eib0995885601

# De la narración del audiolibro a la textualidad verbal y visual del audiotexto: una forma alternativa para la adquisición de conocimientos

Efraín Alfredo Barragán-Perea*
Javier Tarango*

*Artículo recibido:*
*7 de noviembre de 2023*
*Artículo aceptado:*
*30 de enero de 2024*

*Artículo de investigación*

RESUMEN

El acceso a la información a través de la lectura tradicionalmente alude al proceso de percibir y comprender la escritura mediante la vista o el tacto; no obstante, la lectura a través de la escucha se establece como una forma de oralidad terciaria que permite la combinación de la escritura, la imagen y la voz. Tales características la han convertido en una poderosa alternativa para la adquisición de conocimientos para las nuevas generaciones, las cuales, en algunos casos, prefieren escuchar en lugar de leer. Por este motivo, se realizó una investigación de tipo documental de la literatura científica sobre el tema, mediante un estudio descriptivo, con el objetivo de profun-

* Facultad de Filosofía y Letras, Universidad Autónoma de Chihuahua, México
ebarragan@uach.mx jtarango@uach.mx

dizar en el uso de los audiotextos como forma alternativa para la adquisición de conocimientos. Para ello, fueron analizados los conceptos de audiotexto, audiolibro, oralidad terciaria, conversión de texto a voz, voces sintéticas y *deepfake* de voz. Se encontró que el impacto de las tecnologías de la información y la comunicación han hecho posible que los audiotextos se conviertan en una poderosa herramienta para la reivindicación de la palabra hablada y una herramienta complementaria para el proceso de enseñanza-aprendizaje.

**Palabras clave:** Audiotexto; Audiolibro; Oralidad terciaria; Conversión de texto a voz; Voces sintéticas; Deepfake de voz

**From Audiobook Narration to the Verbal and Visual Textuality of the Audiotext: An Alternative for Knowledge Acquisition**

*Efraín Alfredo Barragán-Perea* and *Javier Tarango*

## ABSTRACT

Traditionally, access to information through reading refers to perceiving and understanding writing through sight or touch; however, reading via listening has been established as a form of tertiary orality that allows writing, image, and voice combination. It conforms to a powerful alternative to knowledge acquisition for new generations, who sometimes prefer to listen to books instead of reading them. For this reason, a documentary-type investigation of the scientific literature on the subject was carried out –through a descriptive study– to delve into using audiotext as an alternative way to acquire knowledge. Audiotext, audiobook, tertiary orality, text-to-speech conversion, synthetic voices, and voice deepfake were the concepts analyzed to do this. The impact of information and communication technologies has made it possible for audiotexts to become a powerful tool for spoken word vindication and a complementary device for teaching-learning processes.

**Keywords:** Audiotext; Audiobook; Tertiary Orality; Text-to-Speech Conversion; Synthetic Voices; Voice Deepfake

DOI: http://dx.doi.org/10.22201/iibi.24488321xe.2024.99.58856

## Introducción

El término 'lectura' refiere al proceso de percibir y comprender la escritura a través de la vista, el tacto (sistema de lectura braille) o el oído. La lectura es un hábito comunicativo que permite a las personas desarrollar el pensamiento cognitivo e interactivo. Leer permite construir con facilidad nuevos conocimientos, ayuda al desarrollo y perfeccionamiento del lenguaje, mejora la expresión oral y escrita, contribuye al desarrollo de la imaginación y la creatividad y hace el lenguaje más fluido. La lectura nos mantiene informados, facilita el exponer nuestro pensamiento, potencia la capacidad de observación, de atención y de concentración y libera nuestras emociones (alegría, tristeza, cólera, miedo, sorpresa, amor, entre otras), por lo que es el camino hacia el conocimiento y la libertad.

Leer implica acceder a la información por medio de la utilización de signos, aunque ello no significa que los signos deban ser gráficos, pues leer no implica solamente ver. De hecho, así como con la lectura visual se puede obtener información, lo mismo sucede con la que ingresa al cerebro mediante la audición. De esta manera, la experiencia de escucha de un texto comprende un hecho social de compartición con otros, a diferencia de la lectura, que siempre es un acto individual, aun cuando quien lea esté acompañado (García-Roca, 2020). Lo anterior se debe a que quien escucha un texto no tiene necesidad de realizar la decodificación de los signos a través de la vocalización o subvocalización, definida por Llanga Vargas, Arias Cáceres y Araque Zaldaña (2019) como el mal hábito de los lectores de mover los labios, susurrar o repetir mentalmente cada palabra al leer, lo que ocasiona que estos tengan problemas para mejorar su comprensión y velocidad de lectura.

A decir de Ong (1987), la oralidad se clasifica en oralidad primaria y secundaria. La primera sucede en culturas analfabetas donde no hay contacto con la escritura y el aprendizaje sucede a partir de la imitación de experiencias donde, por medio de la memoria logra mantenerse, recrearse y socializarse el conocimiento acumulado. La segunda tiene que ver con la llegada de los medios de comunicación (la radio, el teléfono, la televisión, entre muchos otros), cuya existencia y funcionamiento depende de la escritura y la impresión. Así también, Vallorani y Gibert (2022) hablan de una oralidad terciaria, propia de nuestro tiempo, la cual permite la integración de soportes tales como la escritura, la imagen, el video, la voz, etcétera.

La oralidad terciaria vino a reafirmarse gracias a los avances en las tecnologías de la información y la comunicación, las cuales permitieron el desarrollo de dispositivos de audio portátiles con gran capacidad de almacenamiento y de gestión de contenidos que, a su vez, habilitaron el surgimiento de nuevas experiencias para los usuarios. Más adelante, la evolución del Internet trajo consigo

la llegada de las primeras plataformas multimedia, tales como YouTube, Vevo, Vimeo, Twitch, Daily Motion, iVoox, entre otras; el perfecto ejemplo de la reivindicación de la palabra hablada. Posteriormente, el fenómeno de los *podcasts* (contenidos multimedia en forma de serie episódica, bajo demanda, que el cliente puede escuchar cuando lo desee) y de los *booktubers* (comunidad de creadores de contenido que realiza y publica vídeos en Internet relacionados con los libros) ha contribuido a la migración de grupos de lectura tradicionales y presenciales a formatos audiovisuales (Paladines y Aliagas, 2023), es decir, a formatos basados en la utilización conjunta del oído y de la vista mediante imágenes y sonidos grabados. Esto demuestra que las nuevas generaciones de usuarios están deseosas de escuchar, en muchos casos, más que de leer. Bajo el contexto descrito, el objetivo del estudio fue profundizar en el uso de la textualidad verbal y visual del audiotexto como forma alternativa para la adquisición de conocimientos.

## Metodología

La presente investigación trata de un artículo de revisión de literatura y no, en sí, de un estudio sistemático o estado del arte que brinde información narrativa, resúmenes, evaluaciones de hallazgos o teorías sobre el tema en bases bibliográficas, especialmente de aquellos documentos que conforman las bases conceptuales del objetivo temático. A través de la literatura revisada, se propone definir y aclarar el problema estudiado, resumir hallazgos de investigaciones anteriores y reconstruir contenidos a través de la identificación de relaciones temáticas, contradicciones, lagunas e, incluso, inconsistencias en la literatura.

Para el logro de esta contribución fue realizada una investigación de tipo documental basada en la revisión de la literatura científica sobre el tema, para lo cual, se plantearon los siguientes elementos que caracterizan su diseño de investigación según el carácter del estudio:

1. de acuerdo a su enfoque paradigmático, se considera un estudio cualitativo;
2. por su naturaleza, trata de una investigación no experimental;
3. según su finalidad, consiste en un estudio descriptivo; y
4. por su temporalidad, es un estudio transeccional.

La investigación toma como criterio de selección de documentos a aquellos artículos originales cuya pertinencia de contenidos estén vinculados a los propósitos de esta propuesta, publicados en bases de datos científicas (SciELO, Redalyc, Dialnet, ScienceDirect, WoS y Scopus) mayormente entre los años 2018 y 2023, en idioma español e inglés. Se partió de la hipótesis que la textualidad verbal y

DOI: http://dx.doi.org/10.22201/iibi.24488321xe.2024.99.58856

visual del audiotexto representan una forma alternativa para la adquisición de conocimientos. Por tal motivo, fueron abordadas las diferencias entre los audiolibros y los audiotextos, el proceso de creación de voces sintéticas para la producción de audiotextos, el uso de la didáctica, audiotexto como estrategia, las desventajas del uso de la tecnología en la transformación de texto a voz, las tendencias futuras de los sistemas de texto a voz en la producción de audiotextos, para finalmente ofrecer las conclusiones del estudio.

## Revisión de la literatura

### *Diferencias entre el audiolibro y el audiotexto*

El audiolibro es la grabación de la narración de un libro completo o parcial que, generalmente, conserva su estructura y contenido original, incluyendo la narrativa, los diálogos y las descripciones. Los audiolibros pueden ser narrados por una persona, o bien dramatizados por un actor que interpreta los personajes y describe las escenas, lo cual los convierte en creaciones con gran aceptación para aquellos usuarios que desean escuchar la narración de una historia en lugar de leerla (Gramajo, Santagada y Paoletta, 2017). Algunas plataformas como Google Play Books, Audible de Amazon, Storytel, entre otras, ofrecen un servicio de audio-entretenimiento digital basado en suscripciones con acceso a una gran selección de audiolibros con costos incluso más bajos que los del texto impreso. Consecuentemente, el audiolibro es un recurso elaborado específicamente como oralidad secundaria.

Por su parte, un audiotexto es una grabación de contenido escrito o textual que no necesariamente proviene de un libro, y que ofrece la posibilidad de seleccionar diferentes tipos de voces para reproducirlo, que ahora son mucho más orgánicas de lo que eran en el pasado. Este puede incluir contenidos como artículos, ensayos, discursos, informes, notas de conferencias, instrucciones, entre otros (Sierra Berrocal, 2022). Los audiotextos no tienen que seguir una narrativa continua ni estar vinculados a una historia específica y con frecuencia son creados con el propósito de hacer que la información escrita sea más accesible para personas que prefieren escuchar en lugar de leer, como sucede con documentos de carácter académico y materiales educativos. Esta es una de las grandes ventajas de esta tecnología, la cual puede resultar más inclusiva y superior en prestaciones y disponibilidad que los audiolibros convencionales. Por lo tanto, el audiotexto debe considerarse dentro de la categoría de oralidad terciaria.

El término 'audiotexto' fue acuñado por Bernstein (1998) para describir la recepción auditiva del habla que se reproduce por medios analógicos o digitales.

Posteriormente, Bjork y Rumrich (2018) lo definieron como un recurso multimodal (presentación del texto de forma auditiva y visual) que permite al usuario leer y/o escuchar y que, igualmente, permite al lector elegir entre varios modos de lectura. De esta manera, tanto la visión como la audición adquieren gran importancia, así como el hecho de que el lector pueda elegir entre cualquiera de las dos formas (o ambas) para consumir contenidos. Así, la audiolectura por el medio visual y auditivo se convierte en una experiencia diferente, pero semejante y complementaria a la tradicional modalidad visual de los textos impresos o a la modalidad auditiva de los audiolibros.

### *La creación de voces sintéticas para la producción de audiotextos*

Un sistema de texto a voz (Text-to-Speech o TTS en inglés) tiene la capacidad de producir una voz con un sonido natural similar al humano a partir de texto escrito, lo que ofrece múltiples oportunidades para la creación de audiotextos. El sistema computarizado usado con este propósito se denomina 'sintetizador de voz', el cual utiliza algoritmos y tecnologías de procesamiento del lenguaje natural (Natural Language Processing o NLP en inglés) para analizar el texto y generar una representación sonora del mismo (Kaur y Singh, 2023). Las voces sintéticas son utilizadas en una variedad de aplicaciones y contextos (*Tabla 1*), por ello, han experimentado avances significativos en los últimos años.

| Aplicaciones y contextos | Descripción |
|---|---|
| Accesibilidad | Ayudan a las personas con discapacidades visuales o del habla a acceder a contenidos digitales por medio de dispositivos habilitados para voz o lectores de pantalla (Sierra Berrocal, 2022). |
| Navegación por GPS | Se utilizan en sistemas de posicionamiento global para proporcionar orientación a los conductores (Pesaru y Goswami, 2021). |
| Asistentes virtuales | Los asistentes de voz virtuales (Siri de Apple, Google Assistant y Amazon Alexa) usan voces sintéticas para interactuar con los usuarios a fin de ofrecer información y ayuda (Balci, 2019). |
| Servicios telefónicos automatizados | Muchas empresas emplean voces sintéticas en sus sistemas de respuesta de voz interactiva (Interactive Voice Response/IVR), ya sea para atender llamadas o proporcionar información a sus clientes (Fauzi *et al.*, 2021). |
| Videojuegos y aplicaciones interactivas | Las voces sintéticas son utilizadas en videojuegos y aplicaciones con el propósito de dar vida a los personajes, proporcionando a los usuarios una experiencia mucho más inmersiva (López Delacruz, 2023). |
| Generación de contenido multimedia | Se usan para crear contenido multimedia mediante la narración de videos, anuncios de radio y podcasts (Maldonado, 2020). |

DOI: http://dx.doi.org/10.22201/iibi.24488321xe.2024.99.58856

| | |
|---|---|
| Traducción y aprendizaje de idiomas | Pueden servir para ayudar a las personas a aprender nuevos idiomas y a mejorar su pronunciación (Ronda Pupo, Cueto Rodríguez y Cougle Iglesias, 2020). |
| Apoyo a la productividad | Las voces sintéticas pueden aprovecharse en aplicaciones de productividad como la conversión de texto en discurso para facilitar la revisión y edición de documentos (Franganillo, 2023). |

*Tabla 1.* Aplicaciones y contextos de las voces sintéticas
Fuente: elaboración propia

Las voces sintéticas tienen múltiples aplicaciones, por lo que desempeñan un papel muy importante en la mejora de la accesibilidad, la interacción con la tecnología y la automatización de procesos en diversas áreas. Es por esto que su uso continuará evolucionando a medida que la tecnología de texto a voz siga mejorando (Rodero y Lucas, 2023).

Los sintetizadores de voz pueden variar en calidad y naturalidad de la voz generada ya que algunos están diseñados para imitar voces humanas con gran fidelidad, mientras que otros pueden enfocarse en la simplicidad y la claridad del discurso (Kuligowska, Kisielewicz y Wlodarz, 2018). Esta tecnología ha avanzado significativamente en los últimos años gracias a los avances en el aprendizaje automático y la inteligencia artificial, lo que ha conducido a la creación de voces sintetizadas cada vez más naturales y realistas. Al proceso inverso se le denomina 'reconocimiento de voz'.

La rápida evolución de las tecnologías de la información y la comunicación y la amplia disponibilidad de equipos de cómputo, sistemas operativos y *software* facilitaron la tecnología de síntesis de voz. Tal condición permitió a los usuarios crear archivos de audio a partir de cualquier texto digital, lo que constituyó un nuevo fenómeno de oralidad (oralidad terciaria) que incidió notablemente en las prácticas lectoras. Sin embargo, dicha conversión corresponde más al concepto de un audiolibro y no al de un audiotexto.

El primer sistema de conversión de texto a voz fue desarrollado en Japón en el año 1968 por Noriko Umeda, pero fue hasta la década de 1980 cuando pudieron lograrse niveles prácticos en el empleo de esta herramienta (Adeyemo e Idowu, 2015). No obstante, hubo que esperar hasta los primeros años del siglo XXI para que fuera alcanzado un buen nivel en cuestión de naturalidad e inteligibilidad, condiciones básicas de la síntesis de voz. La naturalidad determina qué tan parecida es la salida a la voz humana, mientras que la inteligibilidad la determina la facilidad con la que se entiende la salida; el sintetizador de voz ideal generalmente intenta maximizar ambas funciones (Taylor, 2009). Posteriormente, los esfuerzos se centraron en la expresividad de las voces y otros aspectos prosódicos (relacionados con la correcta pronunciación y acentuación) y pragmáticos

del discurso (la adaptación a una determinada situación para que los interlocutores se comprendan entre sí) (Coto Jiménez y Morales Rodríguez, 2020).

Las voces sintéticas son creadas a partir de inteligencia artificial y sirven para leer palabras en voz alta desde un archivo de texto en formato PDF, desde el correo electrónico, o cualquier documento o sitio web. El teléfono móvil es el dispositivo más utilizado para este propósito, ya que al usuario solo le basta activar la función cuando usa los programas de lectura o interfaces con la opción de lectura en voz alta, es decir, la conversión se efectúa automáticamente. Como puede verse en la *Tabla 2*, algunas de las herramientas más utilizadas en la actualidad para la creación de voces sintéticas han sido desarrolladas por algunas de las principales empresas relacionadas con el ámbito tecnológico, como es el caso de Google, Microsoft o Amazon.

| Herramientas Text-to-Speech | Funcionalidades |
|---|---|
| Google Cloud Text to Speech | Es una interfaz de programación de aplicaciones (Application Programming Interface/API) para el procesamiento de voz a través de la tecnología de inteligencia artificial con niveles de comprensión cercanos a los humanos para 73 idiomas. Funciona a partir de algoritmos de redes neuronales de aprendizaje profundo para el reconocimiento automático de voz, esto es, emula cómo el cerebro humano procesa la información (Google Cloud, 2023). |
| Amazon Polly | Servicio de conversión de texto a voz generado por inteligencia artificial. Utiliza tecnologías de aprendizaje profundo para sintetizar el habla asemejándose a una voz humana, lo cual le permite convertir el texto de artículos a voz (Amazon Polly, 2023). |
| Microsoft Azure Speech Services | Este motor usa redes neuronales profundas para hacer que las voces de las computadoras sean casi indistinguibles de las de las personas. Con una articulación clara de las palabras, la conversión de texto a voz neuronal reduce considerablemente la fatiga auditiva cuando los usuarios interactúan con sistemas de inteligencia artificial (Microsoft Azure, 2023). |
| Natural Reader | Aplicación de texto a voz que convierte cualquier texto escrito en palabras habladas. A través de ella es posible pegar textos y documentos para que sean leídos en voz alta con voces naturales o convertirlos directamente a MP3 para escucharlos en cualquier momento y lugar. Es capaz de leer archivos en formato PDF, TXT, DOC, EPUB, ODS, ODT, PAGES, PPT, PNG y JPEG (Natural Reader, 2023). |

DOI: http://dx.doi.org/10.22201/iibi.24488321xe.2024.99.58856

| | |
|---|---|
| Murf AI | Es una plataforma de conversión de texto a voz basada en la nube que permite generar locuciones realistas apoyadas en inteligencia artificial para múltiples casos de uso como *e-learning*, vídeos de YouTube, podcasts, publicidad, sistemas de respuesta de voz interactiva (Interactive Voice Response/IVR), audiolibros, audiotextos, videojuegos, entre otros. Cuenta con más de 120 voces en 20 idiomas diferentes (Murf AI, 2023). |

*Tabla 2.* Herramientas para la creación de voces sintéticas
Fuente: elaboración propia

Como puede observarse, la creación de una voz sintética implica el uso de algoritmos y modelos lingüísticos en ambientes de inteligencia artificial (concepto que no se profundiza en este estudio). Este proceso de creación puede variar dependiendo de la herramienta o tecnología utilizada, pero, generalmente, involucra los siguientes pasos (Alonso *et al.*, 2013):

1. Se recopila un volumen de datos de audio sobre una determinada persona denominados 'datos de entrenamiento' con el fin de crear su voz sintética. Estos datos pueden incluir grabaciones de voz de la persona que son utilizados para preparar el sistema y entrenarlo en cómo debe escucharse la voz sintética resultante. Asimismo, se suman datos de texto para enseñar al sistema la pronunciación correcta de las palabras. Cabe mencionar que, conforme ha ido evolucionando esta tecnología, el volumen de datos requerido es cada vez menor.
2. Adicionalmente debe realizarse un análisis fonético que descompone el texto en unidades fonéticas (sonidos individuales que componen a las palabras). A continuación, el sistema debe tener en cuenta factores como el acento, el ritmo y la entonación de la persona para generar una voz sintética que suene lo más natural posible.
3. Después, se genera el audio mediante la utilización de algoritmos y modelos lingüísticos. La voz sintética resultante es generada y reproducida en tiempo real, o bien, puede ser pregrabada y almacenada para uso posterior.
4. Finalmente, la voz sintética final es sometida a un proceso de mejora mediante la edición de los parámetros de la voz (velocidad de habla, entonación y acento).

En la conversión de texto a voz intervienen también otros dos componentes (además de las voces sintéticas), el *software* lector con la función TTS y los motores de síntesis de voz (sintetizadores) que permiten a las aplicaciones 'hablar' y posibilitan que el texto escrito sea reproducido de forma oral y personalizable con la posibilidad de ver lo que está escuchándose. La calidad de un sintetizador

de voz se mide por la similitud que alcance con la voz humana y su habilidad para ser entendido con claridad.

La conversión de texto a voz presenta importantes ventajas frente al audiolibro, por ejemplo, porque es posible que una lectura pueda iniciarse con una lectura textual (visual) en una computadora o una tableta, para después continuar reproduciéndose con una aplicación de audio en un teléfono móvil. No obstante, Rodero y Lucas (2023) encontraron que aun cuando la síntesis de voz ha experimentado avances considerables, los audiolectores disfrutan más de las historias narradas por una voz humana que por una sintética debido a que la persona crea un mayor número de imágenes mentales, se involucra más a profundidad, presta mayor atención y tiene una respuesta emocional más positiva. Por su parte, Gil y Bergonzi Martínez (2023) descubrieron que la lectura en voz alta y comentada parece no sólo favorecer la comprensión en general, sino también el disfrute de los textos.

Respecto a las condiciones para la lectura literaria en las escuelas, Henkel, Mygind y Svendsen (2021) encontraron que estas están cambiando a medida que los lectores jóvenes tienen cada vez más la opción de alternar entre distintos medios (libro impreso, audiolibro y audiotexto) que ofrecen diferentes atractivos sensoriales y, por lo tanto, brindan experiencias distintas. Asimismo, afirman que el audiotexto facilita la sensación de estar presente dentro de la historia, por lo que tiene la oportunidad de convertirse en un evento sensorial y corporal que puede beneficiar la lectura de libros.

## *Uso del audiotexto como estrategia didáctica*

Las plataformas digitales han creado múltiples oportunidades para practicar, enseñar, aprender y crear métodos interesantes e innovadores para aquellos interesados en nuevos procesos de aprendizaje, por lo que el uso de las tecnologías de la información y la comunicación puede tener efectos positivos en el aprendizaje y desarrollo de los estudiantes. Actualmente, cada vez se utilizan más materiales digitales interactivos para promover la alfabetización oral en diferentes disciplinas, entre ellos se encuentran los relacionados con las herramientas de texto a voz, las cuales pueden tener su origen tanto en aplicaciones de *software*, dispositivos de *hardware*, o la combinación de ambos.

### *Aplicaciones de software*

Una de las primeras aplicaciones del audiotexto estuvo relacionada con la accesibilidad de personas con discapacidad visual, a las que hasta ese momento no les había sido posible acceder a la información escrita con la misma facilidad que al resto de la población (Gramajo, Santagada y Paoletta, 2017). Como muestra de

DOI: http://dx.doi.org/10.22201/iibi.24488321xe.2024.99.58856

ello, encontramos el caso del desarrollo de aplicaciones *web* inclusivas para los visitantes de museos, las cuales mejoran la experiencia de visita para personas no solo con discapacidad visual, sino también auditiva (Paddeu *et al.*, 2019). Otro ejemplo puede encontrarse en el uso de audiotextos como parte de los materiales educativos incorporados por el profesorado en plataformas de *e-learning* como Moodle, Educativa y Chamilo para la atención de estudiantes con necesidades educativas especiales, como la discapacidad visual y auditiva, o bien, con dificultades para leer (Juca Faicán, 2023).

De la misma manera, la conversión de texto a audio posibilita a las personas con discapacidades visuales el acceder a la producción académica y científica contenida en los diferentes repositorios institucionales, los cuales tienen como uno de sus principales objetivos dar mayor visibilidad a sus contenidos y maximizar su impacto en los distintos usuarios (De Giusti *et al.*, 2016). Esto es posible gracias al uso de *apps* TTS por medio de dispositivos móviles que están disponibles para su descarga gratuita, como es el caso de Google Cloud Text to Speech, Amazon Polly, Microsoft Azure Speech Services, Natural Reader, entre otras.

El audiotexto también es empleado para la enseñanza de idiomas en el mejoramiento de la pronunciación y velocidad de lectura (Ronda Pupo, Cueto Rodríguez y Cougle Iglesias, 2020). De igual modo, es usado en la creación de entornos interactivos basados en audio para desarrollar y utilizar la memoria a corto plazo y así ayudar al aprendizaje de matemáticas en niños con discapacidad visual (Sánchez y Flores, 2005) o, incluso, para reducir la distracción mental en estudiantes con dislexia (Bonifacci *et al.*, 2022). A este respecto, estudios en psicología educativa sugieren que las personas aprenden mejor cuando los materiales visuales de aprendizaje van acompañados de explicaciones auditivas en lugar de textuales (Zavgorodniaia *et al.*, 2020).

Según Cahill y Richey (2015), el uso de los audiotextos como parte de las estrategias didácticas en el aprendizaje de los niños da soporte a la lectura en voz alta y, por ende, al aprendizaje de la lectura y la escritura. Gracias a los audiotextos, los niños acceden de forma equitativa a la información, sin importar si pueden leer o no por sí mismos. De la misma manera, contribuyen a que las infancias incorporen referencias culturales vinculadas al entorno que habitan. Asimismo, a través del audiotexto podemos afirmar que, a diferencia de la conversación cotidiana, el lenguaje de los libros presenta una gran diversidad de formas y usos lingüísticos que permiten expresar diferentes significados. De este modo, si el niño ha estado expuesto a situaciones en las que escucha una lectura y, al mismo tiempo, puede seguirla a través del texto, comprenderá las peculiaridades de la lengua hablada y la escrita. Igualmente, podrá establecer la relación entre las formas de la lectura oralizada y las formas gráficas de la escritura, la tipografía, el tamaño de las letras, la función de las imágenes, entre otras.

Actualmente el avance de la tecnología de síntesis de voz es utilizado también en aplicaciones domóticas como las tecnologías de *hardware* y *software* en la automatización de los hogares (Coronado Arjona *et al*., 2017), asistentes de voz, lectores de pantalla, navegación GPS, atención telefónica automatizada, y demás (Pesaru y Goswami, 2021).

### *Dispositivos de hardware*

Otra forma de aprovechamiento de la tecnología de texto a voz como estrategia didáctica para desarrollar habilidades lingüísticas es a través de dispositivos de *hardware*, como es el caso del lápiz de audio o *audio pen*, también llamado lápiz digital, lápiz de lectura o lápiz parlante.

El lápiz de audio es una tecnología de asistencia que funciona como un lápiz convencional ya que puede escribir sobre cualquier papel, aunque en realidad fue diseñado para usarse sobre el que provee el fabricante. Integra una cámara de infrarrojos que detecta en qué parte del soporte estamos escribiendo, de modo que captura digitalmente todo lo que el usuario anota para, después, traspasarlo a una computadora. Además, integra un micrófono que guarda el sonido ambiental mientras el usuario escribe, esta grabación puede recuperarse posteriormente para su escucha con solo pulsar en algún punto de la hoja. Tales particularidades lo convierten en una herramienta ideal para que, por ejemplo, un estudiante en clase pueda tomar apuntes sin dejar pasar nada. Algunos ejemplos de este tipo de dispositivos son ScanMarker Air, Lector C-Pen, IRISPen Air 7, Livescribe SmartPen, entre otros.

En palabras de Tan, Chen y Lee (2013), los lápices digitales ofrecen flexibilidad, portabilidad y familiaridad al permitirle a sus usuarios explotar amplias funciones digitales, al tiempo de mantener la interacción natural, común en las interfaces tradicionales de lápiz y papel. La tecnología de lápiz de audio puede resultar adecuada para los niños más pequeños que aún no saben leer, pues ofrece nuevas oportunidades que amplían el uso tradicional de los libros. Un estudio realizado por Chen, Tan y Lo (2016) reveló que las tecnologías de lápiz de audio ya se han aplicado con éxito en países como Estados Unidos y China en el diseño de entornos de aprendizaje interactivos basados en papel y lápiz digital. Por su parte, Greenwood *et al.* (2016) demostraron que la tecnología de lápiz de audio apoya efectivamente las habilidades de lectura, como lo son el vocabulario y la comprensión.

## *Desventajas en el uso de la tecnología de texto a voz*

Las tecnologías de texto a voz han proporcionado formas que generan voces sintéticas a partir de textos con sólo una pequeña colección de expresiones grabadas.

DOI: http://dx.doi.org/10.22201/iibi.24488321xe.2024.99.58856

Si bien estos nuevos enfoques de la síntesis de voz pueden facilitar experiencias más fluidas, también es verdad que abren la puerta a quienes buscan consumar algún tipo de engaño, por lo que es importante anticipar los daños potenciales e idear estrategias para ayudar a mitigar usos indebidos (Noah *et al*., 2021).

Como ejemplo de lo anterior surge el término '*deepfake* de voz', acrónimo del inglés constituido por las palabras *deep learning* 'aprendizaje profundo' y *fake* 'falsificación'. Se trata de una técnica de inteligencia artificial que posibilita la creación de audios y videos de personas reales diciendo/haciendo cosas que ellos nunca dijeron o hicieron. Para ello, se vale de avanzados algoritmos de aprendizaje no supervisados conocidos como RGA (Red Generativa Antagónica), así como de audios y videos existentes (Masood *et al*., 2023).

Las técnicas de aprendizaje automático han aumentado la sofisticación de la tecnología haciendo que los *deepfakes* sean cada vez más realistas y difíciles de detectar. La tecnología *deepfake* tiene características que permiten una difusión rápida y generalizada que normalmente es empleada de forma maliciosa o para difundir información falsa, como sería el caso de la propagación de versiones incendiarias sobre un tema, publicaciones en audio que desacreditan a una persona, historias que intentan darle sentido a una mentira, entre otros escenarios.

Las preocupaciones son principalmente de naturaleza ética, política, jurídica y tecnológica y se basan en el hecho de que los *deepfakes* destruyen la credibilidad de los documentos audiovisuales (principalmente videos), como medios de información o confirmación de hechos, pues ponen en duda su veracidad y generan riesgos que pasan por la desinformación, la difamación o el chantaje (Bañuelos Capistrán, 2020).

Una de las primeras herramientas para generar *deepfakes* de voz fue el programa Adobe Voco del año 2016. Este utilizaba inteligencia artificial para imitar la voz de una persona a partir de una grabación de su voz. No obstante, el proyecto fue cancelado una vez que se demostró que podía ser utilizado para crear falsificaciones de voz, lo cual representaba un peligro para la sociedad (Rini, 2020).

Desde entonces, el avance en la tecnología de redes neuronales ha hecho posible la creación de *deepfake* de voz cada vez más realistas. Actualmente, existen diversas herramientas que utilizan dicha tecnología para crear falsificaciones de voz, como es el caso de Wavenet, Tacotron y, últimamente, VALL-E de Microsoft (*Tabla 3*). Esta última puede replicar la voz humana a partir de una grabación de escasos tres segundos (Taki y Mastorakis, 2023):

| Herramientas *deepfake* | Descripción |
|---|---|
| WaveNet | Es una red neuronal desarrollada por DeepMind (propiedad de Google desde 2014), la cual es capaz de modular directamente ondas de sonido en lugar de concatenar fragmentos de sonido grabados, como hacen otras tecnologías. Se entrena con una gran cantidad de muestras de voz y aprende las características de diferentes voces, tanto masculinas como femeninas, en distintos idiomas. Incluso, puede generar música y otros sonidos como los de la respiración o los movimientos de la boca. Mediante WaveNet se obtienen voces sintéticas que tienen un sonido más natural en comparación a otros sistemas que imitan la voz humana (Van den Oord *et al.*, 2016). |
| Tacotron | Es un sistema de síntesis de voz desarrollado por Google que utiliza técnicas de aprendizaje profundo para generar un habla similar a la humana a partir del texto. Tacotron aprende a generar voz analizando grandes cantidades de datos provenientes de hablantes humanos para producir un discurso más natural y expresivo. Posee gran capacidad para generar voz con los patrones de acentuación y entonación que dan al lenguaje hablado su característico ritmo y melodía. Puede generar voz con un alto grado de inteligibilidad, por lo que es adecuado para aplicaciones donde la claridad y comprensibilidad son primordiales, como es el caso de los asistentes de voz o los sistemas para la atención de emergencias. El sistema mejora su rendimiento continuamente según se expone a más datos, así que puede adaptarse fácilmente a nuevos idiomas (Ning *et al.*, 2019). |
| VALL-E | Es un sistema de síntesis de voz desarrollado por Microsoft, basado en inteligencia artificial, con la capacidad de imitar cualquier voz humana. Está preparado para sintetizar el audio de una persona una vez que aprende su voz y genera entonaciones para preservar el tono emocional del hablante original. Puede utilizarse en contextos donde sea necesario editar la voz de una persona para cambiarla por el contenido de una nueva transcripción de texto, haciéndola decir algo que la persona originalmente nunca dijo. Puede combinarse con GPT 3 para interactuar en forma de diálogo y proporcionar respuestas que parezcan humanas. Su código no está abierto al público probablemente por el riesgo que implica poner palabras nunca dichas en la boca de otro. Es una situación similar a lo que sucede con los *deepfakes* (Hernández, 2023). |

*Tabla 3.* Herramientas para la creación de falsificaciones de voz
Fuente: elaboración propia

En caso de utilizar alguna de las herramientas descritas para la creación de *deepfake*, y ante los siempre presentes vacíos jurídicos en todas las sociedades del mundo, es necesario establecer límites éticos para hacerlo responsablemente, asegurándonos de contar con el consentimiento de la persona involucrada y tener presente si otros pudieran verse afectados (García-Ull, 2021).

DOI: http://dx.doi.org/10.22201/iibi.24488321xe.2024.99.58856

### *Tendencias futuras de los sistemas de texto a voz en la producción de audiotextos*

La revisión de las tecnologías de síntesis de voz expuesta hasta ahora ha permitido obtener un amplio panorama sobre los métodos, técnicas y aplicaciones de las mismas. Algunas de las tendencias futuras de dichas tecnologías, que actualmente están bajo desarrollo, son las siguientes:

a) Se prevén importantes avances en la conversión de texto a voz neuronal mediante algoritmos de aprendizaje profundo que permitirán analizar e imitar los patrones, la entonación y el tono del habla humana haciendo la experiencia más natural y atractiva (Costa-Jussà y Fonollosa, 2017).
b) El desarrollo de tecnologías de texto a voz en la búsqueda de inclusividad para usuarios con algún tipo de trastorno o discapacidad, como la dislexia o la ceguera (Keelor *et al*., 2020).
c) El perfeccionamiento de procesos de clonación de voz a partir de voces sintéticas como herramienta para anunciantes, cineastas, desarrolladores de video juegos y otros creadores de contenido (Franganillo, 2023).
d) Mayor equidad de género en la creación de voces sintéticas, lo que permitirá que los audiotextos puedan producirse, según sea el propósito, con voces tanto masculinas como femeninas.
e) La capacidad para generar voz en un mayor número idiomas o incluso desarrollar canciones a partir de la tecnología de texto a voz (Nekvinda y Dušek, 2020).
f) Aumentar la participación de los alumnos que estudian de forma remota (*e-learning*) a través de estrategias de gamificación con la implementación de voces con sonido natural. Estas explicarían las instrucciones del juego y los desafíos basados en habilidades, lo que aumenta la atención y la participación del cuerpo estudiantil (Orozco Aguirre y Riego Caravantes, 2019).
g) El enriquecimiento de los cursos en línea alojados en la nube, los cuales supondrían espacios idóneos para que los educadores creen contenido personalizado para los estudiantes (como los audiotextos) mediante herramientas dinámicas de texto a voz (Sierra Berrocal, 2022).

## Conclusiones

Con base en las propuestas teóricas investigadas, a fin de profundizar en el uso de los audiotextos como forma alternativa para la adquisición de conocimientos, se concluye lo siguiente:

a) El término lectura ha venido evolucionado a través del tiempo. Actualmente refiere al proceso de percibir y comprender la escritura a través de la vista, el tacto y también del oído (oralidad terciaria).
b) El desarrollo de las tecnologías de la información y la comunicación ha llevado a la sociedad moderna a transitar hacia la oralidad terciaria mediante la integración de soportes como la escritura, la imagen, el video, la voz, entre otros.
c) La gran disponibilidad de equipos de cómputo, sistemas operativos, *software* y acceso a Internet han facilitado la utilización de la tecnología de síntesis de voz. Tal ha incidido significativamente en la calidad de las creaciones, así como en las prácticas lectoras de las nuevas generaciones.
d) La conversión de texto en voz involucra tres elementos: la creación de voces sintéticas, el *software* lector con la función texto a voz y los motores de síntesis de voz (sintetizadores).
e) Las condiciones para la lectura literaria en las escuelas están cambiando a medida que los lectores jóvenes tienen, cada vez más, la opción de alternar entre los medios del libro impreso, el audiolibro y el audiotexto.
f) El audiotexto se ha convertido en una valiosa estrategia didáctica y de acercamiento al conocimiento, especialmente para aquellas personas con discapacidades visuales, auditivas y que padecen dislexia, tanto para quienes simplemente no desean leer.
g) Las herramientas de texto a voz son cada vez más utilizadas para la creación de materiales digitales interactivos de apoyo a la alfabetización oral en distintas disciplinas; dichas herramientas pueden tener origen tanto en aplicaciones de *software*, dispositivos de *hardware* o en la combinación de ambos.
h) Si bien los nuevos enfoques de la síntesis de voz facilitan experiencias más fluidas, también permite la entrada para los que buscan actuar engañosamente a través del *deepfake* de voz, lo que puede llegar a tener implicaciones de naturaleza ética, política, jurídica y tecnológica. Por esta razón, es importante anticipar los posibles daños potenciales para idear estrategias que mitiguen su uso indebido.

## Referencias

Adeyemo, Olufemi, y Anthony Idowu. 2015. "Development and Integration of Text to Speech Usability Interface for Visually Impaired Users in Yoruba Language". *African Journal of Computing and ICT* 8 (1): 87-94. https://bit.ly/3LNqbHR

DOI: http://dx.doi.org/10.22201/iibi.24488321xe.2024.99.58856

Alonso, Agustín, Iñaki Sainz, Daniel Erro, Eva Navas e Inma Hernaez. 2013. "Sistema de conversión texto a voz de código abierto para lenguas ibéricas". *Procesamiento del lenguaje natural* 51: 169-75.
https://bit.ly/3PZwESH

Amazon Polly. 2023. "¿Qué es Amazon Polly?" Amazon Web Services. Consultado el 20 octubre 2023.
https://docs.aws.amazon.com/es_es/polly/latest/dg/what-is.html

Balci, Erdem. 2019. "Overview of Intelligent Personal Assistants". *Acta Infológica* 3 (1): 22-33.
https://doi.org/10.26650/acin.454522

Bañuelos Capistrán, Jacob. 2020. "Deepfake: la imagen en tiempos de la posverdad". *Revista Panamericana de Comunicación* 2 (1): 51-61.
https://doi.org/10.21555/rpc.v0i1.2315

Bernstein, Charles. ed. 1998. *Close Listening: Poetry and the Performed Word*. Oxford University Press.

Bjork, Olin, y John Rumrich. 2018. "Is There a Class in This Audiotext? Paradise Lost and the Multimodal Social". En *Digital Milton,* editado por David Currell e Islam Issa, 47-76. Palgrave Macmillan.
https://doi.org/10.1007/978-3-319-90478-8_3

Bonifacci, Paola, Elisa Colombini, Michele Marzocchi, Valentina Tobia y Lorenzo Desideri. 2022. "Text to Speech Applications to Reduce Mind Wandering in Students with Dyslexia". *Journal of Computer Assisted Learning* 38 (2): 440-54.
https://doi.org/10.1111/jcal.12624

Cahill, Maria, y Jennifer Richey. 2015. "Audiobooks as a Window to the World". En *The School Library Rocks: Proceedings of the 44th International Association of School Librarianship (IASL) Conference 2015, Volume 1: Professional Papers*, editado por Lourense Das, Saskia Brand-Gruwel, Kees Kok y Jaap Walhout, 92-98. Heerlen: Open Universiteit.
https://www.iaslonline.org/resources/Pictures/IASL2015_Proceedings_Vol12ndEd_ProfPapers.pdf

Chen, Chih-Ming, Chia-Chen Tan y Bey-Jane Lo. 2016. "Facilitating English-Language Learners' Oral Reading Fluency with Digital Pen Technology". *Interactive Learning Environments* 24 (1): 96-118.
https://doi.org/10.1080/10494820.2013.817442

Coronado Arjona, Manuel Alejandro, Víctor Manuel Bianchi Rosado y Juan Alberto Vivas Burgos. 2017. "Evaluación de la usabilidad en aplicaciones domóticas móviles usando el método de recorrido". *Tecnología Educativa Revista CONAIC* 4 (1): 53-63.
https://doi.org/10.32671/terc.v4i1.114

Costa-Jussà, Marta, y José Fonollosa. 2017. "DeepVoice: Tecnologías de aprendizaje profundo aplicadas al procesado de voz y audio". *Procesamiento del Lenguaje Natural* 59: 117-20.
https://www.redalyc.org/pdf/5157/515754427013.pdf

Coto Jiménez, Marvin, y Maribel Morales Rodríguez. 2020. "Tecnologías del habla para la educación inclusiva". *Actualidades Investigativas en Educación* 20 (1): 631-656.
http://dx.doi.org/10.15517/aie.v20i1.40129

De Giusti, Maria Raquel, Ariel Lira, Julieta Paz Rodríguez Vuan y Gonzalo Luján Villarreal. 2016. "Accesibilidad de los contenidos en un repositorio institucional: análisis, herramientas y usos del formato EPUB". *e-Ciencias de la Información* 6 (2): 1-23.
http://dx.doi.org/10.15517/eci.v6i2.23690

Fauzi, Esa, Adri Genta Rahdian, Agustinus Ipan Suryana, Penta Al, Tiara Nastiti Handana Ningtias y Kinanti Dara Nurkhofifah. 2021. "Design and Implementation IVR Outbound Service API Using Text to Speech". *Review of International Geographical Education* 11 (5): 789-96.
https://onx.la/f62f1

Franganillo, Jorge. 2023. "La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos". *Methaodos. Revista De Ciencias Sociales* 11 (2): 1-17.
https://doi.org/10.17502/mrcs.v11i2.710

García-Roca, Anastasio. 2020. "Virtually Digital Reading: The Collective Challenge of Textual Interpretation." *Cinta de moebio* 67: 65-74.
http://dx.doi.org/10.4067/s0717-554x2020000100065

García-Ull, Francisco José. 2021. "Deepfakes: el próximo reto en la detección de noticias falsas". *Anàlisi* 64: 103-20.
https://doi.org/10.5565/rev/analisi.3378

Gil, José María, y Jonás Ezequiel Bergonzi Martínez. 2023. "Lectura en voz alta y comentada para enseñar (y disfrutar) a Borges". *Prometeica-Revista de Filosofía y Ciencias* 26: 143-62.
https://doi.org/10.34024/prometeica.2023.26.14766

Google Cloud. 2023. "IA de Text-to-Speech". Cloud Text-to-Speech. Consultado el 10 octubre 2023.
https://cloud.google.com/text-to-speech

Gramajo, María Cecilia, Miguel Santagada y Anabel Paoletta. 2017. "Una audioteca en la UNICEN". *La Escalera - Anuario de la Facultad de Arte* 27: 123-36.
https://www.ojs.arte.unicen.edu.ar/index.php/laescalera/article/view/567/486

Greenwood, Charles R., Judith J. Carta, Gabriela Guerrero, Jane Atwater, Elizabeth S. Kelley, Na Young Kong y Howard Goldstein. 2016. "Systematic Replication of the Effects of a Supplementary, Technology-Assisted, Storybook Intervention for Preschool Children with Weak Vocabulary and Comprehension Skills". *The Elementary School Journal* 116 (4): 574-99.
http://dx.doi.org/10.1086/686223

Henkel, Ayoe Quist, Sarah Mygind y Helle Bundgaard Svendsen. 2021. "Exploring Reading Experiences of Three Media Versions: Danish 8th Grade Students Reading the Story Nord". *L1-Educational Studies in Language and Literature* 21: 1-29.
https://doi.org/10.17239/L1ESLL-2021.21.02.04

Hernández, Gonzalo. 2023. "VALL-E: así es la IA de Microsoft capaz de simular cualquier voz a partir de una muestra de audio de tan solo tres segundos de duración". Xataka México, 10 enero 2023.
https://cutt.ly/dwnLB0JM

Juca Faicán, Wilmer Adrián. 2023. "Diseño de un entorno virtual de aprendizaje para atender las necesidades educativas especiales de un estudiante con discapacidad visual en la asignatura de Lengua y Literatura". Tesis de maestría, Universidad del Azuay.
https://bit.ly/3PFi4hN

Kaur, Navdeep, y Parminder Singh. 2023. "Conventional and Contemporary Approaches Used in Text to Speech Synthesis: A Review". *Artificial Intelligence Review* 56: 5837-80.
https://doi.org/10.1007/s10462-022-10315-0

Keelor, Jennifer L., Nancy Creaghead, Noah Silbert y Tzipi Horowitz-Kraus. 2020. "Text to Speech Technology: Enhancing Reading Comprehension for Students with Reading Difficulty". *Assistive Technology Outcomes and Benefits* 14: 19-35. https://acortartu.link/mpe4z

Kuligowska, Karolina, Paweł Kisielewicz y Aleksandra Włodarz. 2018. "Speech Synthesis Systems: Disadvantages and Limitations". *International Journal of Engineering & Technology* 7 (2.28): 234-39. https://doi.org/10.14419/ijet.v7i2.28.12933

Llanga Vargas, Edgar Francisco, Tatiana Silvana Arias Cáceres y Francisco José Araque Zaldaña. 2019. "Vicios de la lectura y el aprendizaje". *Revista Atlante: Cuadernos de Educación y Desarrollo*. https://bit.ly/46wGBwj

López Delacruz, Santiago. 2023. "Un vínculo paradójico: narrativas audiovisuales generadas por inteligencia artificial, entre el pastiche y la cancelación del futuro". *Hipertext.net* 26: 31-35. https://doi.org/10.31009/hipertext.net.2023.i26.05

Maldonado, Lucía. 2020. *Tecnología y educación: recursos para personas con dificultades de aprendizaje, limitaciones intelectuales, motoras, visuales y auditivas*. Buenos Aires: Editorial Biblos.

Masood, Momina, Mariam Nawaz, Khalid Mahmood Malik, Ali Javed, Aun Irtaza y Hafiz Malik. 2023. "Deepfakes Generation and Detection: State-of-the-art, Open Challenges, Countermeasures, and Way Forward". *Applied Intelligence* 53: 3974-4026. https://doi.org/10.1007/s10489-022-03766-z

Microsoft Azure. 2023. "¿Qué es Speech Service?". 23 enero 2024. https://rb.gy/kyr1e

Murf AI. 2023. "Go from Text to Speech with a Versatile AI Voice Generator". Consultado 5 octubre 2023. https://murf.ai/

Natural Reader. 2023. "AI Text to Speech". Consultado 5 octubre 2023. https://www.naturalreaders.com/

Nekvinda, Tomáš, y Ondřej Dušek. 2020. "One Model, Many Languages: Meta-Learning for Multilingual Text to Speech". Ponencia presentada en INTERSPEECH 2020 en Shanghai, China: 2972-76. https://doi.org/10.48550/arXiv.2008.00768

Ning, Yishuang, Sheng He, Zhiyong Wu, Chunxiao Xing y Liang-Jie Zhang. 2019. "A Review of Deep Learning Based Speech Synthesis". *Applied Sciences* 9 (19): 1-16. https://doi.org/10.3390/app9194050

Noah, Ben, Arathi Sethumadhavan, Josh Lovejoy y David Mondello. 2021. "Public Perceptions Towards Synthetic Voice Technology". Proceedings of the Human Factors and Ergonomics Society Annual Meeting 65 (1): 1448-52. https://doi.org/10.1177/1071181321651128

Ong, Walter J. 1987. *Oralidad y Escritura*. Ciudad de México: Fondo de Cultura Económica.

Orozco Aguirre, Héctor Rafael, y Gonzalo Ivan Riego Caravantes. 2019. "Un tutor virtual inteligente para apoyar y asistir el proceso de enseñanza-aprendizaje en los primeros tres grados de educación primaria en México". *Pistas Educativas* (134): 524-41. http://hdl.handle.net/20.500.11799/106254

DOI: http://dx.doi.org/10.22201/iibi.24488321xe.2024.99.58856

Paddeu, Gavino, Andrea Devola, Andrea Ferrero y Antonio Pintori. 2019. "Interactive Audio-Text Guide for Museum Accessibility". Poster presentado en la 18th IADIS International Conference WWW/Internet 2019 en Cagliari, Italia, noviembre 2019. http://dx.doi.org/10.33965/icwi2019_201913P027

Paladines, Lenin, y Cristina Aliagas. 2023. "Literacy and Literary Learning on BookTube through the Lenses of Latina BookTubers". *Literacy* 57 (1): 17-27. https://doi.org/10.1111/lit.12310

Pesaru, Swetha, y Tilottama Goswami. 2021. "AI Based Assistance for Visually Impaired People Using TTS (Text to Speech)." *International Journal of Innovative Research in Science and Technology* 1 (1): 8-14. https://acortartu.link/2pew3

Rini, Regina. 2020. "Deepfakes and the Epistemic Backstop". *Philosophers' Imprint* 20 (24): 1-16. https://philpapers.org/archive/RINDAT.pdf

Rodero, Emma, e Ignacio Lucas. 2023. "Voces sintéticas versus voces humanas en audiolibros: el efecto de la intimidad emocional humana". *New Media and Society* 25 (7): 1746-64. https://doi.org/10.1177/14614448211024142

Ronda Pupo, Jorge Carlos, Niurka Cueto Rodríguez y María del Carmen Cougle Iglesias. 2020. "Dimensiones e indicadores para la evaluación de la comprensión auditiva en la práctica integral de la lengua inglesa". *Varona. Revista Científico Metodológica* 70: 98-102. http://scielo.sld.cu/pdf/vrcm/n70/1992-8238-vrcm-70-98.pdf

Sánchez, Jaime, y Héctor Flores. 2005. "AudioMath: Blind Children Learning Mathematics through Audio". *International Journal on Disability and Human Development* 4 (4): 311-16. https://doi.org/10.1515/IJDHD.2005.4.4.311

Sierra Berrocal, Ángel. 2022. "Adaptación de libros hablados digitales mediante síntesis de voz en el Servicio Bibliográfico de la ONCE". *RED Visual: Revista Especializada en Discapacidad Visual* 80: 106-26. https://hdl.handle.net/11162/242234

Taki, Sifat Ut, y Spyridon Mastorakis. 2023. "Rethinking Internet Communication Through LLMs: How Close Are We?". *Journal of Latex Class Files* 18 (9): 1-6. https://arxiv.org/pdf/2309.14247.pdf

Tan, Chia-Chen, Chih-Ming Chen y Hanh-Ming Lee. 2013. "Using a Paper-Based Digital Pen for Supporting English Courses in Regular Classrooms to Improve Reading Fluency". *International Journal of Humanities and Arts Computing* 7: 234-46. https://doi.org/10.3366/ijhac.2013.0073

Taylor, Paul. 2009. *Text to Speech Synthesis*. Cambridge: Cambridge University Press.

Vallorani, Cecilia María, e Isabel Gibert. 2022. "The Audiobook: The New Orality in the Digital Era. Visual Review". *International Visual Culture Review* 12 (2): 1-9. https://doi.org/10.37467/revvisual.v9.3734

Van den Oord, Aäron, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior y Koray Kavukcuoglu. 2016. "Wavenet: A Generative Model for Raw Audio". https://arxiv.org/pdf/1609.03499.pdf

Zavgorodniaia, Albina, Arto Hellas, Otto Seppälä y Juha Sorva. 2020. "Should Explanations of Program Code Use Audio, Text, or Both? A Replication Study". Artículo presentado en la 20th Koli Calling International Conference on Computing Education Research en Koli, Finlandia, 19-22 noviembre 2020. https://doi.org/10.1145/3428029.3428050

*Para citar este texto:*

Barragán-Perea, Efraín Alfredo, y Javier Tarango. 2024. "De la narración del audiolibro a la textualidad verbal y visual del audiotexto: una forma alternativa para la adquisición de conocimientos". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 38 (99): 13-33. http://dx.doi.org/10.22201/iibi.24488321xe.2024.99.58856

DOI: http://dx.doi.org/10.22201/iibi.24488321xe.2024.99.58856