

Extra-linguistic environment

Speech understanding

Generation

Interpreted utterance $\tilde{\mathbf{a}}_{\mathbf{u}}$

Intended response $\mathbf{a}_{\mathbf{m}}$

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

Utterance to synthesise $\mathbf{u}_{\mathbf{m}}$

State update

Action selection

Dialogue state

Dialogue management

Speech recognition

Speech synthesis

input speech signal
(user utterance)

output speech signal
(machine utterance)

User

