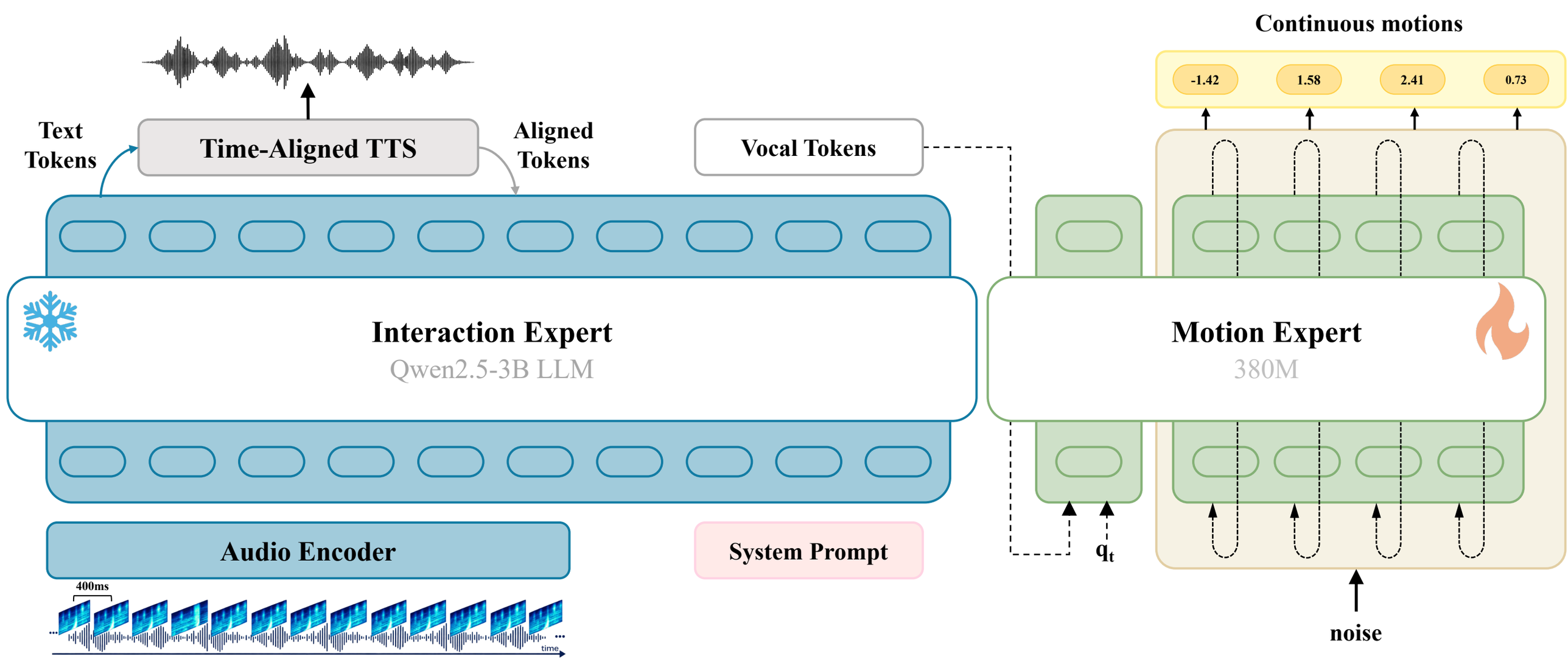

Continuous motions
-1.42
1.58
2.41
0.73
Text Tokens
Time-Aligned TTS
Aligned Tokens
Vocal Tokens
Interaction Expert
Qwen2.5-3B LLM
Motion Expert
380M
Audio Encoder
System Prompt
400ms
time
q_t
noise