🎤 VOICE ENGINE SPECIFICATION - AGGIORNATO

[Sostituisci solo questa sezione nel PROJECT_SPEC.md esistente]

🎯 VOICE ENGINE ARCHITECTURE - DEFINITIVA

📊 VOICE PROCESSING STACK:

STT (Speech-to-Text) - CONFERMATO:

- **Engine:** OpenAI Whisper (completamente locale)
- **Models:** tiny, base, small, medium, large (configurabile)
- **Default:** base (39 languages, 74MB, 1-3s processing)
- **Language:** Italiano (it) con supporto multilingue
- **Hardware:** CPU-only (nessuna GPU richiesta)
- **Privacy:** 100% locale, nessun dato trasmesso

TTS (Text-to-Speech) - AGGIORNATO:

- **Engine:** Piper Neural TTS ✅ **NUOVO** (sostituisce Edge-TTS)
- **Technology:** Neural synthesis completamente locale
- **Models:** Auto-download da Hugging Face (rhasspy/piper-voices)
- **Italian Voice:** it_IT-riccardo-x_low (qualità naturale)
- **Performance:** 500ms-2s generation, alta qualità
- **Privacy:** 100% locale, nessuna API cloud richiesta

Audio Processing:

- **Input:** PyAudio + SpeechRecognition
- **Output:** Sistema audio nativo (Windows/macOS/Linux)
- **Formats:** WAV, MP3 support
- **Sample Rate:** 16kHz (ottimizzato per Whisper)
- **Wake Words:** Custom fuzzy matching algorithm

🔧 IMPLEMENTATION SPECIFICATIONS

📦 Dependencies - Voice Stack:

```txt
```

```
# Speech-to-Text (STT)
openai-whisper==20231117        # Local STT engine
SpeechRecognition==3.10.0       # Audio input wrapper

# Text-to-Speech (TTS) - AGGIORNATO
piper-tts==1.3.0                # Neural TTS locale (NUOVO)

# Audio Processing
PyAudio==0.2.11                 # Audio I/O
pydub==0.25.1                   # Audio manipulation
```

⚙️ Configuration Schema:

```json
{
  "voice": {
    "stt_engine": "whisper",
    "stt_model": "base",
    "stt_language": "it",
    "tts_engine": "piper",                  // AGGIORNATO
    "tts_model": "it_IT-riccardo-x_low",    // AGGIORNATO
    "wake_words": ["jarvis", "ehi jarvis", "hey jarvis"],
    "always_listening": true,
    "voice_activity_detection": true
  }
}
```

🎯 Performance Targets - Validati:

```yaml
```

```
# STT Performance (Whisper)
stt_accuracy: >90%          # Italiano
stt_latency: <3s            # Base model su CPU
stt_languages: 39           # Multilingue support

# TTS Performance (Piper) - AGGIORNATO
tts_quality: "neural"       # Qualità superiore
tts_latency: <2s            # Generation time
tts_voices: "natural"       # Voce italiana naturale
tts_privacy: "100% local"   # Nessuna API cloud

# Wake Word Detection
wake_accuracy: >95%         # Detection rate
wake_latency: <500ms        # Response time
false_positives: <5%        # Accuracy rate
```

🏗️ VOICE MANAGER ARCHITECTURE

📊 Component Interaction:

```
┌─────────────────┐    ┌─────────────────┐    ┌─────────────────┐
│  MICROPHONE     │──►│  WHISPER STT    │──►│  WAKE WORD      │
│  (PyAudio)      │    │  (Local)        │    │  DETECTION      │
└─────────────────┘    └─────────────────┘    └─────────────────┘
                                │
                                ▼
┌─────────────────┐    ┌─────────────────┐    ┌─────────────────┐
│  SPEAKERS       │◄──│  PIPER TTS      │◄──│  CORE SYSTEM    │
│  (System Audio) │    │  (Neural Local) │    │  RESPONSE       │
└─────────────────┘    └─────────────────┘    └─────────────────┘
```

🔧 Voice Manager Class - Updated:

```python
```

```
class VoiceManager:
    """
    Voice Manager con stack definitivo:
    - STT: Whisper (local, accurate)
    - TTS: Piper (neural, local)
    - Audio: PyAudio (cross-platform)
    - Detection: Custom wake word algorithm
    """

    # TECHNOLOGY STACK - FINAL
    whisper_model: str = "base"            # ✅ CONFERMATO
    tts_engine: str = "piper"              # ✅ AGGIORNATO
    audio_backend: str = "pyaudio"         # ✅ CONFERMATO
    wake_detection: str = "fuzzy_matching"  # ✅ CUSTOM

    # PERFORMANCE CONFIGURATION
    max_stt_latency: float = 3.0    # <3s Whisper processing
    max_tts_latency: float = 2.0    # <2s Piper generation
    wake_sensitivity: float = 0.8   # Wake word threshold

    # PRIVACY GUARANTEES
    local_processing: bool = True    # 100% local
    cloud_apis: bool = False         # No cloud dependencies
    data_retention: str = "memory"  # No audio storage
```

🎮 VOICE COMMANDS SPECIFICATION

🎯 Command Categories:

```yaml
```

```yaml
# System Control
system_commands:
  - "Jarvis, apri [applicazione]"
  - "Jarvis, chiudi tutto"
  - "Jarvis, volume [su/giù/[numero]]"
  - "Jarvis, modalità non disturbare"

# Information Queries
info_commands:
  - "Jarvis, che ore sono?"
  - "Jarvis, che tempo fa?"
  - "Jarvis, dimmi le notizie"
  - "Jarvis, cerca [argomento]"

# Productivity
productivity_commands:
  - "Jarvis, crea promemoria [contenuto]"
  - "Jarvis, aggiungi al calendario [evento]"
  - "Jarvis, scrivi email a [persona]"
  - "Jarvis, apri [progetto/file]"

# Configuration
config_commands:
  - "Jarvis, cambia voce in [voce]"
  - "Jarvis, modalità susurro"
  - "Jarvis, velocità parlato [veloce/lento]"
  - "Jarvis, salva queste impostazioni"
```

🔊 Voice Responses - Italian:

```yaml
```

```yaml
# System Status
status_responses:
  startup: "Sistema Jarvis attivo. Come posso aiutarti?"
  standby: "Sono in ascolto..."
  processing: "Un momento, sto elaborando..."
  completed: "Fatto!"
  error: "Mi dispiace, c'è stato un problema..."

# Confirmations
confirmations:
  understood: "Ho capito"
  executing: "Sto eseguendo [azione]"
  completed: "[Azione] completata con successo"
  need_clarification: "Puoi ripetere o essere più specifico?"
```

🔐 PRIVACY & SECURITY SPECIFICATIONS

🛡️ Voice Privacy Guarantees:

- ✅ **No Cloud Processing:** Whisper + Piper = 100% locale
- ✅ **No Audio Storage:** Processing solo in memoria
- ✅ **No API Keys:** Nessuna registrazione richiesta
- ✅ **No Network:** Funziona completamente offline
- ✅ **User Control:** Disattivazione vocale istantanea

📊 Data Handling:

```yaml
# Audio Data
microphone_input: "processed in-memory only"
wake_word_detection: "real-time, no storage"
voice_commands: "transcribed to text, audio discarded"
tts_generation: "created in-memory, no caching"

# Text Data
conversations: "stored locally in SQLite"
user_preferences: "local configuration files"
voice_settings: "local user profile"
system_logs: "local debug files only"
```

[Fine aggiornamento sezione Voice - il resto del PROJECT_SPEC.md rimane invariato]