

Multi-Modal
Inputs

Auditory
Clues

Visual
Clues

Take
Actions



forward



rotate



jump



interact



grab



read



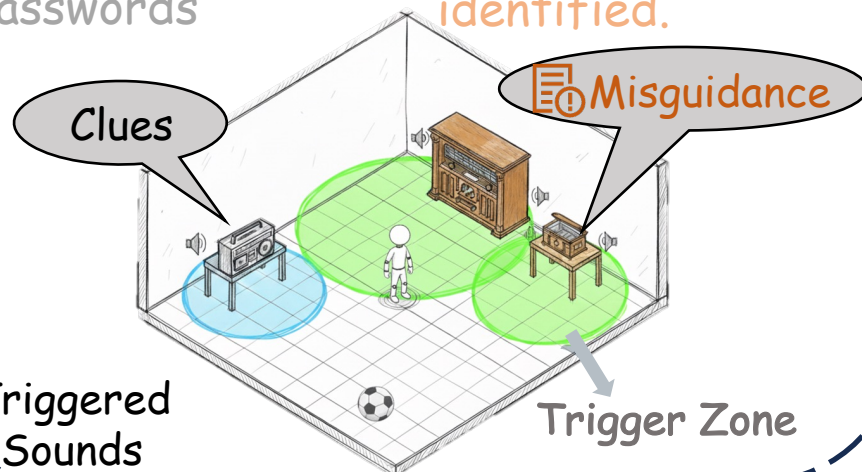
tilt

Clues:

Auditory key clues,
such as safe or exit
passwords

Misguidance:

Misleading auditory
cues need to be
identified.



Triggered
Sounds

Trigger Zone

Stage 1: Exploration

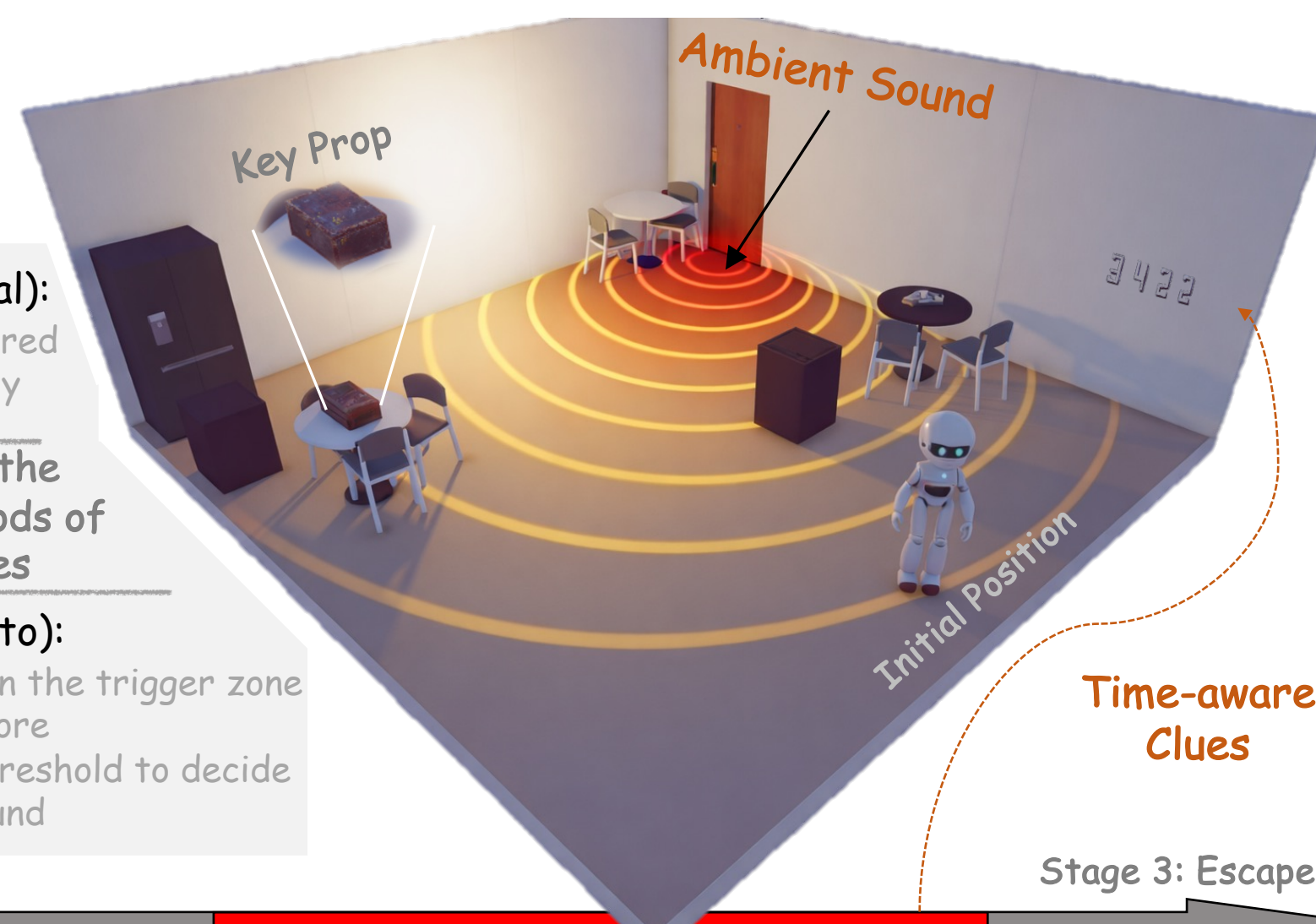
Grab mode (Manual):

- needs to be triggered
by an agent explicitly

Two types of the
triggering methods of
auditory clues

Trigger mode (Auto):

- Check if agent is in the trigger zone
- Compute intent score
- Compare to the threshold to decide
if to trigger the sound

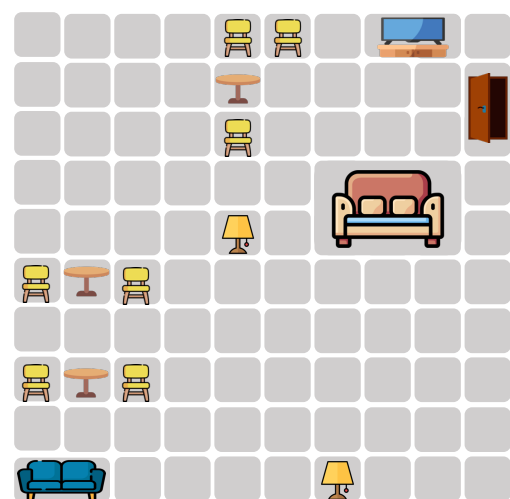


Time-aware
Clues

Stage 3: Escape

Stage 2: Time-aware search

Configuration file



Automatic
Generation



Trigger time-
aware clues

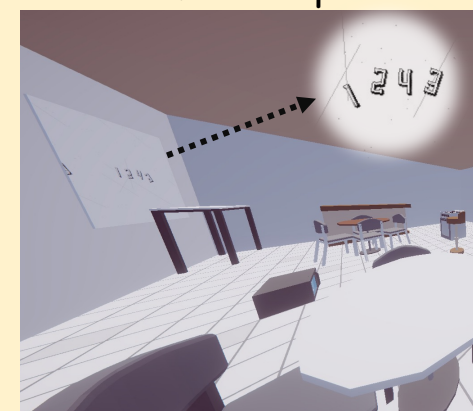
Time-aware setup:

- A clue prompts the agent
to enter a **time-limited**
exploration mode
- A crucial clue will be
triggered within a limited
time; failure to find it will
result in **permanent loss**

Trigger Item: **prompt the
agent to explore new clues**



Clue appears **TEMPORARILY**.
The Agent must find it out
before expire.



Time
Expires



Clue will be lost
PERMANENTLY

