MANUAL TÉCNICO Y BITÁCORA DE SOLUCIONES (EDICIÓN COMPLETA)

Configuración de Mantella v0.14 en Español con XTTS v2 Local (GPU NVIDIA) y Entrada por Teclado

Arquitectura del entorno Miniconda, parcheo de endpoints REST (/speakers y /speakers_list), enrutamiento gratuito en OpenRouter, gestión de voces .wav y configuración de controles por teclado en Skyrim SE.

01 Arquitectura y Especificaciones del Entorno Estable

Para lograr que Mantella v0.14 sintetice voces en español en tiempo real utilizando una tarjeta gráfica NVIDIA bajo Windows —y operando mediante entrada de texto por teclado en lugar de micrófono—, se aisló el motor de síntesis neuronal (Coqui XTTS v2) en un entorno virtual dedicado de Miniconda, exponiéndolo como un servicio HTTP local al que Mantella se conecta directamente.

Componente	Versión / Valor Configurado	Función en el Sistema
Entorno Virtual	Miniconda — env: xtts (Python 3.10)	Aísla PyTorch con soporte CUDA y las dependencias de síntesis de audio.
Motor TTS	coqui-tts 0.24.1	Ejecuta el modelo multilingüe XTTS v2 sobre GPU (--device cuda).
Librerías Núcleo	transformers 4.36.2 | tokenizers 0.15.2 | coqpit-config 0.2.5	Versiones fijadas para evitar conflictos de compilación C++ y deserialización.
Servidor API Local	xtts-api-server 0.9.0 (http://127.0.0.1:8020)	Expone endpoints REST (/speakers, /speakers_list, /tts_to_audio) vía Uvicorn/FastAPI.
Proveedor LLM	OpenRouter — Modelo: openrouter/free	Enrutador automático hacia modelos gratuitos disponibles sin bloqueos por saturación.
Directorio de Voces	C:\XTTS\speakers	Almacena muestras .wav en minúsculas nombradas por VoiceType (ej. MaleYoungEager.wav).
Directorio de Salida	C:\XTTS\output	Almacena temporalmente los archivos de audio generados por el servidor.
Bypass de Interfaz	C:\XTTS\xtts-api-server-mantella.exe (0 bytes)	Archivo ficticio que satisface el selector de ruta de Mantella v0.14.
Cliente y UI en Juego	Mantella v0.14 + SKSE64 + UIExtensions (MO2)	Conecta diálogos multi-NPC en español mediante ventana flotante de texto (teclas Y y G).

02 Bitácora de Errores Encontrados y Soluciones Magistrales

Durante el despliegue y las pruebas en juego surgieron seis bloqueos técnicos en cadena. A continuación se documentan tanto los errores iniciales de instalación y parcheo del servidor como los hallazgos finales de extensión de archivo, saturación de API y configuración de controles en Skyrim:

2.1. Conflictos de compilación C++, dependencias cruzadas y Coqpit

- **Síntoma:** Al instalar paquetes estándar de TTS en Windows, pip intentaba compilar extensiones C++ desde cero y fallaba por incompatibilidades entre versiones recientes de transformers, tokenizers y coqpit.
- **Solución aplicada:** Se fijaron versiones compatibles en el entorno Miniconda xtts (Python 3.10): coqui-tts 0.24.1, transformers 4.36.2, tokenizers 0.15.2 y coqpit-config 0.2.5, aplicando un parche directo en coqpit.py para evitar fallos al deserializar la configuración del modelo XTTS v2.

2.2. Bloqueo en la interfaz de Mantella: el selector 'XTTS Folder' no acepta carpetas

- **Síntoma:** Al configurar XTTS en la pestaña Text-to-Speech de Mantella v0.14 y pulsar Browse... en el campo XTTS Folder, el explorador de Windows filtraba exclusivamente por el ejecutable 'xtts-api-server-mantella.exe (*.exe)', impidiendo seleccionar la carpeta C:\XTTS.

> **SOLUCIÓN MAGISTRAL #1 — Bypass de validación mediante archivo ficticio de 0 bytes**
>
> **Fundamento:** Como el servidor XTTS ya se ejecuta externamente en Anaconda (http://127.0.0.1:8020), Mantella no necesita lanzar ningún ejecutable real, pero su interfaz gráfica exige validar la existencia del archivo xtts-api-server-mantella.exe.
>
> **Procedimiento ejecutado:** Desde una consola cmd se generó un archivo vacío con el nombre exacto esperado mediante 'type nul > C:\XTTS\xtts-api-server-mantella.exe'. Al seleccionarlo con Browse..., la validación de Mantella se completó correctamente.

2.3. Extensión en mayúsculas (.WAV) no reconocida por xtts-api-server

- **Síntoma:** Al colocar la primera muestra de voz como MaleYoungEager.WAV (con extensión en mayúsculas), el servidor arrancaba reportando 'Latents created for all 0 speakers' e ignoraba el archivo de audio.
- **Solución aplicada:** El escáner de archivos de xtts-api-server filtra estrictamente por la extensión en minúsculas (.wav). Al renombrar el archivo a MaleYoungEager.wav, el servidor generó inmediatamente sus latentes ('Latents created for all 1 speakers').

2.4. Error persistente en tiempo de ejecución: AttributeError: 'list' object has no attribute 'get' (en /speakers y /speakers_list)

- **Síntoma inicial:** Al pulsar Start Mantella o iniciar diálogo en Skyrim, la consola de Mantella arrojaba un traceback en src\tts\xtts.py (línea 126, _get_available_speakers): AttributeError: 'list' object has no attribute 'get'.
- **Por qué persistió tras el primer parche:** Al parchar únicamente el endpoint /speakers, el error volvió a ocurrir en la siguiente prueba. Al inspeccionar los logs HTTP de Uvicorn se descubrió que Mantella v0.14 realiza dos consultas consecutivas: primero GET /speakers y luego GET /speakers_list HTTP/1.1 200 OK, esperando que ambos endpoints devuelvan un diccionario JSON con metadatos de cada voz.

> **SOLUCIÓN MAGISTRAL #2 — Parche polimórfico dual en /speakers y /speakers_list de server.py**
>
> **Fundamento:** Mantella v0.14 fue diseñado para un fork de XTTS donde tanto /speakers como /speakers_list devuelven diccionarios indexados por el nombre del hablante (invocando .get() sobre la respuesta), mientras que xtts-api-server 0.9.0 devuelve listas simples en ambas rutas.
>
> **Transformación aplicada:** Se respaldó server.py como server.py.bak y se parcharon ambos endpoints (/speakers y /speakers_list) para convertir dinámicamente la lista de hablantes en un diccionario estructurado con claves voice_id y name. Tras aplicar este segundo parche, Mantella superó _get_available_speakers sin errores.

2.5. Saturación upstream (Error 429) en OpenRouter y captura accidental de ruido ambiental

- **Síntoma:** Una vez conectado XTTS, la consola de Mantella mostraba 'ERROR: LLM API Streaming Error: Error code: 429 - google/gemma-4-26b-a4b-it:free is temporarily rate-limited upstream'. Simultáneamente, el módulo STT sin Push-to-Talk captaba ruido de fondo ('And all that.', 'Wait.') disparando múltiples peticiones seguidas.
- **Solución aplicada:** En la pestaña Large Language Model de Mantella se cambió el modelo principal (y el Custom Summary Model) a 'openrouter/free', permitiendo que OpenRouter enrute automáticamente las peticiones hacia modelos gratuitos con capacidad disponible en ese instante.

2.6. La tecla H no abría el panel de escritura y el micrófono seguía escuchando

- **Síntoma:** Al entrar en Skyrim y pulsar la tecla H frente a un NPC, no se desplegaba ningún panel de texto y el juego seguía intentando escuchar por micrófono.
- **Diagnóstico en el menú MCM de Mantella:** Al revisar Configuración de Mods > Mantella > General dentro del juego, se comprobó que 'Microphone > Enabled' estaba activado, 'Open Text Prompt' estaba asignado a H, pero 'Start Conversation / Add NPC' estaba sin asignar ('???'). Sin iniciar primero la conversación con el NPC, la tecla de cuadro de texto no tiene efecto.

> **SOLUCIÓN MAGISTRAL #3 — Configuración de flujo 100% por teclado en MCM y Push-to-Talk**
>
> **En la web de Mantella (Speech-to-Text):** Se activó la casilla Push-to-Talk para impedir que el micrófono capture audio ambiental automáticamente.
>
> **En Skyrim (MCM > Mantella > General):** Se desmarcó 'Microphone > Enabled', se asignó la tecla Y a 'Start Conversation / Add NPC' (permitiendo iniciar charla e incorporar múltiples NPCs al diálogo) y se asignó la tecla G a 'Open Text Prompt' (desplegando la ventana de UIExtensions para escribir y enviar líneas con Enter).

2.7. Extracción y formato de muestras de voz de Skyrim (.bsa / .fuz vs .wav)

- **Síntoma:** Los audios nativos de Skyrim están empaquetados en contenedores .bsa y comprimidos en formato .fuz (audio .xwm + sincronización labial).
- **Solución aplicada:** Crear clips limpios en formato .wav en minúsculas (PCM, 6 a 12 segundos recomendados, con música y efectos del juego en 0) o extraerlos mediante BAE + Unfuzer, nombrándolos según el VoiceType interno del motor de Skyrim (ej. MaleYoungEager.wav, MaleNord.wav, FemaleNord.wav).

03 Comandos Exactos de Instalación y Parcheo

3.1. Creación del archivo ficticio para validar XTTS Folder en Mantella

Ejecutar una única vez en Símbolo del sistema (cmd):

```
type nul > C:\XTTS\xtts-api-server-mantella.exe
```

3.2. Parcheo del endpoint /speakers en xtts_api_server

Comando ejecutado en el entorno (xtts) de Miniconda para respaldar server.py y convertir /speakers de lista a diccionario:

```
python -c "import shutil, xtts_api_server.server as s; p=s.__file__; shutil.copy2(p, p+'.bak');
txt=open(p, encoding='utf-8').read(); old='@app.get(\"/speakers\")\ndef get_speakers():\n
speakers = XTTS.get_speakers_special()\n    return speakers'; new='@app.get(\"/speakers\")\ndef
get_speakers():\n    speakers = XTTS.get_speakers_special()\n    return speakers if
isinstance(speakers, dict) else {str(i).replace(\".wav\",\"\"): {\"name\":
```

```
str(i).replace(\".wav\",\"\"), \"voice_id\": str(i).replace(\".wav\",\"\")} for i in
speakers}'; assert old in txt; open(p, 'w', encoding='utf-8').write(txt.replace(old, new, 1));
print('OK_SPEAKERS')"
```

3.3. Parcheo del endpoint /speakers_list en xtts_api_server

Segundo parche obligatorio ejecutado en (xtts) para que la consulta /speakers_list de Mantella v0.14 también reciba un diccionario:

```
python -c "import pathlib, xtts_api_server.server as s; p=pathlib.Path(s.__file__);
txt=p.read_text(encoding='utf-8'); old='@app.get(\"/speakers_list\")\ndef get_speakers():\n
speakers = XTTS.get_speakers()\n    return speakers'; new='@app.get(\"/speakers_list\")\ndef
get_speakers():\n    speakers = XTTS.get_speakers()\n    if isinstance(speakers, dict):\n
return speakers\n    out = {}\n    for item in speakers:\n        key =
str(item).replace(\".wav\", \"\").replace(\".WAV\", \"\")\n        out[key] = {\"voice_id\":
key, \"name\": key}\n    return out'; assert old in txt, 'No se encontro /speakers_list';
p.write_text(txt.replace(old, new, 1)); print('OK_SPEAKERS_LIST')"
```

04 Parámetros Configurados en Mantella (Web UI y MCM en Skyrim)

Resumen consolidado de todos los parámetros verificados tanto en la interfaz web de Mantella (http://localhost:4999/ui) como en el menú MCM dentro del juego:

Ubicación / Pestaña	Parámetro	Valor Establecido	Observación Técnica
Web: Game	Game / Spell Mod Path	Skyrim | .../MO2/.../mods/Mantella	Apunta a la carpeta del mod dentro de Mod Organizer 2.
Web: LLM	LLM Service / Model	OpenRouter | openrouter/free	Evita errores 429 rotando entre modelos gratuitos activos.
Web: Text-to-Speech	TTS Service / XTTS Folder	XTTS | C:\XTTS\xtts-api-server-mantella.exe	Usa el ejecutable ficticio de 0 bytes para validar la ruta.
Web: Text-to-Speech	XTTS URL / Device / VRAM	http://127.0.0.1:8020 | cuda | Low VRAM: ON	Inferencia acelerada en GPU NVIDIA junto con Skyrim.
Web: Text-to-Speech	Lip File Generation	Enabled	Genera archivos .lip para sincronización labial de NPCs.
Web: Speech-to-Text	Push-to-Talk	Activado (Checked — Tecla V)	Impide que el micrófono capte ruido ambiental por su cuenta.
Web: Language	Language	es	Configura la generación de diálogos y síntesis en español.
MCM: General	Microphone > Enabled	Desactivado (Unchecked)	Deshabilita la escucha por voz dentro de Skyrim.

Ubicación / Pestaña	Parámetro	Valor Establecido	Observación Técnica
MCM: General	Start Conversation / Add NPC	Tecla Y	Inicia diálogo con el NPC mirado o añade más NPCs a la charla.
MCM: General	Open Text Prompt	Tecla G	Abre la ventana flotante (UIExtensions) para escribir por teclado.

05 Rutina Paso a Paso para Jugar (Orden Obligatorio de Arranque)

Cada vez que vayas a iniciar una sesión de Skyrim con Mantella y XTTS local (o tras agregar nuevas muestras .wav en C:\XTTS\speakers), sigue exactamente esta secuencia:

- **Paso 1 — Abrir Anaconda Prompt y activar el entorno:** Ejecuta 'conda activate xtts' y verifica que el prompt cambie a (xtts) C:\Users\MSI>.
- **Paso 2 — Levantar el servidor XTTS en GPU:** Ejecuta el comando de inicio del servidor y espera hasta ver 'Latents created for all X speakers.' y 'Uvicorn running on http://localhost:8020'. Mantén esta ventana abierta:

```
python -m xtts_api_server --device cuda -sf C:\XTTS\speakers -o C:\XTTS\output
```

- **Paso 3 — Iniciar Mantella:** Ejecuta Mantella.exe, pulsa 'Start Mantella' y verifica en su consola que cargue XTTS sin errores y muestre 'Waiting for player to select an NPC...'.
- **Paso 4 — Ejecutar Skyrim vía SKSE64 (con UIExtensions activo en MO2):** En Mod Organizer 2, verifica que UIExtensions esté marcado y ejecuta SKSE. Carga tu partida, mira a un NPC, pulsa la tecla Y para iniciar la conversación (puedes mirar a un segundo NPC y pulsar Y nuevamente para incluirlo en la charla), y pulsa la tecla G para abrir el cuadro de texto, escribir tu diálogo en español y enviarlo con Enter.

ESTADO DEL SISTEMA: OPERATIVO Y VALIDADO EN JUEGO (MULTI-NPC + TECLADO)

Hito completado: Sistema completamente funcional en Skyrim SE con conversaciones simultáneas entre 2 NPCs en español, entrada de diálogo por teclado (Y para agregar NPCs, G para abrir el panel de texto) y síntesis de voz local en GPU NVIDIA.

Próximas mejoras opcionales (pendientes): 1) Ampliar el banco de voces .wav en C:\XTTS\speakers (MaleNord.wav, FemaleNord.wav, MaleBrute.wav, FemaleYoungEager.wav, etc.); 2) Optimizar latencia de respuesta y parámetros de XTTS (temperature, repetition_penalty, speed, stream_chunk_size) para suavizar el timbre robótico.