

Voice & AI Integrations

Overview

The home runs a fully **local voice pipeline** using HA's Assist framework. All components — wake word detection, speech-to-text, text-to-speech, and LLM conversation — run on local hardware with no cloud dependency. HA Cloud is also configured as a fallback/alternative.

Voice Pipeline Architecture

```
Microphone (HA Voice PE)
    ↓
Wake Word: openWakeWord
    ↓
STT: Whisper (local) or Speech-to-Phrase (local) or HA Cloud
    ↓
Intent / LLM: Home Assistant Assist OR Ollama (qwen3:4b)
    ↓
TTS: Piper (local) or Google Translate or HA Cloud
    ↓
Speaker (HA Voice PE)
```

Hardware

Home Assistant Voice PE

Device: Home Assistant Voice 0919e8 | **Manufacturer:** Nabu Casa | **Model:** Home Assistant Voice PE | **ESPHome:** 2026.3.2 | **Area:** Bedroom

The dedicated voice satellite device for local Assist.

Entity	Notes
<code>assist_satellite.home_assistant_voice_0919e8_assist_satellite</code>	Satellite control
<code>media_player.home_assistant_voice_0919e8</code>	Audio playback
<code>light.home_assistant_voice_0919e8_led_ring</code>	LED ring indicator
<code>switch.home_assistant_voice_0919e8_mute</code>	Microphone mute
<code>switch.home_assistant_voice_0919e8_wake_sound</code>	Chime on wake
<code>event.home_assistant_voice_0919e8_button_press</code>	Physical button
<code>select.home_assistant_voice_0919e8_assistant_2</code>	Pipeline selector

Wake Word Detection

Add-on: openWakeWord v2.1.0 | **Wyoming protocol**

openWakeWord is an open-source wake word detection library. It processes the audio stream locally to detect activation phrases before sending to STT.

Integration	Entry
<code>wyoming</code> (openWakeWord)	<code>01KSAJTKT9G4SPH8F262Y0R656</code>

Entity: `wake_word.openwakeword`

Speech-to-Text (STT)

Two local STT options + HA Cloud:

Option	Add-on	Version	Notes
Whisper	core_whisper	3.2.0	General-purpose neural STT
Speech-to-Phrase	core_speech-to-phrase	1.4.3	Fast, intent-optimized local STT
HA Cloud	—	—	Nabu Casa cloud fallback

Recommended: Speech-to-Phrase for speed; Whisper for accuracy with free-form speech.

Text-to-Speech (TTS)

Option	Add-on	Notes
Piper	core_piper v2.2.2	Local neural TTS (English)
Google Translate	—	Cloud TTS (en-com)
HA Cloud	—	Nabu Casa cloud TTS

Entity: `tts.piper`, `tts.google_translate_en_com`, `tts.home_assistant_cloud`

Conversation / LLM

Home Assistant (Assist)

Entity: `conversation.home_assistant` — built-in intent recognition for HA commands.

Ollama (Local LLM)

Integration: `ollama` | **Host:** `http://192.168.1.85:11434` | **Model:** `qwen3:4b-instruct` | **Firmware:** 4b-instruct

Ollama runs a local LLM on the home server at 192.168.1.85. The `qwen3` model (4B parameters, instruction-tuned) handles natural language conversations and complex commands beyond Assist's built-in intents.

Entity: `conversation.ollama_conversation`

“ The Ollama server at 192.168.1.85 must be running for this conversation agent to be available. HA Assist falls back to the built-in agent if Ollama is unreachable.

Home Assistant Cloud (Nabu Casa)

Integration: `cCloud` | **Status:** loaded

Nabu Casa Cloud provides:

- Remote access (no port forwarding required)
 - Cloud STT/TTS as fallbacks
 - Google Assistant / Amazon Alexa voice control relay (if configured)
 - Webhook relay for IFTTT and mobile notifications
-

Notes / Gotchas

- The `select.home_assistant_voice_0919e8_assistant_2` entity allows switching the Voice PE between different conversation pipelines (e.g., Assist built-in vs Ollama).
- Whisper accuracy improves with higher model variants — the add-on may be configured with a small model for speed; check the add-on config for model selection.
- Speech-to-Phrase is purpose-built for HA commands and significantly faster than Whisper for home control tasks.
- Google Translate TTS requires internet access; Piper works fully offline.

Last Updated

2026-06-24

Revision #1

Created 2026-06-24 21:07:17 UTC by Admin

Updated 2026-06-24 21:07:35 UTC by Admin