



Revista Interamericana de Bibliotecología

ISSN: 0120-0976

ISSN: 2538-9866

Escuela Interamericana de Bibliotecología

Suárez Quiceno, Carlos; Castaño Muñoz, Wilson  
El audiotexto, una forma de oralidad terciaria y una experiencia alternativa de lectura\*  
Revista Interamericana de Bibliotecología, vol. 46, núm. 1, e02, 2023, Enero-Abril  
Escuela Interamericana de Bibliotecología

DOI: <https://doi.org/10.17533/udea.rib.v46n1e347133>

Disponible en: <https://www.redalyc.org/articulo.oa?id=179077513002>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en [redalyc.org](https://www.redalyc.org)

redalyc.org  
UAEM

Sistema de Información Científica Redalyc  
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

# El audiotexto, una forma de oralidad terciaria y una experiencia alternativa de lectura\*

## Resumen

A partir de la investigación “El recurso de conversión de texto a voz más allá de la asistencia”, se plantea que la lectura por medio de aplicaciones de síntesis de voz conlleva una doble relación con el texto: auditiva y visual, por lo que se propone denominarla *lectura audiotextual*. Se trata de un formato textual diferente del tradicional audiolibro, y que debe integrarse dentro de la categoría de oralidad terciaria. En la parte final del artículo se retoman algunos hallazgos de la investigación basados en las percepciones de los *lectores audiovisuales*, término que también se propone en el presente artículo. Estas percepciones aportan a la comprensión de las condiciones en que se emplea la conversión de texto a voz como una alternativa de experiencia lectora.

**Palabras clave:** audiotexto, audiolectura, oralidad terciaria, conversión de texto a voz, experiencia de lectura.

### **Carlos Suárez Quiceno**

Doctor en Diseño y creación, Universidad de Caldas. Magíster en Estética, Universidad Nacional. Licenciado en español y literatura, Universidad de Antioquia. Sociólogo, Universidad Autónoma Latinoamericana. Docente de tiempo completo de la Universidad Católica Luis Amigó. [carlos.suarezqu@amigo.edu.co](mailto:carlos.suarezqu@amigo.edu.co) <https://orcid.org/0000-0002-3578-9174>

### **Wilson Castaño Muñoz**

Magíster en Comunicación digital, Universidad de Antioquia. Especialista en Administración con énfasis en Calidad y Competitividad, Universidad ICESI. Coinvestigador, bibliotecólogo y profesor Escuela Interamericana de Bibliotecología, Universidad de Antioquia [wilson.castaño@udea.edu.co](mailto:wilson.castaño@udea.edu.co) <https://orcid.org/0000-0003-4937-2801>

**Cómo citar este artículo:** Suárez, Carlos; Castaño Wilson (2023). El audiotexto, una forma de oralidad terciaria y una experiencia alternativa de lectura. *Revista Interamericana de Bibliotecología*, 46(1), e347133. <https://doi.org/10.17533/udea.rib.v46nle347133>

Recibido: 2021-06-08 / Aceptado: 2022-18-08

\* Este artículo hace parte del proyecto de investigación La conversión de texto a voz, más allá de la asistencia. Fue ejecutado entre los grupos de investigación Urbanitas, de la Facultad de Comunicación, Publicidad y Diseño, de la Universidad Católica Luis Amigó; radicado 53005, Convocatoria interna de investigación, en conjunto con el grupo Información y Sociedad, de la Escuela Interamericana de Bibliotecología de la Universidad de Antioquia.

# The Audiotext, a Form of Tertiary Orality and an Alternative Reading Experience

## Abstract

Based on the research “The text-to-speech resource beyond assistance”, it is suggested that reading through voice synthesis applications entails a double relationship with the text: hearing and visual. This article proposes the term *audiotextual reading*, which is a textual format different from the traditional audiobook and it must be integrated into the category of tertiary orality. Finally, some research findings based on the perceptions of the audiences are considered. The term *audioreaders* is also suggested in this article. These insights contribute to the understanding of the conditions in which text-to-speech is used as an alternative form of reading experience.

**Keywords:** Audiotext, audioreading, tertiary orality, text to speech, reading experience.

## 1. Introducción

El primer paso de la investigación fue revisar el concepto de *lectura* en relación con el cambio del medio visual al auditivo. Al examinar etimológicamente la palabra *leer* y sus derivaciones, se encontró que está relacionada morfológicamente con términos como *escoger*, *seleccionar*, *elegir*, *elector*. De tal modo, leer implica acceder a la información por medio de un ejercicio de selección de signos. Nótese que esto no implica que los signos deban ser gráficos, leer no es ver. Pero, si el resultado final de la lectura visual es obtener información, también vale preguntarse por la información que ingresa al cerebro a través de la audición. ¿Qué sucede cuando se escuchan los textos? De hecho, la información a través del oído se procesa, por lo general, mucho más rápido que la que ingresa a través de la lectura visual (Gramajo et al., 2017), porque al escuchar no hay necesidad de realizar la decodificación de los signos a través de técnicas como la subvocalización, mientras que, al leer de forma visual, deben representarse los fonemas mentalmente.

Cuando la electrónica analógica hizo posible transmitir la voz y se crearon medios de comunicación como la radio y la televisión, la palabra hablada ganó un nuevo

espacio en la sociedad. Ong (1982) afirmó que a través de estos medios electrónicos se dio un resurgimiento de la oralidad. Pero ya no se trata de una oralidad primaria, la que se vivía antes de la escritura, sino de una oralidad secundaria. Fue también gracias a este tipo de medios que los individuos volvieron a hacer uso masivo de la escucha como instrumento para el entretenimiento y la adquisición de información y conocimiento.

Más adelante, a partir de los avances de la electrónica digital, que permitió el desarrollo de dispositivos de sonido portátiles con gran capacidad de almacenamiento, gestión de contenidos y autonomía, surgieron nuevas experiencias de escucha que reafirmaron este tipo de oralidad. Luego, la expansión de internet permitió la aparición de plataformas como Ivoox, SounCloud, Youtube, Vimeo, Daily Motion, claro ejemplo del protagonismo que gana de nuevo la palabra hablada. El fenómeno de los podcasts y los *booktubers* como influyentes, que han ido migrando sus clubes de lectura a experiencias audiovisuales (García-Canclini, et al., 2015), demuestra que el público está dispuesto a escuchar.

Desde hace poco, ha llegado a ser posible la experiencia de leer en dispositivos electrónicos por medio de programas de síntesis de voz. Esta nueva práctica bien puede ubicarse dentro de una oralidad terciaria, siguiendo la derivación del concepto de *oralidad* empleada por Ángel-Botero y Alvarado-Duque (2016), basada en Logan (2010), cuando hablan de la experiencia de escuchar radio digital por medio de dispositivos móviles. No obstante, la primera mención al concepto de *oralidad terciaria* debe buscarse en las reflexiones acerca de la lingüística computacional presentadas por Joseba Abaitua (1997) en el Simposio de Lingüística Española organizado por el Instituto Cervantes de Mánchester y el Centre for Computational Linguistics de UMIST en mayo de 1995 y luego retomadas en 1997 en el Congreso de Informática y Sociedad: retos para el nuevo milenio, en la Universidad de Deusto. En esa ocasión, Joseba planteó la oralidad terciaria partiendo de los conceptos de *oralidad secundaria* de Ong (1982) y del concepto de *ideografía dinámica* elaborado por Pierre Levy (1998). Es una visión en la que la comunicación oral por medio de la tecnología adquiere otras dimensiones.

Ambas derivaciones se fundamentan en la distinción de Ong (1982) entre oralidad primaria, la que se produce en sociedades que no hacen uso de la escritura, y oralidad secundaria, la que está influenciada por el conocimiento de los sistemas de escritura. Aunque Ong (1982) no lo planteó, a estos tipos de oralidad se les puede sumar la oralidad terciaria, la que se produce por intervención de la tecnología digital.

En un sentido similar, la investigación “El recurso de conversión de texto a voz” planteó que la expansión de la oralidad, que ocurre cuando se emplea dicho recurso, cabe dentro de esa nueva forma de oralidad, y se encontró que es una alternativa válida de lectura en un ecosistema de medios digitales en el que se posibilitan otras prácticas y experiencias lectoras.

## 2. Reflexión

### 2.1 Las tecnologías digitales como generadoras de una nueva oralidad

Con el advenimiento de la era digital y la amplia disponibilidad de equipos de cómputo personales, la tecnología de síntesis de voz estuvo al alcance de cualquier usuario. Los sistemas operativos y el software especializado hicieron posible la conversión automática y la creación de archivos de audio sobre cualquier texto digital. Esta función, denominada *conversión de texto a voz* (CTV) y en inglés *text to speech* (TTS), pertenece a un nuevo fenómeno de oralidad que plantea una especial incidencia en las prácticas lectoras.

La CTV es una nueva forma de oralidad tecnológica digital. Tiene antecedentes remotos, pero solo existe en propiedad desde el año 1968, cuando Noriko Umeda, en Japón (Klat, 1987), desarrolló el primer programa de síntesis de voz aplicado a texto en inglés. Hacia los años ochenta, ya se había alcanzado un nivel funcional para el empleo de esta herramienta; pero habría que esperar hasta las primeras décadas del siglo XXI para que se lograra un buen nivel en las dos condiciones básicas de la síntesis de habla: la inteligibilidad y la naturalidad. Los primeros esfuerzos para realizar síntesis de voz se enfocaron en lograr la inteligibilidad del habla. Cuando esto se consiguió, la nueva meta fue alcanzar la naturalidad de las voces. Actualmente, se produce síntesis de voz con gran naturalidad e inteli-

gibilidad. El nuevo reto está puesto sobre una tercera condición: la variabilidad o expresividad de las voces y otros aspectos prosódicos y pragmáticos del discurso (Bulut et al., 2002).

Entre los primeros beneficiados con los avances de la tecnología de síntesis de voz estuvo la población invidente, que por sus mismas condiciones y limitaciones no había podido acceder a la información escrita de la misma manera que el resto de la población. La función inicial de este tipo de recurso fue sobre todo brindar asistencia y mayor autonomía en el acceso a la información para este tipo de usuario; pero ya ha quedado atrás la idea generalizada de que los programas de síntesis de voz solo se usan para personas en situación de discapacidad. Las aplicaciones domóticas, la navegación terrestre con equipos GPS, los asistentes virtuales y la conversión de texto a voz en programas informáticos de lectura no dejan duda acerca del amplio campo de acción de este recurso.

Para que exista oralidad terciaria, en el caso de la oralidad tecnológica digital, y en especial en la CTV, se requiere un cierto nivel de manipulación del texto y el audio, que los convierta en un producto nuevo, no solo en una reproducción.

### 2.2 El audiotexto, una textualidad verbal y visual

En la primera década de este siglo, se hizo posible que cualquier persona con suficiente conocimiento y habilidad convirtiera un archivo de texto digital en un archivo de voz sintetizada. El archivo obtenido se podía almacenar en un reproductor de música. En la investigación se constató que esos primeros usuarios encontraron en la síntesis de voz la oportunidad de crear sus propios audiolibros de manera automática. Estos usuarios pioneros partían del empleo de algún tipo de software dedicado para obtener un archivo de audio que podían ejecutar en un reproductor de música. Lo que lograron hacer corresponde más al concepto de *audiolibro* que al de *audiotexto*.

Por esta misma época, los avances de la lingüística computacional, que habían hecho posible la reproducción oral de cualquier texto digital por medio de voces sintéticas, se integraron a los programas informáticos de lectura y abrieron paso a un nuevo producto

de audio que aquí se porpone denominar *audiotexto*, para diferenciarlo de los ya conocidos audiolibros. El concepto de *audiotexto* ha sido explorado desde diferentes ámbitos, como aquellos que se enfocan en la interacción hombre-máquina a través del ingreso de información por medio de dispositivos como el teléfono y las respuestas auditivas del sistema (LaRose y Atkin, 1992; Schwartz y Schwab, 1993; DeGroot y Schwab, 1993). Por otra parte, el *audiotexto* también es empleado para la enseñanza de idiomas, ya sea para hablantes nativos o extranjeros (Al Ghafli, 2011; Jones et al., 2008; Mariño-Medina, 2017) como medio para que los aprendices puedan mejorar la pronunciación y velocidad de lectura. Otras áreas que se enfocan en el *audiotexto* son aquellas vinculadas a la accesibilidad de personas con discapacidad visual (Gramajo et al., 2017). Este último estudio, en particular, define específicamente el *audiotexto* como

un texto grabado en voz alta, generalmente en un formato electrónico, que puede reproducirse innumerables veces. La grabación del texto puede ser leída o dramatizada por actores o aficionados, pero también se genera de manera artificial utilizando un software de sonido que permite lograr versiones completas o abreviadas. (Gramajo et al., 2017, p. 126)

No obstante, esta definición no señala la importancia que tiene la presencia visual del texto como soporte y guía de la lectura auditiva, que aquí se considera condición necesaria para que pueda hablarse de *audiotexto*.

Otros autores han puntualizado al *audiotexto* como un recurso enfocado al usuario vidente que tenga el propósito de incrementar su consumo lector o mejorar su comprensión. La definición de *audiotexto* más aproximada a la que se construyó en la investigación se encuentra en Bjork y Rumrich (2018), para quienes el *audiotexto* aparece como recurso multimodal: “An audiotext is multimodal in two senses: as its name implies, it mediates the text in aural and visual modes, enabling the user to read and/or listen; in addition, it allows the reader to choose between va-

rious reading modes”<sup>1</sup> (p. 49). De esta manera, tanto la visión como la audición cobran gran importancia, al igual que la posibilidad de que el lector pueda elegir entre cualquiera de estos sentidos, o los dos simultáneamente, para consumir contenidos. Así, la audiolectura por el medio visual y auditivo ha de ser una experiencia diferente, pero comparable y en cierto modo complementaria a la convencional forma visual de los textos impresos o a la auditiva de los audiolibros.

Bjork y Rumrich (2018) desarrollaron una versión de *audiotexto* para la emblemática obra de Milton, *Lost Paradise* —la misma que en 1965 IBM había codificado en binario usando tarjetas perforadas—, con el objetivo de que su prosa pudiera ser leída y escuchada al mismo tiempo, ya que ellos consideran que la escucha es tan importante como la lectura en la comprensión de poemas. Una idea que está en consonancia con lo que propone Gramajo et al. (2017) cuando afirma que “algunos teóricos aseguran que la comprensión es mayor a través de los oídos” (p. 126).

En este sentido, la reflexión que plantea este artículo encuentra una confirmación y se sostiene en denominar *audiotexto* al flujo automático de la conversión de texto a voz que discurre paralelo a la visualización del texto en pantalla, lo que permite al audiolector, si es de su interés, identificar en todo momento el texto que está siendo verbalizado, de tal modo que puede ubicar sus propiedades tipográficas y editoriales. Esta propiedad del *audiotexto* lo hace idóneo para ejercicios de lectura intensiva, en los que la comprensión exige considerar elementos discriminantes y paratextuales que requieren confirmación visual.

El *audiotexto* se genera en la conversión automática de texto a voz efectuada por medio de software especializado, sumando a esto una condición esencial: el texto continúa siempre visible, aunque no se esté leyendo visualmente. Como tal, el *audiotexto* permite una forma combinada de lectura visual y auditiva. Aunque puede considerarse como una variante de los

1 Un *audiotexto* es multimodal en dos sentidos: como su nombre lo indica, media el texto en modos auditivos y visuales, permitiendo al usuario leer o escuchar; además, permite al lector elegir entre varios modos de lectura (traducción propia).

audiolibros, sus rasgos distintivos y la experiencia de usuario lo diferencian notablemente de aquellos.

En la CTV intervienen voces sintéticas, software lector con la función texto a voz y motores de síntesis de voz, que permiten que el texto escrito sea reproducido automáticamente en forma oral y personalizable, mientras que ofrece la posibilidad de visualizar en todo momento el texto que está siendo escuchado. Por consiguiente, en este sentido, una grabación a partir de la función de texto a voz, como aquellas que realizaban los usuarios pioneros del recurso de conversión, con el propósito de escucharla luego en un reproductor de audio, es más un audiolibro que un audiotexto. Para que se pueda hablar de este último, en la experiencia de usuario debe ser posible interactuar eventualmente con el texto visible y audible sincrónicamente. Por otra parte, el audiotexto posee gran navegabilidad dentro de sus contenidos y es prácticamente ilimitado en la disponibilidad de los contenidos.

Por medio del recurso CTV, los textos digitales no solo se reproducen en las pantallas de los equipos electrónicos, sino en sus altavoces o audífonos, sin necesidad de tener que pasar por cabinas de grabación de audio, ni locutores profesionales ni cadenas comerciales. Así entonces, se hace común la CTV, aplicable a escritos de cualquier extensión, en especial para lecturas lineales prolongadas, realizadas por lectores a quienes se les denomina *audiolectores*, pues están consumiendo información a través de audiotextos.

El recurso CTV ofrece una alternativa para el consumo textual y beneficia no solo a quienes presentan alguna discapacidad de tipo visual, sino a todos aquellos que desarrollan la habilidad de configurar los dispositivos para obtener el mejor provecho de esta función y tienen la disposición de adaptarse al medio verbal generado por los programas de síntesis de voz.

La lectura bajo conversión de texto a voz debe ser vista como una forma alternativa de leer, pero, al fin y al cabo, lectura. Así lo expresaba uno de los colaboradores de la investigación: “Leo mucho, siempre estoy leyendo o escuchando, que para mí es otra forma de leer” (D. Fernández, comunicación personal, 20 de septiembre de 2018). Asimismo, la conversión de texto a voz presenta ventajas notables frente al audiolibro,

así lo expresaron otros usuarios, al manifestar que sus lecturas muchas veces son multidispositivo, de modo que, por ejemplo, empiezan a leer una novela en un *e-reader* u otro dispositivo como celular y luego, cuando no pueden leer el texto visible, emplean el móvil para reproducir el audio a través de las aplicaciones de lectura que integran la función CTV.

### 2.3 Diferencias entre el audiolibro y el audiotexto

La experiencia de escuchar textos ha sido más frecuente de lo que suele suponerse. Solo en los últimos siglos se volvió común la lectura silenciosa, pero antes la lectura en voz alta fue la manera imperante. “En el mundo clásico, en la Edad Media, y hasta los siglos XVI y XVII, la lectura implícita, pero efectiva, de numerosos textos es una oralización, y sus 'lectores' son los oyentes de una voz lectora (Cavallo y Chartier, 2004, p. 20).

En una época todavía reciente en la historia, la electrónica analógica logró la reproducción, almacenamiento y transmisión de contenidos orales. Estos factores hicieron posible la creación de una oralidad de tipo tecnológico, en la que la voz tomó un nuevo valor: empezó a apreciarse por su registro y por la técnica de locución. En ese contexto tecnológico surgió el audiolibro como un producto destinado a la población con dificultades para acceder a los textos. Fue Edison, el inventor del fonógrafo, quien primero grabó en rollos de cera una pieza literaria, pequeños versos de una canción popular en 1877. No le fue posible grabar textos de mayor extensión. Tendrían que pasar varias décadas para que esto fuera posible, pero, al fin, el sueño de Edison y de Uzanne (1894) se hizo posible.

Los audiolibros son productos culturales que siguen teniendo aceptación entre el público lector. Se han comercializado ampliamente durante varias décadas y se han adaptado a los cambios en el ecosistema tecnológico. Desde un punto de vista comunicacional, el medio en los audiolibros sigue siendo el texto, la palabra escrita; pero se trata del texto oralizado, ejecutado en su materialidad fonética, a veces teatralizado, pero siempre subordinado íntegramente a la palabra escrita.

Básicamente, el audiolibro es un recurso que está elaborado específicamente como oralidad secunda-

ria, cuenta con narradores profesionales y entornos ambientados con diseños sonoros que buscan generar una mayor compenetración entre el receptor y el audiolibro. Los audiolibros se diseñan para cierto tipo de contenidos, que generalmente son novelas o relatos. Sin embargo, también existen otro tipo de audiolibros, con menos sofisticación que los anteriores, y son, sobre todo, los que se narran por un locutor ya sea profesional o no, o inclusive por su mismo autor. Un ejemplo de esto son plataformas como Audible de Amazon, que ofrece la posibilidad de adquirir el audiolibro narrado por el autor o un locutor profesional y cuyos costos son mucho más económicos que los del texto digital.

Una de las características de la mayoría de los audiolibros es su precaria navegabilidad, porque, al ser archivos en formato de audio, el usuario no puede ubicar la información fácilmente. También presentan muchas limitaciones en la interoperabilidad, es decir, en el paso de un dispositivo o sistema a otro, lo que afecta la disponibilidad del recurso.

Por su parte, el audiotexto, como se mencionó previamente, implica programas que se encuentran instalados dentro de los dispositivos, que permiten la lectura en voz alta de cualquier texto, con la posibilidad de seleccionar diferentes tipos de voces, mucho más naturales de lo que eran en el pasado, al punto de ser consideradas como amenaza por parte de la industria editorial, que percibe que la calidad de estas voces dentro de los dispositivos como Kindle de Amazon son lo suficientemente competitivas para amenazar la industria del audiolibro. Por esta razón, le solicitaron a Amazon que retirara la función dentro de sus Kindle (Beaumont, 2009).

Los audiotextos le permiten al lector leer de manera visual y auditiva en el dispositivo móvil o elegir entre cualquiera de ellas. Precisamente esta es una de las grandes ventajas de esta tecnología, que puede resultar más inclusiva y superior en prestaciones y disponibilidad que los audiolibros convencionales.

#### 2.4 Síntesis de voz en la experiencia lectora

El teléfono móvil es el dispositivo más empleado para la lectura por medio de voces sintéticas, de modo que

la denominada *cuarta pantalla* va ganando espacio en esta otra esfera de la comunicación. Se debe entender que la conversión se efectúa de modo automático, por lo tanto, al usuario solo le basta activar la función cuando usa los programas de lectura o interfaces con la opción de lectura en voz alta.

Existen diversos programas informáticos de lectura que admiten la función de conversión de texto a voz, así como variedad de voces sintéticas; pero a los usuarios, por lo general, les basta con encontrar un programa y una voz con la que logren adecuarse suficientemente. En la investigación se observó que los usuarios avanzados tienen la tendencia a emplear el programa y la voz con la que han alcanzado suficiente familiaridad.

Las voces sintéticas pueden ser inteligibles sin ser naturales ni expresivas. De hecho, la naturalidad y expresividad puede sacrificar en algo la inteligibilidad, por eso, no siempre se puede juzgar una voz por el mismo criterio. Por ejemplo, si se quiere escuchar un texto en alta velocidad, es preferible una voz inteligible antes que natural.

Las tecnologías digitales, en la medida en que alcanzan mayor perfeccionamiento, logran crear productos indistinguibles de los del mundo real. De hecho, Google ha anunciado recientemente, y ha puesto en demostración, el sistema Tacotron 2, un desarrollo de voces sintéticas que para el oyente poco o nada se diferencian de las voces reales (Shen et al., 2018). Lo más probable es que en poco tiempo se podrá superar el calificativo de *mecánicas* o *robóticas* que suele darse a estas voces y que impone un juicio estético que obstaculiza su aprovechamiento. No obstante, al observar con más detalle las posibilidades de este recurso en su estado actual de desarrollo, resulta claro su enorme potencial para la lectura de textos digitales.

Al hablar de las ventajas y bondades que ofrecen este tipo de tecnologías para el consumo de contenidos textuales, se encuentran las siguientes, extraídas de la investigación:

- Posibilidad de acceder a información cuando no se puede leer visualmente. Sin duda, la característica más importante que ejercen estas tecnologías

es la posibilidad de prolongar el acceso a la información, aunque se estén realizando diferentes tipos de actividades operativas, como oficios domésticos, conducir, caminar, entre otras. De esta manera, las personas admiten que con el uso de estas tecnologías leen mucho más que cuando se encontraban limitados únicamente al texto impreso.

- Posibilidad de lectura audiotextual: los programas lectores que ofrecen la funcionalidad de lectura en voz alta le permiten al usuario leer en el dispositivo móvil de manera visual, pero, cuando lo deseen, pueden alternar o escuchar simultáneamente el texto.
- Disponibilidad de lectura de cualquier tipo de contenido en formato digital: aquí yace una de las diferencias más grandes con respecto a los audiolibros y es que estos últimos dependerán de la actividad del editor que produzca el contenido y lo adapte al formato auditivo. Por lo tanto, la disponibilidad será bastante reducida al momento de escoger los audiolibros que se deseen leer, ya que no se podrá seleccionar cualquier tipo de material, solo el adaptado para audiolibro, mientras que los programas que cuentan con la función de síntesis de voz están en la capacidad de leer casi cualquier tipo de formato textual digital, desde el tradicional PDF hasta documentos ofimáticos y páginas web. Es más, algunas aplicaciones ofrecen la posibilidad de reproducir listas de textos entre artículos, libros o páginas web, de manera que su reproducción sea continua hasta terminar los textos seleccionados.
- Personalización de la experiencia: los sintetizadores de voz cuentan con la posibilidad de personalizar la experiencia auditiva mediante la modificación de funciones como el cambio del mismo sintetizador de voz, cambio de voces, ajuste de la entonación, ajuste de la velocidad y temporizador para apagar el reproductor en caso de somnolencia.
- Creación de archivos en formato de audio: algunos lectores ofrecen la posibilidad de convertir el texto en un archivo de audio en formato MP3 que

puede ser extraído del dispositivo y compartido con otros usuarios.

- Navegabilidad: otra de las grandes diferencias de estos recursos con respecto a los audiolibros está en las prestaciones relacionadas con la navegación y el control de pantalla.
- Recuperación del punto de lectura: la gran mayoría de programas lectores y plataformas de audiolibros permiten retomar la escucha del texto en la parte en la que se dejó. Incluso guardan la ubicación de varios textos de manera automática y simultánea, así que se puede alternar entre diferentes lecturas.

## 2.5 Conversión de texto a voz y experiencias de lectura

La investigación que motiva este artículo se ocupó, entre otros temas, de la pregunta sobre cómo pueden integrarse las aplicaciones de síntesis de voz para mejorar los niveles de acceso a la lectura en la enseñanza universitaria.

Al indagar a los entrevistados acerca de las percepciones sobre la lectura auditiva, todos estuvieron de acuerdo en que escuchar un texto se puede considerar un proceso de lectura, puesto que el objetivo de ambas prácticas radica en la adquisición de información verbal, aunque sea a través de diferentes sentidos. En relación con la comprensión del contenido en las pruebas de lectura realizadas, los participantes consideraron que no tenían necesidad de leer visualmente el texto ya escuchado.

La cuestión clave de la audiolectura radica en los momentos destinados a la adquisición de la información, ya que, cuando se lee visualmente, se debe estar inmerso de lleno en esta actividad, mientras que cuando se escucha, se pueden realizar diferentes tipos de actividades. Obviamente, estas actividades deberán ser altamente operativas, para que no comprometan significativamente la comprensión y la concentración.

Pero la síntesis de voz es una tecnología que aún no cuenta con un alto grado de apropiación. Se hace necesario cumplir ciertas condiciones para que logre

su aceptación mayoritaria, como se observó dentro del público estudiado, estudiantes universitarios en Medellín, Colombia. Se contó con la participación de los estudiantes de Bibliotecología del primer semestre del curso de Informática y TIC. Los participantes fueron 35 estudiantes en el primer semestre y 33 en el segundo. Sus rangos de edades, al ser estudiantes que finalizaron hace poco su bachillerato, oscila entre los 16-19 años en promedio. Los hombres fueron un total del 28 % mientras que las mujeres fueron un 62 %. El contacto se realizó directamente desde la clase y la muestra fue un trabajo que debían presentar como parte de sus actividades académicas. Los datos fueron recogidos por medio de dos instrumentos: el primero fue el cuestionario que diligenció la totalidad de los participantes por medio de la herramienta Google Forms y el segundo instrumento fue el trabajo presentado en el marco del curso ya mencionado, en el que expresaron su experiencia con la herramienta TTS seleccionada para el curso, que fue la aplicación At@Voice. Los estudiantes debían informar sus sensaciones y opiniones acerca de la concentración, los tiempos de lectura, el número de libros, documentos o artículos que escucharon, cómo les parecían las voces, qué tanto ruido les generaron para la comprensión, cómo consideraron la concentración realizando otras actividades y se les indicó que preferiblemente escucharán mientras hacían otras actividades. Por último, se les preguntó si consideraban que estas aplicaciones les ayudaban a consumir más contenido y más información.

Algunas de las razones por las cuales este estudio consideró que la síntesis de voz no es una tecnología intuitiva se exponen a continuación.

#### 2.5.1 Adaptación a las voces

Una de las características que por tradición ha marcado la transmisión de la información es la oralidad del lenguaje, lo cual tiene una implicación psicológica: el individuo se familiariza con determinadas formas dialectales del habla desde su infancia y asume los acentos típicos de los lugares geográficos donde ha crecido.

Por lo tanto, está más familiarizado y es más susceptible a aceptar y comprender la transmisión de

información oral que cuente con las mismas características de entonación, prosodia y acento de su entorno familiar, sobre la de cualquiera otra variación.

Un claro ejemplo de esto se puede observar en el idioma castellano y sus diferentes variantes, como el castellano latino y el castellano peninsular de España. Ciertamente, se presenta un obstáculo cuando los hablantes de alguna de estas regiones son expuestos a mensajes en otra variación dialectal. Esto se observa frente a las películas dobladas al castellano, las cuales generan una marcada preferencia o rechazo por parte de las comunidades receptoras, dependiendo de su adscripción lingüística.

#### 2.5.2 Género y diversidad de las voces

Otra de las características que influyen en la recepción de la información verbal radica en el género de las voces. Estas preferencias obedecen más a cuestiones de atracción. Diferentes estudios han demostrado que los hombres son propensos a prestar mayor atención cuando identifican una voz femenina que se encuentre en un rango similar al de su edad y pueda ser sujeto potencial de un vínculo afectivo. De esta manera, el género de las voces será relevante para atraer y mantener la atención sobre la información escuchada. La característica de este recurso es la inmediatez y universalidad en la reproducción del habla.

- Integrar nuevos hábitos de lectura

La consolidación de hábitos lectores desde la infancia, ya sea mediante el ejemplo o la enseñanza, es un factor decisivo para que muchos lectores estén en capacidad de experimentar en nuevos tipos de formatos y soportes.

Un lector formado aprende a valorar el texto en sí mismo y llega a un punto en el que le da mayor trascendencia al contenido textual que al soporte con el que ha estado habituado. Un buen lector lo es en cualquier formato.

- Alfabetización digital

El uso de los sintetizadores de voz para la lectura requiere mayor promoción y empleo por fuera de la población con dificultades visuales. Esta tecnología que se desarrolló, inicialmente, pensando en ofrecer

accesibilidad, también es ideal para generar una experiencia alternativa y complementaria para todo tipo de lector.

- Entrenamiento técnico

La sola referencia acerca de la existencia del recurso CTV no es suficiente para su aprovechamiento, también se hace necesario conocer los diferentes procedimientos, aplicaciones y programas por medio de un entrenamiento adecuado.

### 3. Conclusiones

La distinción entre audiotexto y audiolibro tiene una función práctica que orienta al lector para identificar las características de cada formato y cuál es el que le viene mejor para sus condiciones particulares. Ambos son parte de las múltiples formas que asume la comunicación escrita, que incorpora en todo momento el desarrollo de la tecnología.

La lectura auditiva tradicional es una práctica que permite acceder al contenido de un texto de forma equivalente a la lectura visual. Sin embargo, mientras que en la lectura visual se está leyendo en dos planos, el espacial de la página y el temporal del discurso escrito, en la auditiva prima la dimensión temporal. Ese rasgo puede hacer que la experiencia de escuchar no solo sea más cómoda, sino que parezca más inmediata e inmersiva, con una percepción más fluida del texto; mientras que el lector visual puede lograr una comprensión más analítica, en la medida en que contempla los elementos paratextuales y los discriminantes presentes en el texto. Ahora bien, el lector audiovisual, aquel que empleando el recurso CTV escucha y visualiza el texto, puede tener una experiencia complementaria, en la que aprovecha ambos medios.

Estas posibilidades de alternar en la lectura que brinda la síntesis de voz en programas informáticos de lectura podrán ser una ventaja dependiendo del tipo de texto al que se enfrente el lector, de las condiciones externas a las que esté expuesto y del conocimiento que espere de la obra, así como de las estrategias que emplee. Es notable la conducta de los lectores que mantienen simultáneamente ambos modos en un mismo texto, lo que les permite orientarse visualmente en el desarrollo

del texto y los paratextos, con el propósito de lograr mayor comprensión del texto. En todo caso, puede hablarse de lectura audiovisual cuando el usuario conserva una referencia visual con el texto, así no siga plenamente su desarrollo visual.

Uno de los usos más frecuentes del sistema CTV se presenta cuando el usuario debe tener los ojos enfocados fuera del texto, cuando está realizando otra actividad o no le resulta conveniente realizar la lectura visual. Aun así, prefieren la síntesis de voz sobre el audiolibro, porque el software lector, además de tener la funcionalidad de leer en voz alta, ofrece la posibilidad de que el usuario lea el texto en un formato redimensionable. Las interacciones entre ambos tipos de lectura no son del todo posibles con los audiolibros, ya que la mayoría de estos limitan sus funciones de navegabilidad, y hacen difícil establecer la ubicación precisa dentro de los textos.

En la época actual, el énfasis, si bien sigue estando en el desarrollo de mejores programas de síntesis de voz, se hace también en la experiencia de usuario para llamar la atención del posible lector sobre la forma de consumir textos de una manera alternativa. Los sistemas operativos presentan diversidad de aplicaciones para conversión de texto a voz. De tal modo, puede verse que es un recurso muy extendido y al alcance de quien lo desee. Este punto es relevante por cuanto en el campo del desarrollo de software y de contenidos la no utilización de ellos conlleva una forma de obsolescencia.

Una de las principales razones para no usar los recursos CTV por los públicos masivos es su falta de conocimiento y alfabetización digital. En pruebas realizadas con estudiantes universitarios y lectores en general, el 100 % afirmó no conocer la existencia de este recurso, aunque el 100 % de ellos contaba con dispositivos móviles. Por lo tanto, se requiere una gran educación y entrenamiento en diferentes fases de la vida educativa de los individuos para exponerlos a este tipo de tecnologías y darles la oportunidad de que también puedan usar los recursos CTV con el fin de consumir más contenidos en momentos en los que no se pueda o no se desee emplear la visión.

Tomar en consideración el audiotexto como un nuevo formato textual y establecer sus características y uso,

tal como se planteó, contribuirá a integrar de modo más cercano la lectura mediada por dispositivos electrónicos en el conjunto de las prácticas lectoras.

## 4. Referencias

1. Abaitua, Joseba (1997). *Ratones en la biblioteca digital. I Jornada sobre Informática y Sociedad: Retos para el nuevo milenio*. Facultad de Filosofía y Letras, Universidad de Deusto.
2. Al Ghafli, Mansour (2011). *The effect of mediated glosses on vocabulary retention and reading comprehension with English language learners in Saudi Arabia*. University of Kansas.
3. Ángel-Botero, Adriana; Alvarado-Duque, Carlos (2016). Oralidad terciaria: mirada ecológica a la radio digital. *Palabra Clave*, 19(2), 473-500. <https://doi.org/10.5294/pacla.2016.19.2.6>
4. Beaumont, Claudine (2 de marzo, 2009). Amazon backtracks over Kindle's 'read aloud' function. *Telegraph*, 1. <https://www.telegraph.co.uk/technology/news/4927034/Amazon-backtracks-over-Kindles-read-aloud-function.html>
5. Bjork, Olin; Rumrich, John (2018). "Is There a Class in This Audiotext?" Paradise Lost and the Multimodal Social Edition. En D. Currell; I. Issa (Eds.), *Digital Milton* (pp. 47-76). Springer International Publishing. [https://doi.org/10.1007/978-3-319-90478-8\\_3](https://doi.org/10.1007/978-3-319-90478-8_3)
6. Bulut, Murtaza; Narayanan, Shrikanth; Syrdal, Ann (2002, 16-20 de septiembre). Expressive speech synthesis using a concatenative synthesizer. *7th International Conference on Spoken Language Processing, ICSLP2002 - INTERSPEECH 2002*. Denver, Colorado, EE. UU.
7. Cavallo, Guglielmo; Chartier, Roger (2004) Introducción. En G. Cavallo; R. Chartier (Eds), *Historia de la lectura en el mundo occidental* (2<sup>a</sup> ed.) (pp. 15-67). Taurus.
8. DeGroot, Jenny; Schwab, Eileen (1993). Understanding time-compressed speech: The effects of age and native language on the perception of audiotext and menus. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 37(3), 244-248. <https://doi.org/10.1177/154193129303700307>
9. García-Canclini, Néstor; Gerber-Bicecci, Verónica; López-Ojeda, Andrés (2015). *Hacia una antropología de los lectores*. Paidós.
10. Gramajo, María Cecilia; Santagada, Miguel; Paoletta, Anabel (2017). Una audioteca en la Unicen. *La Escalera - Anuario de la Facultad de Arte*, 27, 14. <http://www.ojs.arte.unicen.edu.ar/index.php/laescalera/article/view/567>
11. Jones, Greg; Squires, Todd; Hicks, Jeramie (2008). Combining speech recognition/natural language processing with 3D online learning environments to create distributed authentic and situated spoken language learning. *Journal of Educational Technology Systems*, 36(4), 375-392. <https://doi.org/10.2190/ET.36.4.c>
12. Klatt, Dennis (1987). Review of text-to-speech conversion for English. *The Journal of the Acoustical Society of America*, 82(3), 737-793.
13. LaRose, Robert; Atkin, Davis (1992). Audiotext and the re-invention of the telephone as a mass medium. *Journalism Quarterly*, 69(2), 413-421. <https://doi.org/10.1177/107769909206900215>
14. Levy, Pierre (1998). *A ideografia Dinâmica. Rumo a uma Imaginação Artificial?* Edições Loyola.
15. Logan, Robert (2010). *Understanding new media: extending Marshall McLuhan*. Peter Lang.
16. Mariño-Medina, Martha; Álvarez-Amargos, Michelle; Reyna-Suárez, Zaida (2017). Tendencias actuales para el desarrollo de la comprensión auditiva en el proceso de enseñanza de español como lengua extranjera: Coordenadas para su análisis. *Revista Científico-Educacional de la Provincia Gramma*, 13(3), 200-210.
17. Ong, Walter (1982). *Oralidad y escritura. Tecnologías de la palabra*. Fondo de Cultura Económica.
18. Schwartz, Amy; & Schwab, Eileen (1993). Improving menu design for the rapid order audiotext system using cluster analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 37(3), 230-234. <https://doi.org/10.1177/154193129303700304>
19. Shen Jonathan; Pang, Ruoming; Weiss, Ron; Schuster, Mike; Jaityl, Navdeep; Yang, Zongheng; Chen, Zhifeng; Zhang, Yu; Wang, Yuxua; Skerry-Ryan, RJ; Saurous, Rif; Agiomyrgiannakis, Yannis; Wu, Yonghui (2018) Natural TTS synthesis by conditioning wavenet on mel spectrogram predictions. *Arxiv*. <https://arxiv.org/abs/1712.05884>
20. Uzanne, Octave (1894). The end of Books. *Scribner's Magazine*, 16, 221-231. <https://publicdomainreview.org/collection/octave-uzanne-the-end-of-books-1894>