

Voces abiertas de Piper TTS clasificadas: naturalidad, claridad, velocidad y licencias de 1.683 voces

Asad Shehzad
Loudness.ai

Informe técnico LA-RB-2026-003, versión 1.0, 15 de octubre de 2026
<https://loudness.ai/es/research/piper-tts-voices-ranked/>

Resumen

Piper es un motor neuronal de texto a voz muy usado en procesadores corrientes, con un catálogo oficial de 177 archivos de voz en 53 idiomas y otros modelos de la comunidad. Elegir una voz es difícil: el catálogo no ofrece datos comparativos de calidad, varios modelos contienen cientos de hablantes y las licencias cambian de un archivo a otro. Revisamos la licencia de cada archivo de voz oficial y excluimos 102: 76 ajustados a partir de la voz Lessac, cuyas grabaciones solo pueden usarse sin fines comerciales, y 26 con una licencia propia no comercial o solo para investigación. Después probamos todos los hablantes de los 75 archivos oficiales restantes, en 27 idiomas, y de tres modelos de la comunidad entrenados desde cero: 1.683 voces en 32 variantes de idioma. Cada voz leyó tres frases en su idioma y recibió una puntuación de naturalidad del predictor UTMOS22 y otra de claridad a partir de la tasa de error de caracteres de una transcripción de Whisper large-v3, combinadas en una puntuación global y clasificadas dentro de su idioma. Las mejores voces son inglesas: lo son las 793 voces que obtuvieron 90 o más sobre 100. Un modelo de la comunidad, Clean 100, igualó a la mejor voz oficial (96,4 frente a 96,2), con un hablante mediano mejor (93,2 frente a 90,8), y sintetizó 6,2 veces más rápido en la CPU de un portátil. Dentro de un mismo modelo, los hablantes difirieron hasta 46,3 puntos, más que la distancia entre la mayoría de los modelos. Publicamos las puntuaciones, las licencias, las sumas de verificación, enlaces de descarga fijados a la versión probada, muestras de audio, los scripts y un explorador interactivo.

1. Introducción

Piper [1] sintetiza voz con modelos VITS [4] lo bastante pequeños para funcionar más rápido que el tiempo real en unos pocos núcleos de procesador, por lo que es una opción habitual para asistentes locales, herramientas de accesibilidad y servicios web. Su catálogo oficial [2] contiene 177 archivos de voz en 53 idiomas, la mayoría entrenados por voluntarios con corpus de voz públicos, y otros autores publican más modelos [3].

Tres problemas complican la elección de una voz. Primero, el catálogo describe cada archivo por idioma, conjunto de datos y un nivel de calidad aproximado, pero no da ninguna medida de calidad, así que los usuarios escuchan las voces una a una. Segundo, los modelos multihablante esconden muchas voces tras un solo archivo: solo el modelo LibriTTS tiene 904 hablantes, y el hablante 0 predeterminado es el que usa la mayoría de las herramientas. Tercero, las licencias varían: una ficha puede indicar una licencia de datos permisiva aunque el modelo se haya ajustado a partir de otra voz cuyas grabaciones solo pueden usarse sin fines comerciales.

Este informe responde a una pregunta práctica: *¿qué voces de Piper puede usar una empresa y cuáles suenan mejor en cada idioma?* Nuestras aportaciones son:

1. Una revisión de licencias de todos los archivos de voz oficiales de Piper, incluido el linaje de ajuste fino que las fichas no siempre indican, y la lista de archivos excluidos con su motivo.
2. Puntuaciones automáticas de naturalidad y claridad, una puntuación global y la posición dentro del idioma de todos los hablantes de cada archivo oficial restante y de tres modelos de la comunidad: 1.683 voces.
3. Resultados sobre modelos de la comunidad frente a oficiales, variación entre hablantes, niveles de calidad, género y velocidad en CPU.
4. Datos abiertos: puntuaciones por voz, licencias por archivo, sumas de verificación y enlaces de descarga fijados a la revisión probada, muestras de audio, scripts y un explorador interactivo.

2. Voces y licencias

Catálogo. Usamos la revisión `c10ece1` del repositorio oficial de voces [2], la vigente en octubre de 2026, y registramos el tamaño y la suma de verificación de cada archivo. Leímos íntegra la ficha de cada voz, con sus datos de entrenamiento y su licencia.

Linaje. Las voces de Piper suelen entrenarse ajustando una voz existente en lugar de partir de cero, lo que requiere muchos menos datos. El ajuste fino parte de los pesos del modelo de origen, así que la nueva voz deriva en parte de las grabaciones de ese modelo. Muchas voces oficiales descienden, directamente o a través de otra voz, de la voz Lessac, entrenada con grabaciones publicadas para el Blizzard Challenge 2013 [5] con una licencia no comercial. No está resuelto si un modelo ajustado hereda la licencia de los datos de su modelo de origen; por precaución excluimos los 76 archivos oficiales con linaje Lessac, aunque su propio conjunto de datos sea permisivo. Excluimos otros 26 archivos cuya propia licencia es no comercial o solo para investigación. El apéndice B enumera los 102 archivos. Los 75 archivos oficiales restantes abarcan 27 idiomas.

Modelos de la comunidad. Bryce Beattie, autor de varias voces oficiales, publica otros modelos de Piper entrenados desde cero [3]. Añadimos tres: Clean 100 (245 hablantes, entrenado con el subconjunto clean-100 de LibriTTS-R [7], CC BY 4.0), ManyVoice (16 hablantes de inglés a partir de grabaciones de LibriVox de dominio público) y Jenny (inglés de Irlanda; su licencia permite el uso comercial pero exige que la voz se llame «Jenny» en el software que genera audio). Su modelo Norman es idéntico byte a byte al archivo oficial Norman y figura una sola vez; probado dos veces con los dos nombres, obtuvo 89,6 y 89,5.

Clases de licencia. Agrupamos las licencias según lo que piden a un usuario comercial (cuadro 1): *sin condiciones* (dominio público, CC0), *requiere atribución* (CC BY, CC BY-SA, Apache 2.0 y la licencia de Jenny; CC BY-SA pide además que las adaptaciones de la voz conserven la licencia), *copyleft* (licencias de modelo GPL y AGPL) y *poco clara* (la ficha no indica licencia o solo remite a una página web). Este resumen no es asesoramiento jurídico.

3. Método

Frases de prueba. Todas las voces de un idioma leyeron las mismas tres frases: una breve presentación de la voz, una frase sobre una reunión el lunes por la mañana y una pregunta sobre la estación de tren más cercana, traducidas para cada una de las 27 familias de idiomas. No contienen

Clase de licencia	Voces	Licencias
Sin condiciones	51	CC0 1.0, Dominio público
Requiere atribución	1.613	CC BY 4.0, CC BY-SA 4.0, Apache 2.0, CC BY-SA 3.0 ES, Licencia Jenny (Dioco)
Copyleft	2	GPL-2.0, AGPL-3.0
Poco clara	17	Sin indicar

Tabla 1: Voces probadas por clase de licencia. Todas las clases permiten el uso comercial; las voces de licencia «poco clara» requieren comprobar la fuente antes de ese uso.

cifras, abreviaturas ni símbolos, así que la normalización del texto no influye en el resultado (la estudiamos aparte [14]).

Síntesis. Usamos Piper 1.8.0 con la configuración predeterminada de cada modelo (escalas de duración, ruido y anchura de ruido de su configuración), una toma por frase y dos hilos, y convertimos cada clip a 16 kHz mono para puntuarlo. Los modelos multihablante se ejecutaron para cada identificador de hablante de su configuración.

Naturalidad. El predictor UTMOS22 «strong» [9], mediante la versión SpeechMOS [11], predice la puntuación media de opinión (1–5) que darían los oyentes a un clip; fue uno de los mejores sistemas del VoiceMOS Challenge 2022 [10]. La promediamos en las tres frases y la reescalamos a 0–100: $\text{Naturalidad} = 100(\text{MOS} - 1)/4$.

Claridad. Las tres frases se unen con 0,6 s de silencio y se transcriben de una vez con Whisper large-v3 [12] mediante faster-whisper 1.2.1 [13] (int8, decodificación voraz, idioma forzado, sin filtro de actividad de voz). La transcripción y la referencia se normalizan con NFKC, se pasan a minúsculas y se les quitan los signos de puntuación (y los espacios en chino), y $\text{Claridad} = 100(1 - \text{CER})$, donde CER es la tasa de error de caracteres.

Global. $\text{Global} = \text{Naturalidad}/2 + \text{Claridad}/2 + 4 \cdot [22 \text{ kHz}]$. Las voces de Piper de calidad media y alta están grabadas a 22,05 kHz, y las de calidad baja y x_low a 16 kHz; las dos medidas trabajan con audio a 16 kHz y no perciben el ancho de banda adicional, que los oyentes sí perciben, así que las voces a 22 kHz reciben 4 puntos. Las voces se clasifican solo dentro de su variante de idioma: UTMOS22 se entrenó sobre todo con inglés y Whisper es mucho más débil en algunos idiomas, así que las puntuaciones no son comparables entre idiomas.

Dos rondas. La primera ronda probó hasta doce hablantes de cada archivo en un servidor en la nube (onnxruntime 1.31.0). La segunda, en un portátil (onnxruntime 1.19.2), cribó con una frase todos los hablantes de los seis conjuntos multihablante más grandes y de los modelos de la comunidad, y después probó los 1.580 con las tres frases; la naturalidad de una frase predijo la media de tres con $r = 0,967$. Cuando una voz se probó dos veces, conservamos el resultado de la segunda ronda. Sesenta voces medidas en ambas rondas coincidieron con $r = 0,948$ y una diferencia absoluta media de 3,4 puntos, que tomamos como ruido de medida: unos pocos puntos de diferencia entre dos voces no son significativos.

Género. Indicamos el género de una voz solo cuando su conjunto de datos de entrenamiento registra al lector (las listas de hablantes de LibriTTS y LibriTTS-R [6, 7] y el conjunto de datos

de Jenny); nunca se deduce del audio. Por eso los hablantes de MLS [8] y la mayoría de las voces de un solo hablante figuran como «sin registrar». Así quedan 553 voces de mujer, 597 de hombre y 533 sin registrar.

Velocidad. Para cada modelo que tenemos en local cronometramos las tres frases tres veces con dos hilos y damos la mediana del factor de tiempo real (segundos de cálculo por segundo de voz). La velocidad depende del tamaño del modelo, no del hablante.

4. Resultados

4.1. Las mejores voces por idioma

El cuadro 2 y la figura 1 muestran la voz mejor y la mediana de cada variante de idioma. El inglés va muy por delante: las 793 voces con una puntuación global de 90 o más son todas inglesas, y 1.197 voces obtienen 80 o más. Italiano, ucraniano, sueco, euskera, neerlandés, islandés, urdu y español de México tienen cada uno una voz por encima de 85. La mayoría de las variantes no inglesas tienen entre una y tres voces. Telugu, vietnamita, nepalí y armenio obtienen las puntuaciones más bajas, en parte porque el propio reconocedor falla (sus transcripciones en telugu derivan a la escritura canaresa); su clasificación necesita la comprobación de un oyente nativo.

4.2. Un modelo de la comunidad iguala al mejor oficial

El mejor hablante de Clean 100 obtuvo 96,4 y el mejor hablante del modelo oficial LibriTTS de alta calidad 96,2, una diferencia muy inferior al ruido de medida. Clean 100 es más regular: su hablante mediano obtuvo 93,2 frente a 90,8, y el peor 87,1 frente a 78,6 (cuadro 3). Es un modelo de calidad media y sintetizó la voz en 0,091 del tiempo real frente a 0,568 de LibriTTS alta (cuadro 5), 6,2 veces más rápido; en un servidor en la nube con la misma configuración la relación fue de 4,3. Ambos modelos proceden de lectores de LibriTTS, pero Clean 100 se entrenó con el audio restaurado de LibriTTS-R, sin las restauraciones fallidas, lo que podría explicar su menor dispersión.

4.3. El hablante importa más que el modelo

La figura 2 muestra todos los hablantes de los ocho modelos con diez o más hablantes. En MLS francés, los hablantes van de 24,4 a 70,7 (46,3 puntos); en MLS alemán el rango es de 31,0 puntos. No estudiamos por qué algunos hablantes puntúan bajo. El hablante 0, el predeterminado en la mayoría del software, rara vez es el mejor: quien use un modelo multihablante debería elegir el hablante a propósito.

4.4. Niveles de calidad, género y las dos medidas

Los niveles de calidad de Piper describen el tamaño del modelo y la frecuencia de muestreo. El nivel alto tiene la mediana más alta (cuadro 4), pero los niveles se confunden con el idioma y el conjunto de datos: la mayoría de las voces de calidad alta son hablantes ingleses de LibriTTS, y muchas de calidad media son hablantes de MLS de otros idiomas. En inglés, las mejores voces de calidad media igualan a las mejores de calidad alta. Las voces de mujer y de hombre puntuaron igual: entre los lectores de LibriTTS, las mujeres promediaron 90,8 y los hombres 90,1; en Clean 100, 93,6 y 92,5. Naturalidad y claridad se correlacionan en el conjunto de voces ($r = 0,80$), pero no tanto como para sustituirse: una voz puede entenderse bien y sonar poco natural, o al revés.

Idioma	Voces	Mejor voz	Mejor	Mediana
Inglés (Estados Unidos)	1.158	Clean 100 · hablante 118	96,4	91,3
Inglés (mezcla EE. UU. y Reino Unido)	16	ManyVoice · hablante 2	95,0	89,2
Inglés (Irlanda)	1	Jenny	93,5	93,5
Inglés (Reino Unido)	4	Cori (medium)	92,4	90,1
Italiano (Italia)	3	Serena (high)	89,6	86,3
Ucraniano (Ucrania)	7	Tetiana	88,1	84,8
Sueco (Suecia)	2	NST	86,4	86,3
Euskera (España)	2	Maidier	85,6	82,3
Neerlandés (Países Bajos)	55	Alex	85,6	61,3
Islandés (Islandia)	4	Steinn	85,3	83,2
Urdu (Pakistán)	1	Aegis Female	85,2	85,2
Español (México)	2	Claude	85,0	79,5
Esloveno (Eslovenia)	1	Artur	84,6	84,6
Persa (Irán)	1	Gyro	84,2	84,2
Letón (Letonia)	1	Aivars	82,8	82,8
Noruego (Noruega)	10	NVCC · hablante 6	82,0	77,8
Alemán (Alemania)	241	Karlsson	81,8	70,0
Chino (China)	1	Huayan	79,6	79,6
Neerlandés (Bélgica)	3	Nathalie	79,5	76,8
Checo (República Checa)	1	Kasandra	78,1	78,1
Griego (Grecia)	2	Rapunzelina (medium)	77,0	72,2
Kazajo (Kazajistán)	8	ISSAI · hablante 4	77,0	71,1
Finés (Finlandia)	1	Harri	76,8	76,8
Francés (Francia)	129	Gilles	76,8	61,8
Catalán (España)	2	UPC Ona	76,6	76,2
Español (España)	3	MLS 9972	75,8	75,1
Portugués (Brasil)	1	Edresson	74,1	74,1
Polaco (Polonia)	1	MLS 6892	68,8	68,8
Armenio (Armenia)	1	Gor	64,3	64,3
Telugu (India)	2	Padmavathi	56,8	54,1
Vietnamita (Vietnam)	1	25hours single	56,4	56,4
Nepalí (Nepal)	18	Google · hablante 1	55,5	48,0

Tabla 2: Puntuación global mejor y mediana por variante de idioma, con el número de voces probadas.

Voz mejor y mediana de cada idioma

Puntuación global (0-100) de las voces que permiten uso comercial. Compara dentro de un idioma.

● Mejor voz ● Voz mediana

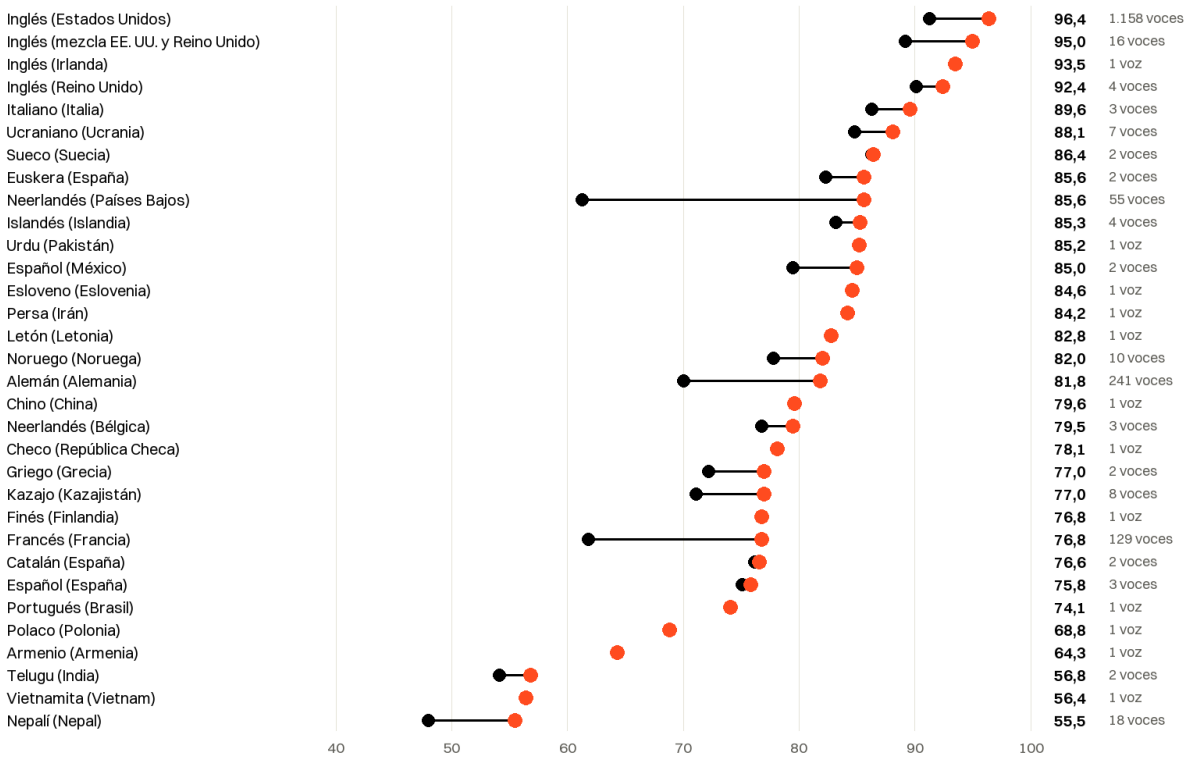


Figura 1: Puntuación global mejor (naranja) y mediana (negro) por variante de idioma.

Modelo	Hablantes	Mejor	Mediana	Peor	Rango
Clean 100 (comunidad)	245	96,4	93,2	87,1	9,3
LibriTTS (inglés, alta)	904	96,2	90,8	78,6	17,6
ManyVoice (comunidad)	16	95,0	89,2	76,1	18,9
NVCC (noruego, media)	10	82,0	77,8	73,0	9,0
MLS (alemán, media)	236	81,2	69,9	50,2	31,0
MLS (francés, media)	125	70,7	61,8	24,4	46,3
MLS (neerlandés, media)	52	69,7	61,2	30,9	38,8
Google (nepalí, x_low)	18	55,5	48,0	43,8	11,7

Tabla 3: Puntuación global de todos los hablantes de los modelos con diez o más hablantes.

Calidad	Voces	Mediana	Mejor
alta	917	90,7	96,2
media	715	72,6	96,4
baja	21	75,8	92,2
muy baja (x_low)	30	52,0	79,6

Tabla 4: Puntuación global por nivel de calidad de Piper.

4.5. Velocidad

El cuadro 5 da el factor de tiempo real de cada modelo que tenemos en local, en un portátil Apple M4 con dos hilos. Los modelos de calidad media y x_low sintetizan la voz entre 10 y 18 veces más rápido que el tiempo real; LibriTTS alta, unas dos veces. En un servidor en la nube con 12 núcleos virtuales AMD EPYC y los mismos dos hilos, las voces de calidad media midieron 0,17–0,19 y LibriTTS alta 0,83, entre 1,5 y 2 veces más lento que el portátil.

5. Cómo usar los resultados

La página web del informe incluye un explorador que filtra las 1.683 voces por idioma, género, clase de licencia, calidad, fuente y puntuación mínima, reproduce una muestra de cada una y enlaza sus archivos de modelo y configuración. Los archivos oficiales enlazan al repositorio en la revisión probada, de modo que coinciden con sus sumas de verificación publicadas; los de la comunidad enlazan al alojamiento de su autor, con sumas SHA-256. Cada voz muestra su licencia, sus datos de entrenamiento, una línea de atribución cuando hace falta y la orden de Piper para su hablante. El apéndice A enumera las tres mejores voces de cada variante de idioma.

De los resultados se desprenden consejos prácticos. En inglés, Clean 100 ofrece muchos hablantes buenos a la velocidad de un modelo medio. En los modelos multihablante, elige el hablante escuchando los primeros de la clasificación en lugar de usar el hablante 0. En los idiomas con pocos recursos, toma nuestra clasificación como punto de partida y consulta a un hablante nativo. Revisa la clase de licencia: la mayoría de las voces necesitan una línea de atribución visible.

6. Limitaciones

Todas las puntuaciones son automáticas; no se hizo ninguna prueba de escucha con personas. UTMOS22 se entrenó sobre todo con valoraciones en inglés, y sus valores en otros idiomas son menos fiables que su orden dentro de un idioma. Las tres frases de cada idioma se escribieron para

El hablante que eliges importa más que el modelo

Puntuación global de todos los hablantes de cada modelo con diez o más; la forma gris muestra cuántas voces obtienen cada nivel.

● Mejor voz | Voz mediana

Clean 100 · inglés (EE. UU.), comunidad

245 hablantes · 87,1-96,4

LibriTTS · inglés (EE. UU.), alta

904 hablantes · 78,6-96,2

ManyVoice · inglés, comunidad

16 hablantes · 76,1-95,0

NVCC · noruego

10 hablantes · 73,0-82,0

MLS · alemán

236 hablantes · 50,2-81,2

MLS · francés

125 hablantes · 24,4-70,7

MLS · neerlandés

52 hablantes · 30,9-69,7

Google · nepalí, x_low

18 hablantes · 43,8-55,5

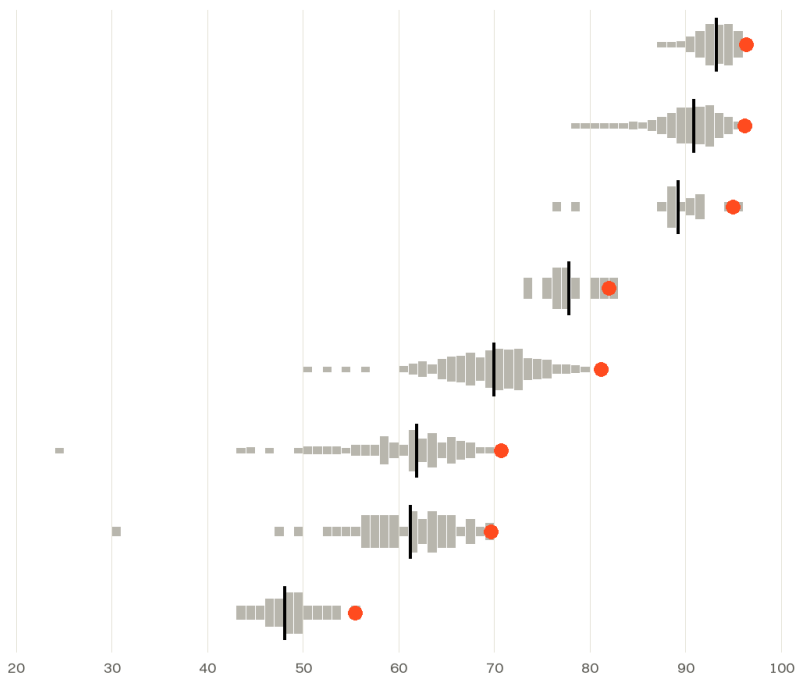


Figura 2: Todos los hablantes de cada modelo con diez o más: distribución (gris), mediana (negro) y mejor voz (naranja).

esta prueba y se tradujeron, no las compusieron hablantes nativos, y cada una se sintetizó una vez; la salida de Piper varía un poco de una toma a otra. Las dos rondas se ejecutaron en máquinas y versiones del entorno distintas. El género solo cubre lo que registran los conjuntos de datos. Los resúmenes de licencia reflejan las fichas en la revisión probada y no son asesoramiento jurídico.

7. Disponibilidad de los datos

El paquete del estudio (puntuaciones, licencias por archivo, sumas de verificación y enlaces, las frases de prueba, 1.683 muestras de audio y los scripts de síntesis, puntuación, clasificación y velocidad) está disponible en la página web del informe. Los datos y el texto se publican con licencia CC BY 4.0 y los scripts con la licencia Apache 2.0; las muestras conservan la licencia de su modelo.

Modelo	Calidad	Frec. (Hz)	FTR
LibriTTS (inglés, alta)	alta	22.050	0,568
MLS (alemán, media)	media	22.050	0,083
MLS (francés, media)	media	22.050	0,086
MLS (neerlandés, media)	media	22.050	0,089
Google (nepalí, x_low)	muy baja (x_low)	16.000	0,055
Clean 100 (comunidad)	media	22.050	0,091
ManyVoice (comunidad)	media	22.050	0,095
Jenny (comunidad)	media	22.050	0,084

Tabla 5: Factor de tiempo real (FTR, menos es más rápido): segundos de cálculo por segundo de voz, mediana de tres pasadas.

Conflictos de intereses y financiación. El autor desarrolla Loudness.ai, cuya herramienta de texto a voz ofrece una selección de estas voces elegida con las mismas puntuaciones. No se recibió financiación externa.

Referencias

- [1] M. Hansen y la Open Home Foundation. Piper: a fast, local neural text to speech system. <https://github.com/OHF-Voice/piper1-gpl>.
- [2] Voces de Piper, revisión c10ece1. <https://huggingface.co/rhasspy/piper-voices/tree/c10ece1aade47bb51c153c893d14e5bf8e5b7117>.
- [3] B. Beattie. Text to speech models. <https://brycebeattie.com/files/tts/>.
- [4] J. Kim, J. Kong y J. Son. Conditional variational autoencoder with adversarial learning for end-to-end text-to-speech. En *Proc. ICML*, 2021.
- [5] S. King y V. Karaiskos. The Blizzard Challenge 2013. En *Proc. Blizzard Challenge Workshop*, 2013.
- [6] H. Zen, V. Dang, R. Clark, Y. Zhang, R. J. Weiss, Y. Jia, Z. Chen y Y. Wu. LibriTTS: a corpus derived from LibriSpeech for text-to-speech. En *Proc. Interspeech*, 2019.
- [7] Y. Koizumi, H. Zen, S. Karita, Y. Ding, K. Yatabe, N. Morioka, M. Bacchiani, Y. Zhang, W. Han y A. Bapna. LibriTTS-R: a restored multi-speaker text-to-speech corpus. En *Proc. Interspeech*, 2023.
- [8] V. Pratap, Q. Xu, A. Sriram, G. Synnaeve y R. Collobert. MLS: a large-scale multilingual dataset for speech research. En *Proc. Interspeech*, 2020.
- [9] T. Saeki, D. Xin, W. Nakata, T. Koriyama, S. Takamichi y H. Saruwatari. UTMOS: UTokyo-SaruLab system for VoiceMOS Challenge 2022. En *Proc. Interspeech*, 2022.
- [10] W.-C. Huang, E. Cooper, Y. Tsao, H.-M. Wang, T. Toda y J. Yamagishi. The VoiceMOS Challenge 2022. En *Proc. Interspeech*, 2022.
- [11] tarepan. SpeechMOS: easy-to-use speech MOS predictors, v1.2.0. <https://github.com/tarepan/SpeechMOS>.
- [12] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey y I. Sutskever. Robust speech recognition via large-scale weak supervision. En *Proc. ICML*, 2023.
- [13] SYSTRAN. faster-whisper, versión 1.2.1. <https://github.com/SYSTRAN/faster-whisper>.
- [14] A. Shehzad. Enseñar a leer a un motor neuronal ligero de síntesis de voz. Informe técnico de Loudness.ai LA-RB-2026-002, 2026. <https://doi.org/10.5281/zenodo.23392104>.

A. Las tres mejores voces de cada variante de idioma

Idioma	N.º	Voz	Natur.	Clar.	Global	Licencia
Alemán (Alemania)	1	Karlsson	66,2	97,4	81,8	Sin indicar
	2	MLS · hablante 162	62,3	92,2	81,2	CC BY 4.0
	3	Kerstin	60,2	100,0	80,2	CC0 1.0
Armenio (Armenia)	1	Gor	56,5	64,2	64,3	GPL-2.0
Catalán (España)	1	UPC Ona	55,5	97,8	76,6	CC BY-SA 3.0 ES
	2	UPC Pau	59,8	91,8	75,7	CC BY-SA 3.0 ES
Checo (República Checa)	1	Kasandra	49,3	99,0	78,1	CC BY 4.0
Chino (China)	1	Huayan	67,5	91,7	79,6	Sin indicar
Esloveno (Eslovenia)	1	Artur	61,5	99,6	84,6	CC BY 4.0
Español (España)	1	MLS 9972	57,0	94,6	75,8	CC BY 4.0
	2	CarlFM	53,0	97,1	75,1	Dominio público
	3	MLS 10246	17,2	90,0	53,6	CC BY 4.0
Español (México)	1	Claude	62,3	99,6	85,0	Apache 2.0
	2	Ald	49,0	99,2	74,0	Sin indicar
Euskera (España)	1	Maider	71,8	91,4	85,6	CC BY 4.0
	2	Antton	57,8	92,3	79,0	CC BY 4.0
Finés (Finlandia)	1	Harri	60,0	93,6	76,8	CC0 1.0
Francés (Francia)	1	Gilles	59,0	94,7	76,8	CC0 1.0
	2	Tom	46,0	99,6	76,8	AGPL-3.0
	3	Siwis	51,5	98,4	74,9	CC BY 4.0
Griego (Grecia)	1	Rapunzelina (medium)	55,8	90,1	77,0	CC0 1.0
	2	Rapunzelina (low)	51,5	83,4	67,5	CC0 1.0
Inglés (Estados Unidos)	1	Clean 100 · hablante 118	84,7	100,0	96,4	CC BY 4.0
	2	Clean 100 · hablante 25	84,8	100,0	96,4	CC BY 4.0
	3	Clean 100 · hablante 108	84,3	100,0	96,2	CC BY 4.0
Inglés (Irlanda)	1	Jenny	78,9	100,0	93,5	Licencia Jenny (Dioco)
Inglés (Reino Unido)	1	Cori (medium)	76,8	100,0	92,4	Dominio público
	2	Cori (high)	78,8	97,2	92,0	Dominio público
	3	Alan	77,2	99,1	88,2	Sin indicar
Inglés (mezcla EE. UU. y Reino Unido)	1	ManyVoice · hablante 2	82,0	100,0	95,0	Dominio público
	2	ManyVoice · hablante 9	80,2	100,0	94,1	Dominio público
	3	ManyVoice · hablante 10	76,8	99,1	91,9	Dominio público
Islandés (Islandia)	1	Steinn	69,8	93,0	85,3	Sin indicar
	2	Salka	65,2	94,7	84,0	Sin indicar
	3	Ugla	64,0	93,0	82,5	Sin indicar
Italiano (Italia)	1	Serena (high)	71,2	100,0	89,6	CC BY 4.0
	2	Serena (medium)	65,5	99,2	86,3	CC BY 4.0
	3	Riccardo	44,0	97,5	70,7	Sin indicar
Kazajo (Kazajistán)	1	ISSAI · hablante 4	55,8	90,2	77,0	CC BY 4.0
	2	ISSAI · hablante 1	51,0	89,4	74,2	CC BY 4.0
	3	ISSAI · hablante 0	53,2	86,0	73,6	CC BY 4.0
Letón (Letonia)	1	Aivars	68,8	88,9	82,8	CC0 1.0
Neerlandés (Bélgica)	1	Nathalie	63,5	95,6	79,5	CC0 1.0
	2	RDH (medium)	46,0	99,6	76,8	CC0 1.0

Idioma	N.º	Voz	Natur.	Clar.	Global	Licencia
Neerlandés (Países Bajos)	3	RDH (x_low)	54,8	97,2	75,9	CC0 1.0
	1	Alex	63,7	99,6	85,6	CC0 1.0
	2	MLS · hablante 51	52,8	78,5	69,7	CC BY 4.0
Nepalí (Nepal)	3	MLS · hablante 37	49,9	80,6	69,2	CC BY 4.0
	1	Google · hablante 1	41,3	69,7	55,5	CC BY-SA 4.0
	2	Google · hablante 5	37,2	70,1	53,6	CC BY-SA 4.0
Noruego (Noruega)	3	Google · hablante 8	37,9	66,8	52,4	CC BY-SA 4.0
	1	NVCC · hablante 6	58,2	97,7	82,0	CC0 1.0
	2	NVCC · hablante 8	60,5	94,5	81,5	CC0 1.0
Persa (Irán)	3	NVCC · hablante 0	62,3	90,9	80,5	CC0 1.0
	1	Gyro	72,5	87,9	84,2	Sin indicar
	1	MLS 6892	43,5	94,2	68,8	CC BY 4.0
Polaco (Polonia)	1	Edresson	49,8	98,4	74,1	CC BY 4.0
Portugués (Brasil)	1	NST	66,2	98,5	86,4	CC0 1.0
Sueco (Suecia)	2	Alma	66,8	97,5	86,2	Sin indicar
	1	Padmavathi	67,8	37,9	56,8	CC BY 4.0
Telugu (India)	2	Venkatesh	59,5	35,2	51,4	CC BY 4.0
	1	Tetiana	69,5	98,7	88,1	Apache 2.0
Ucraniano (Ucrania)	2	Oleksa	67,2	99,6	87,4	Apache 2.0
	3	Ukrainian TTS · hablante 2	67,5	95,2	85,3	CC0 1.0
Urdu (Pakistán)	1	Aegis Female	65,0	97,4	85,2	Sin indicar
Vietnamita (Vietnam)	1	25hours single	25,0	87,8	56,4	Sin indicar

B. Archivos de voz oficiales excluidos

Archivo de voz	Idioma	Motivo	Licencia indicada
ar_J0-kareem-low	Árabe (Jordania)	Basada en Lessac	Ver la fuente
ar_J0-kareem-medium	Árabe (Jordania)	Basada en Lessac	Ver la fuente
bg_BG-dimitar-medium	Búlgaro (Bulgaria)	Basada en Lessac	CC0
bn_BD-google-medium	Bengalí (Bangladés)	Basada en Lessac	Sin indicar
ca_ES-upc_ona-medium	Catalán (España)	Basada en Lessac	CC BY-SA 3.0 ES
cs_CZ-jirka-low	Checo (República Checa)	Basada en Lessac	CC0
cs_CZ-jirka-medium	Checo (República Checa)	Basada en Lessac	CC0
cy_GB-bu_tts-medium	Galés (Reino Unido)	Basada en Lessac	CC BY 4.0
cy_GB-gwryw_gogleddol-medium	Galés (Reino Unido)	Basada en Lessac	Ver la fuente
da_DK-talesyntese-medium	Danés (Dinamarca)	Basada en Lessac	CC0
de_DE-pavoque-low	Alemán (Alemania)	No comercial	CC BY-NC-SA 4.0
de_DE-thorsten-high	Alemán (Alemania)	Basada en Lessac	CC0
de_DE-thorsten-medium	Alemán (Alemania)	Basada en Lessac	CC0
de_DE-thorsten_emotional-medium	Alemán (Alemania)	Basada en Lessac	CC0
el_GR-joy-medium	Griego (Grecia)	No comercial	CC BY-NC 4.0
en_GB-alan-medium	Inglés (Reino Unido)	Basada en Lessac	Ver la fuente
en_GB-alba-medium	Inglés (Reino Unido)	Basada en Lessac	CC BY 4.0
en_GB-arun-medium	Inglés (Reino Unido)	Basada en Lessac	CC BY 4.0
en_GB-jenny_dioco-medium	Inglés (Reino Unido)	Basada en Lessac	Ver la fuente
en_GB-northern_english_male-medium	Inglés (Reino Unido)	Basada en Lessac	CC BY-SA 4.0
en_GB-semaine-medium	Inglés (Reino Unido)	No comercial	CC BY-NC-SA 4.0
en_GB-vctk-medium	Inglés (Reino Unido)	Basada en Lessac	CC BY 4.0

Archivo de voz	Idioma	Motivo	Licencia indicada
en_US-amy-medium	Inglés (Estados Unidos)	Basada en Lessac	Ver la fuente
en_US-arctic-medium	Inglés (Estados Unidos)	Basada en Lessac	Ver la fuente
en_US-hfc_female-medium	Inglés (Estados Unidos)	No comercial	CC BY-NC-SA 4.0
en_US-hfc_male-medium	Inglés (Estados Unidos)	No comercial	CC BY-NC-SA 4.0
en_US-joe-medium	Inglés (Estados Unidos)	Basada en Lessac	CC0
en_US-kusal-medium	Inglés (Estados Unidos)	Basada en Lessac	Ver la fuente
en_US-l2arctic-medium	Inglés (Estados Unidos)	No comercial	CC BY-NC 4.0
en_US-lessac-high	Inglés (Estados Unidos)	No comercial	Lessac (Blizzard 2013)
en_US-lessac-low	Inglés (Estados Unidos)	No comercial	Lessac (Blizzard 2013)
en_US-lessac-medium	Inglés (Estados Unidos)	No comercial	Lessac (Blizzard 2013)
en_US-libritts_r-medium	Inglés (Estados Unidos)	Basada en Lessac	CC BY 4.0
en_US-mike-medium	Inglés (Estados Unidos)	Basada en Lessac	CC0
en_US-reza_ibrahim-medium	Inglés (Estados Unidos)	Basada en Lessac	CC0
en_US-ryan-high	Inglés (Estados Unidos)	No comercial	CC BY-NC-SA 4.0
en_US-ryan-low	Inglés (Estados Unidos)	No comercial	CC BY-NC-SA 4.0
en_US-ryan-medium	Inglés (Estados Unidos)	No comercial	CC BY-NC-SA 4.0
en_US-sam-medium	Inglés (Estados Unidos)	Basada en Lessac	Apache 2.0
es_AR-daniela-high	Español (Argentina)	Basada en Lessac	CC BY-SA 4.0
es_ES-davefx-medium	Español (España)	Basada en Lessac	CC0
es_ES-sharvard-medium	Español (España)	Basada en Lessac	CC BY 3.0
es_MX-ald-medium	Español (México)	Basada en Lessac	Unlicense
et_EE-news-medium	Estonio (Estonia)	Basada en Lessac	CC BY 4.0
fa_IR-amir-medium	Persa (Irán)	Basada en Lessac	CC0
fa_IR-ganji-medium	Persa (Irán)	Basada en Lessac	CC0
fa_IR-ganji_adabi-medium	Persa (Irán)	Basada en Lessac	CC0
fa_IR-reza_ibrahim-medium	Persa (Irán)	Basada en Lessac	CC0
fi_FI-harri-medium	Finés (Finlandia)	Basada en Lessac	CC0
fr_FR-siwis-medium	Francés (Francia)	Basada en Lessac	CC BY 4.0
fr_FR-upmc-medium	Francés (Francia)	Basada en Lessac	CC BY-SA 4.0
he_IL-saspeech-medium	Hebreo (Israel)	No comercial	Ver la fuente
hi_IN-pratham-medium	Hindi (India)	No comercial	CC BY-NC-SA 4.0
hi_IN-priyamvada-medium	Hindi (India)	No comercial	CC BY-NC-SA 4.0
hi_IN-rohan-medium	Hindi (India)	Basada en Lessac	IndicTTS licence
hu_HU-anna-medium	Húngaro (Hungria)	Basada en Lessac	CC0
hu_HU-berta-medium	Húngaro (Hungria)	Basada en Lessac	CC0
hu_HU-imre-medium	Húngaro (Hungria)	Basada en Lessac	CC0
id_ID-news_tts-medium	Indonesio (Indonesia)	Basada en Lessac	Ver la fuente
it_IT-paola-medium	Italiano (Italia)	Basada en Lessac	Ver la fuente
ja_JP-hi-fi_captain-medium	Japonés (Japón)	No comercial	CC BY-NC-SA 4.0
ka_GE-natia-medium	Georgiano (Georgia)	Basada en Lessac	Ver la fuente
ko_KR-kss-medium	Coreano (Corea del Sur)	No comercial	CC BY-NC-SA 4.0
ku_TR-berfin_renas-medium	Kurdo kurmanji (Turquía)	No comercial	Sin indicar
lb_LU-marylux-medium	Luxemburgués (Luxemburgo)	No comercial	CC BY-NC-SA 4.0
lt_LT-reginute1-medium	Lituano (Lituania)	Basada en Lessac	CC BY 4.0
ml_IN-arjun-medium	Malayalam (India)	Basada en Lessac	Ver la fuente
ml_IN-meera-medium	Malayalam (India)	Basada en Lessac	Ver la fuente
mr_IN-google-medium	Maratí (India)	Basada en Lessac	CC BY-SA 4.0
ne_NP-chitwan-medium	Nepalí (Nepal)	Basada en Lessac	CC0
ne_NP-google-medium	Nepalí (Nepal)	Basada en Lessac	CC BY-SA 4.0
nl_BE-nathalie-medium	Neerlandés (Bélgica)	Basada en Lessac	CC0
nl_NL-pim-medium	Neerlandés (Países Bajos)	Basada en Lessac	CC0
nl_NL-ronnie-medium	Neerlandés (Países Bajos)	Basada en Lessac	CC0
no_NO-talesyntese-medium	Noruego (Noruega)	Basada en Lessac	CC0

Archivo de voz	Idioma	Motivo	Licencia indicada
pl_PL-bass-high	Polaco (Polonia)	Basada en Lessac	Sin indicar
pl_PL-darkman-medium	Polaco (Polonia)	Basada en Lessac	CC0
pl_PL-gosia-medium	Polaco (Polonia)	Basada en Lessac	CC0
pl_PL-mc_speech-medium	Polaco (Polonia)	Basada en Lessac	CC0
pt_BR-cadu-medium	Portugués (Brasil)	Basada en Lessac	CC0
pt_BR-faber-medium	Portugués (Brasil)	Basada en Lessac	CC0
pt_BR-jeff-medium	Portugués (Brasil)	Basada en Lessac	CC0
pt_PT-tugão-medium	Portugués (Portugal)	Basada en Lessac	CC0
ro_RO-mihai-medium	Rumano (Rumanía)	Basada en Lessac	CC0
ru_RU-denis-medium	Ruso (Rusia)	Basada en Lessac	CC0
ru_RU-dmitri-medium	Ruso (Rusia)	Basada en Lessac	CC0
ru_RU-irina-medium	Ruso (Rusia)	Basada en Lessac	Sin indicar
ru_RU-ruslan-medium	Ruso (Rusia)	No comercial	CC BY-NC-SA 4.0
sk_SK-lili-medium	Eslovaco (Eslovaquia)	Basada en Lessac	CC0
sq_AL-edon-medium	Albanés (Albania)	Basada en Lessac	CC0
sr_RS-serbski_institut-medium	Serbio (Serbia)	No comercial	CC BY-NC-SA 4.0
sv_SE-lisa-medium	Sueco (Suecia)	Basada en Lessac	Sin indicar
sw_CD-lanfrica-medium	Suajili (República Democrática del Congo)	No comercial	Ver la fuente
te_IN-maya-medium	Telugu (India)	Basada en Lessac	IndicTTS licence
th_TH-tsync2-medium	Tailandés (Tailandia)	No comercial	CC BY-NC-SA 3.0
tr_TR-dfki-medium	Turco (Turquía)	No comercial	CC BY-NC-SA 4.0
ur_PK-fasih-medium	Urdu (Pakistán)	Basada en Lessac	Sin indicar
vi_VN-vaish1000-medium	Vietnamita (Vietnam)	Basada en Lessac	CC BY 4.0
vi_VN-vivos-x_low	Vietnamita (Vietnam)	No comercial	CC BY-NC-SA 4.0
zh_CN-chaowen-medium	Chino (China)	Basada en Lessac	CC0
zh_CN-huayan-medium	Chino (China)	Basada en Lessac	Sin indicar
zh_CN-xiao_ya-medium	Chino (China)	No comercial	DataBaker, non-commercial