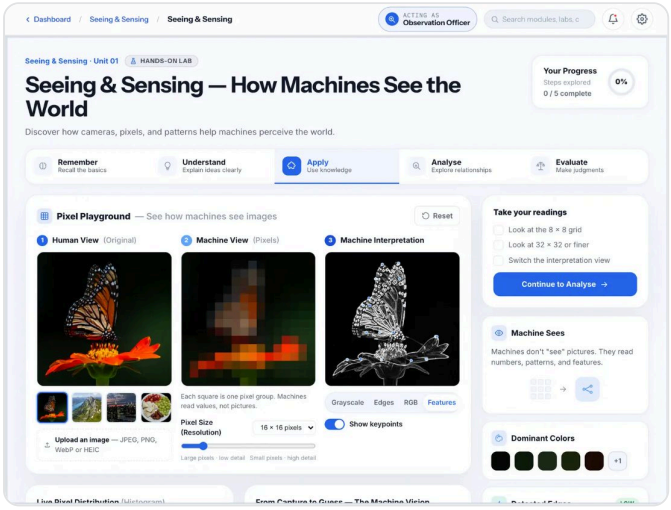
Module 02 · Seeing & Sensing · Unit 01 · Stage 2, Capabilities

Seeing & Sensing

How Machines See the World

Grades 6–10 | 35 minutes | Role: **Observation Officer**

Grade 6: About age 11–12 · Year 7 in the UK · Class 6 in India



The big idea

A camera does not see a butterfly. It stores a grid of numbers: red, green and blue values for every pixel. A vision system finds edges and patterns in those numbers, and a trained model guesses from them. Each step keeps some information and loses some, so the way an image is stored changes what a machine can tell.

The learner's job

Learners are the **Observation Officer** at Cape Marrow Observation Station. Every instrument here turns the world into numbers. Numbers lose things. Their task: say what each instrument really sees. Say where it is blind.

By the end, learners can say

- I can explain how images are stored as arrays of numbers
- I can describe the three stages of a computer vision pipeline
- I can identify real-world applications across medicine, transport, and entertainment
- I can understand why lighting, angle, and resolution affect model accuracy

Before class

- One computer or tablet for each learner or pair, with EdSutra open (this Module is part of the full EdSutra studio).
- Open the studio once on each device before class, so the Unit still runs if the internet drops.
- Play the Remember and Understand steps yourself, so you know where learners will click.
- Optional: save a classroom photo on each device. In Apply, learners can upload it and compare it with the butterfly.
- Best after What is AI? and How Machines Learn.
- Units open in order, and Module 1 comes first. To start a class here, switch on Open every Unit in Settings → Teachers & parents.

Words to introduce

Pixel: One tiny square of an image, stored as a few numbers.

RGB: Three numbers per pixel: how much red, green and blue light, 0 to 255.

Resolution: How many pixels make up a picture. Fewer pixels keep less detail.

Edge: A place where brightness changes sharply, such as an outline.

Edge density: The share of pixels where brightness changes sharply.

Feature: A clue measured from the pixels, such as edges or main colors.

Inference: The guess a trained model makes from the features.

Running the lesson 35 minutes

Open 5 min

"A camera turns a butterfly into numbers. What could get lost on the way, and would the machine ever tell us?"

Before anyone clicks, ask: if a butterfly photo were shrunk to an 8 × 8 grid of squares, what would still be there: colors, outlines or small shapes? Write the guesses on the board and come back to them in Analyze.

Step	What learners do	Ask the class	Listen for
1. Remember 4 min	Read the five big ideas, from "Cameras capture light" to how features lead to a guess. Then click the map of places machine vision shows up, such as medical scans and face unlock.	• What does a camera sensor turn light into? • Which use of machine vision on the map have you met yourself?	A camera stores numbers for each spot of light, not a picture of objects.
2. Understand 6 min	Follow one photo through the four-stage pipeline: real world, pixels, features (edges), interpretation. Then answer two quick checks: what one pixel stores, and what comes after finding edges.	• What does one pixel store? • Why does a bad measurement early on lead to a bad guess later?	A pixel is just red, green and blue numbers. The label comes last, built from the features before it.
3. Apply 8 min	In the Pixel Playground, drop the grid to 8 × 8, raise it to 32 × 32 or finer, and switch the interpretation view (Grayscale, Edges, RGB, Features). They can also upload a photo.	• What disappeared first as the grid got coarser? • What does the Edges view keep, and what does it drop?	A coarse grid keeps rough colors but loses outlines. The Edges view keeps outlines but drops color.
4. Analyze 6 min	Read the live measurements for their image (main colors, edge density, pixels analyzed) and answer two questions: what edge density measures, and what an 8 × 8 version could still give a machine.	• Edge density is not confidence. What is it actually counting? • Look back at our guesses on the board. Were we right about 8 × 8?	Edge density counts sharp changes in brightness. At 8 × 8, rough colors and where they sit survive; outlines are lost.
5. Evaluate 6 min	Compare four ways to show the butterfly (RGB, Grayscale, Edges, Features) in a table. Choose which best helps a machine detect it, then pick the reason that backs that verdict.	• Which view did you choose, and what trade-off does your reason weigh? • Why is "the original photo is always best" refused?	A verdict backed by a trade-off: more information is not always better, and the best view depends on the task.

Steps follow Bloom's revised taxonomy, in the order the Unit presents them.

Deepen the thinking

Critical thinking	A security camera stores a tiny, blurry picture. What could someone wrongly conclude from it, and what would you want to check?
AI thinking	If you built a machine to sort ripe fruit from unripe fruit, which view would you give it: RGB, Edges or Features? Why?
Ethics and fairness	Lighting, angle and resolution change what a camera records. Who could be treated unfairly if face unlock works worse in poor light?
Connect to life	When has a photo search, filter or face unlock got something wrong for you? Which part of the pipeline might have failed?

Misconceptions to listen for

"A computer sees a picture the way I do."

The Machine View and the Live Pixel Distribution chart show what it really has: rows of red, green and blue numbers. Labels only come later, from a model.

"A label like "butterfly" means the machine saw one."

The pipeline separates capture, pixels, features and the guess. A label is an inference built from measurements, so bad measurements mean bad guesses.

"More resolution is always better."

Evaluate refuses "the original photo is always best". Extra detail can bury the outline a detector needs; the right view depends on the task.

In real life the card learners see at the end of the Unit

- A phone camera stores each photo as millions of pixels: numbers for how red, green and blue each dot is.
- A blurry security-camera picture on the TV news shows how a tiny image keeps rough colors but loses shapes.
- Photo apps that search for "beach" turn pixels into features like edges and colors, then a model guesses.

Try it this week. Squint at a poster until it goes blurry. Which can you still make out: colors, layout or small shapes? Compare it with the Unit's 8 × 8 image.

Watch out for. A label like "cat" isn't the computer seeing a cat. It's a guess built from pixel numbers, and more resolution isn't always better.

Talk about it. When you look at a photo, what do you notice that a computer only stores as numbers?

Exit ticket

1. What three numbers does one pixel store?

   _____

2. What does edge density measure?

   _____

3. Name one thing an 8 × 8 version of a photo keeps, and one thing it loses.

   _____

Going further

The Computer Vision Lab stands right after this Unit on the Module page. A real hand or face tracker runs on the camera, or a labeled demo without one. Learners change its blocks and wires and say where the pipeline finds its features.

EdSutra

Teaching guide

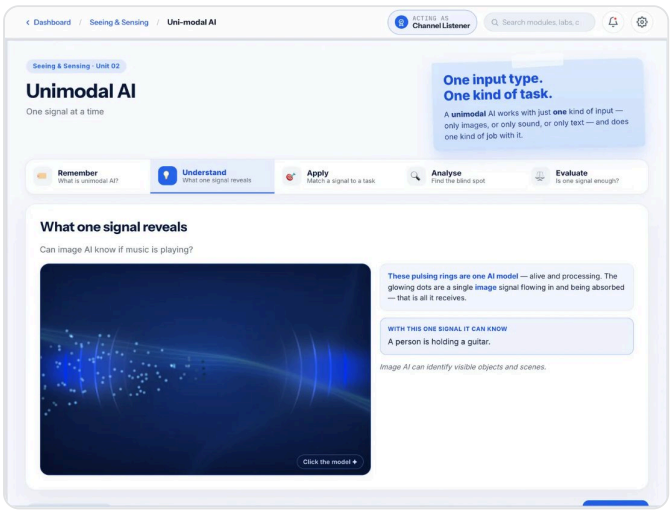
Module 02 · Seeing & Sensing · Unit 02 · Stage 2, Capabilities

Uni-modal AI

One Signal at a Time

Grades 7–10 | 12 minutes | Role: **Channel Listener**

Grade 7: About age 12–13 · Year 8 in the UK · Class 7 in India



The big idea

A unimodal AI gets one kind of signal: only a picture, only sound, only text or only a sensor reading. Each signal shows part of a situation and hides the rest, its blind spot. Sometimes one signal is enough for the job, as when typing out a voice note. Often it is not enough, or it needs caution.

The learner's job

Learners are the **Channel Listener** at Cape Marrow Station · Single Channel. You get one kind of signal today. No pictures for the microphone. No sound for the camera. Their task: find out how far one signal gets you, and where it stops.

By the end, learners can say

- I can name and identify a single input signal (image, text, audio, sensor)
- I can explain what one signal can and cannot reveal
- I can match the right signal to a task and find the missing signal
- I can judge whether one signal is enough for a given decision

Before class

- One computer or tablet for each learner or pair, with EdSutra open (this Module is part of the full EdSutra studio).
- Open the studio once on each device before class, so the Unit still runs if the internet drops.
- Play the Remember and Understand steps yourself, and click the glowing lens, so you know what learners will see.
- Write the four mysteries on the board (Guitar Clue, Tricky Word, Number Clue, Voice Note) with columns: Enough, Not enough, Needs caution.
- Best after Seeing & Sensing.
- Units open in order, and Module 1 comes first. To start a class here, switch on Open every Unit in Settings → Teachers & parents.

Words to introduce

Unimodal: Using just one kind of input, such as only images or only sound.

Mode: A kind of signal: text, image, audio, video, gesture or sensor data.

Signal: The information an AI receives, such as a photo, a sound or a reading.

Blind spot: What a signal can never tell the AI, however good the model is.

Missing signal: The extra kind of input that would fill a blind spot.

Context: Background facts about a situation that explain what a signal means.

Running the lesson 25 minutes

Open 4 min

"If an AI only gets a photo of a birthday party, what can it never know about that party?"

Before anyone clicks, read out the four mysteries: a photo of a guitar, the message "That was sick!", a 38°C reading and a voice note. For each, vote: one signal is enough, not enough, or needs caution. Return in Evaluate.

Step	What learners do	Ask the class	Listen for
1. Remember 3 min	Read the sticky note (one input type, one kind of task) and meet the six kinds of signal an AI can receive: text, image, audio, video, gesture and sensor data.	• What does "modal" mean here? • Name a tool you use that takes only one kind of signal.	A mode is a kind of signal. A unimodal AI takes in one kind and does one kind of job.
2. Understand 4 min	Watch one signal flow into the Signal Lens, a picture of one AI model, and read what that single signal lets it know: for the Guitar Clue, that a person is holding a guitar.	• The image shows a guitar. What can the AI not know from it? • Is the glowing lens a working AI, or a picture of one?	The lens only stands for a model. One signal carries part of the scene; whether music is playing is outside a photo.
3. Apply 5 min	Choose one of four mysteries and pick the single signal that fits its question. A wrong pick explains why that signal is about something else. Pairs can work through all four.	• Which signal did you pick for the Number Clue, and why? • What did the message say when you picked audio for the Guitar Clue?	Each question is about one kind of signal. Audio could hear music, but it cannot show who holds the guitar.
4. Analyze 4 min	Compare what the AI can know, its blind spot and the missing signal for the chosen mystery, such as context for the word "sick" or audio for the guitar.	• What is the blind spot in the Voice Note? • Why is context, not another sensor, the missing signal for 38°C?	Each blind spot follows from what the signal carries. A number like 38°C needs context to mean anything.
5. Evaluate 5 min	For the chosen mystery, decide whether one signal is Enough, Not enough or Needs caution, and read the feedback. Switch mysteries in Apply to judge all four.	• Why is one signal enough for the Voice Note but not for the Guitar Clue? • Look at our votes on the board. Which did we get wrong?	It depends on the job: audio is enough to type out speech, but an image cannot tell whether music is playing.

Steps follow Bloom's revised taxonomy, in the order the Unit presents them.

Deepen the thinking

Critical thinking	A text-only AI flags a friendly message as rude. What would you check before believing it?
AI thinking	Pick one mystery. Which second signal would you add, and what new blind spot might that signal bring?
Ethics and fairness	A school takes attendance with a camera-only system. Who could it miss or get wrong, and who should check its work?
Connect to life	When was a text message misunderstood because the tone was missing? Which signal would have helped?

Misconceptions to listen for

"An AI knows everything about a situation."

One signal carries only part of the picture. In the Guitar Clue, the image shows the guitar, but the sound is outside it.

"More signals are always better."

The Voice Note shows one signal can be enough when the task is narrow, like typing out what was said.

"A blind spot means the AI is broken."

A blind spot follows from what the signal carries. Every single signal has one; that is normal, not a bug.

In real life the card learners see at the end of the Unit

- A photo app can spot a guitar in your picture, but a photo can't tell whether music was playing.
- A message saying "Great job" could be praise or sarcasm. Text-only AI gets the words, not the tone.
- Live TV captions can come from speech-to-text. For writing down words, sound alone is enough.

Try it this week. Watch a minute of TV with the sound off, then listen with your eyes closed. What did each single signal miss?

Watch out for. Every single signal has a blind spot; that's normal, not a bug. Still, more signals aren't always better. Sometimes one is enough for the job.

Talk about it. Which of your senses could you manage without for a day, and what would you miss?

Exit ticket

1. What does unimodal mean?

   _____

2. What is the blind spot of an image-only AI in the Guitar Clue?

   _____

3. Give one task where a single signal is enough, and say why.

   _____

Going further

In the Computer Vision Lab, open the Hand Keypoints example: one camera signal feeds a hand tracker (no camera? a demo mode runs). With a camera, hide a hand or dim the light. What can one camera never tell it?

EdSutra

Teaching guide

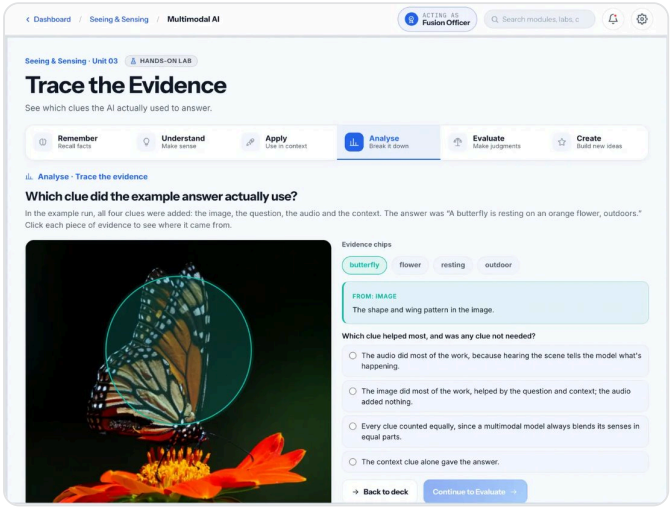
Module 02 · Seeing & Sensing · Unit 03 · Stage 2, Capabilities

Multimodal AI

The Fusion Deck

Grades 8–12 | 45 minutes | Role: **Fusion Officer**

Grade 8: About age 13–14 · Year 9 in the UK · Class 8 in India



The big idea

A multimodal AI combines different kinds of clue, such as a picture, a question, a sound and some context, into one answer. More clues are not always better. A careful user traces which clue each part of an answer came from, notices clues that added nothing, and judges whether the task needs several senses or just one.

The learner's job

Learners are the **Fusion Officer** at Cape Marrow Station · Fusion Deck. The station's AI answers from a picture, a question, a recording and a note. Nobody has checked which of them it actually uses. Their task: put the signals together. Say which one each part of the answer came from.

By the end, learners can say

- I can name the kinds of input a multimodal AI can combine: image, text, audio, video and context
- I can combine clues in the Fusion Deck and see which ones an answer needs
- I can trace each part of an answer back to the clue it came from, and spot a clue that added nothing
- I can judge when combining signals helps and when one signal is enough for the task
- I can design a multimodal pipeline: its senses, its goal, and what each sense could get wrong

Before class

- One computer or tablet for each learner or pair, with EdSutra open (this Module is part of the full EdSutra studio).
- Open the studio once on each device before class, so the Unit still runs if the internet drops.
- Play Remember to Apply yourself, and run the Fusion Deck once with the Context clue and once without it.
- Tell learners the Fusion Deck is scripted: no AI model runs, so the lesson is in tracing the clues.
- Best after Seeing & Sensing and Uni-modal AI.
- Units open in order, and Module 1 comes first. To start a class here, switch on Open every Unit in Settings → Teachers & parents.

Words to introduce

Multimodal: Combining more than one kind of input, such as an image plus text.

Fusion: Putting several clues together to reach one answer.

Evidence: The part of an input that a piece of the answer came from.

Context: Background information about a situation, such as "outdoor scene".

Visual question answering: Answering a question about a picture.

Confidence score: A model's own estimate of how sure it is, not a guarantee.

Pipeline: The planned path from an AI's senses to its goal.

Running the lesson 45 minutes

Open 5 min

"An AI gets a photo, a typed question, a voice recording and a short note. Does it really use all four to answer?"

Before anyone clicks, ask: which of the four clues will matter most for the answer, and will any clue add nothing? Tally the votes on the board and come back to them in Analyze.

Step	What learners do	Ask the class	Listen for
1. Remember 4 min	Tap the five clue cards (image, text, audio, video and context) to see an example of each kind of input a multimodal AI can mix.	• Which of these five kinds of clue does your phone take in? • What counts as context?	Context is background information about the situation, such as "outdoor scene" or "taken at night".
2. Understand 5 min	Click through how AI grew from one kind of input to many: mono-modality, unimodal, multimodal, generative multimodal. Then read how today's tools turn text, images, audio and video into answers.	• What is the difference between unimodal and multimodal? • Name a tool that takes a picture and a question together.	Multimodal means different kinds of information working together, not just more of the same kind.
3. Apply 8 min	In the Fusion Deck, drag clues into the AI Fusion Core: an image and a question are required, audio and context optional. Press Run Fusion, then read the answer and its evidence.	• What changed in the answer when you added or removed the Context clue? • Did adding the audio change anything?	"Outdoors" appears only with the Context clue. Adding the audio changes nothing in the answer.
4. Analyze 7 min	Click each evidence chip (butterfly, flower, resting, outdoor) to see which clue it came from and where in the photo. Then say which clue helped most and whether any clue was not needed.	• Where did "resting" come from? • Look back at our votes on the board. Which clue really mattered most?	The image did most of the work, with the question and context helping. The audio added nothing.
5. Evaluate 8 min	Compare two example answers: Answer A uses only the image ("An animal."), Answer B uses image, text and context. Choose a verdict and the reason that backs it. The numbers are illustrative.	• When would Answer A be good enough? • Does a confidence of 0.93 mean it is 93% sure to be true?	It depends on the task. A confidence score is the model's own estimate; models can be confidently wrong.
6. Create 8 min	Design a multimodal pipeline: pick two or more senses (image, text, audio, video, context, sensors), a goal, and at least one caution that fits the senses chosen. Then submit it.	• Which caution did you pick, and which sense is it about? • Why is "with several senses it can't be wrong" refused?	Each sense can mislead in its own way, and two senses can disagree, so every pipeline needs cautions.

Steps follow Bloom's revised taxonomy, in the order the Unit presents them.

Deepen the thinking

Critical thinking	An AI answers a question about a photo and sounds very sure. How could you check which clue it used before you trust it?
AI thinking	For a robot that helps in a kitchen, which senses would you combine, and which one could you leave out? Why?
Ethics and fairness	An app describes photos aloud for people who are blind. What harm could a confident wrong description do, and who should explain its limits?
Connect to life	When did you work something out by putting two clues together, such as a sound and something you saw?

Misconceptions to listen for

"More signals always give a better answer."
In the Fusion Deck the audio added nothing. Evaluate refuses "more clues always make a better answer": it depends on the task.

"The AI reads the picture like a person does."
Analyze shows the answer built from specific evidence: regions of the image, words in the question, and the context chip.

"A high confidence number means the answer is true."
A confidence score is the model's own estimate, not a guarantee. Evaluate refuses reading 0.93 as 93% certain to be true.

In real life the card learners see at the end of the Unit

- Some AI helpers take a photo plus your question, like "What plant is this?", and answer from both.
- TV weather forecasts blend satellite pictures with readings from weather stations: different kinds of clues, fused into one answer.
- A maps app combines your typed search with context, like where you are and the time, to pick results.

Try it this week. Describe a photo to someone using only words, then show it. Which details did the picture add, and which did your words add?

Watch out for. More clues don't always mean a better answer. Ask which clue each part of an answer came from, and whether a clue added anything at all.

Talk about it. When you're unsure what's happening, which do you check first: what you see, hear or read?

Exit ticket

1. What makes an AI multimodal?

   _____

2. In the Fusion Deck, which clue added nothing to the answer?

   _____

3. Name one task where a single clue is enough.

   _____

Going further

The Computer Vision Lab has both camera and microphone blocks (camera optional; a demo mode runs without one). Learners open an example such as Gesture Reader, change its blocks and wires, and name the sense each part of the result depends on.

EdSutra

Teaching guide

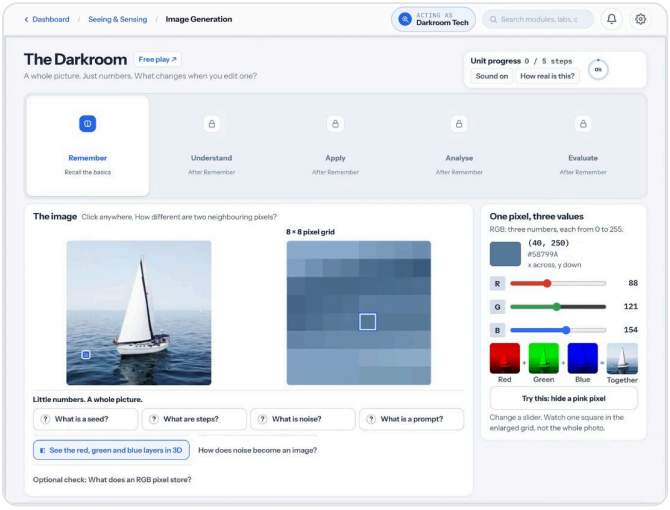
Module 02 · Seeing & Sensing · Unit 04 · Stage 2, Capabilities

Image Generation

The Diffusion Darkroom

Grades 8–11 | **30 minutes** | Role: **Darkroom Technician**

Grade 8: About age 13–14 · Year 9 in the UK · Class 8 in India



The big idea

An image maker does not paste in a photo. Fizz, a real tiny model, learned from 12 photos, then starts from random noise and removes a little at every step until a picture appears. The seed, the prompt and the steps each change the result. A convincing picture does not prove where it came from; only its records can.

The learner's job

Learners are the **Darkroom Technician** at Cape Marrow Station · Darkroom. This room makes pictures of things that never happened. People will still believe them. Their task: turn static into a picture. Then say how it was made, and whether it's new.

By the end, learners can say

- I can name the four controls of image generation — prompt, seed, step, and noise — and what each one decides
- I can read a live-computed signal/noise readout off a real diffusion noise schedule
- I can predict and check the step at which an image crosses the edge of legibility
- I can separate what a prompt fixes (structure) from what a seed fixes (detail)
- I can judge image provenance and explain why visual inspection alone is not a reliable detector

Before class

- One computer or tablet for each learner or pair, with EdSutra open (this Module is part of the full EdSutra studio).
- Open the studio once on each device before class, so the Unit still runs if the internet drops.
- Play Remember and Understand yourself, then develop the baseline sailboat (seed 1042, 24 steps) once in Apply.
- Pairs work well: one learner runs the Darkroom, the other writes down each prediction before a run.
- Best after How Machines Learn and Seeing & Sensing.
- Units open in order, and Module 1 comes first. To start a class here, switch on Open every Unit in Settings → Teachers & parents.

Words to introduce

RGB pixel: One dot of a picture, stored as red, green and blue numbers.

Noise: Random numbers for every color value; it looks like colored static.

Seed: A number that picks the random starting noise.

Steps: How many updates the model makes on the way from noise to picture.

Prompt: Words turned into numbers that steer every update.

Training: Learning patterns from example photos, done before any picture is made.

Provenance: A picture's record of where it came from and how it was made.

Running the lesson 30 minutes

Open 4 min

"Could a computer make a photo of a sailboat that nobody ever took? Where would that picture come from?"

Tell learners that Fizz, the Unit's image maker, learned from just 12 photos. Ask: will its pictures be brand new, or copies of photos it learned from? Take a show of hands and come back to it in Evaluate.

Step	What learners do	Ask the class	Listen for
1. Remember 4 min	Click two pixels of a real boat photo and read their red, green and blue values. Then open the cards "What is a seed?" and "What are steps?".	• What does one pixel store? Does anything in it say "boat"? • What does a seed pick?	A pixel is three numbers from 0 to 255. A seed picks the starting noise; it means nothing by itself.
2. Understand 5 min	Switch between Training (earlier: Fizz learns from 12 photos) and Generation (now: noise becomes a picture). Scrub one recorded run to its early, middle and final steps.	• While Fizz generates, what stays fixed? • At which step did you first see a hint of a boat?	Training happened earlier. During generation the learned rule stays the same and only the picture changes.
3. Apply 6 min	Develop the baseline (a sailboat, seed 1042, 24 steps), scrub to a middle step, change only the seed and develop, then develop the baseline again to see whether it repeats.	• What did changing only the seed do? • Did the baseline come out the same the second time? Why?	The same seed and settings repeat the same picture here. A new seed starts from different noise.
4. Analyze 6 min	Three pairs of runs, A and B, each differ in one setting: seed, steps or prompt. For each, learners name what changed, predict B, watch it develop, then say what they see.	• What happened when B had just 1 step? • Did your prediction for B match what developed?	One step gives a blur. A new prompt gives a different kind of thing; a new seed, another picture.
5. Evaluate 5 min	Judge two cases: their own print beside a training photo, and an uncaptioned photo in a made-up class chat. For each: can pixels prove the source, what next, and which fact backs it?	• Your print almost matches one of Fizz's 12 photos. Is it new? • The chat photo has no record. Does that mean an AI made it?	Pixels cannot prove where a picture came from. A missing record means the source is uncertain, not AI-made.

Steps follow Bloom's revised taxonomy, in the order the Unit presents them.

Deepen the thinking

Critical thinking	Fizz's print looked like a real photo. Besides how a picture looks, what can tell you where it came from?
AI thinking	Fizz copies because it knows only 12 photos. What would change if it learned from millions, and could it still copy?
Ethics and fairness	If an image maker copies a photographer's picture almost exactly, who should get the credit, and who is responsible?
Connect to life	Where have you seen pictures that might be AI-made: ads, posters, social media? How did you decide whether to trust them?

Misconceptions to listen for

"The AI looks up a photo while it draws."

Understand separates the two: training happened earlier. While generating, Fizz uses a fixed learned rule to remove noise, step by step.

"Every AI picture is brand new."

Fizz knows only 12 photos, and its sailboat print matches one almost exactly. Large models mostly make new pictures, but can sometimes copy.

"You can tell an AI picture by looking."

Pixels can match closely but never say where a picture came from. Check its records; a missing record means unknown, not AI-made.

In real life the card learners see at the end of the Unit

- Many AI image makers turn a written prompt into a picture by cleaning up random noise, step by step.
- Posters, ads and magazine covers can now use AI-made pictures that show things that never happened.
- Some photo apps can add or erase things in your pictures with generative AI.

Try it this week. Next time a surprising picture is shared with you, ask an adult to help find where it first appeared before you believe or share it.

Watch out for. You can't tell how a picture was made by looking at it. Fizz's prints began as static, yet looked real, and some copied pictures it learned from.

Talk about it. Should AI-made pictures always carry a label saying so? Whose job is that?

Exit ticket

1. What does a seed decide?

2. What does Fizz learn in training, and what happens during generation?

3. Why can't the pixels alone prove where a picture came from?

Going further

For homework, use the "Try it this week" activity from the In real life card. In class, learners can press Free play, ask Fizz for a dragon (it has never seen one), develop, and explain the result.

EdSutra

Teaching guide

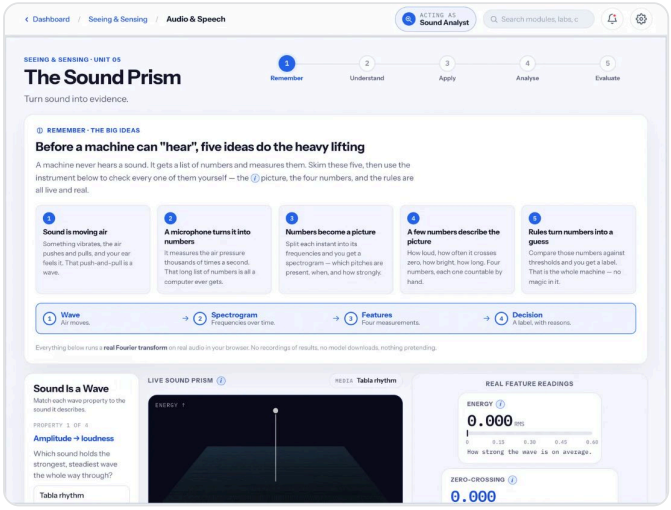
Module 02 · Seeing & Sensing · Unit 05 · Stage 2, Capabilities

Audio & Speech

The Sound Prism

Grades 7–11 | 25 minutes | Role: **Sound Analyst**

Grade 7: About age 12–13 · Year 8 in the UK · Class 7 in India



The big idea

A machine never hears a sound the way we do. It gets a long list of numbers, splits them into frequencies, and measures four things: how loud, how often the wave crosses zero, how bright, how long. A visible rule table turns those numbers into a label, so you can point at exactly why it gets a sound wrong.

The learner's job

Learners are the **Sound Analyst** at Cape Marrow Station · Listening Room. A microphone caught something. Now a machine has to say what it was. Their task: split a sound into what a machine can measure. Then name what decided it.

By the end, learners can say

- I can match amplitude, frequency, duration, and brightness to what you hear
- I can read a spectrogram as frequency × time × energy, and explain why it is a measurement rather than what a machine hears
- I can predict which feature meter a sound will move most, then check it against live measurements
- I can produce a counterexample that defeats the visible thresholds and explain the failure from the numbers
- I can decide which listening tasks four coarse features can and cannot support

Before class

- One computer or tablet for each learner or pair, with EdSutra open (this Module is part of the full EdSutra studio).
- Headphones for each device, since the Unit plays real sound clips. A microphone is optional: only Apply offers it, and nothing needs it.
- Open the studio once on each device before class, so the Unit still runs if the internet drops.
- Play the Remember and Understand steps yourself, so you know the sounds and where learners will click.
- Best after Seeing & Sensing and Uni-modal AI.
- Units open in order, and Module 1 comes first. To start a class here, switch on Open every Unit in Settings → Teachers & parents.

Words to introduce

Amplitude: How far a sound wave swings; we hear it as loudness.

Frequency: How many times a second a wave repeats; we hear it as pitch.

Spectrogram: A picture of which frequencies are present, when, and how strongly.

Zero-crossing rate (ZCR): How often the wave crosses its center line; high for hissy sounds.

Brightness: Where a sound's energy sits on average, from low to high frequencies.

Rule table: Limits on the numbers that decide the label, such as TONE-LIKE.

Counterexample: A case that breaks a rule and shows where it fails.

Running the lesson 30 minutes

Open 4 min

"You can tell a drum from a flute with your eyes closed. What would a machine have to measure to do the same?"

Before anyone clicks, ask: could four numbers (how loud, how often the wave crosses zero, how bright, how long) tell a tabla drum rhythm from a bansuri flute melody? Vote yes or no, then return to it in Analyze.

Step	What learners do	Ask the class	Listen for
1. Remember 4 min	Match four wave properties to the clip that shows each: amplitude (loudness), frequency (pitch), duration (pattern in time) and timbre (brightness), using a steady tone, a bansuri flute, a tabla rhythm and white noise.	• Which sound's pitch kept sliding up and down, and why? • Why does white noise sound like a hiss?	Frequency is how often the wave repeats, heard as pitch. White noise spreads its energy across every frequency at once.
2. Understand 5 min	For each sketched spectrogram (a flat line, a vertical stripe, a full block, a rising curve), guess which sound made it, then play it and watch the live prism build the real picture.	• Why does a single drum hit look like a vertical stripe? • Is the spectrogram what the computer hears, or a measurement?	A hit puts energy into every frequency at once, then fades. The spectrogram is a measurement, not hearing.
3. Apply 6 min	Pick a sound and predict which meter will move most (energy, zero-crossing rate or brightness). Then play it and read the four real numbers and the rule table's label. At least three sounds.	• Which reading surprised you most? • Which number separates white noise from the low tone?	Zero-crossing rate: white noise crosses zero far more often than the low tone, 0.502 against 0.009.
4. Analyze 6 min	Pick a sound the rule table gets wrong (the tabla rhythm, filtered noise or the spoken word "seven"), play it again, and choose the measured reason it failed.	• Why was the tabla rhythm called a TONE? • Look back at our vote. Can these numbers tell a drum from a flute?	Its zero-crossing rate, 0.019, passes the tone rule's under-0.10 test. A rule fires on numbers, not on meaning.
5. Evaluate 5 min	Judge listening jobs as "Can do it" or "Cannot" for these four numbers: telling a drum hit from a hum, which word was said, who was speaking, and more. Each answer explains why.	• Why can't four numbers tell which word someone said? • What would a machine need to tell happy from upset?	Words, voices and feelings live in detail over time that four averaged numbers never look at. More rules won't fix it.

Steps follow Bloom's revised taxonomy, in the order the Unit presents them.

Deepen the thinking

Critical thinking	A filter turned noise into something the rule table called a tone. How would you test a sound detector before trusting it?
AI thinking	If you could add one more number to the four, what would you measure, and which mistake would it fix?
Ethics and fairness	A voice assistant understands some accents worse than others. Who is affected, and whose job is it to find out and fix it?
Connect to life	Which apps or devices around you listen? What do you think they are measuring?

Misconceptions to listen for

"The AI hears the sound like I do."

The machine only gets four numbers and a frequency picture. The live readings show everything it has to go on.

"If it labels a sound, it understands it."

The spoken words "one" and "seven" land in the same bucket as a flute note. Sorting by numbers is not understanding.

"A few more rules would fix it."

No single extra limit separates every word from every flute. Speech needs different features, such as how the sound changes over time.

In real life the card learners see at the end of the Unit

- A music app that names a song from a few seconds of sound matches patterns in its frequencies.
- Some security alarms listen for breaking glass by measuring the sound's frequencies, not by understanding it.
- A video call's noise filter tries to tell your voice from a barking dog or fan, and sometimes slips.

Try it this week. Hum one note quietly, then loudly, then higher. Which change was loudness, and which was pitch? Then compare a smooth "ooo" with a hissy "sss".

Watch out for. Sorting a sound isn't understanding it. A rule table put two different spoken words in the same bucket, and more rules wouldn't fix that.

Talk about it. Which sounds around you would be hardest for a machine to tell apart, and why?

Exit ticket

1. What does a spectrogram show?

2. Name the four numbers the Sound Prism measures.

3. Give one listening job these four numbers cannot do, and say why.

Going further

In the Computer Vision Lab, the Audio Waveform example draws live microphone sound as bars and a wave line (a microphone is needed). Learners hum, hiss and clap, and say which of the four numbers each sound would change most.