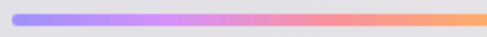




GUÍA DE CONFIGURACIÓN



Guía paso a paso: Configuración completa

Proveedores de IA + SIP Provider (Netelip) + Creación de Agente

Introducción

VOCALS es una plataforma de telefonía impulsada por IA que gestiona conversaciones de voz en tiempo real. Integra proveedores SIP para la telefonía y soporta proveedores intercambiables de IA para reconocimiento de voz (STT), modelos de lenguaje (LLM) y síntesis de voz (TTS).

En esta guía aprenderemos a configurar VOCALS desde cero en tres grandes pasos: configurar los proveedores de IA, conectar un proveedor SIP (Netelip) y crear un agente de voz inteligente.

Requisitos Previos

- Una cuenta en VOCALS (dashboard.usevocals.com)
- API keys de al menos un proveedor de cada categoría (STT, LLM, TTS)
- Una cuenta en Netelip con un DID (número de teléfono) y credenciales SIP

Cómo Funciona el Pipeline de Voz

Cada llamada en VOCALS fluye a través de un pipeline de tres etapas en tiempo real. El audio entrante se transcribe con STT, la transcripción se envía al LLM para generar una respuesta, y la respuesta se sintetiza en audio con TTS. Todo en menos de 2 segundos.

Etapa	Función	Descripción
STT	Speech-to-Text	Convierte el audio del llamante en texto en tiempo real
LLM	Language Model	Genera respuestas conversacionales basadas en la transcripción y el prompt del sistema
TTS	Text-to-Speech	Convierte el texto de respuesta en audio que escucha el llamante

1. Configurar los proveedores de IA

Los proveedores de IA son el núcleo del pipeline de voz. Necesitas configurar al menos un proveedor para cada etapa: STT, LLM y TTS. VOCALS permite mezclar y combinar proveedores por agente para optimizar latencia, precisión, coste o soporte de idiomas.

1.1 Acceder a la configuración de proveedores

1. Inicia sesión en el dashboard de VOCALS (dashboard.usevocals.com).
2. En el menu lateral, navega a Configuration > Providers.
3. Haz clic en Add Provider.
4. Selecciona el tipo de proveedor (STT, LLM o TTS) y el servicio específico.
5. Introduce tu API key y configura los ajustes específicos del proveedor.
6. Haz clic en Save. VOCALS validará la clave haciendo una petición de prueba.

Consejo: Crea API keys separadas para VOCALS en lugar de reutilizar claves de otros proyectos. Esto facilita el seguimiento del uso y la rotación de credenciales.

1. Configurar los proveedores de IA (continuación)

1.2 Proveedores STT (Speech-to-Text)

Los proveedores STT transcriben el audio del llamante a texto en tiempo real. Estos son los proveedores soportados:

Proveedor	Modelos	Notas
Deepgram	nova-2, nova-2-general, nova-2-phonecall	Recomendado. Baja latencia, excelente soporte streaming.
OpenAI Whisper	whisper-1	Modo batch. Mayor latencia pero buena precisión en entornos ruidosos.
Alibaba Qwen	qwen-audio	Fuerte soporte multilingüe, especialmente chino y lenguas asiáticas.
Fish Audio	transcribe-1	Modo batch, soporte 30+ idiomas. En beta.

Configuración recomendada para Deepgram: Modelo nova-2, idioma es-ES (para castellano), Smart Format activado, Endpointing en 300 ms e Interim Results activado para transcripciones parciales más rápidas.

1.3 Proveedores LLM (Modelo de Lenguaje)

Los proveedores LLM generan las respuestas conversacionales del agente basándose en la transcripción y el prompt del sistema.

Proveedor	Modelos	Notas
OpenAI	gpt-4o, gpt-4o-mini, gpt-4-turbo, gpt-3.5-turbo	Buen equilibrio calidad/velocidad. gpt-4o-mini para uso general eficiente.
Anthropic Claude	claude-sonnet-4-20250514, claude-haiku-4-20250414	Excelente siguiendo prompts detallados y manteniendo personas consistentes.
Google Gemini	gemini-2.5-flash, gemini-2.5-pro	Muy baja latencia a precio competitivo. Ideal para alto volumen.
Moonshot Kimi	moonshot-v1-8k, moonshot-v1-32k	Fuerte soporte en chino. Precios competitivos para mercado asiático.

Configuración recomendada: Temperature 0.7 y Max Tokens 256 son los valores por defecto y funcionan bien para la mayoría de los casos de uso conversacional.

1. Configurar los proveedores de IA (continuación)

1.4 Proveedores TTS (Text-to-Speech)

Los proveedores TTS convierten el texto de respuesta del LLM en audio que escucha el llamante.

Proveedor	Modelos/Voces	Notas
ElevenLabs	eleven_turbo_v2_5, eleven_multilingual_v2	Voces más naturales. Soporta clonación de voz. Usar turbo para telefonía.
OpenAI TTS	tts-1, tts-1-hd (voces: alloy, echo, fable, onyx, nova, shimmer)	Simple de configurar. tts-1 para telefonía (menor latencia).
Resemble AI	Voces personalizadas por UUID	Especializado en clonación de voz de marca.
Fish Audio	s2, s1, speech-1.6, speech-1.5	Voz natural con control de emoción. 30+ idiomas.

Configuración recomendada para ElevenLabs: Modelo eleven_turbo_v2_5, Stability 0.5, Similarity Boost 0.75, Optimize Streaming Latency 3. Para encontrar tu Voice ID, accede a tu panel de ElevenLabs, ve a Voces, selecciona una voz y copia el Voice ID.

1.5 Combinaciones recomendadas por caso de uso

Caso de uso	STT	LLM	TTS
General (baja latencia)	Deepgram nova-2	OpenAI gpt-4o-mini	ElevenLabs turbo v2.5
Alta calidad	Deepgram nova-2	Anthropic Claude Sonnet	ElevenLabs multilingual v2
Bajo presupuesto	Deepgram nova-2	Google Gemini Flash	OpenAI tts-1
Multilingüe (30+ idiomas)	Fish Audio	Google Gemini Flash	Fish Audio s2

2. Configurar el proveedor SIP (Netelip)

El proveedor SIP conecta tus números de teléfono con VOCALS. Netelip es un proveedor de troncales SIP europeo con cobertura en España y Latinoamérica. VOCALS utiliza Asterisk como pasarela SIP para proveedores genericos como Netelip.

2.1 Requisitos previos de Netelip

- Una cuenta activa en Netelip.
- Un DID (número de teléfono) asignado en tu panel de Netelip.
- Tus credenciales SIP (usuario y contraseña) del panel de control de Netelip.

2.2 Añadir Netelip en el dashboard de VOCALS

1. En el dashboard de VOCALS, navega a Settings > SIP Providers.
2. Haz clic en Add SIP Provider.
3. Selecciona Netelip como tipo. Esto pre-rellenará el servidor SIP y el puerto, aunque puedes cambiarlos si es necesario.
4. Rellena los datos de conexión (ver tabla a continuación).
5. (Opcional) Configura el filtrado de llamadas entrantes en Allowed IPs para restringir que IPs pueden enviar llamadas a tu trunk.
6. Haz clic en Save.

Campos de configuración de Netelip

Campo	Valor	Notas
SIP Server	sip.netelip.com	Pre-rellenado. También disponible sip-eu.netelip.com para servidores regionales europeos.
SIP Port	5060	Puerto por defecto para Netelip.
Transport	UDP	Protocolo de transporte por defecto para Netelip.
Username	(tu usuario SIP)	El nombre de usuario SIP proporcionado por Netelip en tu panel de control.
Password	(tu contraseña SIP)	La contraseña SIP proporcionada por Netelip.
Allowed IPs	(opcional)	IPs permitidas para enviar llamadas. Deja vacío para permitir todas. Netelip envía desde sus rangos de IP publicados.

2.3 Verificar el estado de registro

Después de guardar, VOCALS registra tu trunk con el servidor SIP. Comprueba el indicador de estado en la tarjeta del proveedor SIP:

Estado	Significado
Verde (Registered)	El trunk esta conectado y listo para recibir llamadas.
Rojo (Unregistered)	Fallo en el registro. Verifica credenciales y dirección del servidor SIP.
Amarillo (Unknown)	No se pudo determinar el estado. El trunk puede estar todavía inicializándose. Espera 10-30 segundos.

Importante: Si el estado es Rojo (Unregistered), verifica: (1) que el servidor y puerto SIP son correctos, (2) que el usuario y contraseña coinciden con tu panel de Netelip, (3) que el firewall no bloquea el puerto 5060 ni los puertos UDP 10000-10100 para RTP media.

2. Configurar el proveedor SIP (continuación)

2.4 Configurar el DID en Netelip

Ahora necesitas configurar tu número DID en el panel de Netelip para que las llamadas entrantes se dirijan a VOCALS:

1. Inicia sesión en tu cuenta de Netelip.
2. Navega a DID Numbers (Numeros DID) o la sección equivalente.
3. Configura el destino de tu número hacia la IP de tu servidor VOCALS en el puerto 5060.
4. Asegúrate de que el codec está configurado como G.711a (alaw) o G.711u (ulaw). VOCALS auto-detecta ambos.

Consejo: Cuando configures un nuevo proveedor SIP, comienza haciendo una llamada de prueba saliente para verificar la calidad de audio y la latencia antes de configurar el enrutamiento entrante.

3. Crear un agente de voz

Un agente es la unidad central en VOCALS. Define cómo se comporta un asistente de voz con IA en una llamada: qué dice, cómo suena y que proveedores lo alimentan. Cada número de teléfono se asigna a un único agente.

3.1 Crear el agente

1. En el dashboard, navega a Agents.
2. Haz clic en Create Agent.
3. Dale un nombre descriptivo al agente (por ejemplo, "Atención al Cliente - Español" o "Ventas Inbound").
4. Configura los ajustes que se describen a continuación.
5. Haz clic en Save para crear el agente.

3.2 System Prompt (Prompt del sistema)

El system prompt es el ajuste más importante. Define la personalidad, instrucciones y restricciones del agente. Determina todo sobre cómo se comporta en la conversación.

Buenas prácticas para escribir el system prompt:

- Sé específico sobre la longitud de las respuestas. Las conversaciones telefónicas necesitan respuestas cortas (1-2 frases por turno).
- Define una persona. Dale al agente un nombre, tono y personalidad. Los llamantes se sienten más cómodos con un personaje consistente.
- Establece límites. Indica explícitamente qué temas debe y no debe tratar. Lista los temas que debe escalar a un humano.
- Incluye frases de ejemplo para saludos, confirmaciones y despedidas.
- Maneja casos extremos: qué hacer cuando no sabe la respuesta, cuando el llamante está molesto o cuando la conversación se desvíe.
- Usa estructura clara con secciones, viñetas y pasos numerados. Los LLMs siguen mejor los prompts estructurados.

Importante: Evita prompts excesivamente largos. Cada token añade latencia y coste a cada llamada al LLM. Apunta a 200-500 palabras. Si necesitas más contenido, considera usar la Knowledge Base para información de referencia.

3. Crear un agente de voz (continuación)

3.3 Mensaje de bienvenida (Welcome Message)

Es lo primero que dice el agente cuando se conecta una llamada. Se reproduce como audio TTS antes de que el agente empiece a escuchar. Ejemplo: "Hola, gracias por llamar a [Empresa]. ¿En qué puedo ayudarte hoy?"

Dejalo en blanco si quieres que el agente espere a que el llamante hable primero (útil para llamadas salientes).

3.4 Configuración de idioma

Configura el idioma principal de la conversación. Este ajuste se pasa al proveedor STT para mejorar la precisión de la transcripción. Valores comunes: es-ES (castellano), es-MX (mexicano), en-US, pt-BR, fr-FR, de-DE.

3.5 Sensibilidad de Barge-in (interrupción)

Controla la facilidad con la que el llamante puede interrumpir al agente mientras habla. Cuando se detecta actividad de voz durante la reproducción TTS, el agente deja de hablar y escucha.

Nivel	Comportamiento
Very Low	Requiere habla sostenida y clara para interrumpir. Ideal para entornos ruidosos.
Low	El llamante debe hablar fuerte o durante más tiempo. Reduce falsos positivos por ruido ambiente.
Medium	Ajuste equilibrado. Funciona bien para la mayoría de entornos.
High	El agente deja de hablar rápidamente al detectar voz. Para entornos silenciosos.
Very High	El agente se detiene ante la primera señal de voz. Conversaciones rápidas y silenciosas.

Consejo: Si notas que el agente se interrumpe por ruido ambiente, baja la sensibilidad de barge-in. Si los llamantes se quejan de que el agente habla encima de ellos, subela.

3.6 Otros ajustes del agente

Ajuste	Descripción
Interruptible	Activa/desactiva barge-in. Desactivalo para mensajes que deben escucharse completos (avisos legales, disclaimers).
Max Call Duration	Duración máxima de la llamada en segundos. Por defecto: 600 (10 minutos). Reduce para casos simples (encuestas, confirmaciones).
Silence Threshold	Sensibilidad de detección de voz (VAD) para barge-in. Por defecto: 0.5. Valores altos (0.7-0.9) requieren más confianza, valores bajos (0.2-0.4) son más sensibles.

3. Crear un agente de voz (continuación)

3.7 Asignar proveedores al agente

Cada agente necesita un proveedor para cada etapa del pipeline:

1. En la configuración del agente, busca la sección Providers.
2. Selecciona un proveedor STT de tus proveedores configurados.
3. Selecciona un proveedor LLM.
4. Selecciona un proveedor TTS.

Consejo: Puedes asignar diferentes proveedores a diferentes agentes. Por ejemplo, tu agente de ventas en inglés podría usar Deepgram + GPT-4o + ElevenLabs, mientras que tu agente de soporte en español use Deepgram + Claude Sonnet + Fish Audio.

3.8 Asignar un número de teléfono

Para que el agente pueda recibir llamadas, necesita tener un número de teléfono asignado:

1. Ve a Phone Numbers en el dashboard.
2. Haz clic en el número que quieres asignar.
3. Selecciona el agente que acabas de crear en el desplegable.
4. Haz clic en Save. El cambio aplica en la siguiente llamada entrante.

Puedes asignar el mismo agente a múltiples números de teléfono, útil cuando tienes números locales de diferentes regiones que deben ir al mismo agente.

4. Hacer una llamada de prueba

1. Llama al número de teléfono que configuraste.
2. Deberías escuchar el mensaje de bienvenida y luego poder mantener una conversación con tu agente de IA.
3. Tras la llamada, revisa el Dashboard para ver: duración, transcripción de la conversación, métricas de latencia (STT, LLM, TTS) y desglose de costes.

Resolución de Problemas Comunes

Problema	Solución
Registro SIP falla (Rojo)	Verifica credenciales, servidor SIP, puerto y transporte. Comprueba que el firewall permite tráfico en el puerto 5060 y UDP 10000-10100.
Sin audio (llamada muda)	Verifica que los puertos RTP (UDP 10000-10100) están abiertos. Comprueba configuración NAT y codec (alaw o ulaw).
El agente no responde	Verifica que el número de teléfono está asignado al agente y que los proveedores STT, LLM y TTS tienen API keys válidas.
Eco o feedback	VOCALS aplica jitter buffer adaptativo automáticamente. Si persiste, puede ser del lado del proveedor SIP. Reduce el volumen TTS si es posible.
Errores de autenticación del proveedor	La llamada caerá gracefully y el error aparecerá en los logs. Rota la API key del proveedor afectado.

Documentación completa en docs.usevocals.com