Voces consistentes en videos con IA

@StephanoBrooke · Palabra clave: TONO

Lo que necesitas

La mayoría todavía no sabe cómo lograr voces consistentes con Seedance 2.5. Este es el flujo completo para que tu personaje suene igual en cada escena.
Necesitas Claude, una cuenta de ElevenLabs con Voice Design, Text to Speech y la pestaña Video (Seedance 2.5) con créditos, y dos imágenes: la hoja de tu personaje y la locación.

Paso 1: Claude describe la voz de tu personaje

Sube la hoja de tu personaje (de frente, de lado y detalles de cara y ropa) a Claude y escribe:

> *"Describe la voz de este personaje."*

Original en inglés: "Describe this character's voice."
Qué pasa después: Claude te da una descripción detallada de la voz. Esta es la que le salió al creador del video original:

> *"Una mujer de casi cuarenta años, con una voz grave y rasposa, ronca por años de polvo, tabaco y licor barato. Acento con toque británico, cadencia victoriana un poco cortada y desgastada por una vida dura. Forma de hablar seca, irónica y sin prisa, con un aire permanente de diversión, como si ya hubiera escuchado cada historia dos veces. Registro cálido de medio a grave, con aire en las orillas y un ligero quiebre en las vocales largas. Habla como alguien que nunca necesita levantar la voz para que la obedezcan. Grabación limpia, micrófono cerca, sin ruido de fondo."*

Tip: entre más detalles tenga tu hoja de personaje (edad, ropa, época), más precisa sale la voz.

La descripción original en inglés

Pégala en inglés si Voice Design te da mejores resultados así:

> *"A woman in her late thirties with a low, gravel-scarred voice, raspy from years of dust, smoke and cheap liquor. British-tinged accent, slightly clipped Victorian cadence worn down by rough living. Dry, wry, unhurried delivery with a permanent edge of amusement, like she has heard every story twice. Warm mid-to-low register, breathy at the edges, with a slight crack on held vowels. Speaks like someone who never needs to raise her voice to be obeyed. Clean recording, close-mic, no background noise."*

Paso 2: crea la voz en Voice Design

En ElevenLabs entra a Voice Design y pega la descripción completa del Paso 1 en el campo Prompt.
Da clic en Generate voice (en el video original marca 350 créditos).
Qué pasa después: salen tres opciones, Voice 1, Voice 2 y Voice 3. Escúchalas y quédate con la que más te guste.
Tip: si ninguna te convence, ajusta una o dos palabras de la descripción (edad, acento, ritmo) y vuelve a generar.

Paso 3: genera el audio de prueba

En Text to Speech elige la voz que guardaste y escribe esta frase, que pronuncia todos los sonidos del inglés:

> *"That quick beige fox jumped in the air over each thin dog. Look out, I shout, for he's foiled you again, creating chaos"*

Da clic en Generate speech y descarga el audio (en el video original queda como voice1.wav).
Qué pasa después: ese audio es la referencia de voz de tu personaje. Lo vas a usar en todas sus escenas.
Extra, si tu personaje habla español (frase con todas las letras del alfabeto):

> *"El veloz murciélago hindú comía feliz cardillo y kiwi. La cigüeña tocaba el saxofón detrás del palenque de paja."*

Paso 4: Claude escribe tu prompt de Seedance 2.5

Regresa a Claude, sube la hoja del personaje, la imagen de la locación y el audio, y escribe:

> *"Escribe un prompt para Seedance 2.5 usando este personaje, esta locación y esta referencia de voz."*

Original en inglés: "Write a Seedance 2.5 prompt using this character, location and voice reference."
Qué pasa después: Claude te da un prompt dividido por tomas, con tiempos, efecto de cámara, luz y la línea que dice el personaje. Abajo está el que le salió al creador, traducido, con la línea ya cambiada por tu CTA.
Tip: si Claude no acepta el archivo de audio, sube solo las dos imágenes y pega la descripción de la voz del Paso 1.

El prompt de Seedance: toma 1

"TOMA 1 (00:00-00:01.2), inserto de la tablet. EFECTO: acercamiento lento (dolly) + brillo de pantalla + polvo flotando en primer plano. Cerrado sobre la tablet maltratada, blindada en latón y cobre, parada en vertical sobre la barra de madera rescatada. Bordes remachados, cables deshilachados colgando del borde de la barra. La pantalla está encendida y muestra una página oscura con un título grande en blanco y bloques pequeños de texto de colores. La cámara se acerca despacio, más o menos 6 a 8 por ciento de escala durante la toma. A la altura de los ojos, un poco arriba de la línea de la barra. Poca profundidad de campo, pantalla nítida, fondo difuminado. Tiempo real, sin cambios de velocidad. Quieto a propósito para crear calma antes de que ella hable. NO hay corte. Sale con un cambio de foco continuo hacia la Toma 2."

El prompt de Seedance: toma 2

"TOMA 2 (00:01.2-00:04.6), la línea. EFECTO: cambio de foco (de primer plano a fondo), ESTE ES EL EFECTO VISUAL PRINCIPAL + cámara fija. El foco sale de la pantalla de la tablet y cae en la cantinera parada detrás de la barra: cabello corto cobrizo y despeinado, piel pálida con pecas, maquillaje de ojos ahumado color ciruela, labios color vino, pequeñas marcas doradas en forma de rombo bajo el ojo derecho, placas remachadas de latón y cobre alrededor del cuello, sombrero de copa de piel verde musgo desgastado con llaves y engranes de latón en la cinta, levita de brocado adamascado verde azulado con hombros abullonados sobre un corsé de satín dorado y puños de encaje crema, guantes de piel café. La tablet sigue en cuadro, ahora desenfocada abajo en primer plano, con el brillo de la pantalla cayendo cálido sobre la barra. Ella apoya una mano en la barra, mira directo a la lente con media sonrisa seca y dice: 'Sígueme y comenta TONO y te mando el flujo completo en PDF.' Cámara totalmente fija durante toda la línea. Sin acercamiento, sin deriva, sin cámara en mano. Tiempo real, sin rampas. Luz cálida de sol bajo entra desde la derecha del cuadro y cruza su hombro, mucho polvo suspendido en el haz, sombra profunda llenando la cabina detrás. Sale en un tiempo sostenido, sin transición."

Cambia la línea entre comillas simples por lo que quieras que diga tu personaje.

El prompt de Seedance: toma 3

"TOMA 3 (00:04.6-00:05.0), cierre. EFECTO: quietud + pulso suave del brillo de pantalla. Ella baja la mirada a la pantalla de la tablet y rompe el contacto visual. Se queda con media sonrisa. Cámara sigue fija. Sin movimiento. Tiempo real. La toma simplemente termina, sin fundido ni adorno. Cuadro final: ella, la barra, la tablet encendida y las dunas ardiendo en blanco por la ventana abierta detrás de ella."

Paso 5: súbelo a Seedance 2.5 y etiqueta el audio

En ElevenLabs ve a Image / Video, pestaña Video, y elige Seedance 2.5. En el video original: 16:9, 720p, 5 s y audio encendido.
Sube la hoja del personaje y la locación como imágenes de referencia, y el audio en Audio refs. Pega las 3 tomas y al final agrega esta línea:

> *"@voice1 es la voz de @character. @voice1 es solo una referencia: iguala el timbre, el tono y el acento de esta voz femenina."*

Original en inglés:

> *"@voice1 is the voice of @character. @voice1 is a reference only, match this female voice's timbre, tone and accent."*

Qué pasa después: sale un clip de 5 segundos del personaje diciendo su línea con la misma voz de tu audio.
Tip: para cada escena nueva repite los pasos 4 y 5 con el mismo audio. Así la voz no cambia entre generaciones.

El prompt de Seedance en inglés (original): tomas 1 y 3

> *"SHOT 1 (00:00-00:01.2) - Tablet Insert. EFFECT: Slow push-in (dolly) + screen bloom + foreground dust drift. Tight on the battered brass and copper armoured tablet standing upright in vertical orientation on the salvaged timber counter. Riveted edges, frayed cabling trailing off the counter edge. Screen is lit, showing a dark document page with a large white heading and small coloured text blocks. Camera pushes in slowly, roughly 6-8% scale over the shot. Eye level, slightly above the counter line. Shallow depth of field, screen sharp, background dissolved. Real time, no speed manipulation. Deliberately still to establish calm before she speaks. Does NOT cut. Exits via a continuous focus pull into Shot 2."*

> *"SHOT 3 (00:04.6-00:05.0) - Resolve. EFFECT: Stillness hold + subtle screen bloom pulse. She glances down at the tablet screen, breaking eye contact. Half smile settles. Camera still locked. No movement. Real time. The shot simply ends - no fade, no flourish. Final frame: her, the counter, the lit tablet, dunes burning white through the open window frame behind her."*

El prompt de Seedance en inglés (original): toma 2

"SHOT 2 (00:01.2-00:04.6) - The Line. EFFECT: Rack focus (foreground to background) - THIS IS THE SIGNATURE VISUAL EFFECT + locked camera hold. Focus travels off the tablet screen and lands on the bartender standing behind the counter: short tousled copper-red hair, pale freckled skin, smoky plum eye makeup, dark wine lipstick, small gold diamond markings under her right eye, riveted brass and copper plating fitted around her throat, weathered mossy-green leather top hat with brass keys and cogs on the band, teal damask brocade tailcoat with puffed shoulders over a gold satin corset and cream lace cuffs, brown leather gloves. The tablet stays in frame, now softly defocused in the lower foreground, its screen glow spilling warm onto the counter. She rests one hand on the counter, looks directly into the lens with a dry half smile, and speaks: 'Comment AI and I'll give you the full workflow as a PDF.' Camera fully locked off for the entire line. No push, no drift, no handheld. Real time, no ramping. Warm low-angle sun rakes in from frame right across her shoulder, heavy dust motes suspended in the beam, deep shadow filling the cabin behind. Exits on a held beat, no transition device."

Las tomas que grabas para el reel

- Toma 1: acostado en la cama, cámara cenital fija. Alguien baja una lámpara sobre tu cara y al final levantas una tablet. Es la mitad Original del hook.
- Toma 2: tu talking head explicando los pasos.
- Toma 3: grabación de pantalla de Claude y ElevenLabs.

La escena final del personaje se genera completa desde referencias, sin toma física. La versión IA de la cama sale de usar tu Toma 1 como referencia de video.

Prompts bonus

Para que Claude te dé la voz lista para Voice Design:

"Describe la voz de este personaje como si fuera un prompt para Voice Design de ElevenLabs: edad, tono, acento, ritmo, actitud y tipo de grabación."

Para la siguiente escena con la misma voz:

"Escribe otro prompt para Seedance 2.5 con este mismo personaje, esta locación y esta referencia de voz. Ahora ella dice: '[TU LINEA]'. Mantén el mismo estilo de tomas y termina con la línea que liga @voice1 a @character."

Para la versión IA de una toma que grabaste:

"Mismo encuadre y mismos movimientos que el video de referencia. Cambia el lugar por [TU LOCACION] y la ropa por [TU VESTUARIO]. Conserva su cara y su voz. @voice1 es la voz de la mujer que le habla fuera de cuadro."

Lo que puedes esperar de verdad

- Cada voz y cada clip gastan créditos de ElevenLabs. Revisa el costo en tu plan antes de generar.
- El resultado cambia en cada generación. Puede que necesites varios intentos para que la voz y la boca coincidan.
- Los menús pueden cambiar: los nombres de esta guía son los que se ven en el video original.
- Voice Design crea una voz nueva desde texto. Si clonas una voz, que sea la tuya o la de alguien que te dio permiso. Y publica con la etiqueta de contenido hecho con IA.

Cómo usarlos

1. Claude describe la voz con la hoja de tu personaje.
2. Voice Design crea la voz y eliges una de tres.
3. Text to Speech te da el audio de prueba.
4. Claude escribe el prompt de Seedance 2.5 con personaje, locación y audio.
5. En Seedance 2.5 subes todo y al final ligas @voice1 a @character.
6. Repite los pasos 4 y 5 por escena con el mismo audio.

Cuéntame qué te salió: coméntalo en el reel o etiquétame en @stephanobrooke.