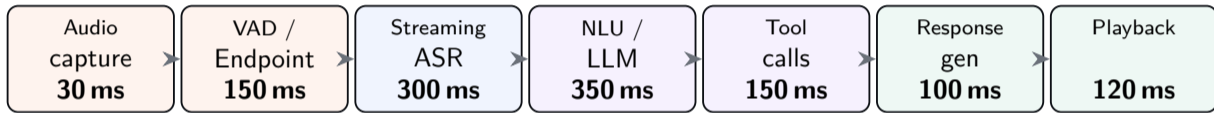


Cascaded Voice-Assistant Pipeline $\approx 1,200$ ms

Each stage contributes to total end-to-end latency. The pipeline is six times the human conversational norm of ~ 200 ms.



Latency budget

