

# Extra-linguistic environment

Understanding

Generation

Speech recognition

Speech synthesis

Interpreted utterance  $\tilde{\mathbf{a}}_u$

Intended response  $\mathbf{a}_m$

Recognition hypotheses  $\tilde{\mathbf{u}}_u$

Utterance to synthesise  $\mathbf{u}_m$

State update

Action selection

Dialogue state

Dialogue management

input speech signal  
(user utterance)

output speech signal  
(machine utterance)

User

