

Voces y modelos de ElevenLabs en español

Datos de esta lección

Fase	F2 — Capa de telefonía y generación de voz
Módulo	M2.2 — Configuración de voz con ElevenLabs
Puntos de profesionalidad al superar el test	35

Objetivo práctico

Al terminar esta lección sabrás elegir una voz en español dentro de ElevenLabs, entender los parámetros que afinan su comportamiento (estabilidad, similitud, expresividad, claridad, latencia) y generar un audio de prueba con NVDA.

Resultado que obtendrás

Una voz elegida y probada, con los parámetros ajustados a un criterio razonado, lista para usarse en la lección final de este módulo (hito de prueba de voz).

Dónde encaja dentro del sistema final

Esta lección configura directamente la pieza VOZ del sistema: la voz que elijas y ajustes aquí será, más adelante, la que escuchen las personas que llamen al agente telefónico.

Conocimientos previos

- Cuenta de ElevenLabs creada (lección anterior).

Herramientas necesarias

- Cuenta de ElevenLabs activa.

Posible coste

Generar audio de prueba consume créditos del plan gratuito (10.000 créditos)

mensuales, verificado el 30 de agosto de 2026); las pruebas de esta lección son breves y no deberían agotar ese saldo.

Modelo mental

Piensa en un estudio de grabación donde, antes de elegir la voz definitiva de un anuncio, se prueban varias voces y se ajustan matices: más pausada, más expresiva, más parecida a una grabación de referencia. ElevenLabs ofrece ese mismo tipo de ajuste fino sobre una voz sintética.

Explicación

Dentro de la biblioteca de voces de ElevenLabs puedes filtrar por idioma (español) y escuchar muestras antes de elegir una. Una vez elegida, varios parámetros permiten afinar cómo suena:

- **Estabilidad:** cuánto se ciñe la voz a un tono constante; valores más bajos producen más variación expresiva, valores más altos producen una entonación más uniforme y predecible.
- **Similitud:** cuánto se parece la generación a la voz de referencia original; valores más altos son más fieles a esa referencia, pero pueden introducir artefactos si se llevan al extremo.
- **Expresividad (styling/estilo):** cuánto énfasis emocional o interpretativo añade el modelo, según el modelo concreto elegido.
- **Claridad:** nitidez percibida de la pronunciación.
- **Latencia:** tiempo que tarda en generarse el audio; especialmente relevante en telefonía en tiempo real (ver lección «La voz del agente: qué es el TTS», módulo 1.2), donde una latencia alta produce silencios incómodos.

No existe una combinación «correcta» universal: para un recepcionista telefónico profesional, una estabilidad media-alta y una latencia baja suelen ser más importantes que una expresividad muy alta, porque priorizan la claridad y la rapidez de respuesta sobre la interpretación emocional.

Vocabulario nuevo

Estabilidad (voz)

Piénsalo así: Lo firme o lo cambiante que suena el tono de alguien al hablar.

Parámetro de ElevenLabs que controla cuánto varía la entonación de la voz generada

entre frases.

Ejemplo: Una estabilidad alta produce una voz más uniforme, adecuada para mensajes informativos breves.

Preparación

Ten tu cuenta de ElevenLabs iniciada.

Procedimiento paso a paso

1. **Contexto:** Vas a elegir una voz en español dentro de la biblioteca de voces.

Acción de teclado: Entra en la sección de voces (Voice Library) y busca un filtro de idioma. Actívalo para español y recorre la lista de resultados con **Tab**.

Respuesta esperada de NVDA: NVDA anunciará el nombre de cada voz conforme la recorras.

Qué significa: Cada elemento de la lista suele incluir un botón para reproducir una muestra.

Acción siguiente: Reproduce al menos tres muestras distintas antes de elegir una.

2. **Contexto:** Has elegido una voz y quieres generar un audio de prueba ajustando sus parámetros.

Acción de teclado: Entra en la sección de generación de texto a voz (Text to Speech), escribe un texto breve de prueba, y recorre con **Tab** los controles deslizantes de estabilidad, similitud y claridad.

Respuesta esperada de NVDA: NVDA debería anunciar el valor actual de cada control deslizante y permitir modificarlo con las flechas del teclado.

Qué significa: Puedes ajustar cada parámetro exclusivamente con el teclado, sin necesidad de arrastrar nada con el ratón.

Acción siguiente: Genera el audio y escúchalo; ajusta los parámetros y repite hasta obtener un resultado que te convenza.

Posibles diferencias de interfaz

Los controles deslizantes de parámetros de voz son, en algunas plataformas, poco accesibles con teclado si no están correctamente etiquetados. Si un control deslizante no responde a las flechas del teclado tras situarte sobre él con Tab, es un problema de accesibilidad real de la plataforma, no un error tuyo: consulta el manual de referencia de ElevenLabs de esta plataforma para alternativas conocidas en el momento en que lo uses.

Problema de accesibilidad y alternativa

Problema de accesibilidad conocido en herramientas de este tipo: los controles deslizantes (sliders) no siempre son accesibles por completo con teclado en todas las plataformas de generación de voz. **Alternativa mediante teclado:** tras situar el foco sobre el control con Tab, prueba las flechas izquierda/derecha o arriba/abajo; la mayoría de implementaciones modernas los soportan.

Alternativa mediante API: si el control web resulta inaccesible, los mismos parámetros (estabilidad, similitud) pueden ajustarse mediante la API de ElevenLabs, que se trata en la siguiente lección, sin depender de ningún control visual.

Errores frecuentes y diagnóstico

La voz generada suena artificial o entrecortada.

Cómo localizarlo: Puede deberse a una estabilidad demasiado baja combinada con un texto de prueba muy corto, o a una similitud llevada al extremo.

Solución: Prueba con valores medios de estabilidad y similitud como punto de partida, y ajusta gradualmente.

Comprobación final

Genera un audio de prueba con el texto «Hola. La voz del agente está configurada correctamente.» y escúchalo completo. Si el audio se genera y se reproduce con claridad, la comprobación es positiva.

Qué acabas de conseguir

Tienes una voz elegida y ajustada, lista para usarse en la lección final de este módulo, donde generarás la prueba de voz definitiva del hito.

Ejercicio práctico

Genera el mismo texto de prueba con dos combinaciones distintas de estabilidad y similitud, y compara cuál suena más adecuada para un recepcionista telefónico profesional.

Resumen de lo imprescindible

- La biblioteca de voces permite filtrar por idioma y escuchar muestras antes de elegir.
- Estabilidad, similitud, expresividad, claridad y latencia son los parámetros clave a ajustar.
- Para un recepcionista telefónico, priorizar claridad y latencia baja suele ser más importante que la expresividad máxima.
- Si un control deslizante no es accesible con teclado, existe alternativa mediante la API (siguiente lección).

No necesitas aprender todavía

Todavía no necesitas clonar una voz personalizada: no forma parte del itinerario obligatorio de este curso.

Estoy preparado para continuar si puedo...

- Elegir una voz en español dentro de la biblioteca de ElevenLabs.
- Ajustar estabilidad y similitud con el teclado.
- Explicar por qué la latencia importa especialmente en telefonía en tiempo real.

Test de comprobación

Para registrar el resultado y obtener Puntos de profesionalidad, realiza este test desde la versión web de la lección. Las opciones se muestran aquí sin señalar la respuesta correcta.

1. Estás configurando la voz de un recepcionista telefónico profesional. ¿Qué combinación de prioridades tiene más sentido según esta lección?
 - a. Priorizar la máxima expresividad emocional posible, aunque aumente la latencia.
 - b. Priorizar claridad y latencia baja, con una estabilidad media-alta, para evitar silencios incómodos y una entonación errática.

- c. La latencia no tiene ninguna relación con la calidad percibida de una llamada telefónica.
2. Generas una prueba de voz y el resultado suena entrecortado y con una entonación muy cambiante entre frases. Según la documentación oficial de ElevenLabs citada en esta lección, ¿qué parámetro ajustarías primero y en qué sentido?
- a. Subir la estabilidad, porque valores bajos introducen un rango emocional más amplio y menos predecible, lo que puede percibirse como una entonación errática.
 - b. Bajar la similitud al mínimo, porque es el único parámetro que afecta a la naturalidad de la voz.
 - c. Aumentar la latencia manualmente, porque una latencia más alta siempre mejora la calidad del audio.
3. Un cliente potencial te pregunta si ElevenLabs puede generar la voz de su recepcionista en español de España. Según la documentación oficial de idiomas de ElevenLabs citada en esta lección, ¿qué puedes responder con confianza?
- a. Que no es posible, porque ElevenLabs sólo admite español de Latinoamérica.
 - b. Que sí, ya que la documentación oficial de idiomas y modelos de ElevenLabs confirma soporte explícito para las variantes de español de España y de México dentro de sus modelos multilingües.
 - c. Que habría que esperar a una futura actualización, porque el español no está soportado todavía.

Fuentes

- [Get voice settings — ElevenLabs \(consultado el 2026-08-30\)](#)
- [What languages do you support? — ElevenLabs \(centro de ayuda oficial\) \(consultado el 2026-08-30\)](#)
- [Models — ElevenLabs \(consultado el 2026-08-30\)](#)

Última actualización de este contenido: **2026-08-30**.