

Maria Fitó-Carreras

Universidad Internacional de Catalunya

@ mfito@uic.es

ID 0000-0002-0500-4006

Montserrat Vidal-Mestre

Universidad Autónoma de Barcelona

@ montsevidalm@gmail.com

ID 0000-0001-6144-5386

Alfonso Freire-Sánchez

Universidad Abat Oliba CEU

@ freire3@uao.es

ID 0000-0003-2082-1212

■ Recibido / Received
18 de octubre de 2024■ Aceptado / Accepted
3 de noviembre de 2024■ Páginas / Pages
De la 179 a la 196

■ ISSN: 1885-365X

Análisis de *softwares* de inteligencia artificial generativa de voz aplicados al *podcasting*

Analysis of voice-generative artificial intelligence software applied to podcasting

RESUMEN:

La inteligencia artificial generativa (IAG) de voz es capaz de generar mensajes en lenguaje humano mediante algoritmos de aprendizaje profundo, como las redes neuronales convolucionales (CNN), que aprenden a imitar los patrones vocales a partir de datos de habla. Ante este contexto, el principal objetivo es ofrecer una radiografía de la IAG de voz aplicada al *podcasting* para responder si la actual oferta tecnológica representa una amenaza para los empleos de los profesionales del audio, en particular para los/as locutores/as. Con este fin, se analizan los principales *softwares* que emplean los creadores de pódcast para la clonación de voz y se establece un marco comparativo. En segundo lugar, se recopilan las percepciones de los creadores acerca de los resultados obtenidos mediante el análisis de 10 títulos. Los principales *softwares* ofrecen herramientas específicas, que pueden mejorar el flujo de trabajo y optimizar los costes de producción. Gracias a los resultados sobre el estado actual de la IAG de voz aplicada al *podcasting*, hemos identificado tanto las oportunidades como las limitaciones que esta tecnología ofrece a los creadores. Se observa que la industria de la IAG de voz está adaptándose a las necesidades del sector, ofreciendo múltiples herramientas a través de plataformas especializadas que permiten clonar la voz, editar grabaciones, publicar pódcast y distribuirlos en varios idiomas. Sin embargo, no se interpreta como una amenaza inmediata debido a la reproducción de una prosodia inexacta y la ausencia de elementos paralingüísticos.

PALABRAS CLAVE:

inteligencia artificial generativa, voz sintética, *podcasting*, automatización, prosodia.

ABSTRACT:

Voice AI is capable of generating human language messages through deep learning algorithms, such as convolutional neural networks (CNN), which learn to imitate vocal patterns from speech data. In this context, the main objective is to provide an overview of voice AI applied to podcasting, aiming to answer whether the current technological offerings pose a threat to audio professionals' jobs, particularly voice-over artists. To this end, the main software used by podcast creators for voice cloning is analyzed, and a comparative framework is established. Secondly,

creators' perceptions of the results are gathered by analyzing 10 titles. The main software provides specific tools that can enhance workflow and optimize production costs. Based on the findings about the current state of voice AI in podcasting, we have identified both the opportunities and limitations this technology offers to creators. It is observed that the voice AI industry is adapting to the sector's needs, offering multiple tools through specialized platforms that allow for voice cloning, editing recordings, publishing podcasts, and distributing them in several languages. However, it is not perceived as an immediate threat due to the reproduction of inaccurate prosody and the absence of paralinguistic elements.

KEY WORDS:

generative artificial intelligence, synthetic voice, podcasting, automation, prosody.

1. Introducción

Existe un amplio consenso respecto a las dimensiones que alcanzará la disciplina de la inteligencia artificial (en adelante IA) en los próximos años dada su capacidad para abordar tareas que hasta hace pocos años requerían la intervención humana (Boden, 2018; Bhargava y Sharma, 2022; Brennen *et al.*, 2022). Este hecho es especialmente notable en la IA generativa (en adelante IAG), una rama del aprendizaje automático capaz de generar nuevo contenido en una gran variedad de formas a partir de la consolidación de datos de diversas fuentes (Dasborough, 2023; Preiksaitis y Rose, 2023). Esta tecnología aprende a «identificar patrones y características del conjunto de datos y textos culturales para generar nuevos datos o textos culturales similares, acordes con un contexto de aplicación» (Aguado-Terrón y Grandío-Pérez, 2024, p. 7). Un ejemplo de ello es su capacidad para crear una herramienta de comunicación tan íntima y personal como la voz.

La IAG de voz es capaz de generar mensajes en lenguaje humano, a través de algoritmos de aprendizaje profundo, como las redes neuronales convolucionales (CNN) (Vaissnave *et al.*, 2024), que aprenden a imitar los patrones vocales a partir de datos de habla (Guzmán y Lewis, 2020). Este proceso es posible gracias a la facilitación a la IAG de ingentes cantidades de datos (Casella *et al.*, 2023): una muestra de voz, cuya extensión puede variar según el *software* empleado. El algoritmo identifica los matices que la hacen única, como el timbre y la cadencia; es decir, el ADN del habla (Álvarez *et al.*, 2022). Este aprendizaje, también denominado entrenamiento, le permite resolver de forma adecuada la tarea que se le encomienda (Rudolph, 2018). Tras el entrenamiento, la IAG puede generar nuevos audios en la voz sintética de destino —voz digital o virtual— a partir del texto que se le proporciona (Sun *et al.*, 2023). Los usuarios, en función de sus necesidades productivas, pueden escoger entre una amplia oferta de voces almacenadas en el *software* —banco de voces—, o facilitar una muestra de su propia voz para que la clone, que, además, puede recrear en varios idiomas.

El alto nivel de realismo alcanzado con la aplicación de esta tecnología en productos multimedia ha generado un amplio debate en torno a sus implicaciones éticas, legales y sociales (Franganillo, 2022; 2023). Entre los principales riesgos que se han identificado, se destacan los peligros asociados al uso fraudulento, como los casos de clonación de voz para la comisión de estafas. Asimismo, la clonación sin el consentimiento de su legítimo propietario ha creado una creciente preocupación por la seguridad y los derechos individuales. Además, la falta de etiquetado claro en los contenidos de audio y video generados por IA desprotege y confunde a la audiencia, situación que podría ser mitigada con la entrada en vigor de la Ley de Inteligencia



Artificial en 2025, la cual establece la obligatoriedad de dicha práctica (European Parliament, 2024). Del mismo modo, el uso de la IAG de voz también despierta preocupaciones en el ámbito profesional debido a que podría incidir en precariedad laboral (Stenbom *et al.*, 2023). Esta situación ha puesto en alerta al sector profesional del doblaje. La *United Voice Artist* (UVA), asociación que representa a los dobladores de películas, series y *spots* de 48 países, desde enero de 2024 exige la cláusula IA en los contratos. Esta prohíbe de forma expresa que la voz o los tonos se utilicen en productos distintos de aquellos para los que los dobladores han sido contratados (UVA, 2024). Con el nuevo escenario que dibuja la IAG de voz, «queda patente el miedo que existe a que esta tecnología disruptiva pueda acabar con los puestos de trabajo de profesionales del audio, especialmente, los/as locutores/as, algo que se achaca a la aplicación de esta tecnología en cualquier tipo de formato» (Chaparro-Domínguez, 2024, p. 135).

Dada la dimensión que está alcanzando el fenómeno y su potencial para transformar la industria, este trabajo persigue obtener una radiografía de la IAG de voz aplicada al *podcasting* para responder si la actual oferta tecnológica se configura como una amenaza para los profesionales del audio, especialmente de locutores y locutoras. Para ello, se combina el análisis del *software* IAG de voz que emplean los *podcasters* en su flujo de trabajo con el estudio de la percepción de los productores de pódcast acerca de los resultados obtenidos mediante el análisis de 10 títulos.

El estudio de IAG de voz ha sido ampliamente abordado por la academia. Destacan los trabajos de Taylor (2024) sobre las posibilidades de la IA aplicada a los pódcast, el de Rime *et al.* (2022) acerca de las innovaciones en el *podcasting*, el de Chaparro-Domínguez (2024) sobre el impacto en los contenidos periodísticos sonoros o el de Ada *et al.* (2024) acerca de las capacidades paralingüísticas de la IAG de voz. Sin embargo, las revisiones de *software* aplicadas al *podcasting* son prácticamente inexistentes, así como trabajos que recojan la percepción de los creadores de pódcast que emplean dicha tecnología para la construcción de la narrativa. Este artículo pretende ampliar la producción académica al respecto, desde la perspectiva profesional, permitiendo una comprensión más profunda del significado y la posible repercusión en términos laborales de esta tecnología.



2. Marco teórico

En pocos años, el *podcasting* se ha consolidado como una fuente relevante para la información y el entretenimiento de la audiencia. El formato de audio contribuye a establecer una relación de intimidad del oyente con la voz que transmite el mensaje (Bottomley, 2015), favorecido por la escucha personal en soledad (Espinosa, 2019). Esta intimidad que genera se traduce en la credibilidad del mensaje (Fitó-Carreras *et al.*, 2023).

Se estima que, para finales de 2024, el número global de oyentes de pódcast alcance los 504,9 millones (Mosby, 2024). La audiencia siente atracción hacia el formato debido a la amplia oferta de títulos disponibles en plataformas de audio y la variedad de temáticas que aborda. Además, también valora la posibilidad de elegir cómo, cuándo y dónde escucharlos, ubicuidad que Kishigami (2004) denomina como la triple A (*any device, any time, anywhere*).

Para Taylor (2024), el uso de la IAG de voz en el *podcasting* puede marcar un nuevo rumbo y redefinir la disciplina. Se observa que, en la oferta de las principales plataformas de distribución de audio, como Spotify o Apple Podcast, empiezan a tener representación títulos gene-

rados con voces sintéticas por los usuarios. Algunas plataformas van más allá, incorporando herramientas propias para la traducción de los *podcast* en varios idiomas. En el caso de Spotify, la tecnología Voice Translation for Podcasters, desarrollada por OpenAI, permite la traducción al inglés, francés, español y alemán. Para Spotify, los resultados de esta IAG «generan una experiencia auditiva más auténtica que suena más personal y natural que el doblaje tradicional» (Spotify, 2023). Sin embargo, a pesar de esta afirmación categórica, la recepción favorable o desfavorable del *podcast* dependerá de la evaluación subjetiva que realice cada individuo, la cual está condicionada por sus rasgos psicológicos y por su contexto (Shin y Biocca, 2018).

La escucha de la voz humana provoca una gran variedad de emociones en el individuo. Russell (1980), en su teoría del modelo circunflejo de las emociones, señala dos aspectos que caracterizan las emociones humanas: la valencia y el *arousal*. La valencia se refiere a cómo se siente el individuo frente a una situación, desde el bienestar hasta el malestar. El *arousal* indica el nivel de activación que se experimenta en respuesta a los estímulos externos. La combinación de estos aspectos genera cuatro emociones básicas: el entusiasmo, la ira, la tristeza y la calma. La voz despierta estas emociones a través de las características idiosincrásicas y la manera de hablar de las personas —lenguaje no verbal— (Alexander y Nygaard, 2008). El lenguaje no verbal proporciona señales socialmente relevantes (Nass y Gong, 2000), el estado de ánimo y la personalidad (Aronovitch, 1976), la pertenencia a grupos sociales y raciales (Nass y Brave, 2005) y el género del individuo (Mullennix *et al.*, 1995). La voz posee una serie de características que Faure-Carvalho *et al.* (2022) describen y sintetizan en diferentes categorías. En primer lugar, la altura o tesitura de la voz, determinada por la frecuencia del sonido, que permite distinguir entre voces graves y agudas. En segundo lugar, el ritmo del habla, que puede asociarse con la edad, la salud y la posición social del hablante y con ciertas emociones. El tercer aspecto es el timbre, que otorga expresividad y singularidad a una voz, facilitando la identificación inmediata con la persona que la genera (Vidal-Mestre, 2018). Por último, la intensidad vocal, que refleja aspectos como la presencia, la cercanía, el secretismo, la intimidad o la confianza, entre otros. Todas estas dimensiones determinan la prosodia, es decir, el estilo de habla único de cada individuo:

La prosodia nos ayuda a hacer el mensaje comprensible, a mostrar la intención que tenemos al hablar y a plasmar lo que sentimos. Los seres humanos utilizamos las combinaciones melódicas de la voz a través de la entonación, la acentuación y los cambios de velocidad y las pausas para dotar de sentido a nuestras palabras (Rodero, 2018, p. 84).

Rodero (2023) señala que, gracias a la suma del significado del lenguaje verbal —palabras— y no verbal, el individuo receptor inicia un proceso cognitivo que genera una imagen mental de la persona que habla.

Esta última cuestión puede completarse con la percepción de la confianza que transmite la voz. Los trabajos llevados a cabo por Atkinson *et al.* (2005) y Walters *et al.* (2008) sobre la percepción de voces sintetizadas, indican que las personas tienden a confiar más en el mensaje cuando las voces suenan más reales, independientemente del contexto. Esta confianza disminuye cuando el discurso no presenta una calidad humana perceptible. Así pues, en nuestro contexto de estudio, cuanto más reales sean las voces más confianza entablará el oyente con el mensaje que transmita.



En cuanto a la percepción humana del habla artificial, Winters y Pisoni (2004) destacan, en primer lugar, que las voces sintéticas son menos comprensibles debido a que la entonación —elemento de alto nivel de la voz— no logra replicar con precisión el habla humana, especialmente cuando las oraciones son largas. En segundo lugar, su decodificación requiere de más esfuerzo cognitivo por parte del oyente para comprenderla. En tercer lugar, mencionan que la familiarización con las voces sintéticas mejora con la experiencia. Finalmente, señalan que las voces con buenas características prosódicas y buena inteligibilidad se perciben como más naturales. En el mismo sentido, el estudio neurocientífico sobre audiolibros elaborado por Media Psychology Lab. (2023), refleja que las narraciones con habla sintetizada generan rechazo en muchos participantes y las narradas con voz humana son más fáciles de recordar.

La síntesis de habla puede imitar acentos, tonos, entonaciones... Pero nunca será capaz de expresar emociones como lo hace una voz real. Aunque es cierto que puede simularlas, estas no van a sonar naturales, al menos a corto plazo. Nuestra voz va cambiando en función de nuestro estado de ánimo, cuando nos reímos o nos enfadamos, cuando nos emocionamos... El habla sintetizada no comunica emociones ni habla de ellas. Por esta razón, no consiguen crear un vínculo con el oyente, ni perduran en el recuerdo como lo hace el habla humana (Rodríguez, 2023).

Chaparro-Domínguez (2024) añade que la IAG de voz no consigue reproducir de manera realista el tono y la respiración de la voz humana. Por su parte, Ada *et al.* (2024), en su estudio sobre la capacidad paralingüística del habla, campo que se ocupa de cómo se dice algo en lugar de lo que se dice, analizan las risas, los tartamudeos y el ritmo de las voces sintéticas que reproduce el *software* de la plataforma ElevenLabs. Los autores concluyen que, para obtener resultados óptimos, es necesario encontrar la transcripción correcta de estas expresiones lingüísticas, lo cual no es una tarea fácil y no siempre produce los resultados esperados. Atribuyen esto a las limitaciones que aún presentan las redes neuronales profundas, en las que se basa la síntesis de voz de la IA.

En función de lo analizado, esta investigación parte de la hipótesis de que, en la actualidad, los resultados proporcionados por los *softwares* de IAG de voz se desvían de la autenticidad de la voz humana, especialmente en lo que respecta a los aspectos prosódicos y paralingüísticos.

3. Metodología

Este trabajo pretende obtener una radiografía de la IAG de voz aplicada al *podcasting* para responder si la actual oferta tecnológica representa una amenaza para los empleos de los profesionales del audio, en particular los/as locutores/as. Con este fin, se analizan los principales *softwares* que emplean los creadores de pódcast para la clonación de voz y se establece un marco comparativo. En segundo lugar, se recopilan las percepciones de los creadores acerca de los resultados obtenidos mediante el análisis de 10 títulos.

Para la recopilación de datos, se lleva a cabo una revisión bibliográfica. Consideramos que este enfoque cualitativo es el más adecuado para la presente investigación. Con el método se «sumarizan diferentes investigaciones y artículos que nos dan una idea sobre cuál es el estado actual de la cuestión a investigar» (Goris, 2015, p. 2) y, además, aporta información acotada a un periodo determinado de tiempo (Hart, 1998).



La recopilación de materiales se realiza entre mayo y julio de 2024 en España. Se realiza la búsqueda en el motor generalista Google debido a que se pretende obtener una radiografía del fenómeno desde la perspectiva de los creadores de *pódcast*, cuya producción se encuentra alojada en páginas web y plataformas de distribución de audio *online*. Este motor se caracteriza por ofrecer la posibilidad de realizar consultas que van desde las más específicas hasta las más amplias, los resultados presentan listados de temas jerárquicos y subtópicos y permite ir refinando la búsqueda (Cohen, 2001).

Las palabras clave que guían la búsqueda son representativas del objeto de estudio, tanto en inglés como en español. Concretamente: ‘*podcast*’, ‘IA voz’, ‘IA generativa de voz’ y ‘voz sintética’, combinando dichos descriptores. Con el objetivo de acotar la búsqueda se establecen los siguientes criterios de inclusión y exclusión:

- Publicaciones entre enero de 2018 y junio de 2024: se elige este periodo temporal debido a que, en esta última década, la IAG de voz es cuando ha experimentado notable crecimiento y consolidación.
- Publicaciones divulgativas especializadas en la materia (web y blog) y portales de *software* de IAG de voz.
- En inglés o español.
- Relación directa con el objetivo de la investigación, que incluyan uno o más términos de búsqueda relacionados con el objetivo planteado.

El establecimiento de estos criterios ha facilitado la depuración de la información obtenida, eliminando aquellos resultados que no estaban lo suficientemente relacionados con la investigación. Una vez obtenidos, se aplican nuevos criterios de inclusión y exclusión sobre los *softwares* de IAG de voz y los títulos de *pódcast* generados mediante esta tecnología. En concreto:

- *Software* de IAG de voz que en su portal hagan referencia explícita a su idoneidad para la creación de *pódcast* o *softwares* mencionados por los *podcasters* o productoras de *pódcast* como herramienta para la clonación de voces en sus flujos de trabajo.
- *Pódcast* creados total o parcialmente con IAG de voz.

Para el análisis de *software* de IAG de voz se proponen las siguientes variables de estudio: la funcionalidad del *software* (clonación de voz y/o banco de voces), los extras que ofrece a los creadores de *pódcast*, el tiempo destinado al entrenamiento de la voz y los planes comerciales que ofrece (en abierto, funcionalidades gratuitas o suscripción). En cuanto a los *pódcast* generados con IAG de voz, se analiza la categoría de los creadores (*podcaster amateur*, profesionalizado independiente o productora), el género narrativo del *pódcast* (ficción narrativa, entrevista, informativo, entre otros), si el *pódcast* clona la voz de los anfitriones y/o emplea voces almacenadas ya entrenadas, si la tecnología se aplica total o parcialmente, y finalmente, si se identifica el *pódcast* como producto de la IA.

4. Resultados

Siguiendo la metodología propuesta, se obtienen los siguientes resultados, que se presentan en dos subepígrafes diferenciados: uno referente a los *softwares* de IAG y sus principales características, y otro dedicado a la percepción de los *podcasters*.



4.1. **SOFTWARES DE INTELIGENCIA ARTIFICIAL GENERATIVA DE VOZ**

La búsqueda permite identificar 36 *softwares* de IAG de voz. Tras aplicar los criterios de inclusión y exclusión, únicamente 11 de estos son empleados comúnmente para la generación de voces en el *podcasting* y valorados positivamente por los creadores de pódcast. Los *softwares* se dividen en tres categorías: aquellos que generan voz clonada a partir de una muestra del usuario, los que funcionan a modo de bancos de voces almacenadas y los que ofrecen una combinación de ambas funcionalidades.

De entre los resultados obtenidos (Tabla 1) destaca, en primer lugar, Revoice, de las desarrolladoras Lyrebird y Finch. El producto se presenta como una solución destinada a mejorar el flujo de trabajo de los creadores de pódcast, desde la grabación y el entrenamiento vocal hasta la creación de perfiles de voz y copias digitales. En su portal, la desarrolladora identifica las fortalezas y las debilidades del producto. Por un lado, destaca que agiliza el proceso de grabación y edición de audio, lo que deriva en un ahorro de tiempo y esfuerzo para el *podcaster*. Sin embargo, también señala que todavía existen diferencias notables entre una voz generada por IA y una voz humana, especialmente en términos de expresión emocional y variaciones en la entonación. Advierte que tienden a sonar más monótonas y uniformes en comparación con las voces humanas. Además, reconoce que, aunque es una herramienta valiosa, no sustituye la habilidad y experiencia de un locutor profesional y aconseja a los usuarios que, para obtener resultados de alta calidad, trabajen con locutores con formación adecuada (Revoice, 2024). Para el entrenamiento, el usuario debe leer 70 frases en inglés que proporciona el programa.

Según el *software* Overdub de la desarrolladora Descript, los usuarios pueden obtener un clon ultrarrealista de su propia voz. La tecnológica asegura que, con la configuración correcta, las voces que genera pueden ser indistinguibles de las reales. Sin embargo, apunta que la calidad de la voz clonada depende de tres factores: la cantidad de datos de entrenamiento de voz que se graban, la calidad del entorno de grabación y la funcionalidad que se le pretende dar. Para entrenar el *software*, el usuario debe grabarse leyendo el guion que se le proporciona. La duración mínima de lectura necesaria es de 10 minutos. No obstante, para garantizar resultados de calidad, recomienda registrar entre 30 y 90 minutos de lectura. El programa ofrece la función adicional Filler Words Pro (traducido al español, palabras de relleno), para eliminar muletillas y palabras repetidas.

Por otro lado, VoiceLab, de Eleven Labs, permite clonar la voz con tan solo un minuto de muestra. Afirma que su tecnología incorpora una conciencia contextual que reconoce los matices del texto, lo que permite crear voces sintéticas con la entonación precisa.

WellSaid Labs, creadores de WellSaid, asegura que su IAG de voz se adapta al estilo, la velocidad y el énfasis que cada proyecto necesita, y que sus voces son adaptables a cualquier tema o estado de ánimo. Dispone de una biblioteca de pronunciación que permite entrenar a la IA para que pronuncie determinadas palabras como solicita el usuario. Finalmente, destaca Veritone Voice de Veritone Inc., que requiere de un entrenamiento de 3 horas. Igual que el resto, también permite realizar ajustes en la velocidad de voz, el tono y el volumen. Además, ofrece un banco de 300 voces.

En cuanto a los *softwares* que únicamente facilitan voces almacenadas, el sector del *podcasting* destaca Lovo.ai, Play.ht, Syntesys, Amazon Polly y Murf.ai. De entre ellas, Murf.ai



Tabla 1. Comparativa de programas de inteligencia artificial generativa de voz

Software	Voces-idioma	Extras	Entreno	Planes
Clonación de voz				
Revoice (Lyrebird & Finch)	N/A	N/I	70 frases inglés	Suscripción
Overdub (Descript)	N/A	Filler Words	10' de audio	Suscripción
Mixtas				
VoiceLab (Eleven Labs)	120 voces 29 idiomas	N/I	1' de audio	Muestra gratis Suscripción
Well Said (Well Said Labs)	120 voces N/I idiomas	Biblioteca pronunciación	N/I	Suscripción
Veritone Voice (Veritone Inc.)	300 voces N/I idiomas	N/I	3 horas de audio	Suscripción
Wondercraft (Wondercraft Inc.)	N/I voces 28 idiomas	Genera textos Subir contenidos	N/I	Suscripción
Banco de voces				
Lovo.ai (Lovo)	150 voces 100 idiomas	25 emociones	N/A	Gratis 15 días Suscripción
Play.ht (Play HT Inc.)	570 voces 100 idiomas	NO	N/A	Suscripción
Synthesys (Synthesis Studio)	370 voces 140 idiomas	NO	N/A	Suscripción
Amazon Polly (Amazon)	36 voces 98 idiomas	NO	N/A	Suscripción
Murf.ai (Murf.ai Inc.)	120 voces 20 idiomas	NO	N/A	Gratis 10' Suscripción

N/A: No aplica. N/I: No informa. *Fuente:* Elaboración propia.

asegura que su *software* «genera voz similar a la humana en tiempo real, no está exento de limitaciones y desafíos» (Murf.ai., 2024). Reconoce que los resultados dependen de la calidad y la cantidad de los datos de entrenamiento que se le proporcione al *software*, y advierte que su uso plantea problemas éticos y legales, como la violación de los derechos de autor y el uso no autorizado de voces en *off*.

Finalmente, destacamos la existencia de dos plataformas enfocadas a la creación de pódcast. Podcastle, que incorpora el *software* de Revoice, se oferta como una plataforma que democratiza el acceso a las herramientas de contenido de audio. Artavazd Yeritsyan, CEO de Podcastle, afirma que la voz clonada que genera no se centra en embellecer la voz ni en mejorar su calidad, sino en lograr una pronunciación precisa de las palabras (Artavazd Yeritsyan en Ashworth, 2023). Por otro lado, Podcast Production Studio de WondercraftAI, ofrece como



ventaja para los *podcasters*, la función de alojar los contenidos generados directamente en las principales plataformas de *podcasting*.

A continuación, se ofrece un resumen de los puntos fuertes y débiles de los *softwares* analizados, según las propias desarrolladoras.

Tabla 2. Argumentos de venta, limitaciones y advertencias de los *softwares* analizados

Revoice	Mejora el flujo de trabajo / Expresión emocional y variaciones de entonación / No sustituye la habilidad de los locutores profesionales
Overdub	Clones ultrarealistas / Solo buenos resultados con la cantidad de datos de entrenamiento y entorno de la grabación
VoiceLab	Crea discursos realistas / Entonación e inflexiones humanas con una fidelidad excepcional / Reconoce los matices del texto
Well Said	Voces hermosas en segundo de profesioales / Creación sin esfuerzos / Confían en nosotros marcas importantes del mundo / Se adapta al estilo, velocidad, estado de ánimo y énfasis
Veritone Voice	Voz verdaderamente realista a una velocidad y escala inigualables / Con la confianza de los líderes de los medios de comunicación, radiodifusión y los deportes / Las voces familiares aportan un valor real, pero programar una grabación en un estudio puede ser un desafío / Voces que cuenta con el consentimiento de sus propietarios
Wondercraft	Facilita la vida a los podcasters, permitiéndoles concentrarse en crear contenidos de calidad / Sonido natural, fluido e hiperrealista / Se adapta al tono y al estilo del contenido
Play.ht	Genera voces indistinguibles de los humanos / Resultados ultrarealistas / Líder en IAG de voz / Descargas ilimitadas gratuitas / Las voces de IA más fluidas y conversacionales / Con la confianza de usuarios y marcas de todos los tamaños / La clonación de voz capta todo acento e idioma
Synthesis	Voces entrenadas con profesionales de la voz / Te ofrecemos voces tan realistas que tendrás que pellizcarte / Con la confianza de cientos de marcas líderes
Amazon Polly	Voces generativas, de formato largo, neuronales y de alta calidad (TTS) de alto rendimiento/ Sintetizan el habla natural con una alta precisión de pronunciación / Admite docenas de voces e idiomas, y ofrece opciones de voz masculina y femenina para la mayoría de los idiomas / La confianza, privacidad y seguridad de nuestros usuarios son nuestras máximas prioridades
Murf.ai	Genera voces similares a la humana en tiempo real / Los resultados dependen de la calidad y cantidad de datos del entrenamiento / Su uso plantea problemas éticos y legales. Solicita al usuario un uso responsable
Lovo.ai	Generador de voz hiperrealista que cautiva a tu audiencia / Generador galardonado / Genera voces sin esfuerzo / Los resultado conservan el acento y matices de la voz original / Perfecto para profesionales y creadores de contenido que necesitan voces de alta calidad



Argumentos de venta / **Limitaciones** / **Advertencias**. N/I: No informa. Fuente: Elaboración propia a partir de los portales de las desarrolladoras.

4.2. PÓDCAST GENERADOS CON INTELIGENCIA ARTIFICIAL GENERATIVA DE VOZ Y PERCEPCIÓN DE LOS *PODCASTERS*

Tras la búsqueda y aplicación de los criterios de inclusión y exclusión establecidos, emana un total de 10 títulos que cumplen con los requisitos definidos para el estudio.

Relatos Sintéticos (2022) es el primer podcast español producido por Prodigioso Volcán con el propósito de ilustrar el potencial de la IA para impulsar la creatividad, mejorar la eficiencia y explorar nuevas posibilidades en la comunicación y la creación de contenidos (Prodigioso Volcán, 2022). Se trata de una serie de cuentos de terror narrados por una voz masculina y una femenina generados artificialmente. Para su generación, los creadores colaboran con Monoceros Labs, especialistas en tecnologías del habla. La productora Prodigioso Volcán subraya que, para lograr una prosodia natural, los textos que se proporcionan a la IAG requieren de una puntuación muy compleja, pero su aplicación en el *podcasting* impulsa la creatividad y mejora la eficiencia (reduce costes de tareas complejas).

Otro título creado con el propósito de demostrar las capacidades de la IA, según su autor a modo de experimento, es *Joe Rogan AI Experience* (2023). En este podcast de entrevistas se replica el podcast original *Joe Rogan AI Experience*, en que el presentador clonado entrevista a invitados igualmente clonados, como Sam Altman, CEO de OpenAI, y el expresidente Donald Trump. Su creador, quien solo ha revelado su nombre, Hugo, es consciente de que no cuenta con el consentimiento para usar las voces clonadas, y afirma que en el futuro las plataformas de *podcasting* podrían prohibir este tipo de suplantación. Respecto a los resultados, sostiene que IAG clona las voces de manera casi perfecta; sin embargo, presenta problemas de cadencia, lo que le obliga a editar las voces, incluyendo muletillas o interjecciones como 'ah' y 'um' para transmitir un proceso de pensamiento genuino y escalonado, es decir, para que suene como una conversación humana normal (Launum, 2023). Por este motivo, el *podcaster* termina invirtiendo más tiempo en la producción que si trabajara con voces humanas (Hugo en Khalid, 2023).

Hacker News Recapt (2023) es un podcast diario que ofrece información sobre la industria tecnológica, con texto y audio generados mediante IA. Se trata de un proyecto de Wondercraft.ai para demostrar a los usuarios el potencial de la plataforma y atraer a creativos independientes que deseen convertir entradas de su blog en audio, pero que carezcan de tiempo o dinero para contratar a un locutor. Dimitris Nikolaou, CEO de la *startup*, opina que los resultados convincentes que se obtienen podrían fidelizar al oyente (Nicolau en Knibbs, 2024).

El ingeniero de sonido israelí, Lior Sol, ha creado el podcast *Myself, I Am and That* (2023) mediante el *software* ElevenLabs. Sol clona su voz dos veces para mantener conversaciones consigo mismo. El autor no busca grandes audiencias, sino una forma para divertirse jugando con las nuevas tecnologías (Sol en Knibbs, 2024). La música del podcast también ha sido creada con IAG mediante Beatoven.ai y Synthesizer V.

Synthetic Stories (2023), de Andy Durrant, es otro ejemplo de podcast de ficción generado con voces sintéticas. El guion y el diseño sonoro también han sido generados por IAG. Durrant afirma que «están orgullosos del podcast como experimento, pero como obra creativa, enseña a dar cuenta de las limitaciones» (Podnews, 2023).



El boletín informativo *WP a day* (2023) sobre el universo de WordPress, creado por el *podcaster* Antonio Cambronero, aplica la IAG en todo el proceso productivo, desde la creación del guion hasta la generación de la voz. Emplea el *software* Amazon Polly, específicamente la voz de Lucía. Cambronero describe el proyecto como un experimento para explorar los alcances y los límites de la IA. Reconoce que la voz aún carece de naturalidad. Considera que «para llegar a conseguir el proceso totalmente automático, es preciso realizar un trabajo de codificación (o entrenamiento) laborioso y que implica mucho esmero». Añade que «las herramientas de IA en la actualidad no son más que una ayuda y muchas veces cometen fallos e imprecisiones» (Cambronero, 2024). Sin embargo, considera que tiene ventajas incuestionables como el ahorro de tiempo y la reducción de costos.

Finalmente, destaca David Metlzer en español (2022), una producción del *coach* y empresario David Meltzer. Se trata de una traducción íntegra de su pódcast en inglés *The Playbook*, para la cual emplea el *software* Veritone Voice. Señala que antes de la IA, la traducción se realizaba con locutores profesionales en un estudio, por lo que su uso implica un ahorro considerable de tiempo y recursos (Metlzer en Rivera y Rivera, 2022).

Tabla 3. Pódcast de la muestra elaborados con inteligencia artificial generativa de voz

Título Año	Autor Categoría	Software	Uso IAG de voz	Género	Se informa naturaleza IAG
Biotopía (2020)	Manuel Bartual Independiente	No informa	Parcial	Ficción	Sí
X-Rey (2021)	The Story Lab Productora	Vicomtech	Parcial	Documental	Sí
Relatos sintéticos (2022)	Prodigioso Volcán Productora	Monoceros Labs	Íntegro	Ficción	Sí
David Meltze en español (2022)	David Meltze Independiente	Veritone Voice	Íntegro	Formativo	Sí
Joe Rogan IA Experience (2023)	Hugo Amateur	Chat GTP	Íntegro	Entrevistas	Sí
Myself, I am and that (2023)	Lior Sol Amateur	Eleven Labs	Íntegro	Entrevistas	Sí
Synthetic Stories (2023)	Andy Durrant Independiente	No informa	Íntegro	Ficción	Sí
WP a day (2023)	Antonio Cambronero Independiente	Amazon Polly	Íntegro	Informativo	Sí
Hacker News Recapt (2024)	Dimitris Nikolau Independiente	Wondercraft	Íntegro	Informativo	Sí

Fuente: Elaboración propia.



En cuanto a los *podcast* que emplean la IAG de voz de forma parcial, destaca *Biotopia* (2020), una ficción creada por Manuel Bartual, donde se presenta un falso boletín de noticias que relata las vivencias de una comunidad científica. Se utiliza la IAG de voz para caracterizar a Amaia, la inteligencia artificial desarrollada por esta comunidad científica. Para Bartual:

Lo interesante aquí es que más allá de elegir entre varios tipos de voces y a qué velocidad quieres que hablen no tienes mucho más control sobre cómo entonan cada frase, de modo que lo que hice fue un proceso de «casting» previo para ver qué voces me funcionaban mejor para cada personaje, y a partir de ahí reescribo algunas de sus frases si cuando las grabo no me gusta la forma en la que las pronuncian. Aunque, sorprendentemente, la mayoría de las veces me encuentro con interpretaciones que te hacen dudar sobre si quien está al otro lado es humano o no (Bartual en Gil, 2020).

En esta categoría también se localiza *X-Rey* (2021). El *podcast* producido por The Story Lab, dirigido y narrado por Álvaro de Cózar, es un documental sonoro que explora la vida del rey emérito Juan Carlos I. Utiliza la IAG de voz para recrear la voz de ciertos personajes históricos, como la del general Franco. Para esta tarea, han contado con la colaboración del centro tecnológico Vicomtech. Según la productora, uno de los mayores desafíos fue localizar audios en buen estado, sin ruidos y con el estilo narrativo deseado para la clonación. Esto resultó en que el modelo de aprendizaje tuviera únicamente 6 horas de grabaciones disponibles para su entrenamiento (Panorama Audiovisual, 2020).

Por último, en lo que respecta a la identificación de su naturaleza artificial, la totalidad de los *podcast* de la muestra la indican en su descripción textual.



Tabla 4. Percepción de los *podcasters* de la inteligencia artificial generativa de voz

Podcaster productora	Oportunidades	Limitaciones
Manuel Bartual	Realismo de las voces	No observa
The Story Lab	Impulsa la creatividad	Entrenamiento laborioso
Prodigioso Volcán	Impulsa la creatividad Reducción de costes	Requiere puntuación compleja
David Meltze	Ahorro de recursos (tiempo y costes)	No observa
Hugo	Interesante para experimentar	Cadencia poco realista Cuestionabilidad ética (usurpación identidad)
Lior Sol	Interesante para experimentar	Problemas con la cadencia
Andy Durrant	Interesante para experimentar	No observa
Antonio Cambroneró	Ahorro de tiempo Reducción de costes	Carencia de naturalidad Entrenamiento muy laborioso Fallos e imprecisiones
Dimitris Nikolau	Resultado convincente	No observa

Fuente: Elaboración propia.

5. Discusión y conclusiones

El análisis realizado evidencia un sesgo entre los resultados proclamados por las empresas tecnológicas de los *softwares* analizados y la percepción de los creadores de pódcast sobre los resultados obtenidos tras su aplicación. Exceptuando a tres empresas tecnológicas, la mayoría sostiene que su *software* genera voces hiperrealistas, capaces de capturar el habla natural con alta precisión en la pronunciación. Además, aseguran que el proceso de clonación es ágil y requiere un mínimo esfuerzo por parte del usuario. Estos argumentos de venta entran en conflicto con la percepción de los creadores de pódcast, quienes opinan que la IAG de voz aún no produce resultados suficientemente realistas para transmitir la esencia de la prosodia humana (solo dos de nueve creadores destacan haber obtenido resultados convincentes). Aunque no se especifica si estos creadores utilizaron las versiones *premium* de los *software*, que ofrecen funcionalidades más avanzadas, en general, coinciden en que la escucha de los títulos resulta monótona y aburrida debido a la falta de matices emocionales inherentes a la voz humana, un aspecto previamente señalado por la academia (Winters y Pisoni, 2004; Chaparro-Domínguez, 2024; Ada *et al.*, 2024). De este modo, sus opiniones se alinean con las de autores como Atkinson *et al.* (2005), Walters *et al.* (2008) y Rodero (2018), quienes consideran que una buena prosodia es esencial para mejorar la atención y comprensión del oyente, y para dotar al mensaje de credibilidad (Fitó-Carreras *et al.*, 2023).

No obstante, más allá de la capacidad actual de la IA para recrear fielmente el habla humana y generar confusión entre la artificialidad y la realidad, la cuestión anterior puede discutirse cuando la clonación es realizada por laboratorios tecnológicos especializados. En este contexto, la exploración revela la existencia de dos categorías de *software* de IAG de voz: por un lado, *software* accesible para cualquier usuario, que permite obtener clonaciones rápidas a través de una interfaz intuitiva, comúnmente utilizado por *podcasters amateurs* o profesionales independientes que publican de forma periódica; y, por otro lado, *software* desarrollado por centros tecnológicos especializados, que no están disponibles de manera pública. De este modo, la calidad de los resultados está condicionada por estas variables, y en el caso del *software* de acceso libre, es importante considerar si el creador ha utilizado la versión gratuita o la *premium*, así como su competencia en el uso de la tecnología. Esta idea se conecta con la concepción experimental del empleo de la IAG de voz por parte de la mayoría de los *podcasters amateurs* o profesionales independientes en la muestra, cuyos pódcast han sido creados con el objetivo de observar y/o demostrar el potencial de la IAG de voz en el ámbito del *podcasting* (tres de cuatro abordan la producción desde un enfoque experimental).

En cuanto al argumento de venta relacionado con la agilización del flujo de trabajo de los creadores, solo dos de los nueve destacan este aspecto como un punto fuerte de la IAG de voz. Los creadores afirman que, para obtener resultados de calidad, es necesario invertir un tiempo considerable en el entrenamiento de la voz y la edición del audio, con el fin de mejorar la cadencia e incorporar elementos paralingüísticos (como muletillas, risas y tartamudeos) que se asemejen a la fluidez real del habla humana. Esto se debe a la limitada capacidad de los *softwares* para replicar estos matices, una observación que coincide con lo señalado por Ada *et al.* (2024).

Finalmente, el discurso de los creadores de pódcast no refleja ningún temor ante los posibles riesgos laborales que la tecnología podría suponer para el futuro de la profesión; por



el contrario, la perciben más como una oportunidad que como una amenaza. Sin embargo, estas reflexiones se enmarcan en un contexto en el que la tecnología aún no ha alcanzado la capacidad para replicar con precisión la prosodia humana. Esto sugiere que la percepción de la tecnología podría cambiar a medida que esta avance, en línea con lo apuntado por Jiménez Peña *et al.* (2024) en su investigación sobre las diferencias entre voces naturales y artificiales en el ámbito de la investigación judicial forense. Los autores anticipan que, a medida que el desarrollo de la inteligencia artificial progrese y se diseñen nuevos algoritmos, es probable que las voces generadas artificialmente mejoren en calidad y se acerquen cada vez más a las voces humanas.

Este análisis del estado actual de la IAG de voz aplicada al *podcasting* ha permitido identificar tanto las oportunidades como las limitaciones que esta tecnología ofrece a los creadores. Se destaca que la oferta actual no logra reproducir con precisión la prosodia de la voz humana y sus elementos paralingüísticos, debido a las limitaciones que aún presentan las redes neuronales profundas, en las que se basa la síntesis de voz de la IA. Por este motivo, la tecnología no es percibida como una amenaza inmediata que ponga en peligro los puestos de trabajo. Sin embargo, los futuros avances en la IAG de voz, que generan gran expectación, podrían presentar un reto significativo, lo que subraya la necesidad de que el sector tome medidas proactivas para prepararse ante los desafíos que esta tecnología podría plantear.

Por otro lado, se observa que la industria de la IAG de voz está adaptándose a las necesidades del sector, ofreciendo múltiples herramientas a través de plataformas especializadas que permiten clonar la voz, editar grabaciones, publicar pódcast y distribuirlos en varios idiomas. No obstante, al contraponer los argumentos de venta con la percepción de los creadores, se evidencia que la mayoría de las empresas tecnológicas analizadas no son completamente transparentes respecto a la calidad de los resultados que prometen.

Esta investigación ha proporcionado una aproximación al uso de la IAG de voz en el *podcasting*, basada en la literatura disponible en el momento del estudio. Sin embargo, los resultados no pueden ser generalizables, dado que este es un campo en constante evolución. Es posible que durante el transcurso del estudio hayan surgido nuevos *softwares* de clonación de voz, se hayan incorporado nuevas voces a los bancos de datos, y se hayan publicado más títulos de pódcast. Por lo tanto, a medida que la tecnología avance, será necesario realizar estudios adicionales que contrasten y amplíen los hallazgos aquí presentados. Asimismo, consideramos que los recientes avances de la IAG de voz también afectarán a las investigaciones sobre el simbolismo y las formas del sonido y la voz (Spence, 2012), y revisiones sistemáticas que se han realizado sobre el *audiobranding* y el *branding* sonoro (Vidal-Mestre *et al.*, 2022).

6. Referencias bibliográficas

- Ada, Jørgensen; Stina Hasse, y Fritsch, Jonas (2024). Cultures of the AI paralinguistic in voice cloning tools. En *Companion Publication of the 2024 ACM Designing Interactive Systems Conference* (pp. 249-252). <https://doi.org/10.1145/3656156.3663708>
- Aguado-Terrón, Juan Miguel, y Grandío-Pérez, María del Mar (2024). Hacia una ecología mediática de la IA generativa: la obra creativa en la era de la automatización. *Palabra Clave*, 27(1), 1-23. <https://doi.org/10.5294/pacla.2024.27.1.8>



- Alexander, Jessica D., y Nygaard, Lynne C. (2008). Reading voices and hearing text: talker-specific auditory imagery in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 34(2), 446. <https://doi.org/10.1037/0096-1523.34.2.446>
- Álvarez Ramírez, Anel; Anzures-García, Mario, y Huerta Rangel, José Alejandro (2022). Construcción de una red neuronal replicadora de datos y una aplicación para clonación de voz. *Res Comput Sci*, 151(5), 15-30. <https://bit.ly/4dWARQO>
- Aronovitch, Charles D. (1976). The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. *The Journal of social psychology*, 99(2), 207-220. <https://doi.org/10.1080/00224545.1976.9924774>
- Ashworth, Boone (2023, 20 de abril). Artificial intelligence can clone the voice of your favorite podcast announcer. *Wired.com*. <https://bit.ly/3wT5qq3>
- Atkinson, Robert K.; Mayer, Richard E., y Merrill, Mary Margaret (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*, 30(1), 117-139. <https://doi.org/10.1016/j.cedpsych.2004.07.001>
- Bhargava, Cherry, y Sharma, Pardeep Kumar (2022). *Artificial intelligence: fundamentals and applications*. CRC Press.
- Boden, Margaret A. (2018). *Artificial intelligence: a very short introduction*. Oxford, Oxford University Press.
- Bottomley, Andrew J. (2015). Podcasting: A decade in the life of a "new" audio medium: Introduction. *Journal of radio & audio media*, 22(2), 164-169. <https://doi.org/10.1080/19376529.2015.1082880>
- Brennen, J. Scott; Howard, Philip N., y Nielsen, Rasmus K. (2022). What to expect when you're expecting robots: Futures, expectations, and pseudo-artificial general intelligence in UK news. *Journalism*, 23(1), 22-38. <https://doi.org/10.1177/1464884920947535>
- Cambroner, Antonio (2024, 11 de marzo). *El porqué de hacer un podcast con inteligencia artificial (y el cómo)*. Blogpocket.com. <https://bit.ly/4dRPAN9>
- Cascella, Marco; Montomoli, Jonathan; Bellini, Valentina, y Bignami, Elena (2023). Evaluating the feasibility of ChatGPT in healthcare: An analysis of multiple clinical and research scenarios. *Journal of medical systems*, 47(1), 1-5. <https://www.doi.org/10.1007/s10916-023-01925-4>
- Chaparro-Domínguez, María Ángeles (2024). Capítulo 5. El impacto de la IA en los contenidos periodísticos sonoros. *Espejo de Monografías de Comunicación Social*, (25), 119-139. <https://doi.org/10.52495/c5.emcs.25.p108>
- Cohen, Laura B. (2001). 10 Tips for Teaching How to Search the Web. *American Libraries*, 32(10), 44-46. <http://www.jstor.org/stable/25646112>
- Dasborough, Marie T. (2023). Awe-inspiring advancements in AI: the impact of ChatGPT on the field of organizational behavior. *Journal of organizational behavior*, 44(2), 177-179. <https://www.doi.org/10.1002/job.2695>
- Edwards, Chad; Edwards, Autumn; Stoll, Brett; Lin, Xialing, y Massey, Noelle (2019). Evaluations of an artificial intelligence instructor's voice: Social Identity Theory in human-robot interactions. *Computers in Human Behavior*, 90, 357-362. <https://doi.org/10.1016/j.chb.2018.08.027>
- Espinosa, Iván (2019). La teoría sobre el sonido y los estudios de radio: una propuesta de divulgación. *Lógoi: revista de filosofía*, (35), 38-50. <https://bit.ly/3SR343D>
- European Parliament (2024, 13 de marzo). *La Eurocámara aprueba una ley histórica para regular la inteligencia artificial*. <https://bit.ly/3QYhugB>



- Faure-Carvalho, Adrien; Calderón-Garrido, Diego, y Gustems-Carnicer, Josep (2022). *Escuchar el cine* [Documento docente]. Universitat de Barcelona. <http://hdl.handle.net/2445/182308>
- Fitó-Carreras, María; Méndiz-Noguero, Alfonso, y Vidal-Mestre, Montserrat (2023). The podcast as a sound experimentation tool for brands: The immersive narrative in Endesa's *Sonidos que nos transforman*. *Cuadernos.info*, (56), 293-312. <https://doi.org/10.7764/cdi.55.62819>
- Franganillo, Jorge (2022). Contenido generado por inteligencia artificial: oportunidades y amenazas. *Anuario ThinkEPI*, 16. <https://doi.org/10.3145/thinkepi.2022.e16a24>
- Franganillo, Jorge (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *Methaodos, revista de ciencias sociales*, 11(2), 15. <http://dx.doi.org/10.17502/mrcs.v11i2.710>
- Gil, Felipe G. (2020, 15 de julio). *Biotopía: el podcast de ciencia ficción dirigido por Manuel Bartual donde todo es posible*. El diario.es. <https://bit.ly/3AndPDK>
- Goris, Silmani J., y Adolf, G. (2015). Utilidad y tipos de revisión bibliográfica. *Revista Ene de Enfermería*, 9(2). <https://dx.doi.org/10.4321/S1988-348X2015000200002>
- Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human-machine communication research agenda. *New media & society*, 22(1), 70-86. <https://doi.org/10.1177/1461444819858691>
- Hart, Chris (1998). *Doing a literature review*. Sage Publications.
- Jiménez, Miguel (2024, 30 de marzo). *Open AI lanza una herramienta de audio capaz de clonar las voces humanas*. El pais.com. <https://bit.ly/4dWezyU>
- Jiménez Peña, John; Torres Castillo, Fernando Aarón, Cueva Sánchez, Óscar Esaul (2024). Comparación forense de voces: un estudio preliminar sobre las diferencias entre una voz natural y una voz artificial para la investigación judicial. *Revista Oficial Del Poder Judicial*, 16(21), 53-81. <http://doi.org/10.35292/ropj.v16i21.881>
- Khalid, Amrita (2023, 3 de mayo). *The Joe Rogan AI Experience shows how hard it is to replace a real podcaster*. The Verge. <https://bit.ly/44VVCYU>
- Kishigami, Jay (2004). Triple A (Any device, Anytime, Anywhere): services in ubiquitous networks and their impacts on the architecture and systems. En *All Star Network Access Workshop* (pp. 2-4).
- Knibbs, Kate (2023, 24 de mayo). *Generative AI Podcasts Are Here. Prepare to Be Bored*. Wired.com. <https://bit.ly/3UOi9IG>
- Launum, Nikolas (2023, 12 de abril). *AI-generated Joe Rogan podcast stuns social media with 'terrifying' accuracy: 'Mind blowingly dangerous'*. Foxnews.com. <https://bit.ly/4cmNm6g>
- Media Psychology Lab. (2023). *Estudio neurocientífico sobre audiolibros: qué formato transmite mejor la historia*. UPF. <https://bit.ly/4ethaAv>
- Mosby, Albert (2024, 20 de febrero). *Podcast Statistics (2024 Data) – Total Listeners & Country Wise*. Yaguara.com. <https://www.yaguara.co/podcast-statistics/>
- Mullennix, John W.; Johnson, Kate A.; Topcu-Durgun, Meral, y Farnsworth, Lynn M. (1995). The perceptual representation of voice gender. *The Journal of the Acoustical Society of America*, 98(6), 3080-3095. <https://doi.org/10.1121/1.413832>
- Murf.ai (2023). *AI enabled real people's voices*. <https://murf.ai/>
- Nass, Clifford, y Brave, Scott (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. MIT press.
- Nass, Clifford, y Gong, Li (2000). Speech interfaces from an evolutionary perspective. *Communications of the ACM*, 43(9), 36-43. <https://dl.acm.org/doi/fullHtml/10.1145/348941.348976>



- Parratt-Fernández, Sonia; Mayoral-Sánchez, Javier, y Mera-Fernández, Montse (2021). The application of artificial intelligence to journalism: an analysis of academic production. *Profesional de la información*, 30(3), 1-12. <https://doi.org/10.3145/epi.2021.may.17>
- Podnews (2023, 21 de marzo). *A groundbreaking podcast series entirely designed and created by Artificial Intelligence*. Podnews. <https://podnews.net/press-release/synthetic-stories-ai>
- Preiksaitis, Carl, y Rose, Christian (2023). Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. *JMIR medical education*, 9, e48785. <http://doi.org.10.2196/48785>
- Prodigioso Volcán (2022). *Relatos sintéticos: un pódcast y tres IA*. <https://bit.ly/3yKfwKg>
- Rime, Jemily; Pike, Chris, y Collins, Tom (2022). What is a podcast? Considering innovations in podcasting through the six-tensions framework. *Convergence*, 28(5), 1260-1282. <https://doi.org/10.1177/13548565221104444>
- Rivera, Melvin, y Rivera, Aracely (2022, 10 de agosto). *Crean otro nuevo podcast con inteligencia artificial*. Viapodcast.com. <https://bit.ly/4buNXn2>
- Rodero, Emma (2018). El peso creciente de la voz y el sonido para comunicar en la era digital: el protagonismo de la oralidad. *Anuario AC/E de cultura digital*, 80-94. <https://bit.ly/4bwUNZ1>
- Rodero, Emma (2023). L'expérience émotionnelle d'écouter une voix. *Hermès, La Revue*, 92(2), 30-37. <https://www.cairn.info/revue--2023-2-page-30.htm>
- Rodríguez, María (2023, 27 de abril). *La síntesis del habla y sus limitaciones emocionales*. Textinnova.com. <https://bit.ly/4buO7uE>
- Rudolph, Rudolph (2018). *Redes Neuronales: Guía Sencilla de Redes Neuronales Artificiales*. CreateSpace Independent Publishing Platform.
- Russell, James A. (1980). A circumflex model of affect. *Journal of Personality and Social Psychology*, 39, 1161-1178. <https://doi.org/10.1037/h0077714>
- Shin, Donghee, y Biocca, Frank (2018). Exploring immersive experience in journalism. *New media & society*, 20(8), 2800-2823. <https://doi.org/10.1177/1461444817733133>
- Spence, Charles (2012). Managing sensory expectations concerning products and brands: capitalizing on the potential of sound and shape symbolism. *Journal of consumer psychology*, v. 22, n. 1, pp. 37-54. <https://doi.org/10.1016/j.jcps.2011.09.004>
- Spotify (2023, 25 de septiembre). *Spotify's AI Voice Translation Pilot Means Your Favorite Podcasters Might Be Heard in Your Native Language*. <https://bit.ly/3VgmaRi>
- Stenbom, Agnes; Wiggberg, Mattias, y Norlund, Tobias (2023). Exploring communicative AI: Reflections from a Swedish newsroom. *Digital Journalism*, 11(9), 1622-1640. <https://doi.org/10.1080/21670811.2021.2007781>
- Sun, Chengzhe; Jia, Shan; Hou, Shewei; AlBadawy, Ehab, y Lyu, Siwei (2023). Exposing ai-synthesized human voices using neural vocoder artifacts. *arXiv preprint*. <http://dx.doi.org/10.48550/arXiv.2302.09198>
- Taylor, Lara (2024). Tune in to tomorrow: AI-driven podcasting and its possibilities. *Public Services Quarterly*, 20(1), 46-53. <https://doi.org/10.1080/15228959.2023.2290288>
- United Voice Artist (2024). *UVA position paper on the artificial intelligence act*. <https://unitedvoiceartists.com/eu-ai-act/>
- Vaissnave, V.; Nandhini, S.; Davamani, K. A.; Malathi, P., y Pothumani, S. (2024). *Advancements in Deep Learning Algorithms*. Magestic Technology Solutions (P) Ltd. ISBN: 978-93-92090-47-9. <https://doi.org/10.47716/978-93-92090-47-9>



- Vidal-Mestre, Montserrat (2018). *Branding Sonoro. Sonokey®: el método de impulso emocional y mnemotécnico para las marcas*. UOC.
- Vidal-Mestre, Montserrat; Freire-Sánchez, Alfonso; Calderón-Garrido, Diego; Faure-Carvalho, Adrien, y Gustems-Carnicer, Josep (2022). Audio identity in branding and brand communication strategy: a systematic review of the literature on audio branding. *Profesional de la Información*, 31(5). <https://doi.org/10.3145/epi.2022.sep.04>
- Walters, Michael L.; Syrdal, Dag Sverre; Koay, Kheng Lee; Dautenhahn, Kerstin, y Boekhorst, Renete (2008). Human approach distances to a mechanical-looking robot with different robot voice styles. En *ROMAN 2008, The 17th IEEE international symposium on robot and human interactive communication* (pp. 707-712). IEEE. <http://doi.org/10.1109/ROMAN.2008.4600750>
- Winters, Stephen J., y Pisoni, David B. (2004). Perception and comprehension of synthetic speech. *Research on spoken language processing report*, 26, 95-138.

