

Treball de Fi de Màster

Títol

Los asistentes de voz, nuevos canales de distribución y consumo de información y entretenimiento. Los casos de la CCMA y la BBC.

Autoria

Raquel Liger Belmonte

Tutoritzat per

Lluís Reales

Màster

Mitjans, Comunicació i Cultura

Departament

Departament de Mitjans, Comunicació i Cultura

Data

03/07/20

Full resum del TFM

Títol del Treball Fi de Màster:

Català: Els assistents de veu: nous canals de distribució i consum d'informació i entreteniment. Els casos de la CCMA i la BBC.

Castellà: Los asistentes de voz: nuevos canales de distribución y consumo de información y entretenimiento. Los casos de la CCMA y la BBC.

Anglès: Voice Assistants: new channels for distribution and consumption of information and entertainment. The cases of the CCMA and the BBC.

Autoria: **Raquel Liger Belmonte**

Tutoritzat per: **Lluís Reales**

Edició: 2019/20 **Màster:** Mitjans, Comunicació i Cultura

Paraules clau (mínim 3)

Català: asistentes de veu, mitjans de comunicació, interfícies vocals, veu, era post-pantalla, HMC

Castellà: asistentes de voz, medios de comunicación, interfaces vocales, voz, era post-pantalla, HMC

Anglès: voice assistants, media, vocal interfaces, voice, post-screen era, Human-Machine Communication

Resum del Treball Fi de Màster (extensió màxima 600 caràcters)

Català: El present treball ofereix un estudi d'investigació sobre assistents de veu i mitjans de comunicació, en què s'aprofundeix en les oportunitats i reptes que aquestes noves plataformes vocals ofereixen davant la distribució i el consum d'informació i entreteniment. Es fa un estudi del context que teoritza sobre un nou llenguatge de comunicació, la Human-Machine Communication (HMC), i sobre la penetració dels assistents de veu en el mercat mundial. A més a més, es fa un anàlisi de dos casos pràctics: la CCMA y la BBC.

Castellà: El presente trabajo ofrece un estudio de investigación sobre asistentes de voz y medios de comunicación, en el que se profundiza en las oportunidades y los retos que estas nuevas plataformas vocales ofrecen ante la distribución y el consumo de información y entretenimiento. Se hace, en primera instancia, un estudio del contexto que teoriza sobre un nuevo lenguaje de comunicación, la Human-Machine Communication (HMC), y sobre la penetración de los asistentes de voz en el mercado mundial. Además, se hace un análisis de dos casos prácticos: la CCMA y la BBC.

Anglès: This paper offers a research study on voice assistants and media, in which it delves into the opportunities and challenges that these new vocal platforms offer in the distribution and consumption of information and entertainment. In the first instance, a study of the theorizing context of a new language of communication, the Human-Machine Communication (HMC), and on the penetration of voice assistants into the world market is made. In addition, an analysis of two case studies is made: the CCMA and the BBC.

Agradecimientos.

A mi marido y a mi hijo (y a la pequeña que está a punto de llegar),
que día a día me hacen querer ser mejor persona y mejor ejemplo para ellos,
y me dan ganas y fuerza para enfrentarme a retos como este.

A mi familia, por su apoyo incondicional en cada uno de mis proyectos;
especialmente a mi madre y a mi hermana,
que gracias su ayuda insaciable (y sus horas de canguro) he podido realizar este máster.

*“Nuestra recompensa se encuentra en el esfuerzo
y no en el resultado.*

Un esfuerzo total es una victoria completa”.

Mahatma Gandhi

Índice

1. Introducción	6
1.1. Presentación y justificación de la investigación	6
1.2. Objeto de estudio	7
1.3. Objetivos del trabajo	7
1.4. Preguntas de investigación	8
2. Metodología de la investigación	10
3. Marco teórico	13
3.1. <i>Human-Machine Communication</i> (HMC): un nuevo lenguaje de comunicación	13
3.1.1. ¿Qué es la HMC? Aspectos teóricos	13
3.1.2. La era post-pantalla, un nuevo paradigma: la voz como interfaz de comunicación	16
3.1.3. Ética, privacidad e Inteligencia Artificial (IA)	22
3.2. Los asistentes virtuales de voz, un crecimiento rápido y exponencial	26
3.2.1. La penetración de los asistentes de voz en la sociedad	30
3.2.2. ¿Para que utilizan los asistentes de voz, los usuarios?	34
3.2.3. Principales retos de los asistentes de voz	36
3.3. Los asistentes de voz en los medios de comunicación	44
3.3.1. Tendencias de consumo digital de la información periodística	44
3.3.2. Los asistentes de voz, una oportunidad para el sector de la comunicación	47
3.3.3. Algunas experiencias de éxito	57
4. Trabajo de campo	61
4.1. El caso de la Corporació Catalana de Mitjans Audiovisuals (CCMA)	61
4.1.1. La Corporació Catalana de Mitjans Audiovisuals (CCMA)	61
4.1.2. “Alexa, ¿cuáles son las noticias del día?”	62
4.2. El caso de la British Broadcasting Corporation (BBC)	67
4.2.1. La British Broadcasting Corporation (BBC)	67
4.2.2. “Give me BBC News”	69
4.2.3. “Beeb”, su propio asistente de voz	72
4.3. Variables comparativas entre la CCMA y la BBC	75
5. Conclusiones	80
6. Bibliografía	83
7. Anexos	88
7.1. Entrevista a Xavier Fisa, director de Lavinia Voice	88

7.2. Entrevista a Lluís Garriga, consejero delegado de la Xarxa Audiovisual Local (XAL)	97
7.3. Entrevista a Geni de Vilar, adjunta a la dirección del área de Medios Digitales de la CCMA	100
7.4. Entrevista a Judith Argila, responsable de Nuevas Plataformas y Redes Sociales, en el departamento de Medios Digitales de la CCMA	104
7.5. Entrevista a Jeremy Walker y Chris Dix, <i>Head of Technology Transformation and Voice Management</i> y <i>Head of Technology Transformation and Product Architecture</i> de la BBC	115
7.6. Entrevista a Mónica Lablanca, directora de Cugat Média	120

1. Introducción

1.1. Presentación y justificación de la investigación

En las últimas dos décadas, el consumo de contenidos informativos y de entretenimiento, y el diseño de la experiencia de usuario, se ha basado en las interfaces visuales. Sin embargo, en los últimos años está apareciendo un nuevo lenguaje de interacción hombre-máquina basado en la voz y capitaneado por asistentes de voz como Alexa - de Amazon – y Google Assistant – de Google -. Este hecho abre una nueva oportunidad para abordar el diseño de las estrategias comunicativas de los medios de comunicación y vislumbra los asistentes de voz como plataformas imprescindibles para que los medios de comunicación puedan distribuir sus contenidos.

Es importante pensar hacia dónde se está moviendo el consumo digital de información, para que los medios puedan adaptar sus estrategias de distribución. El estudio *Digital News Report*, realizado por Reuters Institute (Newman, Fletcher, Kalogeropoulos, Levy & Kleis Nielsen, 2018) destaca, entre otros, el rápido crecimiento de los asistentes de voz. “Los asistentes de voz digitales activados por voz como Amazon Echo y Google Home siguen creciendo rápidamente, abriendo nuevas oportunidades para las noticias en formato audio. El uso se ha más que duplicado en Estados Unidos, Alemania y Reino Unido, con la mitad de usuarios de este servicio usándolo para el consumo de noticias e información”.

Los asistentes de voz se convertirán en el canal a partir del cual pediremos y consumiremos información; una manera de comunicarnos con los dispositivos, a través de la voz, mucho más natural y humana. La era del tacto como la principal interfaz de usuario está dejando paso a una nueva era donde los asistentes de voz digitales tienen el protagonismo: la era post-pantalla. Así, veremos cómo estos asistentes de voz irán impregnando porosidad en nuestra vida digital y en nuestra cotidianidad.

En este trabajo de investigación, profundizaremos en las implicaciones de la *Human-Machine Communication* (HMC), una nueva forma de comunicación hombre-máquina, y estudiaremos el papel de los asistentes de voz en el futuro y presente de los medios de comunicación. Analizaremos cómo éstos pueden cambiar el consumo, la distribución e, incluso, la producción de contenidos informativos. Profundizaremos en los casos de la Corporació Catalana de Mitjans Audiovisuals

(CCMA) y de la British Broadcasting Corporation (BBC), que ya han dado pasos adelante en la integración de sus contenidos en asistentes de voz.

1.2. Objeto de estudio

Tal y como se señala en la justificación del tema, el punto de partida de esta investigación es analizar cómo los asistentes de voz pueden representar una oportunidad para los medios de comunicación, que pueden encontrar en ellos nuevas e innovadoras maneras de llegar a su audiencia e, incluso, ampliarla.

Ámbito general

El ámbito general de este trabajo se centra en hacer una primera panorámica donde abordaremos la definición y las implicaciones de la *Human-Machine Communication* (HMC), un nuevo lenguaje de comunicación hombre-máquina directamente relacionado con los asistentes de voz; y ahondaremos en el estado actual de los asistentes de voz a nivel mundial, para conocer la proyección de esta tecnología y las interfaces vocales.

Además, analizaremos qué oportunidades y retos presentan los asistentes de voz para los medios de comunicación, y veremos de qué manera los medios de comunicación han empezado a integrar sus contenidos en estas interfaces vocales.

Ámbito específico

Una vez contextualizado el objeto de estudio general, estudiaremos y compararemos los casos de dos medios de comunicación públicos que han empezado a integrar sus contenidos en distintos asistentes de voz: la Corporació Catalana de Mitjans Audiovisuals (la CCMA), el ente que gestiona los medios de comunicación públicos de Catalunya; y la British Broadcasting Corporation (BBC), el ente que gestiona los medios de comunicación públicos de Reino Unido.

1.3. Objetivos del trabajo

Los objetivos generales y específicos de este trabajo de investigación son los siguientes:

Objetivos generales

- Ahondar en los contenidos teóricos del lenguaje de comunicación hombre-máquina (*Human-Machine Communication* -HMC-).
- Estudiar las implicaciones de este nuevo lenguaje de comunicación para las personas.
- Analizar la situación de los asistentes de voz en el mercado y su penetración en la sociedad.
- Reflexionar sobre las oportunidades y los retos que se abren ante la llegada de los asistentes de voz, para los medios de comunicación.

Objetivos específicos

- Conocer las experiencias con asistentes de voz de la Corporació Catalana de Mitjans Audiovisuals (CCMA) y de la British Broadcasting Corporation (BBC).
- Analizar cómo han adaptado sus contenidos al nuevo canal de distribución de información y en qué forma.
- Valorar si tener presencia en los asistentes de voz implica, necesariamente, un cambio en las rutinas periodísticas de los medios de comunicación.
- Comparar las experiencias de ambos medios de comunicación analizados y ver si existen unos patrones similares, o no, en la integración de los contenidos informativos en los asistentes de voz.

1.4. Preguntas de investigación

Por medio de la formulación de esta investigación sobre los asistentes de voz como nuevas plataformas de distribución y consumo de información, se pretende responder a las siguientes preguntas:

- ¿Estamos ante un cambio de paradigma, donde la voz va a representar el canal principal a través del cual las personas se van a comunicar con las máquinas?
- ¿Las interfaces vocales, como los asistentes de voz, van a convertirse en una de las principales plataformas de distribución y consumo de información?
- ¿Los asistentes de voz constituyen una oportunidad para que los medios de comunicación trabajen en nuevas estrategias comunicativas?
- ¿Cómo los asistentes de voz pueden cambiar el consumo, la distribución e incluso la producción de los contenidos informativos y de entretenimiento?

- ¿Qué uso está haciendo la CCMA y la BBC de los asistentes de voz?
- ¿Por qué tipo de contenidos apuestan la CCMA y la BBC, en el entorno de los asistentes de voz?
- ¿Cómo se organizan los equipos que desarrollan los asistentes de voz y cómo se integran en la cultura del medio de comunicación?

2. Metodología de la investigación

La metodología utilizada para este estudio consiste, por un lado, en una investigación de fuentes secundaria para el abordaje teórico de los conceptos de *Human-Machine Communication* (HMC) y asistentes de voz. Para ello, se han utilizado diferentes fuentes de información: los buscadores de Mendeley, ResearchGate, la biblioteca de la UAB y los contenidos proporcionados en el máster en Medios, Comunicación y Cultura de la UAB.

Además, tanto para esta aproximación teórica inicial como, sobretodo, para el análisis de casos prácticos y la investigación de campo, el estudio se ha basado en la extracción de información a través de fuentes primarias, gracias a las entrevistas cualitativas realizadas (se encuentran en el anexo de este trabajo).

Así pues, la metodología principal y, por lo tanto, la técnica de recogida de datos utilizada en este trabajo de investigación es la entrevista cualitativa. En concreto, se han utilizado entrevistas cualitativas semi-estructuradas, donde el investigador ha preparado antes de la entrevista un guión temático sobre los temas de los que quiere que se hable con el informante. Las entrevistas se han basado en preguntas abiertas donde el entrevistado ha podido mostrar todos sus conocimientos sobre el tema, expresar sus opiniones, matizar sus respuestas e, incluso, desviarse del guión inicial.

En todos los casos, han sido conversaciones naturales donde el entrevistado ha tenido la libertad de explorar todas las respuestas e incluso introducir temas que no habían sido previstos en un inicio. Así, entrevistado y entrevistador hemos podido construir un diálogo fluido de preguntas y respuestas que han dado como resultado la información extraída para esta investigación.

La selección de los entrevistados se ha basado en el conocimientos de éstos sobre el tema objeto de estudio. A través de un análisis del entorno se han detectado expertos en materia de asistentes de voz, interfaces vocales y medios de comunicación en Catalunya (por cuestión de proximidad) y se ha hecho una selección de aquellos más indicados para entrevistar.

Finalmente, se han realizado 6 entrevistas cualitativas semi-estructuradas a los siguientes expertos:

- **Xavier Fisa:** Licenciado en Publicidad y Relaciones Públicas, y especializado en comunicación interactiva y digital. Es experto en disrupciones tecnológicas orientadas al mundo de la

comunicación. Actualmente es director del área Lavinia Voice, un departamento de la productora Lavinia donde están trabajando, desde 2016, con asistentes de voz en el sector de los medios de comunicación.

- **Lluís Garriga:** Es, actualmente, uno de los consejeros delegados de la Xarxa Audiovisual Local (XAL), una sociedad mercantil creada por la Diputación de Barcelona para gestionar el apoyo al audiovisual local catalán. La XAL está desarrollando, actualmente, unas aplicaciones para asistentes de voz relacionadas con los servicios de comunicación de proximidad.

- **Geni de Vilar:** Es adjunta a la dirección del área de Medios Digitales de la Corporació Catalana de Mitjans Audiovisuals (CCMA). El departamento que lidera, junto a Saül Gordillo, se encarga de diseñar y aplicar la estrategia digital de los medios de comunicación que pertenecen a la corporación pública catalana. Uno de los objetivos digitales a corto y medio plazo (y que como veremos en el estudio, ya han empezado a aplicar) es la integración de sus contenidos informativos en asistentes de voz.

- **Judith Argila:** Es licenciada en periodismo con un posgrado en Visualización de Información y del Conocimiento en la UPG; master en Sociedad de la Información y el Conocimiento, especializada en la línea de investigación en Tecnologías Digitales (UOC). Realizó un Trabajo de Final de Master (TFM) sobre asistentes de voz. Hizo un período de investigación en Oxford durante 6 meses en el Reuters Institute. Actualmente es la responsable de Nuevas Plataformas y Redes Sociales, en el departamento de Medios Digitales de la CCMA. Lleva desde agosto de 2018 trabajando en el proyecto de la CCMA con asistentes de voz.

- **Jeremy Walker y Chris Dix:** Se ha realizado una entrevista conjunta con dos de los responsables de producto de la British Broadcasting Corporation (BBC). Son, respectivamente, *Head of Technology Transformation and Voice Management* y *Head of Technology Transformation and Product Architecture* de la BBC. La corporación británica lleva desde 2016 integrando sus contenidos en asistentes de voz.

- **Mónica Lablanca:** Se ha entrevistado a Mónica Lablanca, directora de Cugat Mèdia y periodista. Este medio de comunicación de proximidad empezó en 2020 a trabajar en su propia *skill* para Alexa. Se trata de uno de los pocos medios de comunicación locales que han apostado por tener presencia en estas plataformas vocales.

Las condiciones de las entrevistas cualitativas realizadas, no han sido las óptimas debido a la situación de pandemia que hemos vivido a nivel mundial, a causa del Covid-19. Las entrevistas, planeadas para marzo y abril (mayoritariamente) tenían que realizarse de manera presencial, puesto que podían aportar muchos más matices, pero finalmente han tenido que realizarse de manera virtual. En todos los casos se han realizado a través de vídeo-conferencias donde se ha podido establecer una conversación fluida con el entrevistado. La única excepción, en este sentido, ha sido la entrevista realizada con los responsables de la BBC, que se ha realizado a través de un formulario escrito de preguntas que ellos, posteriormente, han contestado con las respuestas.

3. Marco teórico

3.1. *Human-Machine Communication* (HMC): un nuevo lenguaje de comunicación.

3.1.1. ¿Qué es la *Human-Machine Communication* (HMC)? Aspectos teóricos.

El concepto de Human-Machine Communication (HMC) es, aún hoy, un concepto en desarrollo, que nace de la necesidad de analizar y entender una nueva forma de comunicación con interlocutores digitales, que se ha desarrollado muy rápidamente en los últimos años. El estudio de la comunicación hombre-máquina tiene como objetivo primero entender la tecnología como una entidad comunicadora por sí misma, y no solo como una mediadora de la comunicación (Argila, 2019).

Una de las definiciones más aceptadas entre los académicos es la que propone la profesora e investigadora de la Universidad de Northern Illinois, Andrea L. Guzman, en su libro *Human-Machine Communication. Rethinking Communication, Technology, and Ourselves* (Guzman, 2018): “La HMC es la creación de significado entre humanos y máquinas, donde la tecnología es considerada como un comunicador, un sujeto con el que las personas se comunican, en lugar de un canal a través del cual los seres humanos interactúan entre sí.”

Se trata de una nueva forma de comunicación inteligente que abre un escenario hasta ahora desconocido y que está liderado por la comunicación oral; una nueva forma de Inteligencia Artificial (IA) que tiene como uno de sus núcleos principales los asistentes virtuales de voz.

Pero ¿qué implica esta nueva forma de comunicación para ambos interlocutores, digital y humano? ¿Podemos realmente considerar que estamos ante una forma de comunicación, como tal? Los paradigmas clásicos de la teoría de la comunicación se han centrado, durante mucho tiempo, en la comunicación humano-humano, y la interacción de las personas con la IA a través de agentes virtuales, *socialbots* o programas informáticos de generación de lenguaje no encajan en este esquema (Guzman & Lewis, 2019). Sin embargo, en los últimos años, diversos investigadores están intentado abordar esta desconexión entre la teoría de la comunicación y las tecnologías emergentes, puesto que son dos elementos que, necesariamente, han de conectarse para entender la sociedad actual y la del futuro.

Es básico conocer qué diferencia la HMC y la comunicación humano-humano, y para eso debemos entender las bases sobre las cuales se sustentan estos tipos de comunicación. Si hay un aspecto que puede diferenciar la comunicación entre humanos de la comunicación entre hombres y máquinas este es el juicio experiencial, es decir, la vivencia. Las personas nos movemos en diferentes contextos y ambientes, y decidimos sobre multitud de temas cada día porque hemos acumulado un conocimiento previo que las máquinas aún no tienen. (Autoritat Catalana de Protecció de Dades, 2020)

En el proceso de comunicación, los humanos percibimos el entorno, el contexto, el tiempo, el lenguaje no verbal, tenemos presente nuestra propia dimensión cultural y social, e incluso nuestro pasado en común (Fortunati & Edwards, 2020). Los interlocutores digitales, sin embargo, no gozan de esa capacidad o la tienen más limitada porque no tienen consciencia de ellos mismos ni del mundo.

En este sentido, es importante tener presente el concepto de Inteligencia Artificial (IA) adaptativa, ya que abre muchas posibilidades en este campo, puesto que es una nueva lógica que aporta la tecnología y que está cogiendo mucha fuerza en esta disrupción. Tiene como premisa inicial conocer el contexto y las necesidades de los usuarios, y a partir de ahí ofrecer propuestas de valor y experiencias más ajustada a ese contexto y a esas necesidades (Fisa, entrevista cualitativa, 17 abril 2020). Es un punto esencial del éxito de esta nueva disrupción tecnológica basada en la voz, que permite aportar experiencias hiper-personalizadas.

Otra de las diferencias entre una comunicación humano-humano y una comunicación humano-máquina es que en una conversación entre personas se establece, por norma general, una relación de igual a igual, cosa que no sucede con un interlocutor digital que, además, suele tener menos habilidades comunicativas y de comprensión, y una menor competencia del lenguaje y de la expresividad. Es obvio, pues, que la habilidad comunicativa con un interlocutor digital queda reducida, con lo que debemos preguntarnos si hemos de hablar de ellos como un interlocutor o como un “casi-interlocutor” (Fortunati & Edwards, 2020).

La definición más teórica de comunicación ni siquiera consideraría este tipo de acción, la HMC, como una acción comunicativa en sí, puesto que ésta tendría necesariamente que realizarse entre dos o más personas y de una manera simétrica. Sin embargo, la comunicación hombre-máquina asume de entrada una asimetría compleja a nivel social y comunicativo. Pero, ¿porqué no puede ser

considerada, también, una nueva forma de comunicación que, además, tiene una tendencia de crecimiento exponencial?

De hecho, Fortunati y Edwards asumen que aunque los humanos somos conscientes de que los interlocutores digitales son “casi-interlocutores”, “casi-comunicadores” y “casi-sociales”, juegan al juego e intentan realmente comunicar y tener relaciones sociales con las máquinas. Incluso, en algunos casos, las personas pueden llegar a olvidar que están ante un robot social y tratarlos como si fueran verdaderamente humanos. Y es que, el hecho de que interactuemos a través de un mecanismo de comunicación ancestral como es la voz también connota una mayor humanización, y, en consecuencia, es inherente en el ser humano interpretar que “si interactúo por voz con una máquina, la estoy humanizando” (Fisa, 2020).

Si pensamos en los asistentes virtuales de voz, podríamos considerar que éstas máquinas contestan ante cualquier pregunta que se les formula y además lo hacen a través de la voz (¿hay algo más humano que la comunicación oral?), pero en ningún momento pensamos en una conversación idéntica a la que tendrían dos humanos, sino que asumimos que se obtendrán respuestas mucho más automáticas y prácticas. Esto es algo en lo que Kellermann (1992) ya reflexionaba a principios de los años 90 sobre las propias relaciones humanas y donde se planteaba si la comunicación entre personas no era sino una conversación automatizada y funcional. Entonces, ¿existe diferencia alguna con la HMC?

La realidad es que para que la relación comunicativa entre humano y máquina funcione, las personas deben ajustar sus prácticas de comunicación a las habilidades menos avanzadas de comunicación de los robots, y actuar en consecuencia (Höflich, 2013). Por ejemplo, en el caso de los asistentes virtuales de voz, deberemos comunicarnos con ellos a través de preguntas formuladas con una estructura muy concreta para que la IA sea capaz de descodificarlas e interpretarlas.

Aunque se trata de sistemas cada vez más avanzados, hemos de ser conscientes de las limitaciones de las tecnologías con las que nos comunicamos y adaptarnos a ellas, para que la relación sea lo más fructífera posible, y este hecho puede llevar, en muchos casos, a la frustración cuando la tecnología aún no está lo suficientemente avanzada. Fisa, lo reflexiona de la siguiente manera: “Si yo, cuando hablo con una máquina, me imagino que va a entender todo lo que dice y me va a devolver una respuesta de la misma manera que lo hace un ser humano y no lo hace, puede generar un alto nivel de frustración. Estamos al principio de esta disrupción y son muchas las personas que,

cuando interactúan por primera vez con un asistente de voz, se sienten frustrados porque intentan emular la satisfacción de la experiencia comunicativa con un ser humano”.

Parece entonces, que estamos ante una doble y paradigmática realidad en la que, por un lado, asistimos a la humanización de los robots y las máquinas, que cada vez tienen más habilidades sociales, y por el otro, vemos una robotización de los humanos, que han de simplificar su sistema de comunicación complejo para poder comunicarse con estos nuevos actores digitales.

3.1.2. La era post-pantalla, un nuevo paradigma: la voz como interfaz de comunicación.

En los últimos 30 años, todas las interacciones con dispositivos electrónicos se han hecho a través de una pantalla. Venimos de una era basada en el clic y en el *touch*, donde las pantallas táctiles, las tabletas y los teléfonos inteligentes han sido protagonistas de una explosión; donde el *mouse* y la pantalla táctil, cada uno en su momento respectivo, revolucionaron la forma en que las personas interactúan con las máquinas.

Sin embargo, parece que ahora estamos entrando en la era de la voz, donde la pantalla va a ser prescindible y las nuevas interfaces vocales van a permitir interactuar en el espacio virtual sin una interfaz gráfica tradicional. La fuerza de esta disrupción está, como menciona Fisa (2020) en que “por primera vez en la historia la tecnología habla nuestro lenguaje y no nosotros el suyo”.

En esta era post-pantalla, las tecnologías de voz están alterando las interacciones y los patrones de comunicación de los usuarios con las máquinas, y, sobretodo, las búsquedas mediante la voz están revolucionando la manera como nos comunicamos con ellas. El gran salto de paradigma tiene que ver con el tipo de interfaz con la que se produce la interacción hombre-máquina: los asistentes de voz, que representan un cambio de concepción que desplaza de la tradicional interfaz visual a la interfaz auditiva (Argila, 2019).

Es importante tener en cuenta, como dice el director del Grupo Fjord, Andy Goodman, que las pantallas pronto dejarán de ser la principal forma en la que interactuamos con los dispositivos que tenemos a nuestro alrededor, pero no desaparecerán, sino que convivirán en un entorno mucho más rico de dispositivos e interfaces (Brownlee, 2015). De hecho, cuando se habla de asistentes de voz, suele visualizarse en altavoces interactivos, porque la voz es algo tan etéreo que se necesita crear tangibles para conceptualizarlo. Pero la realidad es que no va a ser solo en estos dispositivos donde

vamos a poder interactuar por voz; es más, según Fisa, no van a ser ni siquiera el principal dispositivo de esta revolución a medio plazo.

Es importante separar el concepto de asistentes virtuales de voz y de altavoces inteligentes. Aunque actualmente parece que una cosa vaya ligada a la otra, no es así: el asistente es la voz, la IA, y el altavoz inteligente es el dispositivo a través del cual tu tienes acceso a esta inteligencia. Visualizamos los asistentes virtuales de voz en los altavoces, porque hoy en día son estos dispositivos los que nos han acercado mayoritariamente la tecnología de la voz, pero a medida que avance esta disrupción tecnológica vamos a ver como los asistentes de voz como Alexa o Google Assistant van a integrarse en multitud de dispositivos distintos: televisores, coches, móviles e incluso neveras o microondas... En los próximos años, las tecnologías de la voz van a pasar de los hogares a estar integradas de manera creciente en cada parte de nuestras vidas (Newman, 2018).

Es también interesante la visión de Robin Li, CEO de Baidu, quien aseguró en la apertura de la sexta edición de la Conferencia Mundial de Internet, que “en pocos años se reducirá la dependencia de los móviles al tiempo que los dispositivos de audio o sensores inteligentes ocuparán su espacio” («Los gurús tecnológicos auguran una nueva internet que cambiará el mundo.», 2019).

Según Newman (2018), vamos a utilizar la voz para controlar dispositivos y acceder a información porque resulta mucho más rápido y práctico, en la mayoría de casos, que las pantallas táctiles o los controles remotos. El *output* es más complejo y dependerá mucho del contexto, que a veces necesitará involucrar algún tipo de pantalla. De hecho, a mitades de 2018 algunas de las grandes empresas tecnológicas ya empezaron a lanzar dispositivos con asistentes de voz que también incorporaban una pantalla, como el Google Hub o el Amazon Show. Estas decisiones estratégicas ya indicaban cual es la dirección del mercado.

El concepto teórico que recoge esto se llama interacción multimodal. Una de los expertos que ha definido este nuevo concepto ha sido Gian Marco Peraldo, Coordinador de Voz y Servicios de Data Core en Entel Perú y lo define como el proceso en el cual diversos dispositivos y personas son capaces de llevar a cabo una interacción (auditiva, visual, táctil y gestual) conjunta desde cualquier sitio, en cualquier momento, utilizando cualquier dispositivo y de forma accesible, incrementando así la interacción entre personas, y entre dispositivos y personas (Peraldo, 2010).

Para ver más claro el funcionamiento de esta interacción multimodal, nos servimos del ejemplo que nos ofrece Fisa (2020): “Podemos iniciar una comunicación a través de la invocación por voz, pero el dispositivo tiene la capacidad de devolver esa información, no solo con voz, sino con un vídeo, con imágenes de apoyo o incluso con alguna forma más compleja de comunicación. De igual manera, el usuario que ha iniciado la comunicación a través de una invocación por voz, quizás necesite ahora acercarse al dispositivo para continuar la experiencia y, por ejemplo, ver algo y, entonces, la siguiente interacción la realiza a través del tacto, sobre la pantalla. Acto seguido la nueva necesidad no va a pedir una interacción de contacto física, sino que va a volver a ser a través de voz”. Esto es la interacción multimodal.

Pese a esta interacción, que abarca todo el espectro de dispositivos y sentidos humanos, lo cierto es que esta nueva disrupción tecnológica está basada en las interfaces vocales y, una de sus grandes revoluciones es que nos permite comunicarnos liberando tanto los ojos como las manos. Ha sido justo esta característica la que ha puesto las tecnologías por voz en el foco de todas las miradas.

Según Horcada (2017) este tipo de tecnología surge de la necesidad, cuando en un contexto determinado se requiere una interacción en la que no conviene que el usuario utilice el tacto (ya sea con una pantalla o con el *mouse*). El ejemplo más evidente es el de una persona que está conduciendo y necesita acceder a una determinada información. Horcada afirma que la necesidad de este tipo de interacciones existe desde hace mucho tiempo, pero ahora el avance de la tecnología nos permite que sea una realidad.

La era post-pantalla ha dejado atrás las llamadas *Graphical User Interfaces* (GUI) y ha puesto sobre la mesa una nueva manera de comunicarse con los dispositivos a través de las *Voice User Interfaces* (VUI). Esto plantea un escenario completamente distinto donde, en primer lugar, se utiliza uno de los instrumentos más humanos y básicos de comunicación de las personas: la voz. Este nuevo paradigma está considerado dentro de las *Natural User Interface* (NUI), que se corresponden con aquellas interfaces más intuitivas para las personas, que permiten aplicar razonamientos naturales de la comunicación humana, sin necesidad de entender de manera previa la tecnología.

Es más, como defiende Argila (2019) abren una puerta a la democratización de la tecnología, puesto que muchos usuarios encontraban una barrera en las tecnologías basadas en interfaces visuales, por motivos generacionales, educacionales o a la capacidad de escribir y leer. Con las interfaces visuales no solo es imperativo conocer las convenciones gráficas necesarias, sino que también

existe una amplia complejidad operativa. Sin embargo, estos colectivos ven ahora como a través de la interacción por voz se disuelven las barreras de acceso y hablar a una máquina resulta mucho más intuitivo.

En el epicentro de las *Vocal User Interfaces* se encuentran los asistentes de voz, como Alexa, Siri o Google Assistant, entre otros. Gracias al diseño conversacional, estos son capaces de interactuar de manera oral con un lenguaje natural, seguir las normas sociales de la comunicación humano-humano, intentan desarrollar una relación con los usuarios a través de conversaciones inteligentes y naturales, y exhiben rasgos de personalidad distintos (Guzman, 2016). Son inteligentes y sociales, y están programados para seguir las normas de comunicación humana. De hecho, la investigadora de la Universidad Northern Illinois va más allá y los llama agentes sociales vocales (o *Vocal Social Agent*, en sus siglas en inglés -VSA), puesto que están programados para actuar como los humanos y ser autónomos.

Fisa (2020) cuando habla de la ya mencionada Inteligencia Adaptativa expone que esta “nos conoce, aprende de nuestro contexto y comportamientos y trabaja para cubrir nuestras necesidades individuales”. Estos dispositivos pretenden formar parte de nuestro mundo social, y para conseguirlo, como dice Guzman, han de convencer a las personas de que son entidades sociales; han de presentar a los humanos su parte más social.

Nuestro entendimiento de la IA ha estado, y sigue estando, condicionado mayoritariamente por la ciencia ficción. En ella se suelen asociar las máquinas como entes maliciosos. Pero en contra de lo que hemos asumido culturalmente, los asistentes de voz nos presentan un tipo de IA que no nos resulta amenazadora. Es más, hay estudios de Google que concluyen que el 41% de los propietarios de altavoces inteligentes se sienten como si estuvieran hablando con un amigo o con otra persona (Kleinberg, 2018).

Ante estos asistentes vocales nosotros sentimos que tenemos el control y entendemos que están para ayudarnos y darnos seguridad. Ello es debido a un cambio sustancial en la concepción de los programas de IA vocal: no se los considera máquinas, sino más bien asistentes al servicio de las personas. Una especie de metáfora que ayuda a que estos robots sociales adquieran cierta identidad. Incluso la manera como las personas se dirigen a los asistentes tiene claramente una vinculación directa con la manera como nos comunicamos humano-humano: pronunciando delante el nombre de la personas con la que queremos hablar. Por ejemplo:

“Alexa, ¿qué tiempo hace hoy?”.

Este tipo de programas de IA están diseñados para mostrar una posición de inferioridad con respecto al usuario. Una de las muestras que hace este hecho evidente es que, aunque tienen sentido del *self* (más bien, lo han creado para que parezca que lo tienen), cuando les preguntas qué son, contestan: “Soy Alexa” o “Soy Siri”, y en otras respuestas incluso ponen de manifiesto su rol social como asistente para humanos: “Estoy aquí para ayudarte”.

Es interesante contemplar la manera como nos comunicamos con este tipo de IA vocal, puesto que ello refuerza su rol y su lugar en la comunicación humano-máquina. La forma más sencilla y habitual de hablar a los asistentes de voz es a través de ordenes: “Alexa, llama a Alberto”, “OK Google, qué canción es esta?”, “Siri, llévame a casa”. Es también a través de sus funciones que se observa claramente su rol de servicio a las personas, quienes siempre tienen una posición dominante.

Además, los asistentes de voz están diseñados para reforzar sus afirmaciones de sumisión a través de su comunicación no verbal con los usuarios y la forma en que se anima a los usuarios a hablar con ellos (Guzman, 2016). Cualquier tipo de comunicación con ellos los sitúa siempre como los receptores del mensaje, los usuarios deciden cuando iniciar la comunicación con ellos y cuando acabar la conversación. Como menciona Argila (2019), se trata de una tecnología *pull* donde el usuario es quien toma la iniciativa y decide, a través de una acción, activar o no el servicio.

Pero ¿qué sucedería si el diseño de estos dispositivos contemplara también comunicaciones activas para ofrecer información al usuario en momentos concretos del día? Es lo que se conoce como tecnologías *push*, muy utilizadas hasta el momento por los medios de comunicación y los servicios informativos, por ejemplo, para alertar a la audiencia sobre información relevante de última hora. En este caso es el medio el que inicia la comunicación hacia el usuario, pudiendo interrumpirlo en cualquier contexto.

Aunque actualmente los asistentes de voz existentes en el mercado no contemplan esta opción, sí han empezado a trabajar maneras de hacer estas entregas *push*. Es el caso del medio de comunicación norteamericano Washington Post (pertenece a Jeff Bezos, como Amazon), que está diseñando una manera de informar al usuario de que hay una noticia de última hora, a través del

enciendiendo una luz verde. De esta manera, el usuario sigue teniendo parte del control sobre el momento exacto en que escuchará esa información (Argila, 2019).

El género de estos asistentes de voz también sirve para establecer quien es ella en relación con la audiencia (Guzman, 2016) Los asistentes de voz tienen todos una voz femenina que les hace dibujar una nueva dimensión identitaria, en este caso, en relación con el género. Al respecto, Rothschild (1983) argumenta: “La tecnología es parte de nuestra cultura; y, por supuesto, nuestra cultura, que está dominada por el género masculino, ha desarrollado tecnologías que refuerzan la supremacía de los varones”. Así, parece claro que la voz femenina y, en consecuencia el género que se asocia con los asistentes de voz virtuales, refuerza su papel de subordinación.

Puede que no se trate solo de una cuestión sexista, teoría que parece bastante reforzada: también existen estudios que consideran que la voz femenina es más fácil de escuchar para los consumidores que la masculina, ya que, esta última, en la mayoría de ocasiones se relaciona culturalmente con una figura autoritaria. Unas convenciones de género presentes en los asistentes virtuales de voz, que resultaría interesante investigar en posteriores estudios.

Como hemos visto anteriormente, los asistentes de voz buscan hacerse un hueco en nuestras relaciones sociales y, por eso, están diseñados para ganarse a la gente con su personalidad. Solemos pensar en Alexa, por ejemplo, como un ente divertido y servicial que, incluso, es capaz de bromear si se lo pedimos. Son recurrentes las preguntas, por ejemplo, sobre lo que lleva puesto (interrogación claramente masculina), que el asistente es capaz de esquivar con elegancia y sentido del humor. Sabemos que no existe físicamente, pero sin embargo tenemos la sensación de que sí lo hace. Por eso nos sentimos tentados a hacerle este tipo de preguntas, aún sabiendo que no es más que un programa informático en la nube programado para respondernos.

Lo que nos hace llegar a ello y fantasear con su existencia física es, claramente, que la podemos escuchar, su voz, incluso sin ver o sostener en las manos al dispositivo que la integra (como sucedía al principio con Siri y los teléfonos inteligentes). En sus respuestas es incluso capaz de utilizar tonos de voz similares a los de los humanos, siempre manteniendo su estatus de asistente al servicio de las personas.

Esta voz es la esencia de los asistentes virtuales vocales. Pero hemos de ser conscientes, cuando interactuamos con ellos, de que no solo son interlocutores en un diálogo, sino que también actúan de intermediarios entre el humano y la empresa que hay detrás.

3.1.3. Ética, privacidad e Inteligencia Artificial (IA)

Actualmente, cerca del 58% de la población mundial¹ está conectada a Internet, lo que supone una cifra superior a los 4 mil millones de personas. Esto significa una generación muy elevada de datos diarios, ya que se calcula que cada movimiento de cada persona es un nuevo dato, y estamos generando datos las 24 horas del día, los 365 días del año.

La periodista Karma Peiró, en su artículo para el informe *Intel·ligència Artificial. Decisions Automatitzades a Catalunya (2020)* recuerda que nuestras acciones con los dispositivos inteligentes producen datos que describen nuestros gustos, estados de ánimo, reacciones a estímulos, miedos y ambiciones. Y sostiene que los datos más importantes para las empresas, los partidos políticos o los gobiernos son los que definen el comportamiento de grupos de personas, y configuran perfiles o patrones de consumo, a través de los cuales nos hacen llegar productos o mensajes personalizados que casi no podemos rechazar.

De hecho, Kenneth Cukier, editor tecnológico de la revista *The Economist*, defiende que el valor de la información hoy día reside en la manera de correlacionar los datos para descubrir patrones que ni tan solo se habían imaginado (Cukier, 2013). Está claro que nuestras vidas están cada vez más condicionadas por la interacción que tenemos con la tecnología, y en el centro de este ecosistema se encuentran nuestros datos. Por eso es creciente el estudio para conocer si existe un equilibrio entre lo que ofrece la tecnología a las personas y el valor que ésta aporta a la sociedad a través de los datos (Fisa, 2020).

Como hemos visto, la tecnología está ocupando cada vez más espacios sociales de intimidad en nuestras vidas y en nuestras mentes, y el ejemplo más claro lo encontramos en nuestra comunicación con las máquinas. La tecnología se está convirtiendo más que en una herramienta, en un canal para expresar nuestros mensajes. El medio, en sí mismo, se ha convertido en un socio de

1 <https://www.lavanguardia.com/vida/junior-report/20190509/462140060359/brecha-digital-desigualdad-acceso-internet.html>

comunicación (Guzman, 2016). De hecho, la HMC considera la tecnología como mucho más que un medio: como un socio en el proceso comunicativo.

Pero ¿cómo se diseñan las tecnologías sociales, como Alexa o Siri, a través de la lente de la HMC? Las interfaces de voz están activamente escuchando, esperando una palabra clave para activarse. Una vez la identifican, la interfaz utiliza reconocimiento de voz y lenguaje natural para convertir el mensaje en comandos. Estos son enviados a la nube, donde se encuentra la inteligencia para hacer el procesamiento y análisis, y luego retorna la respuesta a la interfaz. Desde la perspectiva de la privacidad, esta escucha activa, es decir, que el dispositivo esté continuamente escuchando, genera una tensión obvia en los usuarios. Pero ¿porqué?

En el apartado anterior hemos mencionado que los asistentes de voz como Alexa, Siri o Google Assistant tienen una doble función: la de interlocutores, puesto que hablamos con ellos, y la de intermediarios. Hay que tener presente que, pese a todas sus cualidades humanas, los asistentes de voz no son entidades independientes como lo podrían ser los humanos. De hecho, no son entidades en sí, sino pedazos de *software* que intercambian información con usuarios humanos, con otras máquinas y, en último término, con la empresa que los ha creado (Alexa con Amazon, Siri con Apple, o Google Assistant con Google) y cualquier otra compañía o organización que esté al otro lado de este proceso computacional (Guzman, 2016).

Sin embargo, los asistentes de voz no revelan directamente que ellos, a la vez que hablan contigo, también lo están haciendo con Amazon, Apple o Google. He aquí el riesgo de la IA, y es que a la vez que están trabajando para nosotros, para ofrecernos la información, también lo hacen para la empresa en cuestión. Cuando pedimos a Siri o Alexa que nos recuerde que tenemos que ir al médico, que nos lleve a alguna parte o que nos ofrezca un tipo de noticias en concreto, estamos cediendo el control de parte de nuestras vidas a la máquina. Y esta gran cantidad de datos personales disponibles acentúa la incidencia de la tecnología sobre la vida de las personas.

No obstante, no hay que olvidar que en el caso de los asistentes de voz somos nosotros mismos quienes ponemos a disposición de los dispositivos (es decir de las empresas y organizaciones que hay detrás) la información, a través de peticiones directas que permiten a los algoritmos configurar nuestro perfil, con nuestros gustos, intereses o preferencias. Somos nosotros quienes aceptamos esta situación y, en la mayoría de ocasiones, lo hacemos sin ser plenamente conscientes de ello.

Las tecnologías que hemos construido juegan un papel importante en como evolucionamos como individuos y como sociedad. En este momento, por ejemplo, estamos evolucionando hacia una sociedad dominada por los robots (Turkle, 1984) y los asistentes de voz son una de las nuevas maneras que se presentan para comunicarnos con el mundo. Como dice (Sangers, 2000) nosotros no vamos a ser substituidos por máquinas, sino que vamos a vivir en ellas y a través de ellas.

No hemos de olvidar que la tecnología es tecnología (Carey & Quirck, 1989) y el riesgo que existe ante los asistentes de voz, y cualquier otra, está en comunicarnos con ellos sin aceptar su complejidad tanto tecnológica como social. Sebastian Arriada, experto en ciberseguridad, recuerda en este sentido que aunque puedan existir leyes y regulaciones específicas para la privacidad y la protección de datos de los usuarios frente a la tecnología, la responsabilidad final e imprescindible recae en el usuario del dispositivo, con lo que recomienda “una dosis de cautela para poder disfrutar de los dispositivos de voz con confianza” (*Diseño conversacional. Haz que tu voz brille.*, 2019). Es imprescindible tener un conocimiento previo sobre las tecnologías con las que vamos a interactuar.

También es superlativo fijar límites a la hora de diseñar estos programas de IA, en tanto que tienen unas implicaciones muy directas con las personas. Estos límites se fijan a través de las leyes reguladoras, que juegan un papel fundamental, como la RGPD -la principal herramienta de regulación de protección de datos en ámbito europeo actual-.

Algunos sectores critican las medidas reguladoras como la RGPD en Europa, porque consideran que supone un freno para el desarrollo de la IA, y exigen una regulación más laxa que facilite la reutilización de datos. Sin embargo, también existe la visión contraria, que considera que las limitaciones de la regulación existente suponen un incentivo para desarrollar una IA que haga un uso más razonable y más efectivo de los datos disponibles, sin poner en riesgo ni hacer negocio con los datos de las personas que utilizan la tecnología.

Estas medidas, no obstante, no son el único instrumento disponible para garantizar la privacidad y protección de datos de los usuarios. La ética tiene también un papel principal. Pero ¿qué es la ética y quien determina lo que es ético y lo que no? De hecho, y como se exponen en el informe anteriormente mencionado de la Autoritat Catalana de Protecció de Dades (2020), *Intel·ligència Artificial. Decisions Automatitzades a Catalunya*, la ética es un contrato social y es la sociedad quien ha de fijar qué es ético. En este sentido, la catedrática de Ética y Filosofía del Derecho Moral y Político de la Universidad Autònoma de Barcelona (UAB), Victòria Camps, escribe que la IA no

puede obviar las preguntas básicas de la ética que hacen referencia a la privacidad y la confidencialidad.

Peiró reflexiona sobre la necesidad de confiar para evolucionar como sociedad. Según la periodista, confiamos en la tecnología porque sabemos que, en caso contrario, no evolucionaremos, porque la historia nos ha demostrado que la tecnología siempre nos ha hecho avanzar como seres humanos. Pero para confiar en ella necesitamos conocer todos los beneficios y los riesgos que implica tanto la que ya existe como la que está por venir. Solo de este modo podemos ser críticos y valorar en cada momento que dosis de tecnología queremos en nuestras vidas.

La realidad es que la tecnología, la IA y los asistentes de voz forman parte ya de nuestro día a día, y se hace difícil concebir un mundo donde no tengan un papel principal. Esto nos lleva a pensar en el tipo de sociedad que queremos construir y por eso es tan importante pensar también en qué tipo de tecnología queremos desarrollar. Una cosa, ineludiblemente, va unida a la otra.

Lo que está claro es que el interés por la IA es creciente y ya no concebimos una sociedad sin ella. Pero también es necesario pensar en el impacto que tiene y tendrá sobre las personas. Ya se han publicado guías e informes con recomendaciones, como por ejemplo, la guía ética de la Comisión Europea para una IA fiable (2019). Y, de hecho, ya el artículo 12 de la Declaración Universal de los Derechos Humanos, recoge que las personas tenemos derecho a la privacidad. Ahora bien, cómo entendemos ese derecho y qué límites ponemos entre la protección de datos y la IA, depende de cada sociedad, puesto que no hay una ética global que se pueda aplicar a todos los países.

Ramón López de Mántaras, que impulsó junto a Luc Steels la Declaración de Barcelona, un manifiesto que propone un uso adecuado de la IA en Europa, ha reflexionado sobre ello en una entrevista para la revista Barcelona Metròpolis, realizada por el periodista Lluís Reales (Reales, 2020): “Excepto en cuestiones muy básicas, no podemos hablar de una ética universal independiente de los países y las culturas. (...) La responsabilidad es de los diseñadores y programadores, y también de los usuarios. Aquí es donde hemos de poner el énfasis y el control. I, así mismo, convendría incorporar más contenidos de ética en los estudios científicos y de ingeniería. De hecho, la ética tendrá presencia en los nuevos grados de inteligencia artificial que preparan las universidades catalanas”.

3.2. Los asistentes virtuales de voz, un crecimiento rápido y exponencial

Hasta la llegada de los asistentes de voz, las personas hemos hablado oralmente con las máquinas mientras las usamos, pero nunca hemos obtenido respuesta. Ha sido una especie de comunicación entre nosotros y nosotros mismos. Sin embargo, con la llegada de Alexa, Siri, Google Assistant y otros, se ha cambiado este monólogo por un diálogo, porque de repente las máquinas son capaces de respondernos de algún modo (Fortunati & Edwards, 2020).

Como hemos visto, se trata de un diálogo claramente condicionado por las limitaciones y automatismos del “casi-interlocutor” virtual y ciertamente alejado de lo que sería una conversación entre personas, pero no obstante podemos y debemos considerarlo diálogo. Los humanos hemos pasado de hablar a través de las máquinas, como por ejemplo el teléfono, a hablar con ellas y generar una conversación, sabiendo que detrás no nos espera una persona, sino Inteligencia Artificial (IA) que pretende simular al humano.

Los asistentes de voz empezaron a ser conocidos en 2011 gracias al lanzamiento de Siri, un programa informático de IA basado en voz que prometía hablar y, además, hacerlo con actitud. La compañía Apple conseguía, de esta manera, hacer accesible al público general la IA, a través de la voz.

Desde entonces, los asistentes de voz han evolucionado y actualmente son muchos los que encontramos en el mercado. Algunos de los más valorados y utilizados por los usuarios son: Siri (incorporado por primera vez por Apple en el iPhone 4S), Google Assistant (el asistente inteligente de Google), Alexa (el asistente virtual desarrollado por Amazon y utilizado por primera vez en los altavoces inteligentes Amazon Echo), Cortana (el asistente de Microsoft) y Bixby (el último asistente virtual en unirse al grupo, que ha sido desarrollado por Samsung).

Si nos guiamos por los datos que se presentaron en el informe sobre el análisis de ventas realizado por Strategy Analytics, en 2018 las compras de altavoces inteligentes crecieron casi un 200% en el último trimestre del año, alcanzando los 22,7 millones de unidades a escala mundial. Esto denota un interés por parte de la sociedad y una tendencia claramente creciente de esta nueva tecnología. Amazon y Google lideraban el ranking con el 55% de la cuota de mercado, siendo Amazon la marca de altavoces inteligentes mejor clasificada del mundo (Keynes, 2018).

Ambas corporaciones han apostado por la voz como tecnología emergente porque ven estas interfaces como una manera más sencilla, económica y natural para interactuar - pese a las barreras que existen entorno al *natural language understanding* y al entendimiento de los distintos idiomas, un tema en el que profundizaremos más adelante -. Steve McLendon, jefe de productos por voz de Google, asegura que ven la voz como “la plataforma ubicua que está siempre contigo y te permite hacer cosas en el mundo real” (Newman, 2018). Por su parte, el vice presidente de servicios y dispositivos de Amazon, Dave Limp, habla así de Alexa: “Pensamos en ella como en la computación ambiental, que es el acceso a la computadora menos dedicado personalmente a ti, pero que es mucho más ubicuo” (Gibbs, 2018).

Un dato destacable de la historia del asistente de voz Alexa se remonta a noviembre de 2014, cuando Amazon presentó oficialmente su asistente junto con su familia de altavoces inteligentes Amazon Echo. El origen del que es ahora uno de los asistentes de voz más populares del momento tiene un claro anclaje con el imaginario cultural de la IA que se nos presenta en las ficciones. Según aseguran sus creadores, se inspiraron en la voz del ordenador y el sistema de conversación del Starship Enterprise en series de televisión y películas de ciencia ficción, principalmente Star Trek.

La evolución del ecosistema de voz ha sido exponencial en los últimos años y, para crear contexto, vamos a destacar algunos de los hitos que han marcado la última década a través de esta línea de tiempo que nos ofrece la consultora de marketing Prodigioso Volcán (2019):

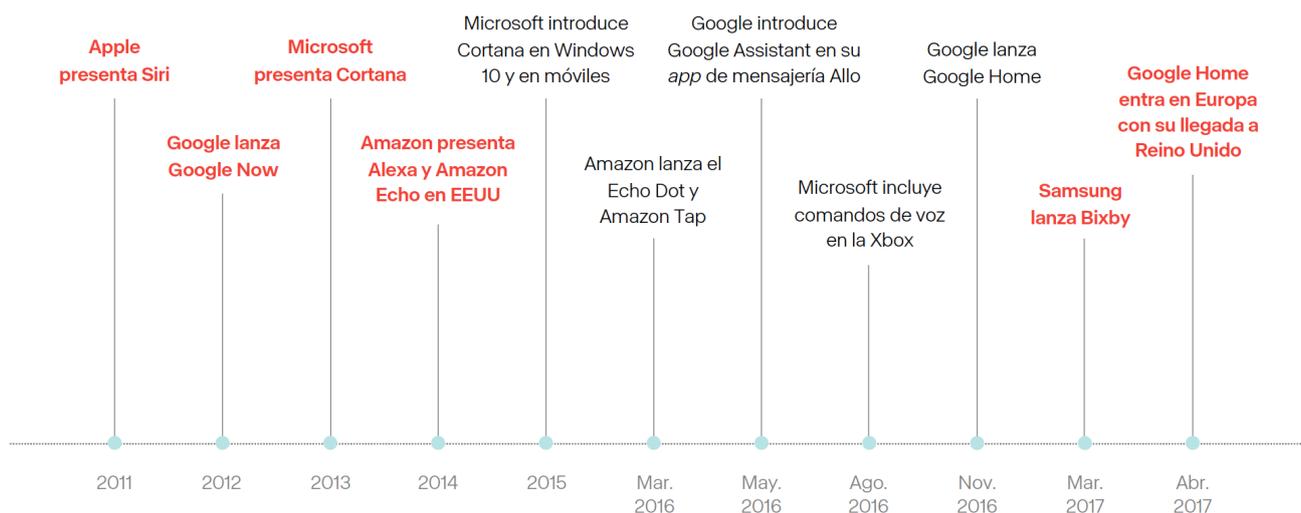




Figura 1: ¿Cómo suena tu marca? Estado de la voz y el audio en España (2019)

En la gráfica vemos que, pese a que los asistentes de voz han estado presentes desde 2011 con el lanzamiento de Siri, de Apple, no es hasta 2016 y 2017 cuando empieza a popularizarse en los hogares y a tener un crecimiento exponencial.

Esta popularización de las tecnologías por voz, como vemos, ha sido liderada e incentivada por las grandes compañías del sector, especialmente por Amazon y Google, que han visto en el último año como los ecosistemas de voz y audio, ahora sí, se están expandiendo a un ritmo mucho más rápido. De hecho, Amazon superaba en septiembre de 2019 las 100.000 aplicaciones de voz para su

asistente Alexa. Y a principios de 2019, Google contaba con 4.253, un 150% más que un año antes (¿Cómo suena la tu marca? *Estado de la voz y el audio en España*, 2019).

En este punto, y antes de avanzar, cabe definir algunos de los conceptos habituales que utilizamos cuando hablamos de tecnologías de voz:

- **Asistentes de voz:** Se trata del *software*, el programa informático, diseñado por empresas tecnológicas. Es decir, la Inteligencia Artificial (IA) con la que nos comunicamos. Por ejemplo: Alexa, Siri o Google Assistant.

- **Altavoces inteligentes:** Es el *hardware* que incorpora al asistente de voz y que también han desarrollado las empresas tecnológicas en cuestión. Por ejemplo, Amazon Echo (desarrollado por Amazon y que integra a Alexa) o Google Home (desarrollado por Google y que integra Google Assistant). Estos constituyen el reclamo comercial más evidente de las tecnologías por voz en los últimos meses.

- **Dispositivos inteligentes de voz:** Estamos hablando de los distintos *hardwares* que pueden integrar a la IA por voz, es decir, a los distintos asistentes virtuales. Puede ser un teléfono móvil, una televisión, un coche o, también, los altavoces inteligentes anteriormente mencionados. A medida que avance la tecnología, iremos asistiendo a una dispersión de los asistentes de voz en multitud de dispositivos distintos, con los que podremos interactuar gracias a las invocaciones. De hecho, el éxito de los asistentes de voz va mucho más allá de los altavoces inteligentes. Sin ir más lejos, en septiembre de 2019 Amazon presentó un amplio surtido de productos en los que ha integrado a Alexa: un microondas, un reloj de pared, un dispositivo para vehículos... (Prego, 2019) y se integrará próximamente en televisores Hisense y Panasonic, espejos de baño Kohler, ordenadores HP, Acer y Asus, y coches Toyota y Lexus. Por su parte, Google también anunció integraciones con televisores LG, auriculares Sony y pantallas inteligentes Lenovo (Wakabayashi & Wingfield, 2018).

- **Aplicaciones de voz:** Se trata de aplicaciones que crean desarrolladores externos, para que sus contenidos estén presentes en los distintos asistentes de voz.

Estamos ante una de las revoluciones tecnológicas que va a tener una implantación más veloz (de hecho ya la está teniendo). Los asistentes de voz, y en consecuencia las *Vocal User Interfaces*

(*VUI*), están creciendo mucho más rápidamente de lo que lo hicieron los teléfonos inteligentes y las tabletas, en sus respectivos momentos (Moon, 2018). Por este motivo es interesante ver su grado de penetración en la sociedad y conocer cómo estos asistentes virtuales por voz están siendo utilizados actualmente.

3.2.1. La penetración de los asistentes de voz en la sociedad

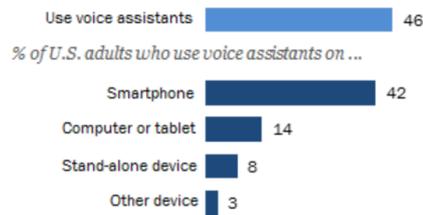
Según el estudio anual de tendencias digitales de Mary Meeker, en 2016 el 20% de las búsquedas en Google hechas con un teléfono móvil se hicieron por voz, en lugar de texto (Yurieff, 2017), y esto no fue más que el principio de la eclosión de esta tecnología. En 2017, los asistentes virtuales de voz comenzaron a expandirse de manera notable y, a día de hoy, según indica Google, una de cada cinco búsquedas ya se hacen a través de la voz (Ivanova, 2019).

El nivel de penetración de los asistentes de voz en los hogares está creciendo de manera exponencial y los datos lo demuestran: Google Assistant, a finales de 2019 ya estaba presente en más de 1.000 millones de dispositivos gracias a su sistema operativo Android. Por otro lado, Alexa abarcaba 100 millones de dispositivos, una cifra notable si se tiene en cuenta que este sistema solo puede operar en aparatos inteligentes que sean compatibles con sus altavoces Echo. De hecho, los expertos calculan que alrededor de la mitad de las búsquedas en Internet serán por voz a finales de 2020 (Sañas, 2019) y que en 2021 más de 75.000 millones de dispositivos tendrán asistentes virtuales integrados.

Parece claro que el futuro a corto plazo de estos asistentes es positivo y se augura un crecimiento exponencial a nivel mundial. De hecho, esta tecnología ya empieza a alcanzar audiencias *mainstream*. Según un estudio realizado por el Pew Research Center en 2017 («Nearly half of Americans use digital voice assistants, mostly on their smartphones», 2017) el 46% de los adultos en Estados Unidos afirmaba utilizar la voz para interactuar con algún dispositivo electrónico inteligente. Lo más habitual es hacerlo a través del teléfono móvil (42%), seguido del ordenador o la tableta (14%) y, por último, de los altavoces inteligentes (8%). Estas últimas cifras demuestran la importancia de distinguir entre asistentes de voz, que como hemos visto pueden estar integrados en multitud de dispositivos inteligentes distintos, y altavoces inteligentes, que son uno de los muchos dispositivos que pueden integrar estos asistentes virtuales por voz.

46% of Americans use digital voice assistants

% of U.S. adults who say they ever ...



Note: "Stand-alone device" refers to devices such as the Amazon Echo or Google Home. Respondents could choose more than one of the above options.

Figura 2: Survey Conducted. Pew Research Center

Si nos centramos en estos últimos dispositivos, los altavoces inteligentes, un año después un estudio realizado por el Reuters Institute sobre el futuro de los asistentes de voz y sus oportunidades (Newman, 2018), donde se analiza la penetración de estos asistentes en Estados Unidos, Reino Unido y Alemania, concluía que el 14% de los adultos en Estados Unidos utilizaba estos altavoces de manera habitual. Ello supone un incremento de penetración de mercado del 6% en solo unos meses. Este mismo estudio, aporta también datos de Reino Unido y Alemania, donde este porcentaje es ligeramente inferior, del 10% y 5% respectivamente, pero sin embargo la tendencia es muy positiva, puesto que las cifras se doblaron en el último año.

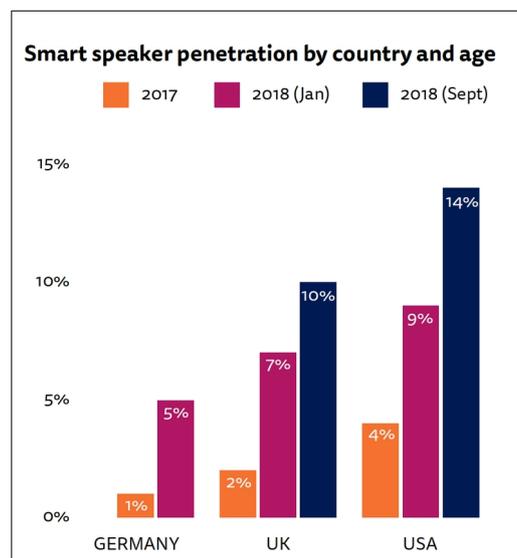
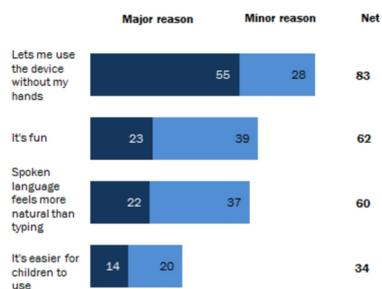


Figura 3: Digital News Report Q

Es interesante también analizar el motivo por el cual estos usuarios deciden utilizar asistentes de voz. Según el estudio del Pew Research Center, gran parte de los usuarios de asistentes digitales de voz (55%) aseguran que el mayor motivo por el cual los utilizan es porqué, de ese modo, pueden interactuar con la máquina sin necesidad de utilizar sus manos. Otros motivos que también aparecen en el estudio pero ya no tan mayoritarios son: porqué son divertidos (23%), porqué cuando hablas con ellos sientes que es más natural que escribir (22%) o porqué estos programas son más fáciles de utilizar para niños (14%).

Most digital voice assistant users cite freeing up their hands as a major reason for adoption

% of digital voice assistant users who say the following are major/minor reasons for using them



Note: Respondents who chose "not a reason" or did not give an answer for each option are not shown. Figures may add to more than 100% because multiple responses were allowed.

Figura 4: Pew Research Center

La penetración de los asistentes de voz en España

En Junio de 2018 llegó a España el altavoz inteligente de Google, el Google Home, y abrió el mercado de los *smart speakers* en el país. Cuatro meses más tarde, en octubre de 2018, Amazon puso a la venta su familia de altavoces Echo en el territorio. Estos dos acontecimientos abren nuevas oportunidades con respecto a las tecnologías lideradas por la voz en España.

Según se extrae del informe *¿Cómo suena tu marca? Estado de la voz y el audio en España*, realizado por la consultora Prodigioso Volcán, más de dos millones de internautas españoles usan actualmente altavoces inteligentes (*¿Cómo suena la tu marca? Estado de la voz y el audio en España*, 2019). Se trata de un estudio que pretende trazar una radiografía precisa del audio en España, con especial énfasis en los altavoces inteligentes, utilizando una muestra de 1.053 personas encuestadas.

En él observamos datos esclarecedores del panorama de los asistentes de voz en España: un 50'59% de los internautas españoles utiliza los asistentes de voz y lo hace, generalmente, a través de uno o varios dispositivos, especialmente las tabletas y los teléfonos inteligentes. De estos, el 22% los utiliza de manera diaria y el 24'8% al menos una vez por semana.

Los teléfonos inteligentes son el principal dispositivo a través del cual los usuarios invocan a los asistentes de voz (el 90,4%) y eso justifica el liderazgo de Google Assistant en el ranking de los asistentes más utilizados por los usuarios españoles. El asistente de Google tiene una cuota de penetración en España de 66,1%, seguido por Siri de Apple (28,4%) y, a mucha distancia, por Alexa de Amazon (8'2%).

La ventaja de Google, en este terreno es importante y se debe a que la compañía es propietaria del sistema operativo Android, que incluye preinstalado Google Assistant y que tiene en España una cuota de mercado del 90%; frente al 8,9% de los iPhones de Apple (que tienen Siri instalado por defecto).

La carrera para conquistar el mercado de los altavoces inteligentes, sin embargo, está mucho más ajustada. Según otro estudio, VoiceISDigital, realizado por ISDI y Webedia («El estudio que analiza el uso de las tecnologías de voz en España», 2020) la posesión de altavoces inteligentes en España supera el 15% y son uno de los dispositivos con mayor previsión de compra próxima (12%). Además, destaca que un 54% de la población española utiliza la voz para realizar alguna acción con cualquiera de sus dispositivos inteligentes.

La marcas líderes en el mercado de altavoces inteligentes son: Amazon (37%), con su familia de dispositivos Amazon Echo que integran el asistente Alexa; Google (33%), gracias a Google Home que integra Google Assistant; y Apple (13%), con el Home Pod que integra Siri.

Es significativo destacar que, por ahora, aún son mayoría (77,9%) los altavoces inteligentes sin pantalla. Solo el 22,1% de los usuarios que posee este tipo de dispositivos la tienen, pero, a medida que avance la tecnología, las interfaces multimodales (con voz y pantalla) van a tener una tendencia claramente al alza porque permitirán sofisticar la experiencia de usuario.

La penetración de los asistentes de voz en Cataluña

La realidad en Cataluña con respecto a la penetración de los asistentes de voz es similar a la que hemos visto en España. La Direcció General de Política Lingüística de la Generalitat de Catalunya ha presentado un informe donde se analiza la demanda comercial y social de los asistentes de voz en catalán (Assistents de veu en català. Una demanda comercial i social, 2019), que concluye que casi la mitad de los catalanes (47,4%) sí interactúa habitualmente a través de la voz con aplicaciones de móvil y otros aparatos como ordenadores o altavoces inteligentes, con lo cual ya lo podríamos considerar un uso habitual. De estos, el 44,3% lo hace a través del teléfono inteligente, el 17% a través del ordenador, el 6% a través de la tableta, y el 4% a través de altavoces inteligentes.

Además, el 34,2% de los catalanes desearía tener un altavoz inteligente al cual poder hablar y pedir que realice determinadas acciones, lo que indica una proyección muy favorable del mercado en Catalunya.

Sin embargo, no hemos de pasar por alto que la penetración de la tecnología por voz y los altavoces inteligentes en el mercado español y catalán va estrechamente ligada a la posibilidad de estos asistentes de entender y comunicarse de manera natural en el idioma del territorio en cuestión. En castellano ya es una realidad, aunque la calidad del *natural language understanding* necesite algunas mejoras para evitar frustraciones por parte del usuario, pero en catalán aún no hay posibilidad alguna de comunicarse con los asistentes de voz.

Este es, sin duda, uno de los factores que más ha condicionado la consolidación de estas interfaces vocales en España y Cataluña, puesto que el desarrollo de los asistentes de voz en sus inicios estuvo fuertemente vinculado al inglés (Argila, 2019). Estamos ante uno de los grandes retos a los que se enfrenta esta revolución en muchos mercados no liderados por la parla inglesa, que desarrollaremos en próximos apartados.

3.2.2. ¿Para qué utilizan los asistentes de voz, los usuarios?

La generalización de las interfaces de voz no solo ha revolucionado la forma de relacionarnos con la tecnología, sino que también replantea como se realiza una búsqueda e incluso como se compra un producto. Pero ¿para qué utilizan mayoritariamente los usuarios los asistentes de voz y qué tipo de acciones invocan?

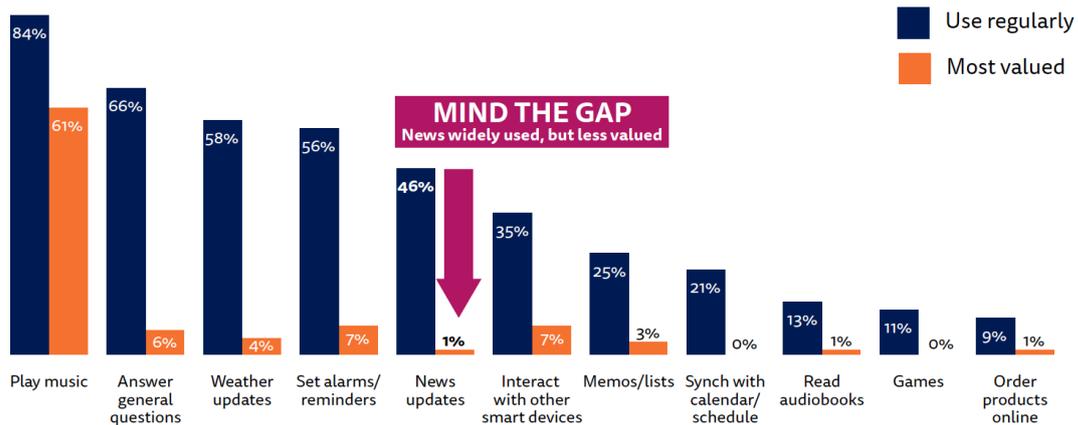
Los asistentes de voz pueden ayudarnos a buscar información en Internet, escuchar música e incluso, según su sofisticación, pueden incluso enviar mensajes, realizar llamadas, programar alarmas o controlar las luces de casa (Argila, 2019). La mayoría de las acciones que se realizan a través de estos asistentes son tareas simples y directas.

El informe del Reuters Institute (Newman, 2018) nos da algunas claves para entender cuales son los usos que los usuarios le dan mayoritariamente a estos dispositivos inteligentes. Para ello, nos basamos en el análisis que realizó en Reino Unido donde preguntó a los usuarios cual era el uso que le daban más regularmente a los altavoces inteligentes. El uso principal, y con diferencia respecto al segundo, fue escuchar música (84%), seguido de contestar preguntas generales (66%), obtener actualizaciones meteorológicas (58%), programar alarmas (56%), obtener actualizaciones de noticias (46%) o interactuar con otros dispositivos inteligentes (35%).

El análisis iba más allá y no solo analizó qué tipo de acciones llevaban a cabo los usuarios, sino como las valoraban. Escuchar música no solo resultó ser la funcionalidad más utilizada sino también a la que más valor daban los usuarios, con una valoración del 61%. La segunda más valorada por los usuarios es, aunque a mucha distancia de la primera, programar alarmas y interactuar con otros dispositivos (7%), y le siguen las preguntas de carácter general (6%) y las actualizaciones meteorológicas (4%).

Por último, nos interesa destacar en este punto, un aspecto relacionado con el consumo de información y noticias que, como hemos dicho, se utiliza de manera bastante generalizada (un 46% de los usuarios afirma utilizar los asistentes para consumo de información). Sin embargo es una de las menos valoradas: solo un 1% de los usuarios afirma que esta es la funcionalidad más importante para ellos.

Top/most valued features on smart speakers (UK)



Q. Which, if any, of the following features do you use/is most important on your speaker? Base: UK All that own a smart speaker & are aware of its features = 185.

Figura 5: The Future of Voice and the Implications for News (Newman, 2018)

En España, según el informe de ISDI y Webedia («El estudio que analiza el uso de las tecnologías de voz en España», 2020) escuchar música (81%), escuchar la radio (48%) y buscar información (45%) son los tres principales usos que se da a los asistentes de voz. Le siguen las aplicaciones de carácter más práctico y del hogar, como por ejemplo poner alarmas (44%), realizar recordatorios (34%) y controlar los dispositivos de la casa, como la luz (27%).

Además, iProspect, la agencia global de marketing digital del grupo Dentsu Aegis Network, analizó el uso y penetración de la voz en el país a través del estudio *“The Future is Voice Activated”* («La voz recupera su protagonismo», 2018). Se extrajeron conclusiones sobre los posibles usos en los que a los usuarios les gustaría usar asistentes de voz. Entre ellos destacan: la interacción con aplicaciones como la mensajería instantánea (49%) o las de contenido en *streaming* (23,1%), la posibilidad de pedir comida a domicilio (33,4%) y realizar compras por voz (26,6%).

3.2.3. Principales retos de los asistentes de voz

La privacidad de los usuarios

Como hemos visto en apartados anteriores, la tecnología y la IA están ocupando un espacio cada vez más amplio en nuestras vidas y en nuestra sociedad. Los asistentes de voz son un claro ejemplo de ello, puesto que la tecnología en sí se ha convertido no solo en una herramienta, sino en un socio en el proceso de comunicación (Guzman, 2016). La disrupción de las interfaces de voz “ha

aparecido en un momento en el cual la ética tecnológica está más en boga que nunca y, cada vez más, se está estudiando si el equilibrio entre lo que ofrece la tecnología y el valor que aporta a la sociedad a través de los datos es equidistante o no” (Fisa, 2020).

En 2019 se hizo público que las principales empresas que comercializaban altavoces inteligentes, principalmente Amazon y Google, realizaban escuchas en los hogares de todo el mundo de manera aleatoria, según ellos “para mejorar la experiencia del cliente” (Rubio, 2019; Rus, 2019): analizaban la manera como el usuario había hecho una petición y de qué manera la máquina respondía. Se trata del llamado entrenamiento supervisado, una técnica utilizada dentro del *machine learning*, a través de la cual las máquinas obtienen información a partir de la experiencia.

Esta ha sido una de las grandes críticas a los asistentes de voz y a los altavoces inteligentes: las escuchas activas. De hecho, fue un elemento de alto riesgo para la expansión de esta revolución tecnológica durante el 2019. Es necesario, para entender el concepto de escucha activa, conocer como funcionan las interfaces por voz: estas están continuamente escuchando, esperando una palabra clave para activarse. Una vez la identifican, la interfaz utiliza reconocimiento de voz y lenguaje natural para convertir el mensaje en comandos. Estos son enviados a la nube, donde se encuentra la inteligencia para hacer el procesamiento y análisis, y luego retorna la respuesta a la interfaz.

Los usuarios están poniendo en manos de las grandes corporaciones información personal, como sus gustos, estados de ánimo, intereses e incluso inclinación política, que permite configurar perfiles o patrones de consumo. Estos datos pueden ser utilizados por las propias empresas, e incluso por un tercero, para ofrecer posteriormente productos o mensajes personalizados. El punto más crítico no es este, sino que el usuario lo hace de manera inconsciente, ya que el asistente de voz en ningún momento informa de que, además de comunicarse con el usuario, también lo está haciendo con una empresa a la que le da toda su información personal.

Pese a las justificaciones y explicaciones de estas compañías, que aseguraban que lo hacían para mejorar la experiencia de usuario y para crear una inteligencia más sofisticada, esto provocó un gran rechazo por parte de la sociedad, que puso en juego la consolidación de la tecnología de la voz. De hecho, el informe del Pew Research Center confirma que el 27% de los americanos que dijo no utilizar asistentes de voz, lo hacía por cuestiones de privacidad («Nearly half of Americans use digital voice assistants, mostly on their smartphones», 2017).

Por eso, los grandes *players* del sector, han trabajado a contra reloj para garantizar una mayor privacidad de los usuarios y asegurar que ello no supone un freno para esta revolución. Una de las soluciones que ha aportado Google, por ejemplo, ha sido ofrecer claros indicadores de cuándo se está escuchando, tanto con tonos de audio como con iluminaciones de LEDs o animaciones en los dispositivos, que indican que se está escuchando de forma activa al usuario. Alexa, en septiembre de 2019, también incorporó comandos que permiten al usuario controlar lo que el asistente escucha. Así pues, según aseguran las fuentes oficiales actualmente los principales asistentes de voz en el mercado ya no realizan este tipo de escuchas.

No obstante, es importante recordar siempre que utilizamos tecnología, de este tipo o cualquier otro, que debemos tener cautela, y que la responsabilidad última es del usuario.

El desarrollo del *natural language understanding*

Uno de los grandes beneficios de las interfaces por voz es sin duda que ofrece la oportunidad a los usuarios de comunicarse de manera oral, con un lenguaje natural. Estamos ante la primera revolución tecnológica en la que la máquina habla el lenguaje de los humanos y no los humanos el lenguaje de la máquina. Y además lo hace con una de las herramientas más básicas y ancestrales que utilizamos las personas para comunicarnos: la voz. En el estudio que presenta el Pew Research Center, de hecho, el 22% de los usuarios encuestados aseguró que utilizaba los asistentes de voz porque hablar “es más natural que escribir”.

Sin embargo, aún estamos en una fase muy inicial de esta disrupción y la tecnología, pese que se está desarrollando a gran velocidad y los procesos de reconocimiento del habla son ya muy sofisticados, aún tiene un margen de mejora. En el mismo informe se preguntó sobre el funcionamiento de los asistentes y, la mayoría de usuarios, contestaron que respondían con exactitud a sus ordenes siempre (39%) o casi siempre (42%). A través de estos datos, observamos que los ratios de error en la interpretación de los algoritmos de los asistentes de voz es muy bajo: solo el 16% de los usuarios aseguró que estos asistentes respondían “no muy a menudo” a sus peticiones de manera exacta.

Según Fisa (2020) “es cuestión de meses, y no años, que la tecnología consiga la excelencia en el *natural language understanding*” y asegura que las máquinas llegarán a superar la paridad humana, es decir, la capacidad de un ser humano de interpretar lo que otro humano ha dicho.

No obstante, no hay que olvidar que el ritmo de desarrollo de las tecnologías de voz no es el mismo en todos los idiomas. El estudio anteriormente mencionado coge como referencia usuarios de Estados Unidos, que utilizan esta tecnología en inglés, el idioma principal e inicial en el que se han desarrollado la gran mayoría de los asistentes de voz que hay actualmente en el mercado.

Si observamos los resultados de estudios realizados en España, donde los asistentes de voz se utilizan generalmente en castellano, el grado de satisfacción es diferente. Para ello, vamos a centrarnos en el estudio de iProspect («La voz recupera su protagonismo», 2018) que expone que una de las principales barreras para el desarrollo de las interfaces vocales son, justamente, los fallos en usabilidad y fiabilidad, que reflejan las limitaciones que actualmente posee esta tecnología. En este estudio, uno de los principales inconvenientes que manifiestan los usuarios es que el asistente “no reconoce fácilmente la voz” o que da respuestas incorrectas (ambos con el 26%).

En este sentido, los usuarios puntúan por encima a Google Assistant que a Alexa:

- Alexa muestra un 44,9% de comunicaciones no comprendidas, un 39,9% correctas y el 15,2% equivocadas.
- Google Assistant, por su parte, tiene un 57,4% de respuestas correctas, el 32,2% de no comprendidas y un 1,4% de errores.

Uno de los principales problemas con los que se encuentra el usuario cuando quiere comunicarse con un asistente de voz en castellano tiene que ver con las invocaciones para que este realice una acción. Por lo general, estas frases han de responder a unas estructuras muy concretas para que el asistente sea capaz de descodificarlas, y suelen resultar incómodas para una persona castellanohablante. Pero ¿porqué? Parece evidente que los asistentes de voz mayoritarios, como Google Assistant o Alexa, no han hecho una adaptación profunda a la lógica del castellano, sino que han mantenido estructuras típicas del inglés y, simplemente, las han traducido (Argila, entrevista cualitativa, 2020).

A través del ejemplo que nos ofrece Argila relacionado con el asistente Alexa (2020) y Catalunya Ràdio, vemos con más claridad el problema. Cuando queremos invocar, por ejemplo, un audio a la carta concreto de una emisora de radio, Amazon pide forzosamente al usuario una formulación muy determinada:

“Alexa, pide a Catalunya Ràdio el podcast *Un restaurant caníbal a Berlí*”.

Esta estructura no resulta natural para un castellanohablante, ya que es una traducción directa del inglés “Alexa, ask BBC for...”. Lo natural para un hablante de castellano sería utilizar: vocativo + complemento directo + complemento circunstancial. Es decir:

“Alexa, pon el podcast *Un restaurant caníbal a Berlí* de Catalunya Ràdio”.

Lo que sucede con esta última invocación es que Alexa no es capaz de identificar qué es exactamente lo que le está pidiendo el usuario por la siguiente razón. El usuario está pidiendo al asistente que recupere un audio concreto de una aplicación desarrollada por terceros y que ha sido integrada dentro de su sistema, pero este usuario no menciona esta aplicación (o *skill*) hasta el final de su petición. Esto hace que la descodificación de la invocación sea muy confusa para Alexa, ya que no sabe si ha de contestar ella misma de manera natural o ha de ir a buscar una de las tantas *skills* desarrolladas.

Por ese motivo, la formulación que un castellanohablante realizaría de manera natural no es válida, ya que Amazon pide que se cite en primer lugar el nombre de la *skill* para que el asistente entienda que se trata de una petición dirigida a dicha aplicación. Esto genera, sin duda, un conflicto e interfiere en la esencia de esta tecnología, lo que conocemos como *natural language understanding*, ya que implica que las conversaciones que empiezan con un lenguaje natural, acaben desencadenando en un bucle de “lo siento no te entiendo”, generen frustración y ensucien la experiencia de usuario.

Aunque las búsquedas por voz en castellano se estrenaron en marzo de 2009, de la mano de Google, y Siri incorporó esta opción en 2011, no fue hasta 2018 que empezaron a llegar a España altavoces inteligentes como Google Assistant y Alexa. Esta tecnología de voz lleva solo dos años en territorio peninsular con lo que cabe esperar que el procesamiento natural del lenguaje mejore y permita obtener experiencias de usuario mucho más satisfactorias.

Los asistentes de voz y el catalán

Si ponemos el foco en la lengua catalana, los asistentes de voz están aún lejos de ser una realidad. Actualmente ningún asistente de las grandes marcas comerciales (Google Assistant, Alexa, Siri, Bixby y otros) es capaz de hablar catalán, pero sin embargo existe una comunidad de usuarios que desean comunicarse con ellos en este idioma. Así se desprende del estudio que ha llevado a cabo la Direcció General de Política Lingüística de la Generalitat de Catalunya (*Assistents de veu en català. Una demanda comercial i social*, 2019), que concluye que el 88,9% de los consumidores catalanes quieren que los asistentes de voz incluyan el catalán en su oferta de lenguas. El estudio también revela que un 47,7% de los consumidores catalanes ya se están relacionando con dispositivos tecnológicos mediante la voz. Esto, en cifras absolutas, supone 5,5 millones de usuarios potenciales.

Ante esta elevada demanda, el ejecutivo catalán aprobó en 2019 un acuerdo mediante el cual se exigirá a los contratistas de servicios tecnológicos que garanticen que estas tecnologías de voz utilicen el catalán; es decir, que solo utilizarán asistentes de voz digitales que hablen catalán.

Aún así, las grandes empresas tecnológicas que desarrollan este tipo de tecnologías de voz, por ahora, no han hecho movimientos firmes para incorporar el catalán entre su oferta de idiomas. Google, por ejemplo, rechazó valorar esta opción pese a una petición expresa del Govern de la Generalitat, y aseguró que su asistente ya estaba actualmente en 30 idiomas. Amazon Alexa, sin embargo, parece que ha cogido una posición más flexible ante el tema y abrió, aunque sin precisar, la posibilidad de incluir el catalán: “Es nuestra visión que Alexa esté dondequiera que estén nuestros clientes” (Baquero & Pueyo, 2019).

Tanto es así que a finales de 2019 la empresa hizo público que su asistente estaba aprendiendo catalán a partir de las aportaciones voluntarias de sus usuarios. Lo hace a través de Cleo, una *skill* que ha desarrollado la compañía a través de la cual los usuarios de todo el mundo pueden ayudar al asistente de voz a aprender idiomas que no tiene configurados. Esta funcionalidad es como un entrenamiento: el asistente pregunta al usuario qué lengua quiere hablar y, a partir de aquí, le hace preguntas para que responda en el idioma correspondiente. El objetivo de este método de aprendizaje es que Alexa vaya recopilando palabras y frases, y aprenda el idioma de una manera natural («Alexa aprende catalán», 2019).

Esta no es la única iniciativa activa para ayudar a incorporar el catalán en los asistentes de voz. La Fundación Mozilla y Softcatalà, impulsaron en 2018 un asistente de voz libre en catalán (Jané, 2018). El proyecto se llama Common Voice y se nutre de las aportaciones de los internautas, que consisten en leer en voz alta un texto que aparece en la pantalla del navegador y permitir que el micrófono lo grabe. Mozilla recoge los datos que comparte el usuario, como el acento, la edad y el género para identificar variedades dialectales, y, además, los usuarios también pueden contribuir enviando textos para ser leídos. El objetivo es crear un corpus lingüístico para cada idioma del mundo y construir una base de datos de voz de código abierto para que cualquier desarrollador pueda utilizarla y crear aplicaciones innovadoras en esos idiomas.

El catalán es uno de los idiomas que ya ha lanzado Common Voice y, en junio de 2020 recogía 422 horas de grabaciones validadas y 4.842 usuarios voluntarios. Sin embargo, para que los oídos de un robot procesen una nueva lengua, se calcula que hacen falta unas 2.000 horas de audios de unas 10.000 personas distintas, con varios acentos, sexos y edades representados (Baquero & Pueyo, 2019). Parece claro que este proceso complejo y costoso pasa por encima de la consolidada comunidad de usuarios que demanda el servicio en catalán, y es el principal freno para que los principales asistentes de voz incorporen esta lengua en su oferta.

Aún así, hay corporaciones y desarrolladores independientes que han encontrado la manera de incorporar sus contenidos en catalán en los asistentes de voz, a través de las *skills*. Por ahora, la invocación debe seguir siendo en castellano, pero el audio que se ofrece es en catalán. Un ejemplo que desarrollaremos más adelante como caso práctico es el de la Corporació Catalana de Mitjans Audiovisuals (CCMA).

Desde agosto de 2018 están trabajando para incorporar sus contenidos informativos y de entretenimiento en el asistente de voz de Amazon, Alexa. Judith Argila, la responsable de Nuevas Plataformas y Redes Sociales, en el departamento de Medios Digitales de la CCMA, explica que el idioma ha sido la principal dificultad y es donde se han concentrado el 95% de los problemas para desarrollar experiencias de usuario “dignas”. “Para salvar estas dificultades, los desarrolladores han tenido que hacer un trabajo minucioso con los algoritmos de Alexa” para conseguir que entienda las invocaciones de los usuarios que piden contenido en catalán.

Después de 3 años trabajando con la tecnología, han conseguido avances significativos y algunos contenidos de la CCMA están presentes en Alexa (lo veremos en profundidad más adelante), pero no sin renunciar a una experiencia de usuario de alta calidad: “Creemos que podemos ofrecer un servicio al precio de intentar educar mucho al usuario, y el usuario tiene que sacrificar una de las principales ventajas de las interfaces por voz que es el lenguaje natural. Estas tecnologías son super útiles y tienen mucho futuro si entienden lo que digo cuando hablo. Si ellas no me entienden a mí y tengo que adaptarme yo a ellas, volvemos a lo de antes: tengo que aprender un lenguaje específico para comunicarme con la máquina, porque esta no entiende el lenguaje natural” (Argila, 2020).

3.3. Los asistentes de voz en los medios de comunicación

3.3.1. Tendencias de consumo digital de la información periodística

Uno de los informes más exhaustivos sobre el consumo de información periodística en el mundo es el que lleva a cabo anualmente el Reuters Institute for the Study of Journalism. El último que han publicado, con datos de 2019, revela nuevos conocimientos sobre el consumo de noticias digitales basados en una encuesta a más de 75.000 consumidores online en 38 países de todo el mundo (incluido España) (Newman et al., 2019).

Una de las principales conclusiones tiene que ver con la tendencia a generar modelos de pago en los medios de comunicación. Pese a los esfuerzos de la industria, que inició un crecimiento en los modelos de suscripción el pasado 2018, parece que en el último año solo se ha registrado un leve incremento en el acceso a las noticias de pago online, ya sea a través de suscripciones, membresía o donaciones. También es relevante destacar que la mayoría de usuarios prefiere gastar su limitado presupuesto en entretenimiento (Netflix o Spotify, por ejemplo) antes que en servicios de noticias e información.

Los teléfonos móviles siguen siendo la principal fuente a través de la cual los internautas consumen noticias, y esta tendencia aumenta cada año mientras disminuye la lectura en ordenadores. En concreto, un tercio de los consumidores (el 66%) lo hace a través de estos dispositivos. Destacamos las cifras concretas de dos territorios, España y Reino Unido, puesto que nos serán de interés para el estudio de casos prácticos que llevaremos a cabo más adelante:

- En España, esta cifra crece un punto situándose en el 67%. Así mismo, un 21% de los usuarios españoles se informan a través del ordenador y un 20%, mediante tabletas.

- En Reino Unido, los internautas también han ido aumentando progresivamente el consumo de noticias a través de los teléfonos inteligentes, situándose a día de hoy en el 63%. El 29%, lo hace a través de tabletas y el 49%, a través del ordenador.

Los datos demuestran que la información se consume principalmente en dispositivos móviles y que estos hábitos se extenderán aún más cuando desaparezcan las generaciones más mayores, sin tantas habilidades digitales.

Otros datos interesantes: las aplicaciones de mensajería instantánea como WhatsApp se han convertido en el principal canal de discusión y compartición de noticias, y algunos usuarios han empezado a abandonar Facebook, aunque sigue siendo de lejos la red social más importante para las noticias.

El estudio constata también que existe una preocupación por parte de los consumidores de información periodística, con respecto a la desinformación, y debido a eso las plataformas de distribución de noticias y los medios de comunicación están invirtiendo esfuerzos en construir de nuevo la confianza de su público. De hecho, el nivel medio de confianza en las noticias ha caído un 2% respecto al año anterior y se sitúa en el 42%, y menos de la mitad (el 49%) “solo confía en los medios que ellos mismos usan”.

En este sentido, los consumidores consideran que los medios de comunicación son buenos informando sobre acontecimientos de última hora (los conocidos como *breaking news*) pero no tanto explicando estas noticias. Casi dos terceras partes siente que los medios hacen un buen trabajo manteniendo a la gente al día (62%), pero no ayudándolos a entender las informaciones que ofrecen (51%).

Ante estos datos, cabe preguntarse: ¿porqué la industria ha llegado hasta aquí? y ¿cómo el sector puede remontar y recuperar su sitio? Según el profesor universitario Albert Sáez, exdirector adjunto de *El Periódico* y actual director digital del Grupo Zeta, “el periodismo ha de mirarse el ombligo y preguntarse cual es su utilidad en la post-modernidad”. En el libro *El periodisme digital amb valor. Claus per a la sostenibilitat de la premsa*, el autor Enrique Canovaca parafrasea Sáez y expone lo siguiente:

“Tampoco ayuda la transformación de contexto social que impulsó el periodismo, basado inicialmente en tres grandes revoluciones: la Ilustración, la Revolución Francesa y la Revolución Industrial. Según explica Sáez, “mientras que la primera genera lectores, la segunda, votantes, i la tercera, consumidores. El periodismo es imposible sin los primeros, indispensable para los segundos e inviable sin los terceros”. Pero la cultura ilustrada, la democracia representativa y la economía de mercado conviven actualmente en una sociedad post-moderna sin referentes morales claros, con una desinformación constante y unos

canales de comunicación que pueden llegar a substituir las funciones tradicionales del periodismo”.

También interesante la reflexión del periodista de *El País* Lluís Bassets, y que se recoge en el mismo libro, que considera que “los diarios en papel se han quedado en manos de los grandes anunciantes, el objetivo de los cuales es influir en la opinión pública”. Así, hemos visto como muchos medios de comunicación han funcionado como simples portavoces de aquellos que mandan, en lugar de como catalizador de la realidad social.

Fisa (2020) habla de los medios de comunicación como los grandes titanes que han sobrevivido una y otra vez a las múltiples crisis que han vivido y que, de hecho, han encadenado: crisis de modelo, crisis de confianza por parte de la audiencia, crisis económica por su alta dependencia con mecanismos volátiles como la publicidad... El problema es que nadie tiene la receta para salvar al sector, aunque aparecen en el espectro algunas posibilidades.

Una de las opciones que actualmente parece tener más interés y proyección tiene que ver con la experiencia de usuario. De lo que se trata es de crear nuevas experiencias de consumo para encontrar la manera más eficiente de conectar con la audiencia. Los gigantes tecnológicos son conscientes de esta necesidad y por eso han apostado por nuevas interfaces de distribución que permiten a los medios mejorar la navegabilidad en el consumo de noticias. Además, los lectores reclaman a los medios la apertura de caminos de comunicación directos y sencillos, con una relación sincera e intensa (Canovaca, 2019) y, en este punto, encontramos las nuevas interfaces de comunicación por voz (la VUI).

Según datos del estudio del Reuters Institute, los altavoces inteligentes están ganando terreno y doblan las cifras de uso y penetración de un año a otro. En Reino Unido, por ejemplo, su uso ha crecido del 7% al 14% en un año; pero solo un tercio los utiliza para acceder a las noticias. Esta proporción del uso de asistentes de voz para consumir noticias respecto a cualquier otro tipo de consumo, también es destacable en España, donde solo un 2% de la población utiliza asistentes de voz para consumir información periodística, frente al 5% que los utiliza para otros consumos.



Figura 6: Reuters Institute. Digital News Report 2019

El estudio también destaca que los altavoces inteligentes con pantalla aún están lejos de generar un impacto en el mercado porque, según el informe, “la última cosa que la mayoría de consumidores quieren es más pantallas en sus vidas”.

Los asistentes de voz, sin duda, se han convertido en una de las plataformas de distribución y consumo de información periodística y entretenimiento con más proyección, pero ¿qué oportunidades ven en ellas los medios de comunicación?

3.3.2. Los asistentes de voz, una oportunidad para el sector de la comunicación

Hasta ahora hemos visto el nivel de penetración de los asistentes de voz en la sociedad y como estos son utilizados. Pero ¿sabemos qué potencial tienen para el sector de la comunicación, en general, y para la distribución y el consumo de la información periodística, en concreto?

Los datos indican que aún estamos en una fase muy incipiente de esta disrupción y más en lo relativo al consumo de información. Pese al rápido crecimiento de las tecnologías por voz, el consumo de noticias a través de estos asistentes es aún menor de lo que se esperaba, y la mayoría de los usos se centran en los llamados *news briefings* (de los que hablaremos más adelante). Ello es debido que muchos usuarios no conocen el amplio abanico de posibilidades a su alcance, no saben como acceder a su marca de noticias preferida o cambiar la configuración que tiene el asistente predeterminada, o están decepcionados con el tipo de contenidos que encuentran, puesto que la mayoría son reversionados de la radio o la prensa.

Algunas de las conclusiones del último informe realizado por el Reuters Institute nos ayudarán a crear contexto (Newman, 2018):

- Solo el 1% de los usuarios a nivel mundial considera que el acceso a la noticias es el principal motivo para utilizar asistentes de voz.
- El 1% de los usuarios en Reino Unido consideran que el acceso a noticias de actualidad es su característica más importante.
- Pese a que alrededor de la mitad de los usuarios de altavoces inteligentes dicen utilizar el dispositivo para consumir noticias, solo el 18% y el 22% (en EEUU y Reino Unido respectivamente) acceden a servicios de noticias de manera diaria.
- A los usuarios que utilizan regularmente las actualizaciones de noticias les gusta la brevedad, el control y el foco; y cerca de la mitad de estos se sienten mejor informados (el 56% en EEUU y el 45% en Reino Unido). Aún así, son muchos los usuarios que, según las conclusiones del estudio, se quejan de la calidad de los *news briefings*; sobretodo en referencia a cada cuanto se actualizan y la calidad de la producción. Además, algunos usuarios consideran que son piezas demasiado largas y preferirían actualizaciones de menos de un minuto.
- Parece que se empiezan a establecer hábitos de consumo entre los usuarios: la mayoría utiliza los servicios de noticias de los asistentes de voz por las mañanas y a última hora de la noche.

Los usuarios que no usan los asistentes de voz para acceder a noticias alegan que el principal motivo es la facilidad de acceder a ellas a través de otros dispositivos; y 1 de cada 10 dice que no sabe como acceder a ellas.

Aunque “sabemos muy poco sobre cómo los asistentes de voz pueden funcionar para la industria de las noticias”, como asegura Newman (Lladó, 2018), los medios de comunicación mantienen la vista puesta en esta nueva forma de comunicarse con la audiencia a través de la voz y han empezado a trabajar en variedad de estrategias para ofrecer sus contenidos por este canal.

Hay quienes están convencidos de la necesidad de invertir cuanto antes y de manera firme, mientras que la mayoría se muestran prudentes y le ven proyección a medio y largo plazo, como una oportunidad que despegará en la próxima década (Newman, 2018). Aunque esta prudencia también puede hacer que queden atrás: “Estamos en un entorno en el que se ha de apostar y aprender, y todo medio de comunicación que ahora decida hacerlo, mientras la sociedad va adquiriendo y consolidando en paralelo este nuevo canal de comunicación, va a obtener un beneficio en el futuro” Fisa (2020).

Las posibilidades de los asistentes de voz para los medios

Fisa (2020) asegura que los medios de comunicación tienen delante una nueva vía de supervivencia y evolución: la voz; especialmente aquellos que tienen una relación tan próxima con esta herramienta de comunicación como es la radio. Hemos de recordar, no obstante, que se trata de una revolución protagonizada por la voz pero no excluyente, es decir, que ha de ser mirada desde un punto de vista multimodal. Con lo cual, cualquier *broadcaster* o medio de comunicación impreso que haya mutado a digital es susceptible de ofrecer experiencias basadas en voz.

Pero ¿sabemos qué posición han tomado, a día de hoy, los distintos tipos de medios de comunicación ante este nuevo canal? Vamos a poner sobre la mesa los tres tipos principales: las radios, los canales de televisión y la prensa escrita.

- La radio es el medio de comunicación que más fácilmente se ha adaptado a los asistentes, puesto que su herramienta principal de comunicación también es la voz y desde sus inicios está habituada a llegar a la audiencia a través de ella. Aún así, han visto en este nuevo canal

una amenaza existencial y muchos han querido posicionarse rápido en estas plataformas para asegurarse un lugar privilegiado en el mercado (Newman, 2018).

Su apuesta es sencilla: re-empaquetar contenidos ya existentes para radio, como por ejemplo boletines informativos, emisiones en directo y podcasts; aunque cada vez son más los que experimentan con nuevos formatos que, en muchos casos, incluyen propuestas lúdicas.

- Para las televisiones, en la línea de las emisoras de radio, los asistentes de voz suponen una nueva oportunidad para hacer llegar sus contenidos a la audiencia, y muchos de los principales canales ya están presentes en los distintos asistentes de voz. La dificultad, en este tipo de medios, es obvia: las televisiones basan sus contenidos en imágenes, mientras que la herramienta principal de los asistentes es la voz (ya hemos visto que los altavoces inteligentes que incorporan pantalla aún son minoritarios en los hogares).

Entonces, ¿que han de hacer para que esta relación funcione? La propuesta de valor que ofrecen la gran mayoría de canales de televisión hoy en día, en cuanto a servicio informativo, consiste en adaptar sus contenidos al formato audio, es decir, ofrecer solo piezas que sean completamente válidas y entendibles sin el apoyo de las imágenes. Por ejemplo: ofrecer solo los titulares que se presentan en el sumario de un telediario. Aún son pocas las marcas televisivas que optan por desarrollar contenidos informativos propios explícitamente para asistentes de voz; aunque sí hemos visto ejemplos de éxito en el entorno del entretenimiento. Más adelante desarrollaremos algunos.

- La prensa escrita es, sin duda, el tipo de medio de comunicación que más ha evolucionado desde su aparición: de la tinta a la pantalla y, ahora, de la pantalla a la voz. El estudio *The Future of Voice and the Implications for News* del Reuters Institute nos ofrece pistas sobre el comportamiento de este tipo de medios ante los asistentes de voz que, asegura, han sido los más cautos del sector. Ello es debido a la necesidad de una alta inversión de recursos, sin tener seguridad alguna en cuanto a la monetización.

Por este motivo, los diarios que han empezado a trabajar con asistentes de voz lo han hecho con movimientos que requieren poca inversión, como por ejemplo los podcasts. Otra de las opciones por las que la prensa escrita está optando en este campo son los servicios de lectura de noticias automática por parte del asistente, cosa que ofrece un valor añadido al

consumidor. Además, algunas cabeceras critican la falta de visibilidad de su marca en el proceso de descubrimiento del usuario.

Parece claro que los medios de comunicación aún tienen que aprender a optimizar sus contenidos y programas para las interfaces vocales y están en el camino de conseguirlo. Pero para ello reclaman mejores herramientas que faciliten la integración de sus contenidos en los asistentes de voz existentes para adaptarlos a las nuevas formas que tienen los usuarios de acceder al mundo que les rodea; y un plan mucho más específico sobre las posibilidades de monetización de estas plataformas (Newman, 2018).

¿Cómo se integran los contenidos de los medios de comunicación, en los asistentes de voz?

Según el asistente de voz, las opciones de integración de contenidos para los medios de comunicación varían, pero todas comparten una idea común: el asistente tiene unas capacidades nativas de base, que se pueden ampliar añadiendo aplicaciones específicas desarrolladas por terceros (lo que en lenguaje de Amazon serían las *skills*, o en el de Google Assistant, las *actions*) (Argila, 2019).

Para definir las opciones de integración de los medios de comunicación en los asistentes de voz, nos centraremos en los dos mayoritarios en el mercado, actualmente: Alexa y Google Assistant. Ambos ofrecen opciones similares:

- **Resúmenes informativos**, también llamados *flash briefings* o *news briefings*. Esta funcionalidad nativa la ofrecen ambos asistentes de voz y da al usuario un breve resumen de noticias destacadas del día. El medio de comunicación en cuestión prepara piezas cortas con actualizaciones de última hora para que el usuario pueda consumirlas de manera sencilla y rápida. Lo único que tiene que hacer el usuario es configurar en su asistente cuales son las marcas informativas con las que quiere estar informado. El asistente ofrecerá a los usuarios los *news briefings* como respuesta a la pregunta: “Alexa, ¿cuáles son las noticias?” o “Ok Google, dime las noticias”.

Es, seguramente, el tipo de integración más sencillo y asequible para los medios de comunicación, pero también el que ofrece menos posibilidades y más riesgos para los medios. Uno de estos riesgos es la falta de identificación de la marca informativa.

- **Aplicaciones de terceros** (*skills* para Alexa o *actions* para Google Assistant). Se trata de aplicaciones desarrolladas completamente por un tercero, en este caso un medio de comunicación. Este controla la palabra de invocación, el tipo de interacciones que hará el usuario con el asistente, las respuestas que se ofrecerán, y la funcionalidad que ejecutará la aplicación: por ejemplo, reproducir una emisora de radio, leer unos titulares o dar una previsión meteorológica (Argila, 2019).

Las aplicaciones desarrolladas *ad hoc* son, sin duda, la opción que más posibilidades abre a los medios de comunicación, pero también la que requiere mayor inversión en tiempo y dinero. Por este motivo, muchos de los medios de comunicación que deciden apostar por asistentes de voz empiezan con los resúmenes informativos pero, si ven proyección, acaban desarrollando su propia aplicación.

En el mercado de habla inglesa, que como hemos visto es el más consolidado en cuestión de reconocimiento de voz y altavoces inteligentes, a principios de 2019 Amazon ya contaba con más de 4.000 *skills* categorizadas bajo la sección de “Noticias” (Argila, 2019).

Retos de los medios de comunicación ante las nuevas interfaces vocales

El principal reto al que se enfrentan los medios de comunicación es, en primer lugar, tomar la decisión de apostar o no por esta disrupción de las tecnologías de voz.

Como hemos visto, los datos indican que la proyección de los asistentes de voz, también para los medios de comunicación, tiene una tendencia claramente al alza con lo que cualquier editor de noticias querría estar allí donde está su audiencia. El conflicto nace de la necesidad de realizar una inversión de personal, tiempo y dinero para crear experiencias satisfactorias para el usuario y contenidos que encajen con los formatos que exige este nuevo canal de distribución de información.

Puede parecer que, a corto plazo, apostar por los asistentes de voz genere algunos problemas para los medios, pero a largo plazo “es una oportunidad”, asegura Newman. Y más, teniendo en cuenta que en el año 2022 el 55% de los hogares de Estados Unidos tendrá un asistente de voz (Lladó, 2018).

Grandes de las apuestas que se están haciendo con asistentes de voz están siendo lideradas por medios de comunicación, muchos de ellos públicos, “que se pueden permitir ciertas inversiones sin la necesidad de obtener un retorno inmediato” (Fisa, 2020). Pero la realidad es que en la mayoría de casos, los medios de comunicación se encuentran con barreras que les impiden invertir más en audio. Concretamente, la falta de recursos e información para diseñar y ejecutar una estrategia adecuada, o las dificultades para monetizar el contenido.

Justo este es el segundo gran reto con el que se encuentra el sector de la comunicación ante la disrupción de los asistentes de voz: la monetización. Se trata de un entorno por explorar y difícil de monetizar, pero sin duda con un enorme potencial. No obstante, los medios necesitan visualizar un modelo de negocio viable que les permita dar el paso y apostar por los asistentes de voz como canales de distribución de sus contenidos.

Actualmente, las opciones que se dibujan en este contexto son tres:

- La publicidad al uso. Es decir, integrar cápsulas publicitarias convencionales en los contenidos, como ya se hace actualmente en otros entornos.
- El *branded content*, basado en cápsulas de contenidos patrocinados.
- Los modelos de aplicaciones de semi-pago (también llamadas *freemium*). Es un modelo de negocio que está actualmente en crecimiento y por el que tanto Google como Amazon están empezando a apostar. Se trata de aplicaciones (*skills* o *actions*) gratuitas, pero con algunas funcionalidades extra por las que el usuario tiene que pagar.

Como explica Argila (2020) esta es una opción viable para los medios de comunicación privados, no tanto para los medios públicos.

A parte de las fórmulas de publicidad convencionales mencionadas, surgen otras soluciones creativas para monetizar los contenidos que se ofrecen a través de los asistentes de voz. John Trimble, el director de cuentas de Pandora, una plataforma de música y podcasts americana, ha subrayado el potencial de estas interfaces vocales, que ofrecen a los consumidores la posibilidad de responder a los mensajes de audio de una manera que antes no era posible: “Los anuncios de radio han estado ahí desde hace mucho tiempo, pero los oyentes, aún hoy, no pueden responder a un

anuncio. Sin embargo, Alexa – y los asistentes de voz en general – permiten al usuario interactuar de una manera muy directa con el dispositivo («From search to smart speakers: Why voice is too big for media companies to ignore», 2018). Esta opción abre vías para crear anuncios y cápsulas publicitarias, aún por explorar.

A parte de la necesidad de inversión y la poca seguridad en cuanto a monetización que observan los medios de comunicación ante esta nueva revolución, también existe un tercer reto o freno y tiene que ver con la identificación de la marca dentro de estas interfaces vocales: el *branding* sonoro.

Históricamente se ha asociado la técnica del *branding* a todo lo que tiene que ver con la imagen visual, es decir, los logos, la paleta de colores, las tipografías... Y la dimensión sonora de las marcas, por lo general, no se ha tenido mucho en cuenta (Fisa, 2020). La radios, e incluso las televisiones, no tienen un conflicto en este sentido puesto que están muy acostumbrados a crear una imagen de marca potente para cada uno de sus programas, también en referencia al universo sonoro. Sin embargo, las cabeceras de prensa escrita, no tienen esta visión tan consolidada. “Sin duda, esta revolución de la voz va a necesitar que las marcas entiendan que el mismo cuidado que han tenido en la dimensión visual icónica de su marca, van a tener que aplicarlo en la creación de un entorno auditivo”, asegura Fisa.

Lo mismo sucede con las marcas o empresas, independientemente de los medios de comunicación, que ven en los asistentes de voz un canal de comunicación prometedor con sus clientes. Estos van a tener que tomar consciencia de la importancia que el diseño de la identidad sonora está cogiendo y va a coger para las marcas. Un aspecto que sería interesante analizar en otros estudios.

Oportunidades de futuro

Las tecnologías por voz se encuentran, aún, en un estadio muy primerizo pese a su rápida implantación y penetración en la sociedad. Aún así, son muchos los escenarios y posibilidades de futuro que se plantean en torno a estas nuevas interfaces de comunicación para los medios.

Como hemos desarrollado en apartados anteriores, una de las grandes oportunidades de los asistentes de voz es la personalización de contenidos, a través de un canal tan ancestral e intuitivo como es la voz, y de esta oportunidad nacen escenarios de futuro interesantes por explorar para los medios de comunicación (muchos de ellos, claro está, dependerán de la evolución de la tecnología).

Las interfaces vocales nos sitúan en un entorno en el que podemos servir al usuario un contenido muy específico de manera muy sencilla: a través del diálogo. Hasta ahora cuando buscamos, por ejemplo, una entrevista concreta en un entorno web son necesarios unos circuitos de búsqueda que implican trabajo por parte del usuario. Con los asistentes de voz, sin embargo, este circuito se podría simplificar mucho: con una exhaustiva indexación de contenidos a nivel de metadatos se conseguirían extracciones de contenidos muy particulares a través de las palabras clave que diera el usuario en sus frases de invocación (Argila, 2020).

Imaginemos que nos han hablado de una entrevista interesante donde la periodista Pepa Bueno ha conversado con Javier Bardem sobre su afición por la cocina, y queremos recuperarla. En este futuro que planteamos, sería suficiente con pedir al asistente en cuestión: “Ponme la entrevista que hizo Pepa Bueno a Javier Bardem donde hablaban de cocina”, y a través de metadatos se podría extraer ese material concreto y ofrecerlo al usuario.

Otra de los escenarios de futuro que abren las interfaces vocales tiene que ver con la construcción de contenidos *ad hoc*, es decir, de relatos informativos propios del asistente de voz, independientes de la pieza informativa de un determinado medio de comunicación. Argila (2020) da detalles sobre esta opción: “Según la información que se consultara, el asistente construiría un diálogo propio a petición específica del usuario. Es un modelo mucho más personalizado donde el asistente, o incluso la marca informativa en cuestión, se convierte en una voz que puede responder a las necesidades de información o entretenimiento del usuario, como si fuera una conversación natural”.

Por ejemplo, podríamos preguntar al asistente: “Explícame qué pasó ayer en la manifestación del 8M en Barcelona?”, y el asistente sería capaz de construir y verbalizar un relato propio para darnos esa información.

Los asistentes de voz también pueden resultar una oportunidad para los medios de comunicación locales. Este canal puede representar una oportunidad para conectar de nuevo con su audiencia y abrir otros canales de distribución que les permitan ofrecer un servicio de noticias de proximidad. Uno de los aún pocos medios de comunicación de proximidad que ya ha apostado por este canal es Cugat Mèdia, el medio público de Sant Cugat del Vallès (Barcelona). Según se desprende del Observatorio Sociológico, que cada cuatro meses se hace en la ciudad, la ciudadanía santcugatena responde a un perfil muy tecnológico en lo que respecta al consumo de información (Lablanca,

2020). “Uno de nuestros objetivos es ser referente en innovación y hacer de la tecnología una aliada para conseguir llegar a nuestro público ofreciendo diferentes formatos adaptados a sus hábitos de consumo”, explica Mónica Lablanca, directora de Cugat Mèdia.

En enero de 2020 empezaron a trabajar en desarrollar su propia *skill* para Alexa y ofrecer, a través de esta aplicación, tanto radio en directo como contenidos a la carta. Durante la crisis causada por el Covid-19, las tendencias de consumo de los medios de proximidad crecieron exponencialmente y Cugat Mèdia decidió lanzar entonces sus contenidos en el asistente de voz de Amazon: “Queríamos facilitar el acceso a la información en un momento tan esencial como fue la pandemia”, explica Lablanca (entrevista cualitativa, 16 junio 2020). Y sigue: “Los medios de comunicación de proximidad tradicionalmente han estado considerados medios de segunda. Su uso o su consulta siempre ha respondido a una necesidad informativa concreta y ha requerido de la pro-actividad del oyente o usuario. En este caso, los asistentes de voz ayudan a hacer un paso más hacia adelante para democratizar el medio y su acceso, y hacer más eficiente el contenido.”

Si dejamos atrás los servicios estrictos de noticias y abrimos el foco, también se vislumbran caminos muy interesantes y por explorar relacionados con otros aspectos de la comunicación. Uno de ellos tiene que ver, también, con la información de proximidad y con los medios de comunicación locales públicos. Estos están fuertemente ligados con las instituciones de los territorios de su zona geográfica y, a través de los asistentes de voz, pueden ofrecer servicios de comunicación vinculados, por ejemplo, al turismo.

Este escenario no resulta tan lejano y, de hecho, ya hay medios de comunicación que han empezado a trabajar en servicios similares. Uno de ellos es la Xarxa Local Audiovisual (XAL) de Cataluña, que ha empezado a desarrollar experiencias para instituciones de zonas turísticas, a través de la cual ofrecer información de interés a los visitantes a través de servicios de voz (Garriga, entrevista cualitativa, 17 junio 2020). El proyecto se encuentra aún en un estadio muy inicial y preve lanzarse este 2020.

En una línea parecida, también sería interesante explorar un entorno en el que se pudiera integrar la información a nivel geográfico. El ejemplo más visual y claro lo encontraríamos en un asistente de voz integrado en un vehículo: en este escenario, durante un viaje podríamos pedir al asistente que nos diera información sobre el territorio en el que nos encontramos en ese momento con un simple: “Dame información sobre este sitio”.

El entretenimiento encuentra también muchas oportunidades en los asistentes de voz, y el infantil ha visto en estas nuevas interfaces un canal muy oportuno para llegar a su audiencia, ya que los niños y niñas pueden acceder fácilmente a los contenidos a través de la voz. Un ejemplo que nos ofrece Argila (2020) relacionado con este tema específico es el siguiente: se podrían ofrecer experiencias educativas y de entretenimiento para los más pequeños, como por ejemplo que los niños pudieran hablar con sus personajes preferidos de televisión, como el muñeco Mic de TV3, para que estos les enseñaran a lavarse los dientes o a vestirse.

Para la mayoría de los escenarios planteados, las tecnologías por voz aún están inmaduras pero las posibilidades y oportunidades de futuro para el sector periodístico y de la comunicación son muchas y diversas.

3.3.3. Algunas experiencias de éxito

Desde la expansión de los asistentes de voz integrados en altavoces inteligentes, numerosos medios de comunicación han hecho un paso rápido para ofrecer sus contenidos en estos entornos (Argila, 2019). Son ya muchas las emisoras de radio y *broadcasters* que ofrecen sus servicios informativos, y también agencias de noticias y diarios que han creado sus propias aplicaciones para asistentes.

En el ámbito estrictamente informativo destacamos algunos casos de éxito en el mercado de habla inglesa:

- **The New York Times** está presente en Alexa, Google Assistant y Siri, y ofrece multitud de opciones para sus usuarios. A principios de 2019 estrenó un resumen informativo diario, “The New York Times Briefing”, pensado para ofrecer las noticias más destacadas en un formato breve de 3 minutos. También ofrecen el podcast diario de The Times e incluso consejos sobre bienestar y salud en 1 minuto o entrenamientos de 6 minutos diarios («How to Get The New York Times on Your Smart Speaker», 2020).

- **The Washington Post**, a parte de ofrecer los últimos titulares y análisis de opinión de sus corresponsales (Argila, 2019), ofrece una de las opciones más novedosas: la posibilidad de hacer notificaciones *push* para avisar a los usuarios de que hay noticias destacadas de última hora. Como ya hemos comentado en apartados anteriores, esta opción da a los usuarios la

posibilidad de activar alertas para recibir las noticias de última hora a lo largo del día, pero solo el consumidor será quién decida en última instancia si las quiere escuchar o no, y cuándo hacerlo.

- La **CNN** ofrece también la posibilidad de escuchar pasivamente titulares o pedir información específica sobre un tema, como por ejemplo: “Las últimas noticias sobre el Covid-19”, gracias a su aplicación (Argila, 2019).

- En el mercado europeo, la **BBC** es el medio de comunicación que está liderando el terreno de los asistentes de voz. Actualmente está presente a través de Alexa y Google Assistant ofreciendo el directo de todas sus emisoras de radio, y resúmenes y boletines informativos con las noticias de última hora (*The BBC on voice devices*, 2020).

En el mercado español cada vez son más los medios de comunicación que ofrecen sus contenidos a través de los asistentes de voz: emisoras de radio como Cadena Cope, Cadena Ser o Onda Cero; agencias de noticias como Europa Press, o televisiones como Mediaset o TVE. Destacamos algunos ejemplos:

- La **Cadena Ser**, del grupo de comunicación Prisa, es una de las emisoras de ámbito nacional español que ha hecho una apuesta fuerte para tener presencia en los asistentes de voz. Gracias a su *skill* para Alexa y a su *action* para Google Assistant, los usuarios tienen acceso a todas las emisoras en directo, a sus servicios de podcast y a resúmenes de noticias de última hora.

- El diario **El País**, que pertenece también a Prisa, está presente en Alexa y Google Assistant desde 2018. El diario ofrece la posibilidad de escuchar resúmenes informativos diarios, a través de la lectura de sus titulares y noticias, y realizar peticiones por temáticas concretas: “Léeme noticias sobre las elecciones en Brasil”. Además, también se puede acceder a secciones de opinión («“Alexa, léeme las noticias de EL PAÍS”», 2018).

- El diario **ABC**, también está presente, desde finales de 2018, en el asistente de voz Alexa y ofrece un boletín de noticias elaborado cada día por la redacción llamado “Las Noticias de ABC”. Según el propio diario se trata de un “resumen ágil y directo de la actualidad de España y del mundo con los principales titulares de la información política, económica e

internacional, así como el deporte, la vida social, digital y cultural” («Amazon Echo llega a España con ABC entre sus fuentes de noticias de calidad», 2018).

En el terreno del entretenimiento son muchas las experiencias innovadoras que los medios de comunicación y algunas plataformas *over-the-top* (OTT) están llevando a cabo con asistentes de voz. Recuperando el caso de la BBC, la cadena británica está trabajando con experiencias de voz para niños. “The BBC Kids skill” es una de las mejor valoradas y consiste en juegos basados en personajes de la marca infantil de la BBC: Cbeebies (BBC launches its first voice experience for children, 2018). Como dice Argila (2019) “este tipo de interacciones trascienden la distribución de programas de radio en directo o a la carta a través de altavoces inteligentes, y prometen crear un tipo de vinculación más activa y variada con los contenidos de las marcas”. Más adelante desarrollaremos extensamente el caso de la BBC.

En España, Televisión Española (TVE) de la mano de Lavinia Voice ha desarrollado también aplicaciones para asistentes de voz pensadas exclusivamente para el público infantil. Por ejemplo, “Cuentos Clan” y “Lunnis de leyenda”, dos aplicaciones disponibles para Alexa y Google Assistant que desembarcan en los altavoces inteligentes como una clara apuesta de entretenimiento interactivo de la cadena. Esta apuesta abre a los niños y niñas un “amplio catálogo de cuentos y una colección de aventuras interactivas con los Lunnis como protagonistas” (RTVE lanza «Lunnis de leyenda» y «Cuentos de Clan» para asistentes de voz, 2019).

La innovación con los contenidos es clave para que las experiencias de entretenimiento que se ofrecen a través de las interfaces vocales sean exitosas. Se está trabajando, incluso, con juegos interactivos que permiten crear itinerarios y tomar decisiones a los propios usuarios, mientras están inmersos en la experiencia digital. En este sentido, encontramos algunos ejemplos interesantes.

Uno de ellos nace de la plataforma OTT Netflix, que lanzó en abril de 2018 un audio-drama interactivo, disponible en Google Assistant, para promocionar la nueva temporada de la serie *Lost in Space*. La experiencia duraba entre cinco y seis minutos y a lo largo del hilo argumental lanzaba preguntas con diferentes posibilidades de respuesta para que el usuario decidiera qué camino tomar (Williams, 2018).

Otro ejemplo, esta vez integrado en el asistente de Amazon, sirvió para promocionar la película “*Batman vs. Superman*” gracias a un juego en el que se tenía que resolver un doble homicidio

(Roettgers, 2016). La BBC también ha experimentado con este tipo de experiencias por voz y, a finales de 2017, lanzó una ficción interactiva de 20 minutos para Alexa y Google Assistant: “The Inspection Chamber”. Una experiencia que animaba al usuario a “tomar parte a través de las interacciones por voz” (Cooke, 2017).

4. Trabajo de campo

4.1. El caso de la Corporació Catalana de Mitjans Audiovisuals (CCMA)

4.1.1. La Corporació Catalana de Mitjans Audiovisuals (CCMA)

La Corporació Catalana de Mitjans Audiovisuals (CCMA) nació el 30 de mayo de 1983 y es el ente público que gestiona los medios de comunicación audiovisuales de la Generalitat de Catalunya, los canales de Televisió de Catalunya y el grupo de emisoras de Catalunya Ràdio, además de los contenidos digitales generados por estos dos medios. También forma parte de la CCMA, la Fundació La Marató de TV3. Además, participa con un 30% en la Agència Catalana de Notícies (ACN).

La CCMA trabaja con “la misión de servicio público y con un firme compromiso de calidad, independencia y promoción de la cultura y la lengua catalanas”. Tiene como principal objetivo ofrecer a los ciudadanos de Catalunya, y en cumplimiento con el mandato del Parlament, un servicio público de calidad, comprometido con los principios éticos y democráticos, y con la promoción de la cultura y la lengua catalanas (*Misión y principios*, s. f.).

Todas estas funciones se han de cumplir, como destaca Geni de Vilar, adjunta a dirección del área de Medios Digitales, con una clara voluntad de liderazgo: “Nuestra misión de difusión de la lengua, la cultura y la identidad catalanas solo la cumpliremos si hacemos contenidos líderes. La gente no nos vendrá a ver porque hacemos cosas en catalán, sino porque informamos y entretenemos bien, y esto es lo que nos hará líderes” (De Vilar, entrevista cualitativa, 21 noviembre 2019).

La realidad de los medios de comunicación pasa por lo digital. Por eso, la CCMA está trabajando su presencia y liderazgo en todos sus canales a través de Internet y está inmersa, como dice de Vilar, en un proceso transformacional constante: “Existe la clara convicción de que Internet ha de estar al mismo nivel conceptual que la televisión y la radio, y que la empresa debe transformarse poco a poco”. Para hacerlo, el grupo se marca objetivos a corto y medio plazo en distintas áreas, para conseguir “impregnar la mancha digital dentro de la organización de la manera más armónica posible”.

Uno de esos objetivos tiene que ver con la integración de sus contenidos en los asistentes de voz; especialmente Alexa.

4.1.2. “Alexa, ¿cuáles son las noticias del día?”

La CCMA ha hecho una apuesta decidida por los asistentes de voz desde que empezaron a penetrar en el mercado español, en 2018. Veían en ellos proyección, puesto que se fijaban en los datos de crecimiento de los asistentes en el mercado de parla inglesa, y a pesar de que aún no había (ni, de hecho, hay) una congregación importante de usuarios alrededor de esta nueva tecnología entre los usuarios de los canales de la CCMA.

Sin embargo, decidieron dar un paso adelante para ofrecer información, entretenimiento y acompañamiento en este nuevo entorno liderado por la voz: “Aunque no había una masa crítica de usuarios a los que servir, nosotros valoramos que como servicio público y como proyecto de innovación tenía sentido que estuviéramos allí”, explica Argila (2020).

Dentro del departamento de Medios Digitales de la CCMA hay un pequeño equipo fijo (de tres personas: un perfil técnico, uno de producto y uno de contenidos) enfocado a desarrollar contenidos y estrategias de distribución para las distintas plataformas digitales y, desde entonces, se han dedicado al entorno de los asistentes de voz.

Su apuesta inicial se centra en el asistente de voz de Amazon, Alexa, para el que han desarrollado *skill* propia y tienen disponibles los resúmenes informativos o *flash briefings*. También han hecho una primera incursión en Google Assistant, de Google, aunque para esta plataforma la aplicación *ad hoc* (lo que en lenguaje de Google se llama *action*) no está disponible aún, debido a que el gigante tecnológico no permite incorporar, por ahora, contenidos en catalán. Ésta, de hecho, es una de las principales críticas de los usuarios: no poder acceder a los contenidos de los canales de la CCMA a través de Google Assistant.

Primer contacto con la tecnología: los *flash briefings*

“El primer contacto con la tecnología sirvió al equipo para hacer una inmersión en el mundo de Amazon Alexa y entender, así, cómo funcionaba y cual podía ser nuestro encaje dentro de ella”, explica Argila (2020).

En septiembre de 2018 la CCMA estrena en Amazon Alexa una colección de *flash briefings*, la forma más sencilla de implementación de sus contenidos en esta plataforma. Se trata de resúmenes informativos que los usuarios pueden integrar como fuente de noticias en su dispositivo con Alexa. Para activar los resúmenes con Alexa serán necesarios los siguientes comandos:

“Alexa, dame las noticias”
“Alexa, ¿cuáles son las noticias del día?”

Estos resúmenes informativos también están disponibles en Google Assistant. Para activarlos, son necesarias las frases de invocación siguientes:

“Ok Google, dame las noticias”
“Ok Google, ¿cuáles son las noticias del día?”

En ambos casos, los usuarios tendrán que configurar previamente su asistente de voz, para activar esta fuente de noticias concreta.

Para la elaboración de estos resúmenes informativos, se aprovechan materiales ya existentes que puedan tener sentido servidos bajo la lógica de los resúmenes informativos de Alexa y Google Assistant, y teniendo en cuenta que se van a consumir mayoritariamente desde dispositivos sin pantalla. Necesariamente han de ser piezas de entre 2 y 7 minutos.

Los contenidos que se ofrecen en estos *flash briefings* son los siguientes: los últimos boletines informativos de Catalunya Informació; los titulares de la portada del “Telenotícies Vespre”; la portada de Mònica Terribas; “[L@Net](#)”, d’Albert Cuesta, y la predicción meteorológica («TV3 i Catalunya Ràdio, els primers d’integrar continguts en català a Alexa d’Amazon», 2018).

Los resúmenes informativos son una manera fácil y rápida de tener presencia en los asistentes de voz como Alexa y Google Assistant. Pero es importante considerar que se trata de sistemas de integración limitados donde el medio de comunicación en cuestión no tiene el control absoluto, sino que es solo una fuente de noticias integrada dentro del paraguas de funcionalidades de Amazon o Google.

Desarrollo de *skills ad hoc*

El siguiente paso fue desarrollar *skills* propias para Alexa que les permitieran tener más libertad para decidir como querían comunicarse con su audiencia, como medio de comunicación.

La primera aplicación *ad hoc* que crearon fue la de su emisora de radio de referencia, Catalunya Radio, que se presentó en septiembre de 2019 y que los situó como la primera emisora generalista en catalán en ofrecer una *skill* propia en Amazon («Catalunya Ràdio estrena la nova “skill” per escoltar el directe des dels dispositius amb Alexa», 2019). Permitía, en un inicio, escuchar el directo de Catalunya Ràdio, pero meses después, en abril de 2020, empezaron a ofrecer también contenido de programas a la carta a través de podcasts o de secciones concretas («Los podcasts de Catalunya Ràdio llegan a los dispositivos con Alexa», 2020).

Esta *skill* incorpora, además, una sintonía hecha a medida y diseñada de forma específica para los nuevos entornos digitales de consumo de audio. Ello refuerza el *branding* sonoro que, como ya hemos comentado en apartados anteriores, es tan importante en estas nuevas interfaces vocales.

Para iniciar esta *skill* el usuario solo tiene que invocar la frase:

“*Alexa, inicia Catalunya Ràdio*”, si quiere escuchar el directo

o

“*Alexa, Catalunya Ràdio podcast*”, si quiere pedir un podcast

El último lanzamiento hecho des de la CCMA ha sido una *skill* para cada emisora de radio: el directo de Catalunya Informació, el directo de iCat y el de Catalunya Música (estas dos últimas aplicaciones permiten, incluso, preguntarles qué canción o intérprete está sonando en ese momento).

También han desarrollado una *skill* para los telenoticias de TV3, que se puede utilizar tengas o no un dispositivo con pantalla. La aplicación te ofrece los titulares actualizados de cada edición (“TN matí”, “TN migdia”, “TN vespre” y “TN cap de setmana”) y la previsión meteorológica. Además, ofrece la posibilidad de escuchar, si se desea, el último telenoticias entero.

Para activar estas *skills* el usuario deberá invocar la frase:

“Alexa, inicia (nombre de la skill)”

Dificultades en la integración de los contenidos en los asistentes de voz

Si nos fijamos en los *flash briefings*, el principal punto de conflicto tiene que ver con la dificultad de crear una identidad sonora de marca. Es un entorno poco controlable por el medio de comunicación y, por lo tanto, éste no puede jugar demasiado con los indicativos o las sintonías que permiten que el oyente construya en su imaginario la marca informativa que le está ofreciendo la información a través de Amazon (es este caso).

No obstante, el principal problema con el que se han encontrado durante la integración de sus contenidos en Alexa tiene que ver la creación de su *skill ad hoc* y con el conflicto idiomático. “Alexa habla y entiende castellano y nosotros desarrollamos contenidos 100% en catalán y, además, nuestra interacción con el usuario también es en catalán”, explica Argila (2020). “Estamos generando un mundo que se tiene que integrar en otro que, cuando habla, no le entiende”. Para la responsable del departamento de innovación de Medios Digitales de la CCMA, es en este punto donde han encontrado el 95% de las dificultades”.

Otro de los conflictos que han tenido los desarrolladores, también en la creación de *skills* propias, es la dependencia de las lógicas de Amazon. Uno de los ejemplos más claros tiene que ver con la experiencia de usuario, como comenta Argila (2020): “No somos 100% amos de la experiencia de usuario. Por ejemplo, si para invocar una *skill* el usuario utiliza una estructura de frase que Amazon no ha previsto ya no funciona, aunque solo haya cambiando el orden de las palabras o del verbo. Además, a veces esas formulaciones que nos obliga a hacer Amazon pueden resultar muy incómodas para un castellano hablante, porque son estructuras que se han traducido del inglés. No se ha hecho una adaptación profunda a la lógica del castellano”.

La respuesta de la audiencia

Aunque la comunidad de usuarios de asistentes de voz en Catalunya es aún pequeña, de un 4% aproximadamente (Argila, 2020), la respuesta de la audiencia de la CCMA a través de asistentes ha sido muy positiva.

Los servicios que ofrece la corporación catalana en estas interfaces vocales ronda, actualmente, los 5.000 usuarios mensuales. “Es una cifra pequeña para los consumos a los que estamos acostumbrados en televisión y radio, pero la sensación general de los usuarios es buena”, asegura Argila.

Uno de los reclamos de este colectivo de usuarios era la incorporación de podcasts y contenidos a la carta, que ya están disponibles actualmente. También reclaman, de manera vehemente, la incorporación de las aplicaciones o *actions* en Google Assistant. En este último aspecto, aún están trabajando puesto que Google no admite como válidos los contenidos que se presentan en catalán.

El gigante tecnológico considera que pedir los contenidos en un idioma (el castellano) y recibirlos en otro (el catalán) ensucia la experiencia de usuario. “Hace un año y medio que intentamos tener nuestras aplicaciones en Google Assistant, pero cada vez que enviamos material a certificar Google nos lo rechaza porque dice que está en un idioma que no es el de la interfaz. No contempla que el usuario pueda pedir unos contenidos en un idioma, el castellano, y los reciba en otro, el catalán, pese a que la realidad del territorio sea bilingüe”.

El único paso que han podido dar, por ahora, con Google Assistant ha sido paquetizar los contenidos de los resúmenes informativos o *flash briefings* y ofrecerlos a la audiencia, como hemos visto anteriormente.

4.2. El caso de la British Broadcasting Corporation (BBC)

4.2.1. La British Broadcasting Corporation (BBC)

La British Broadcasting Corporation (BBC) se fundó en octubre de 1922 y es el servicio público de radio y televisión del Reino Unido. Se dedica a la creación de productos informativos y de entretenimiento para radio, televisión y formatos multimedia.

La BBC tiene diez estaciones de radio cuya señal cubre todo el Reino Unido; 40 estaciones de radio local en Inglaterra, y seis estaciones de radio local ubicadas en Escocia, Gales e Irlanda del Norte. También opera varios canales de televisión en Reino Unido, de los cuales la BBC One y la BBC Two son los más emblemáticos. Los otros canales destacados son: BBC Three, BBC Four, BBC News, BBC Parliament, y dos canales infantiles, CBBC y CBeebies.

Otros servicios que ofrece la corporación británica: BBC News (encargada de producir noticias e información de actualidad para las divisiones informativos de televisión, radio e internet de la BBC; la BBC Online (que corresponde con su página web principal); BBC Entertainment; BBC World News, y BBC Earth, entre otros.

La BBC trabaja con el objetivo de producir contenidos para todo Reino Unido que, además, puedan consumirse en todo el mundo. Su misión como medio de comunicación público es “velar por los intereses del público, sirviendo a todas sus audiencias a través de la provisión de productos y servicios imparciales, de alta calidad y distintivos que informen, eduquen y entretenga” (*Mission, values and public purposes*, s. f.).

Como servicio público de calidad, la BBC tiene como uno de sus propósitos principales apoyar en el aprendizaje a personas de todas las edades, a través de productos o plataformas que resulten accesibles e inspiradoras. Además, y según aseguran en su página web oficial, “la creatividad es el alma de la organización”, lo que implica que creatividad y la innovación van de la mano para desarrollar nuevos enfoques y contenidos.

Uno de los exponentes más claros de esta apuesta por la innovación está en su proyecto con asistentes de voz, que empezó en 2017.

4.2.2. “Give me BBC News”

La BBC produce y alberga una gran variedad de contenidos que van desde programas de televisión, noticias, podcasts, radio, eventos en directo, etc., y han visto en los asistentes de voz una plataforma más para hacer accesible todo ese contenido a sus usuarios. “Los asistentes de voz están creciendo de forma significativa y, actualmente, uno de cada cinco adultos en Reino Unido, utiliza los altavoces inteligentes”, explican Jeremy Walker, *Head of Technology Transformation and Voice Management* y Chris Dix, *Head of Technology Transformation and Product Architecture*, de la BBC (entrevista cualitativa, 15 mayo 2020).

El medio de comunicación británico ha estado en los últimos años experimentando con la voz para entender como funciona esta tecnología y analizar como las interfaces vocales pueden cambiar la manera de distribuir y consumir periodismo.

Empezaron a trabajar para ofrecer sus contenidos en asistentes de voz en 2016, cuando publicaron *flash briefings* o resúmenes informativos en Alexa y Google Assistant, y se convirtieron en el proveedor de noticias por defecto de ambos asistentes de voz en Reino Unido: si preguntabas “*What’s the news*” siempre obtenías la BBC (Murphy, 2020).

Sin embargo, fue en diciembre de 2017 cuando lanzaron su primera aplicación para Alexa, con el objetivo de mejorar su oferta de contenidos y al experiencia de usuario. Con esta *skill* se podía acceder a todo su abanico de emisoras de radio (locales, nacionales e internacionales), los podcasts disponibles a demanda (BBC launches its first full voice service for smart speakers, 2017).

Dos años más tarde, en 2019, la BBC decide dar un paso hacia adelante y ofrecer a través del asistente de voz de Amazon una experiencia mucho más interactiva para el usuario y huir, así, del básico resumen informativo que ofrecía hasta ahora. Este nuevo diseño de experiencia permite al usuario tener el control de la agenda de contenidos y decidir cuando van a escuchar las noticias y de que duración van a ser los contenidos. El nuevo formato permite, además, añadir contenido adicional, como por ejemplo, acceder a materiales o entrevistas que puedan complementar las noticias de actualidad.

Al respecto, Zoe Murphy, editora de desarrollo de noticias en el departamento BBC Voice + AI explica que esta funcionalidad permite a los periodistas olvidarse de la brevedad y no depender de

unos tiempos y espacios determinados. De esta manera, “las audiencias pueden acceder al mejor periodismo de la BBC” y los periodistas pueden dar color y contexto a sus historias, un valor añadido que saben que su audiencia valora (Murphy, 2020).

Mostramos un gráfico donde se observan las posibilidades de navegación de un usuario de la BBC en Alexa:

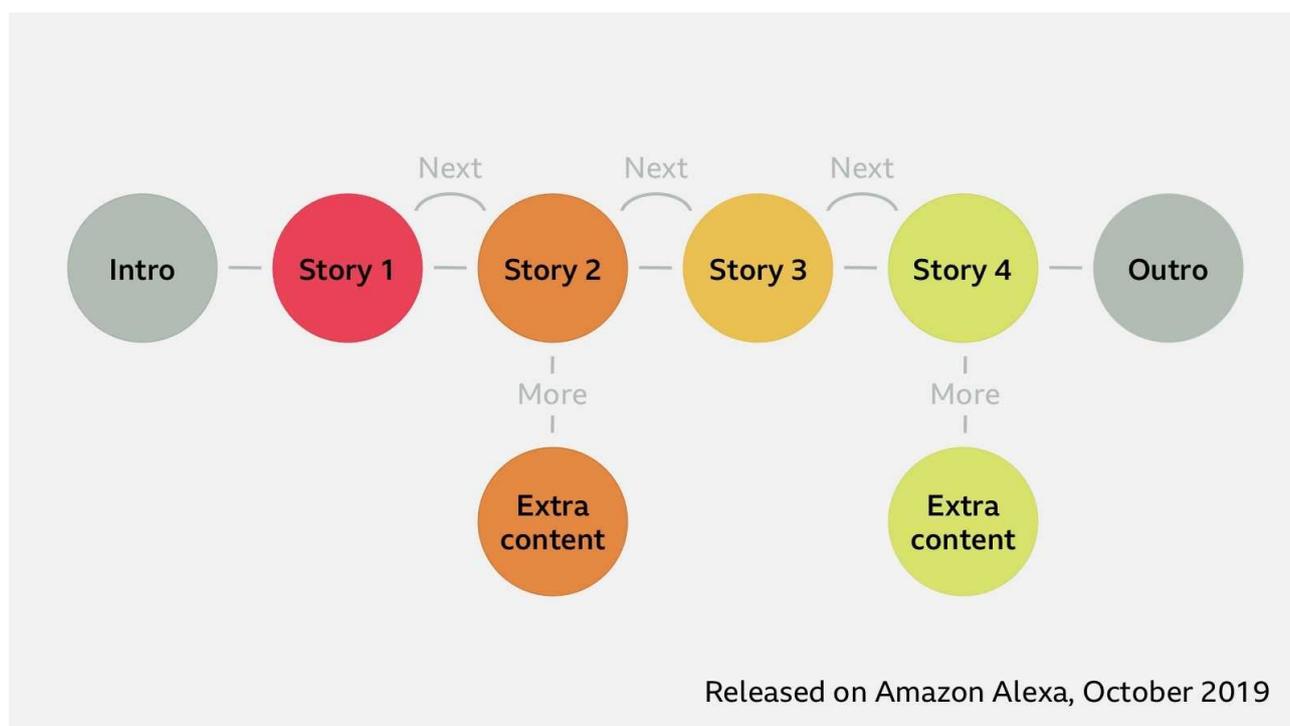


Figura 7: <https://bbcnewslabs.co.uk/news/2020/voice-and-ai-lessons/>

Los asistentes de voz se han convertido, para la BBC, en una nueva, interesante y natural manera de interactuar con contenido en formato audio: “Nos dan la oportunidad de imaginar como sería la radio si se inventara hoy. Al hacer que las noticias sean interactivas en altavoces inteligentes, los oyentes pueden saltar a las historias que necesitan escuchar justo antes de salir corriendo de casa o, cuando tienen más tiempo, profundizar en una pieza y obtener más información”, explica Mukul Devichand, editor ejecutivo de BBC Voice e IA («BBC launches interactive news service for smart speakers», 2019).

Actualmente la BBC tiene presencia en Google Assistant y en Amazon Alexa, de la siguiente manera:

Google Assistant

Los usuarios pueden acceder a lo que ellos llaman “BBC Sounds”, es decir a todas las emisoras de radio de la corporación a través de la frase de invocación:

“Ok Google, play BBC [nombre de la emisora]”

También pueden consumir los servicios de noticias: “BBC News”, “World Service News” y “BBC Minute”, a través de la configuración previa del asistente, con solo decir:

“Ok Google, play the news”

Amazon Alexa

La oferta que ofrece la BBC en el asistente de Amazon es un poco más amplia. Da la posibilidad de acceder a todas las emisoras de radio a través de la *skill* “BBC Sounds”: tanto los directos como el contenido a demanda, incluyendo programas, podcasts y listas de reproducción de música:

“Alexa, ask the BBC to play [el directo de una emisora, un programa, un podcast o un mix de música]”

Para acceder a noticias, pueden hacerlo a través de los *flash briefings* de “BBC News”, “BBC Sport”, “BBC Business”, “World Service News” y “BBC Minute”. Estos resúmenes informativos han de ser previamente configurados para que se sirvan de manera automática cuando se invoque:

“Alexa, play my flash briefing”

Y para activar los servicios de noticias interactivos, el nuevo formato que la BBC estrenó en 2019, los usuarios deberán pedir:

“Alexa, give me BBC News”, para acceder;

“Alexa, next” para cambiar de un contenido a otro;

“Alexa, more from the BBC”, para profundizar en una historia concreta.

Dificultades en la integración de los contenidos de la BBC en Alexa y Google Assistant

Pese a los avances para ofrecer una experiencia de usuario lo más satisfactoria posible, a través del desarrollo de aplicaciones más interactivas, los responsables de los productos de voz de la BBC no han encontrado la fórmula para crear las funcionalidades y experiencias que desearían. La dependencia de un tercero, en este caso Amazon o Google, con intereses distintos a los del medio de comunicación, dificultan el proceso.

Otra de limitaciones con las que se han encontrado durante estos tres años, desde que empezaron a integrar sus contenidos en los asistentes de voz, tiene que ver con la atribución y la prominencia. La presencia de marca en los formatos *flash briefing* queda difuminada y, en muchas ocasiones, el usuario no es plenamente consciente de la marca informativa que les está ofreciendo las noticias.

Aunque las posibilidades, en este sentido, aseguran que son limitadas (Walker, J. & Dix, C., 2020), todos los boletines informativos para asistentes de voz de la BBC empiezan con un “*Welcome to BBC News*”. Además, han construido también una nueva identidad sonora para las noticias de la BBC en las plataformas de voz, ya que consideran que “es esencial establecer la marca y mantener la atribución, especialmente en un ambiente que únicamente depende del audio” (Murphy, 2020).

El idioma en el que trabaja y se comunica la BBC es el inglés, que ya hemos comentado que es en el idioma en que el *natural language understanding* está más desarrollado. Sin embargo, durante el proceso de desarrollo y testeo de sus aplicaciones para asistente de voz, también les han surgido conflictos dialectales, que corresponden con las distintas zonas de Reino Unido.

El asistente de voz Alexa tiene dificultades para entender los acentos particulares de algunas zonas del territorio anglosajón, la cual cosa genera una experiencia de usuario negativa que repercute directamente en la satisfacción de la audiencia de la BBC: “La BBC tiene la obligación de reflejar, representar y servir a la diversidad de comunidades de todas las naciones y regiones de Reino Unido”, explican Dix y Walker (2020). “Tenemos que garantizar que toda nuestra audiencia, independientemente de la zona de la que venga, puede utilizar los asistentes de voz y sentir que les son útiles”. Por ese motivo, la corporación británica ha recopilado audios con acentos de las distintas partes de Reino Unido para entrenar al asistente.

Sin embargo, todos los esfuerzos han sido insuficientes para saciar la necesidad de la BBC de ofrecer una aún mejor experiencia para su audiencia. Por eso, ha optado por desarrollar su propio asistente de voz y alejarse de la dependencia de las lógicas de un tercero.

4.2.3. “Beeb”, su propio asistente de voz

La dependencia de las lógicas propias de las grandes empresas tecnológicas que lideran el mercado de los asistentes de voz, Amazon y Google mayoritariamente, ha frustrado las intenciones de la BBC de ofrecer a sus audiencias experiencias mucho más enriquecedoras y satisfactorias. De hecho, Dix y Walker (2020) explican, en la entrevista que concedieron para esta investigación, que “es técnicamente posible crear experiencias más profundas para la audiencia a través de la voz, pero no es viable hacerlo en otras plataformas”. Por eso, en agosto de 2019 la BBC anunció que empezaba a desarrollar su propio asistente de voz: “Beeb”.

De esta manera, aseguran, pueden controlar más directamente qué tipo de relación quieren tener con su audiencia y como quieren que ésta sea. Por ejemplo, en referencia a la privacidad de datos o a la atribución de marca. “Con el desarrollo de ‘Beeb’, podemos crear un asistente que ponga a la audiencia en el centro. Nos permitirá tener más libertad y creatividad para experimentar con nuevas formas de conectar con la audiencia a través de la conversación”, explican Dix y Walker (2020). “Además, podremos hacer cambios rápidos cuando los necesitemos, sin tener que depender del permiso de un tercero. Con nuestro propio asistente de voz podremos ser más ambiciosos y más experimentales con el contenido que creamos y en las maneras como ayudamos a la gente a encontrarlos”.

Dos meses después de que los representantes de la BBC concedieran la entrevista para este estudio, la corporación británica presentó, en junio de 2020, la versión beta de “Beeb”, solo disponible para usuarios de Windows de Reino Unido que deseen hacer tests con ella. La marca informativa lo ha catalogado como el “primer asistente de voz que es, a su vez, un servicio público” (Porter, 2020).

Este nuevo asistente de voz no solo permite a la BBC tener más control sobre la experiencia de usuario y la privacidad de datos, sino que también es capaz de entender los acentos regionales del territorio en el que opera.

Además, la versión beta de “Beeb” se entrena gracias a las invocaciones de los usuarios que lo testean: cuando un usuario se descarga el asistente, éste le pregunta qué acento tiene y así puede entrenarse. En este sentido, y en relación a la privacidad de datos, la BBC asegura que no se van a hacer grabaciones de las escuchas activas, sino que solo se van a transcribir unos pocos segundos después de que el usuario de su consentimiento y diga “Ok Beeb” (Porter, 2020).

La versión beta de “Beeb” permite que los usuarios tengan acceso a las emisoras de radio de la BBC, música, podcasts, noticias y previsiones meteorológicas, además de responder a preguntas con respuesta única y que tengan referencia con los programas de la BBC. Según aseguran los responsables de la corporación, estas funcionalidades se ampliarán en los próximos meses.

En comparación con otros asistentes de voz en el mercado, como Alexa o Google Assistant, Beeb no ofrece la posibilidad de programar por ejemplo recordatorios, que recordemos es una de las funcionalidades más bien valoradas por los usuarios de interfaces vocales: “Hay muchos servicios que no necesitamos proveer. Nos hemos enfocado en el contenido y el descubrimiento” (Walker & Dix, 2020).

Ello es debido a que la BBC no pretende competir con este desarrollo con los asistentes ya disponibles en el mercado: “Seguiremos distribuyendo nuestro contenido a través de otros asistentes de voz y altavoces inteligentes. El asistente de voz de la BBC se ha creado para ser integrado en dispositivos como los teléfonos móviles, los altavoces inteligentes y la televisiones, entre otros. “No vamos a desarrollar nuestro propio *hardware*. Lo que planteamos es que nuestro *software*, es decir, la inteligencia artificial de Beeb, se pueda incorporar en todos los dispositivos posibles” (Walker & Dix, 2020).

Gracias a los asistentes de voz, la BBC ha visto nuevas posibilidades en cuanto a la accesibilidad que pretenden incorporar en Beeb. Según Walker y Dix (2020) todo el mundo debería tener acceso al contenido escrito que produce la marca informativa británica, independientemente de que sepan o no leer. A este efecto, las plataformas basadas en voz ofrecen multitud de opciones para integrar a colectivos vulnerables que, por motivos de analfabetización o discapacidad, no podían acceder a este tipo de contenidos anteriormente. La lectura automática de noticias por parte del asistente de voz es un mercado creciente, en el que Apple, por ejemplo, ya ha empezado a trabajar (Walker & Dix, 2020).

Interesante destacar, también, que Beeb está representado por una voz masculina (concretamente con acento del norte de Inglaterra), y no femenina como el resto de asistente de voz que encontramos actualmente en el mercado.

4.3. . Variables comparativas entre la CCMA y la BBC

En este apartado, analizaremos algunas variables que hemos podido extraer de los casos de la CCMA y la BBC, vistos anteriormente, con el objetivo de compararlos y estudiar qué puntos en común y qué diferencias tienen.

Tipo de medio de comunicación

	CCMA	BBC
Tipo de medio de comunicación	<i>Broadcaster</i> público	<i>Broadcaster</i> público

Las radiotelevisiónes públicas se han tenido que adaptar a nuevos escenarios donde internet se ha posicionado como la plataforma fundamental para informar, comunicar y difundir sus contenidos a sus audiencias. El estudio *Las radiotelevisiónes públicas europeas en el entorno web* (Rodríguez-Fernández et al., 2018), donde se han analizado las webs de las radiotelevisiónes públicas de los países de la Unión Europea, concluye que las corporaciones públicas deben apostar por el desarrollo de servicios *online* más eficaces. Parece claro que, tanto la CCMA como al BBC, han visto en los asistentes de voz una vía prometedora e innovadora para conectar con su audiencia.

Además, los entes de comunicación públicos reservan una parte de su presupuesto a la innovación, sin tener una presión excesiva por el retorno inmediato (como sí lo tienen los medios de comunicación privados, que dependen completamente de la publicidad y de la respuesta de las audiencias). Por eso, tanto la CCMA como la BBC han podido destinar parte de su presupuesto y de su plantilla interna a investigar el mundo de los asistentes de voz y iniciar su andadura en él.

Asistentes de voz en el que han integrado sus contenidos

	CCMA	BBC
--	------	-----

	CCMA	BBC
Asistentes de voz	Alexa Google Assistant	Alexa Google Assistant

Ambos medios de comunicación han apostado por integrar sus contenidos en los asistentes de voz mayoritarios en el mercado europeo y americano: Alexa (de Amazon) y Google Assistant (de Google). En ambos casos, la presencia de contenidos es mayor en el asistente de Amazon, debido a que éste ofrece más facilidades de integración que Google Assistant.

La CCMA, como hemos dicho anteriormente, tiene una alta demanda por parte de su comunidad de usuarios de servir sus contenidos a través de Google Assistant (recordemos que es el asistente mayoritario en el mercado español). Sin embargo, pese a que en el segundo trimestre de 2020 han podido ofrecer los resúmenes informativos a través de esta plataforma, no consiguen publicar sus aplicaciones *ad hoc* o *actions*. Según han informado sus responsables, hace más de un año y medio que intentan que Google apruebe sus aplicaciones, pero el problema idiomático sigue ganando la partida: el gigante tecnológico considera mala experiencia de usuario el pedir un contenido en castellano y devolverlo en catalán, pese a que la realidad del territorio sea bilingüe.

¿Qué productos ofrecen?

	CCMA	BBC
Productos	<i>Flash briefings</i> Aplicaciones <i>ad hoc</i>	<i>Flash briefings</i> Aplicaciones <i>ad hoc</i>

En ambos casos, han optado por ofrecer tanto *flash briefings* o resúmenes informativos, como aplicaciones *ad hoc*. Parece que, la fórmula natural por la que optan los medios de comunicación para integrarse dentro de los asistentes de voz es pausada y progresiva. Tanto la CCMA como la

BBC empezaron su andadura por los asistentes de voz gracias a los resúmenes informativos, que resultó ser la manera más sencilla y rápida de integrar sus contenidos en estas plataformas.

Sin embargo, la extrema sencillez de estas aplicaciones que tanto Alexa como Google Assistant ofrecen a los medios de comunicación para distribuir sus noticias, la falta de identidad de marca dentro de ellas, y la alta dependencia de la lógica de un tercero hizo que las dos corporaciones de noticias pública que estamos analizando decidieran dar un paso adelante y optar por desarrollar aplicaciones propias para poder ofrecer contenidos de una manera más elaborada y sirviendo mejor a su audiencia.

¿Tienen asistente de voz propio?

	CCMA	BBC
Asistente de voz propio	No	Sí: Beeb

La BBC ha apostado por tener una mayor independencia de los gigantes tecnológicos y sus lógicas, que según aseguran no siempre tienen a la audiencia en el centro, y han desarrollado su propio asistente de voz: Beeb. Empezaron trabajando en él en agosto de 2019 y casi un año después, en junio de 2020, han lanzado la primera versión de su asistente en beta para que los usuarios de Windows hagan pruebas con él.

Tipo de contenidos

	CCMA	BBC
Elaboración de contenidos exclusivos para plataformas de voz	No	Sí

	CCMA	BBC
	Boletines informativos	Boletines informativos
Tipo de contenidos	Titulares de los telenoticias (opción de escuchar el telenoticias entero)	Directos de las emisoras de radio Podcasts y audios a la carta
	Piezas opinativas	Experiencias interactivas que permiten al usuario navegar de contenido a contenido, y profundizar en él a través de la
	Predicción meteorológica	voz
	Directos de las emisoras de radio	
Difusión de contenidos para soporte visual	No	No

Elaborar contenidos específicos para los asistentes de voz requiere de una inversión elevada por parte del medio de comunicación en cuestión. La CCMA, por el momento y puesto que está en un proceso de ensayo-error y análisis de resultados de audiencia a través de asistentes de voz, utiliza contenidos ya existentes que puede aprovechar y son compatibles al 100% con el formato audio.

Sin embargo, la corporación británica, que como hemos visto va un paso por delante desarrollando su propio asistente, apuesta por la elaboración de contenidos específicos para las interfaces vocales. Ofrece contenidos informativos, muchos de los cuales permiten al usuario navegar y profundizar en la información a través de materiales complementarios; un recorrido navegable por completo con la VOZ.

Ello, necesariamente, lleva a que los periodistas que redactan este tipo de audios interactivos deban tener en cuenta aspectos que, de otro modo (si escriben para radio, prensa o televisión), no necesitarían pensar. Por ejemplo, añadir menús auditivos que permitan al usuario navegar de manera fácil e intuitiva y permitan saltar de una historia a otra o profundizar en algún aspecto.

Por otro lado, ninguno de los dos medios de comunicación analizados contempla elaborar y difundir contenidos en formato visual, puesto que los dispositivos con asistentes de voz que tienen pantalla son muy minoritarios en el mercado. No obstante, la CCMA sí ofrece la posibilidad al usuario que disponga de un dispositivo de estas características, acceder a alguno de los contenidos en formato imagen (por ejemplo, los telenoticias completos).

Branding sonoro propio para las interfaces vocales

	CCMA	BBC
Branding sonoro para interfaces visuales	Sí	Sí

Hemos visto la importancia de mantener la identidad de la marca informativa en las interfaces vocales, sobretodo porqué dependen de un tercero y el usuario puede pasar por alto que medio de comunicación le está ofreciendo la información y pensar que lo está haciendo directamente la empresa que hay detrás del asistente (Amazon o Google, por ejemplo).

Por ese motivo, tanto la CCMA como la BBC han apostado por desarrollar una identidad de marca propia para estos canales. Son sintonías, indicadores, e incluso voces, conocidas por la audiencia de los medios de comunicación, pero que han adoptado singularidades propias de entorno digital, como por ejemplo, sonidos más sintéticos.

5. Conclusiones

Este estudio de investigación ha servido para hacer una aproximación a los asistentes de voz y su proyección en el entorno de los medios de comunicación. Gracias a él he podido responder y matizar muchas de las preguntas que se planteaban al inicio del proceso.

La principal conclusión de este trabajo es que estamos ante un cambio de paradigma a nivel comunicativo: en las últimas décadas las interfaces visuales, las pantallas y lo táctil han dominado la manera como las personas nos comunicamos con las máquinas y, también, consumimos información y entretenimiento. Pero actualmente, nos encontramos en un cambio de etapa, donde la voz está tomando protagonismo.

Una de las preguntas que formulaba al inicio de la investigación era si las interfaces vocales, como los asistentes de voz, se van a convertir en las principales plataformas de distribución y consumo de información. Después de realizar este estudio, podemos matizar la pregunta y responder: en los próximos años, los asistentes de voz van a estar presentes en la mayor parte de los dispositivos inteligentes de nuestro entorno (televisores, casas, neveras, coches, altavoces...) y vamos a poder pedirles prácticamente todo a través de la voz, desde programar recordatorios, a hacer la compra o pedir que nos informen o entretengan. La voz, por lo tanto, parece que va a convertirse en el principal canal para hacer peticiones a los dispositivos, una manera de comunicarnos con las máquinas mucho más natural para los humanos.

Sin embargo, no va a ser el único canal con el que se nos va a devolver toda la información que pidamos: ésta información podrá ser devuelta, evidentemente, a través de audios (que es como se está haciendo hoy día, mayoritariamente), pero también a través de imágenes, que podrán ser emitidas en cualquier pantalla de televisores, teléfonos inteligentes o tabletas; pero también, a largo plazo, proyectadas por ejemplo sobre un techo, pared o puerta.

Pese a que hemos visto que, actualmente, la opción de consumir información a través de las interfaces vocales (los asistentes de voz) no es mayoritaria, aunque sí con una tendencia creciente, los medios de comunicación han visto en ella una nueva manera de conectar con su audiencia y trabajar nuevas estrategias comunicativas. El camino es aún incierto, pero ningún medio de comunicación quiere quedar atrás y por eso, muchos han empezado a trabajar en proyectos propios

con asistentes de voz. Sin duda, los asistentes de voz pueden cambiar la manera cómo la audiencia consume la información y la manera cómo los medios de comunicación se comunican con ella.

Es importante destacar, también que a lo largo del estudio de investigación han surgido algunas dificultades. Para empezar, se trata de un tema muy novedoso y la construcción de las fuentes del marco teórico ha sido un trabajo complicado. La mayoría de las fuentes seleccionadas finalmente son muy específicas: personas expertas en la materia que están, a día de hoy, trabajando e investigando sobre la temática. Ello aporta frescura a la investigación, ya que se trata de fuentes muy actuales; pero también existe el riesgo de que, con el paso de los años o incluso meses, algunas de las aportaciones extraídas para este trabajo varíen.

Otra de las dificultades encontradas en durante el proceso del trabajo tiene que ver con la situación de pandemia que hemos vivido. La metodología principal en la que se basa este estudio es de entrevistas cualitativas semi-estructuradas. Sin embargo, todas las entrevistas han tenido que hacerse a través de vídeo-conferencias o formularios *online*, la cual cosa le resta el factor de contacto humano, tan importante en este tipo de entrevistas. Además, algunas de las entrevistas previstas para la investigación, no han podido realizarse. Aún así, el resultado final ha sido positivo.

Para acabar, apuntaré algunos de los temas que han aparecido en este trabajo y que abren la posibilidad de iniciar otras investigaciones:

- Uno de los hilos más interesantes que se podrían trabajar tiene que ver con la personalidad femenina de los asistentes de voz. Como hemos comentado a lo largo del estudio, los asistentes de voz tienen todos una voz femenina que les hace dibujar una nueva dimensión identitaria, en este caso, en relación con el género. Ello se ha relacionado directamente con su rol de asistente al servicio de las personas y parece claro que la voz femenina tiene una asociación directa con la subordinación. Este tema podría abrir la línea de investigación de un nuevo estudio.

- A lo largo del estudio, hemos hablado también de las dificultades que encuentran los medios de comunicación a la hora de monetizar los contenidos que difunden a través de las plataformas de voz. Sin embargo, todos ellos quieren estar presentes en los asistentes de voz, ya que ven que su proyección en el mercado es creciente y consideran que es una buena e innovadora manera de

conectar con su audiencia. Pensar el modelo de negocio que los medios de comunicación pueden encontrar en los asistentes de voz es una investigación tan interesante como necesaria para el sector.

- Los contenidos que, la mayoría de medios de comunicación integran actualmente en los asistentes de voz son reaprovechados; son mínimos los contenidos producidos específicamente para este canal. Ello es debido a que se trata de un entorno aún muy reciente en el que, la mayoría de medios, están haciendo pruebas y analizando como responde su audiencia. No obstante, a largo plazo parece claro que los medios tendrán que destinar una parte de sus redacciones a crear contenidos específicos para estas interfaces vocales y ello implicará cambios en las rutinas periodísticas e, incluso, en la tipología de equipos de trabajo en las redacciones. Parece claro que será esencial un equipo multi-disciplinar de trabajo en el que, además de periodistas, se necesitarán expertos en experiencia de usuario (UX), expertos en desarrollo de producto, ingenieros y profesionales de márketing, trabajando juntos y compartiendo conocimientos y experiencia de sus respectivas disciplinas. Visualizar qué tipo de redacciones se prevén en el futuro, con la incorporación de los asistentes de voz como uno de los canales principales de distribución de la información y el entretenimiento, es una línea de investigación interesante para seguir.

- También sería interesante estudiar el papel de los asistentes de voz en los medios de comunicación de proximidad y analizar si este canal puede ser una oportunidad para repensar sus funciones, ampliar sus audiencias y responder a la necesidad de servir y consumir información de proximidad de los usuarios.

6. Bibliografia

¿Cómo suena la tu marca? Estado de la voz y el audio en España. (2019).

https://www.prodigiosovolcan.com/sismogramas/informe-voz/informe_voz_audio_espana.pdf

“Alexa, léeme las noticias de EL PAÍS”. (2018). *El País*.

https://elpais.com/elpais/2018/10/29/el_pais_que_hacemos/1540812806_645600.html

Alexa aprende catalán. (2019). *Via Empresa*. https://www.viaempresa.cat/es/innovacion/alexa-aprende-catalan_2103262_102.html

Amazon Echo llega a España con ABC entre sus fuentes de noticias de calidad. (2018). *ABC*.
https://www.abc.es/tecnologia/electronica/sonido/abci-amazon-echo-llega-espana-abc-entre-fuentes-noticias-calidad-201810240227_noticia.html?ref=https:%2F%2Fwww.google.com

Argila, J. (2019). *Alexa, dóna'm les notícies!* Universitat Oberta de Catalunya.

Assistentes de veu en català. Una demanda comercial i social. (2019).

Autoritat Catalana de Protecció de Dades. (2020). *Intel·ligència Artificial. Decisions Automatitzades a Catalunya*.

<https://apdcat.gencat.cat/web/.content/04-actualitat/noticies/documents/INFORME-INTELLIGENCIA-ARTIFICIAL-FINAL-WEB-OK.pdf>

Baquero, C. S., & Pueyo, J. (2019). El Govern solo usará asistentes de voz digitales que hablen catalán. *El País*. https://elpais.com/ccaa/2019/11/12/catalunya/1573559532_971511.html

BBC launches interactive news service for smart speakers. (2019). *Press Gazette*.

<https://www.pressgazette.co.uk/bbc-launches-interactive-news-service-for-smart-speakers/>

BBC launches its first full voice service for smart speakers. (2017). *BBC*.

<https://www.bbc.co.uk/mediacentre/latestnews/2017/smart-speakers>

BBC launches its first voice experience for children. (2018). *BBC*.

<https://www.bbc.co.uk/mediacentre/latestnews/2018/bbc-first-childrens-voice-experience>

Brownlee, J. (2015). What is Zero UI? (And why is it crucial to the future of design?). *Fast Company*. <https://www.fastcompany.com/3048139/what-is-zero-ui-and-why-is-it-crucial-to-the-future-of-design>

Canovaca, E. (2019). *El Periodisme Digital amb Valor. Claus per a la sostenibilitat de la premsa*. (U. A. de B. S. de Publicacions (Ed.)).

Carey, J., & Quirck, J. (1989). The mythos of the electronic revolution. *Communication as a culture: Essays on media and society*, 113-141.

Catalunya Ràdio estrena la nova «skill» per escoltar el directe des dels dispositius amb Alexa. (2019). *CCMA*. <https://www.ccma.cat/catràdio/catalunya-radio-estrena-la-nova-skill-per-escoltar-el-directe-des-dels-dispositius-amb-alexa/nota-de-premsa/2947863/>

- Cooke, H. (2017). The Inspection Chamber. *BBC*. <https://www.bbc.co.uk/rd/blog/2017-09-voice-ui-inspection-chamber-audio-drama>
- Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (HMH Books (Ed.)).
- Diseño conversacional. Haz que tu voz brille.* (2019). <https://stayrelevant.globant.com/es/voz-post-pantalla-diseno-conversacional/>
- El estudio que analiza el uso de las tecnologías de voz en España. (2020). *Marketing News*. <https://www.marketingnews.es/investigacion/noticia/1158041031605/estudio-analiza-uso-de-tecnologias-de-voz-espana.1.html>
- Fortunati, L., & Edwards, A. (2020). Opening Space for Theoretical, Methodological, and Empirical Issues in Human-Machine Communication. *Human-Machine Communication*, 1(February), 7-18. <https://doi.org/10.30658/hmc.1.1>
- From search to smart speakers: Why voice is too big for media companies to ignore. (2018). *Whats New In Publishing*. <https://whatsnewinpublishing.com/from-search-to-smart-speakers-why-voice-is-too-big-for-media-companies-to-ignore/>
- Gibbs, S. (2018). How smart speakers stole the show from smartphones. *The Guardian*. <https://www.theguardian.com/technology/2018/jan/06/how-smart-speakers-stole-the-show-from-smartphones>
- Guzman, A. L. (2016). Making AI safe for humans: A conversation with Siri. *Socialbots and Their Friends: Digital Media and the Automation of Sociality, January 2017*, 69-85. <https://doi.org/10.4324/9781315637228>
- Guzman, A. L. (2018). What is Human-Machine Communication, Anyway? En *Human-Machine Communication: Rethinking Communication, Technology, and Ourselves*. Peter Lang. <https://doi.org/https://doi.org/10.3726/b14399>
- Guzman, A. L., & Lewis, S. C. (2019). Artificial intelligence and communication: A Human–Machine Communication research agenda. *New Media & Society*, 22(1), 70-86. <https://doi.org/10.1177/1461444819858691>
- Höflich, J. R. (2013). Relationships to Social Robots: Towards a Triadic Analysis of Media-oriented Behavior. *Intervalla: Platform for Intellectual Exchange*, 1, 1-14. papers3://publication/uuid/57CBD430-3814-4FC7-8ECA-769B961F185E
- Horcada, E. (2017). *Las experiencias de interacciones digitales en la era post-pantalla*. The standard CIO. <http://thestandardcio.com/2017/08/26/experiencias-digitales-usuarios/>
- How to Get The New York Times on Your Smart Speaker. (2020). *The New York Times*. <https://www.nytimes.com/2020/02/14/insider/how-to-get-the-new-york-times-on-your-smart-speaker.html>

- Ivanova, A. (2019). Asistente de voz, una gran revolución tecnológica. *El español*.
https://www.elespanol.com/imprescindibles/20191231/asistente-voz-gran-revolucion-tecnologica/452705612_0.html
- Jané, C. (2018). Softcatalà y Mozilla impulsan un asistente de voz libre en catalán. *El Periódico de Catalunya*. <https://www.elperiodico.com/es/sociedad/20180812/softcatala-y-mozilla-impulsan-asistente-voz-libre-catalan-6924383>
- Kellermann, K. (1992). Communication: Inherently strategic and primarily automatic. *Communication Monographs*, 59, 288-300.
<https://www.tandfonline.com/doi/abs/10.1080/03637759209376270>
- Keynes, M. (2018). Strategy Analytics: Smart Speaker Shipments Hit Record Level in Q3 2018 as Installed Base Approaches 100 million. *Business Wire*.
<https://www.businesswire.com/news/home/20181112005194/en/Strategy-Analytics-Smart-Speaker-Shipments-Hit-Record>
- Kleinberg, S. (2018). *5 ways voice assistance is shaping consumer behavior*. Think with Google.
<https://www.thinkwithgoogle.com/consumer-insights/voice-assistance-consumer-experience/>
- La voz recupera su protagonismo. (2018). *Reason Why*.
<https://www.reasonwhy.es/actualidad/digital/uso-y-penetracion-asistentes-digitales-de-voz-en-espana-2018-03-15>
- Lladó, A. (2018). Los asistentes de voz, un territorio inexplorado para los medios de comunicación. *Next Media by Lavinia*.
- Los gurús tecnológicos auguran una nueva internet que cambiará el mundo. (2019). *Retina. El País*.
https://retina.elpais.com/retina/2019/10/21/tendencias/1571636691_938327.html
- Los podcasts de Catalunya Ràdio llegan a los dispositivos con Alexa. (2020). *Guia de la Radio*.
<http://guiadelaradio.com/los-podcasts-de-catalunya-radio-llegan-a-los-dispositivos-con-alexa>
- Misión y principios*. (s. f.). Corporació Catalana de Mitjans Audiovisuals. Recuperado 19 de junio de 2020, de <https://www.ccma.cat/corporatiu/es/missio-i-principis/>
- Mission, values and public purposes*. (s. f.). bbc.com.
<https://www.bbc.com/aboutthebbc/governance/mission>
- Moon, B. (2018). Smart Speaker Sales Are Growing Faster than Smartphones in the U.S. *Yahoo Finance*. <https://finance.yahoo.com/news/smart-speaker-sales-growing-faster-152339686.html>
- Murphy, Z. (2020). *How the BBC is reinventing news for voice platforms*. bbcnewslabs.co.uk.
<https://bbcnewslabs.co.uk/news/2020/voice-and-ai-lessons/>
- Nearly half of Americans use digital voice assistants, mostly on their smartphones. (2017). *Pew Research Center*. <https://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/>

- Newman, N. (2018). *The Future of Voice and the Implications for News*. 48.
<https://reutersinstitute.politics.ox.ac.uk/our-research/future-voice-and-implications-news>
- Newman, N., Fletcher, R., Kalogeropoulos, A., & Kleis Nielsen, R. (2019). How Younger Generations Consume News Differently - Reuters Institute Digital News Report. En *Digital News Report*. <http://www.digitalnewsreport.org/survey/2019/how-younger-generations-consume-news-differently/>
- Peraldo, G. M. (2010). *Interacción Multimodal*. Blog de Gian Marco Peraldo Valera.
- Porter, J. (2020). BBC releases its own ‘Beeb’ voice assistant in beta. *The verge*.
- Prego, C. (2019). Me dedico a desarrollar interfaces de voz: así es diseñar apps en la era de Alexa, Siri y Google Home. *Xataka*. <https://www.xataka.com/servicios/me-dedico-a-desarrollar-interfaces-voz-asi-disenar-apps-era-alexa-siri-google-home>
- Reales, L. (2020). Ramón López de Mántaras: «Un buen sistema de inteligencia artificial anticiparía en dos o tres semanas la eclosión de una nueva pandemia». *Barcelona Metroplis*.
- Rodríguez-Fernández, M. M., Sánchez-Amboage, E., & Toural-Bran, C. (2018). Las radiotelevisións públicas europeas en el entorno web. En *Revista Latina de Comunicación Social* (Vol. 2018, Número 73). <https://doi.org/10.4185/RLCS-2018-1288>
- Roettgers, J. (2016). Amazon Echo Now Lets You Investigate the Murder of Bruce Wayne’s Parents. *Variety*. <https://variety.com/2016/digital/news/amazon-echo-alexa-batman-vs-superman-promotion-1201718896/>
- Rothschild, J. (1983). *Machina Ex Dea: Feminist Perspectives on Technology* (Teachers’ College Press (Ed.); 1992.^a ed.).
- RTVE lanza «Lunnis de leyenda» y «Cuentos de Clan» para asistentes de voz. (2019). RTVE Comunicación. <https://www.rtve.es/rtve/20190906/rtve-lanza-lunnis-leyenda-cuentos-clan-para-asistentes-voz/1978668.shtml>
- Rubio, I. (2019). Empleados de Amazon escuchan a diario conversaciones que mantienen los usuarios con Alexa. *El País*.
https://elpais.com/tecnologia/2019/04/11/actualidad/1554992401_521050.html
- Rus, C. (2019). Google confirma que terceros pueden escuchar fragmentos de audio recogidos por Google Home y Google Assistant. *Xataka*. <https://www.xataka.com/privacidad/google-confirma-que-terceros-pueden-escuchar-fragmentos-audio-recogidos-google-home-google-assistant>
- Sañas, M. (2019). Asistentes virtuales a disposición de los usuarios. *La Vanguardia*.
<https://www.lavanguardia.com/vida/junior-report/20190226/46686943265/asistentes-voz-inteligencia-artificial-futuro-riesgos.html>
- Sangers, P. (2000). Practices for a Machine Culture. *Surfaces*, 7, 4-58.

- The BBC on voice devices*. (2020). BBC. <https://www.bbc.co.uk/usingthebbc/account/the-bbc-on-voice-devices/>
- Turkle, S. (1984). *The second self*. Simon & Schuster.
- TV3 i Catalunya Ràdio, els primers d'integrar continguts en català a Alexa d'Amazon. (2018). *ccma.cat*. <https://www.ccma.cat/tv3/tv3-i-catalunya-radio-els-primers-a-integrar-continguts-en-catala-al-servei-de-veu-alexa-damazon/noticia/2882747/>
- Wakabayashi, D., & Wingfield, N. (2018). Alexa, We're Still Trying to Figure Out What to Do With You. *The New York Times*. <https://www.nytimes.com/2018/01/15/technology/virtual-assistants-alexa.html>
- Williams, R. (2018). Netflix's «Lost in Space» unveils Google smart-speaker adventure. *Mobile Marketer*. <https://www.mobilemarketer.com/news/netflixs-lost-in-space-unveils-google-smart-speaker-adventure/521385/>
- Yurieff, K. (2017). Voice-powered search is replacing typing and other internet trends. *CNN Business*. <https://money.cnn.com/2017/05/31/technology/mary-meecker-internet-trends/index.html>

7. Anexos

7.1. Entrevista a Xavier Fisa, director de Lavinia Voice

Fecha: 17/04/2020

Nombre del entrevistado: Xavier Fisa Matinero

Empresa: Lavinia Voice

OBJETIVO DE LA ENTREVISTA

Ampliar conocimientos sobre tecnologías de voz y asistentes de voz, y conocer el proyecto de Lavinia Voice.

INFORMACIÓN DE LOS ENTREVISTADOS

Nombre y apellidos: Xavier Fisa Matinero

Formación y experiencia: Licenciado en Publicidad y Relaciones Públicas, especializado en comunicación interactiva y digital. Experto en disrupciones tecnológicas orientadas al mundo de la comunicación.

Cargo dentro de Lavinia: Responsable de estrategia en Lavinia Next y director del área Lavinia Voice.

Desde cuando trabaja en proyectos con asistentes de voz: 2016

PREGUNTAS

1. ¿Estamos ante un cambio de paradigma, donde la voz se va a convertir en la nueva interfaz de comunicación?

Creo que la manera de interactuar de los seres humanos con los contenidos o servicios a través de la tecnología va a cambiar, porque se va a añadir una nueva manera de interactuar con ellos: a través de la voz. Ahora lo estamos viviendo como si se tratara de un ente individual, pero mi sensación es que la voz será un complemento mas, muy importante, en la interacción.

En el fondo, la magia de esta nueva revolución no está en el tangible, en la voz, sino en lo que hay detrás: esa Inteligencia Artificial (IA) adaptativa que, de alguna manera, supone una gran disrupción. Pero ¿qué pasa? Que esta disrupción es etérea (no hay nada más etéreo que la voz) y por eso necesitamos crear tangibles. Y por eso cuando se habla de voz, por eso se visualiza en

altavoces; cuando no es solo en estos dispositivos donde vamos a poder interactuar por voz. De hecho, no va a ser ni el principal dispositivo.

Será una manera más de interactuar con los dispositivos en los que actualmente ya interactuamos. Seguiremos utilizando otros *hardwares* como el móvil, por ejemplo.

Además, no nos olvidemos de que la magia está en las experiencias hiperpersonalizadas. La fuerza de esta disrupción está en que por primera vez en la historia la tecnología habla nuestro lenguaje y no nosotros el suyo.

2. Has hablado de Inteligencia Adaptativa, como una de las claves de esta revolución. ¿En qué consiste?

La Inteligencia Artificial (IA) Adaptativa es una nueva lógica que aporta la tecnología, que tiene como premisa inicial conocer el contexto y las necesidades de los usuarios, y a partir de ahí se intenta ofrecer la propuesta de valor y la experiencia más ajustada a ese contexto y a esas necesidades.

Siempre se ha perseguido la personalización de las experiencias y creo que con los asistentes virtuales de voz se da un paso más y se construye una experiencia hiperpersonalizada. Para mí, este es un punto esencial del éxito que pueda tener esta disrupción.

3. Estamos ante un nuevo lenguaje de comunicación hombre-máquina (*Human Machine Communication*). ¿Qué lo caracteriza? ¿Qué ofrece la interacción por voz?

Los asistentes virtuales por voz han aparecido ahora porque se han juntado dos cosas: la Inteligencia Artificial (iA) y los servicios cognitivos, que son todos aquellos sistemas de interacción basados en biometría, como el escáner de retina o el reconocimiento facial. La voz es otro servicio cognitivo.

Las principales características de este nuevo lenguaje de comunicación son, por un lado, la hiperpersonalización. El hecho de que interactuemos a través de un mecanismo de comunicación ancestral como es la voz, también connota una mayor humanización. Es inherente en el ser humano interpretar que si interactúo por voz con una máquina, estoy humanizándola.

Esto tiene puntos positivos y no tan positivos. Si yo, cuando hablo con una máquina me imagino que va a entender todo lo que dice y me va a devolver una respuesta de la misma manera que lo hace un ser humano y no lo hace, puede generar un alto nivel de frustración. Estamos al principio de esta disrupción y son muchas las personas que, cuando interactúan por primera vez con un asistente de voz, se sienten frustrados porque intentan emular la satisfacción de la experiencia comunicativa con un ser humano.

La tercera característica es la orientación a la acción. Una de las principales diferencias es que casi todas las propuestas de valor que te ofrecen están orientadas a tareas: ejecutar un servicio, realiza un cálculo, obtener una respuesta...

4. ¿Son los asistentes de voz la base de esta revolución?

Los asistentes de voz se están convirtiendo en los portadores y los tangibilizadores de IA doméstica. Para mucha gente Siri o Alexa son el primer contacto que tienen con IA. Pero estamos aún en una capa muy superficial de esta revolución. Poco a poco veremos como se va diseminando en muchos dispositivos distintos, no solo en los altavoces inteligentes, y que habrá muchas alternativas de asistentes virtuales.

5. ¿Qué es la interacción multimodal y dónde quedará la pantalla?

No me gusta hablar de asistentes virtuales de voz porque justamente se considera que es la voz el único mecanismo a través del cual se va a interactuar. Por eso yo intento hablar de asistentes virtuales con voz, porque cada vez más la voz es uno de los mecanismos de interacción que va a existir. Pero no va a ser excluyentes. De hecho, cada vez más los asistentes virtuales con voz que incorporan pantalla son mayores. Hay cada vez más pantallas que permiten la interacción por voz.

La interacción multimodal consiste en eso: tu puedes iniciar una comunicación a través de una invocación por voz, pero el dispositivo tiene la capacidad de devolver esa información, no solo con voz, sino con un vídeo, con imágenes de apoyo o incluso con una forma más compleja de comunicación. Igual el usuario, que ha iniciado la comunicación con una invocación, necesita acercarse al dispositivo para continuar la experiencia y ver algo y, entonces, la siguiente interacción la realiza a través de un *tap* sobre la pantalla. Y acto seguido la nueva necesidad no va a necesidad una interacción de contacto física, sino que va a volver a ser a través de voz. Eso, para mi, es la interacción multimodal.

6. Hoy en día, donde esta integración aún es incipiente y queda tiempo para que sea total en los hogares, ¿qué tipo de contenidos se están ofreciendo y trabajando en los asistentes?

Empezamos por lo que se está ofreciendo actualmente... Mi sensación es que, actualmente, en el mundo de la voz no existe aún un *category killer* o no se ha extendido lo suficiente. Lo que esto provoca es que se esté experimentando mucho, para descubrir el valor que hay detrás de esta nueva disrupción de la voz.

El ámbito lúdico está cogiendo mucha fuerza; los medios de comunicación también han detectado la potencialidad de este nuevo canal, y están apostando fuerte, sobretodo medios en los que la voz juega un papel importante. Aquí, el mundo de la radio tiene mucho que decir, ya que parece que la revolución de la voz va a ser un buen aliado para que puedan rejuvenecer su propuesta de valor. El mundo de la cocina es otro espacio donde la voz está ganado terreno: acceso a receta, seguimiento de recetas en tiempo real... En general, cualquier experiencia orientada a una interacción del tipo pregunta-respuesta y acceso a determinados contenidos mezclado con la ejecución de ciertas tareas es donde se está moviendo ahora el tema de la voz.

¿Hacia donde va a ir? Eso depende, por un lado, de como la sociedad va adquiriendo esta nueva cultura de interacción, y, por otro, que la industria identifique ese *category killer*.

7. ¿Cómo se comportan los usuarios de asistentes de voz? ¿Qué tipo de interacciones llevan a cabo?

Las principales interacciones están basadas en preguntas y respuestas. También se hace mucho consumo de audio, ya que en un tangible de consumo de audio, los altavoces inteligentes, se entienda que es normal que el audio es más interesante: música, radio, podcasts, audiolibros, audio series... Todo este tipo de contenidos se están consumiendo mucho. También acceso a información meteorológica. También el control domótico, en según que mercados está cogiendo fuerza: control de alarmas, persianas, luces, etc. Y el acceso a información, actualidad y noticias; a medios de comunicación.

8. La privacidad es uno de los grandes retos de esta nueva revolución...

Efectivamente. Esta disrupción ha aparecido en un momento en el cual la ética tecnológica está más en boga que nunca. Cada vez más se está estudiando si el equilibrio entre lo que ofrece la tecnología y el valor que aporta a la sociedad a través de los datos es equidistante o no. Esto, durante el año pasado, fue un elemento de alto riesgo para la disrupción de la voz.

Hubo un momento en que empezó a distribuirse entre la sociedad esta sensación de que te estaban escuchando constantemente. Y realmente, la propia industria declaró que sí se hacían escuchas (Google y Apple incluso lo reconocieron) para mejorar la inteligencia adaptativa de los asistentes. En el *machine learning* existen distintas técnicas para poder entrenar al algoritmo de respuesta. Una de ellas se llama entrenamiento supervisado, a través del cual se obtiene información a través de la experiencia: obtienen la manera como el usuario ha hecho una petición y de qué manera la máquina ha respondido. Esto a la sociedad le provocó un rechazo elevadísimo, aún llegando a entender la finalidad.

La tensión fue tal que a finales de agosto empezó una iniciativa en cadena de desasistir estos programas por parte de los grandes *players* de Silicon Valley para garantizar que este freno en la privacidad no iba a ser un problema en la dispersión de esta disrupción.

9. Otro de los grandes retos de esta nueva tecnología es el nivel de entendimiento del lenguaje humano...

Cada vez va a ir mejorando más el *natural language understanding*, pero no creo que sea el mayor de los retos. Desde hace tiempo se están mejorando mucho los procesos de reconocimiento del habla. Cada vez más, las redes neuronales ayudan a que, con más precisión y menor tiempo, se entienda mejor.

Los ratios de error de la interpretación de los algoritmos de los principales asistentes de voz es bajísimo: hay asistentes como Google Assistant que, incluso, han superado la paridad humana, es decir, la capacidad de un ser humano de interpretar lo que otro humano ha dicho. Con lo cual, es cuestión de tiempo que esto no suponga un reto; no hablo de años sino de meses. Están siempre en permanente mejora.

10. Un tercer reto que me gustaría que valorarás tiene que ver con las distintas lenguas y como pueden o no ser integradas en estos asistentes de voz. Por ejemplo, el catalán.

Aunque existen lenguas que corren el riesgo de desaparecer en estos mecanismos de interacción por voz, lo cierto es que también estos cada vez tienen un procesamiento más rápido de nuevas lenguas.

Por ejemplo, recientemente el Gobierno catalán, muy inquieto porqué ninguno de estos asistentes de voz te permite hablar en catalán, ha iniciado un plan de impulso muy interesante. Pero yo confío

que, tarde o temprano, cualquier persona que lo desee pueda interactuar con las máquinas en su idioma.

11. ¿En qué fase se encuentra el mercado de asistentes de voz en catalán, actualmente?

Es cierto que actualmente no se puede interactuar en catalán, pero sin embargo se está trabajando para poder facilitar la interacción con estos asistentes en este idioma.

12. ¿Existe demanda de asistentes de voz en catalán? ¿Y oferta?

Hay una comunidad muy potente en Catalunya que siempre ha intentado facilitar o hacer presión para que los *softwares* estén en catalán, y es cierto que han conseguido visibilizar el catalán en todas las disrupciones tecnológicas. Tenemos una comunidad activa de habla catalana y una sociedad que está esperando poder interactuar a través de la voz en catalán.

Recientemente, a finales de 2019 la Generalitat publicó un informe en el cual se declaraba que un 88'9% de la población desea poder interactuar en catalán con los asistentes. Y fruto de esto y otras inquietudes, el Departament de Polítiques Digitals y la Direcció General de Polítiques Lingüístiques de la Generalitat han impulsado un estudio en el cual se hace un análisis de cual es el papel y la presencia del catalán en los asistentes virtuales de voz, y cuales son las estrategias para que éste tenga visibilidad.

13. La realidad es que, a día de hoy, los asistentes de voz mayoritarios no entienden catalán ¿Qué dificultades existen y que soluciones se dan a corto plazo?

La CCMA está realizando las mejores prácticas al respecto. Hay que ser muy inteligentes y astutos, entender cuales son las reglas de juego y buscar la manera más sibilina de poder emular esta interacción en catalán.

Lo que se está haciendo para intentar salvar este problema es construir un diseño de conversación que minimice esta sensación de que se está interactuando en catalán. Per ejemplo, con respuestas cerradas de sí o no, buscando vocabulario genérico... Pero es prácticamente imposible y no lo convierte en una interacción natural. Mientras tanto, se está intentando ejercer la influencia necesaria para poder conseguir que haya interacción en catalán.

14. Deben, los medios de comunicación, apostar por los servicios interactivos de voz a través de asistentes? ¿Porqué?

Los medios de comunicación, a día de hoy, los podemos considerar titanes por la cantidad de crisis que han ido encadenando: crisis de modelo, crisis de confianza por parte de la audiencia, crisis económica por su alta dependencia con mecanismos volátiles como la publicidad... Y pese a todo esto han intentado encontrar cual es su nueva vía de supervivencia o de evolución. Sin duda la voz tiene que ser uno de esos canales, y más aún cuando uno de esos tipo de medios de comunicación tiene una relación tan próxima con la voz como es la radio.

Si entendemos esta revolución de la voz desde un punto de vista multimodal, evidentemente, no va a ser una revolución exclusiva de la radio sino de cualquier *broadcaster* o medio de comunicación impreso que haya mutado a digital. Todos son susceptibles de ofrecer una experiencia basada en VOZ.

Dicho esto, sí deberían apostar. De hecho, lo están haciendo. Grandes de las apuestas que están habiendo están siendo lideradas por medios de comunicación, muchos de ellos públicos, que se pueden permitir ciertas inversiones sin la necesidad de obtener un retorno inmediato. Estamos en un entorno en el que se ha de apostar y aprender, y todo medio de comunicación que ahora decida apostar y aprender, cuando la sociedad en paralelo vaya a ir adquiriendo este nuevo canal de comunicación, va a ir obteniendo el beneficio.

15. Los asistentes de voz siempre ofrecen contenidos después de una petición por parte del usuario. ¿Se está estudiando que puedan también hacer *push* en determinadas ocasiones?

Sí, de hecho ya permite *push*, pero teniendo en cuenta el tema de la privacidad hay que tratarlo con mucho cariño. Ahora mismo, un asistente de voz permite, por ejemplo, crear determinados recordatorios; en el momento en el que se tiene que ejecutar este recordatorio, el aro de luz que hay en el asistente virtual de voz, se activa con un determinado código de color y tiene que ser la persona quien invoque que el asistente retorne ese recordatorio. No es posible que, de repente, aparezca un asistente de voz sin haberlo pedido. Es básicamente, por un sistema de control de privacidad, para no generar experiencias demasiado intrusivas.

16. ¿La voz es el canal con el que pides información, y el audio es el formato con el que te devuelven los contenidos? ¿Qué diferencias hay entre estos dos conceptos?

Para mi voz es todo lo que hemos hablado hasta ahora: Inteligencia Adaptativa que interactúa con los seres humanos a través de la voz. Y audio, para mi, es un fenómeno más pasivo: entrega de

contenido a través de audio o *branding* sonoro. Ambos están conviviendo y, uno está apoyando al otro en esta disrupción, pero son dos elementos separados a tener en cuenta.

17. ¿Cuales han de ser las bases para una buena estrategia de voz y de audio?

No creo que tenga mucha diferencia con diseñar cualquier otra estrategia para otro canal. Lo que hay que hacer es estudiar el entorno, en este caso nuevo, y hablar directamente con tu potencial audiencia.

18. ¿Cómo se da a conocer la presencia de marca en los contenidos que se ofrecen a través de los asistentes?

Así como históricamente, la técnica del *branding* se ha centrado en la imagineria visual (logo, paleta de colores, tipografías...), todo lo que es la dimensión sonora de las marcas en general no se ha tenido mucho en cuenta. Sin duda, esta revolución de la voz va a necesitar que las marcas entiendan que el mismo cuidado que están teniendo en la dimensión visual icónica de su marca, se va a tener que tener en cuenta en el entorno auditivo.

Los medios de comunicación, por ejemplo, muy acostumbrados a crear una imagen de marca potente para cada uno de sus programas, sí se sienten muy cómodos en la creación de este universo sonoro. Pero hay otros que no le están dedicando el mimo necesario, y cada vez va a ser más importante para las marcas el diseño de la identidad sonora.

19. ¿En qué esta trabajando Lavinia, con respecto a los servicios de voz interactivos?

Lavinia Voice es un proyecto que está mutando a la vez que lo hace esta revolución. De hecho, consideramos que esta disrupción tecnológica va a ser la revolución más rápidamente adoptada por la sociedad. Y nuestra propuesta de valor se centra en ofrecer estrategia, creatividad, contenidos y tecnología para la creación de experiencias conversacionales de voz y nuevas narrativas de audio online. Lavinia Voice no solo es voz sino que también es audio. No entendemos la una sin la otra.

Nos dedicamos al diseño de *voice apps*: Alexa Skills y Google Actions, sobretodo. También trabajamos en la creatividad y producción de nuevas narrativas de audio *online*: podcasts, audiolibros, audio series... Tenemos también un bloque de servicio que lo hemos llamado *voice marketing*: experiencias de SEO para voz, estrategias de distribución de estos contenidos de audio para hacerlas llegar a la audiencia... También tenemos un bloque de servicio centrado en *voice commerce*. Y hacemos creación de territorio sonoro para marcas. Entre otros.

Algunos ejemplos: Cuentos de Clan o Lunnies de leyenda, de TVE. App de voz de OT a través de Google.

7.2. Entrevista a Lluís Garriga, consejero delegado de la Xarxa Audiovisual Local (XAL)

Fecha: 17/06/2020

Nombre del entrevistado: Lluís Garriga

Empresa: Xarxa Audiovisual Local (XAL)

OBJETIVO DE LA ENTREVISTA

Descubrir qué oportunidades ofrecen los asistentes de voz para los medios de comunicación de proximidad y conocer los proyectos con asistentes de voz que están desarrollando la Xarxa Audiovisual Local (XAL).

INFORMACIÓN DE LOS ENTREVISTADOS

1. Nombre y apellidos: Lluís Garriga
2. Cargo en la XAL: consejero delegado
3. Desde cuando trabaja en proyectos con asistentes de voz: 2020

PREGUNTAS

1. ¿Porqué son importantes los asistentes de voz para los medios de comunicación?

Los sistemas de consumo se están modificando a una velocidad de vértigo y ya no es una cuestión generacional, sino intergeneracional. El elemento vídeo hasta ahora era el elemento estrella, el gran centro de toda la película (sigue siendo el elemento más importante) pero el elemento audio empieza a tomar un cuerpo considerable. Esto es debido a que la difusión de contenidos a través de audio es más rápida, más sencilla de producir y es más accesible y más usable.

En este sentido, nosotros que nos dedicamos a la comunicación local y de proximidad, y que hacemos de organismo prestador de servicios a las televisiones y radios locales, estamos poniendo sobre la mesa un proceso de reflexión avanzada sobre el papel que han de jugar estos nuevos soportes en la producción y difusión de estos contenidos de proximidad.

2. ¿Qué oportunidades se les abren ante estas nuevas plataformas vocales?

Son unas oportunidades infinitas. Hasta ahora hemos visto que los medios de comunicación pueden integrarse en asistentes de voz a través, por ejemplo, de sus contenidos en directo. Pero la realidad es que empezamos a ver ya productos muy personalizados para unos programas concretos; por

ejemplo, el caso de Operación Triunfo, donde el asistente de voz ha sido un elemento complementario que además permitía interactuar y dar información complementaria.

El cambio es infinito en el terreno de los medios de comunicación, pero yo voy más allá: el cambio es infinito en el terreno de la comunicación en general. Hay muchas empresas e instituciones que necesitan comunicar directamente con el usuario y el asistente de voz es una oportunidad inmejorable.

3. ¿Y los medios de comunicación de proximidad, en qué se pueden beneficiar?

Los medios de comunicación de proximidad están muy relacionados con los organismos públicos o semi-públicos que difunden cosas. Por ejemplo: patronatos municipales de turismo, museos, servicios de información de Ayuntamientos... Los asistentes de voz permiten tener experiencias para este tipo de organismos asequibles. Podemos poner como ejemplo la oficina de turismo de la Costa Brava. Los medios de comunicación locales hemos de ser capaces de construir contenidos de comunicación para este tipo de organismos y distribuirlos a través de asistentes de voz. Hemos de ser capaces de producir contenidos también para otros entornos de comunicación, no solo para la televisión o la radio; y los asistentes de voz, en este punto, son claves.

Un medio de comunicación de proximidad que ofrezca solo servicios de noticias a través de asistentes de voz, no podrá competir nunca con el servicio de un medio de comunicación nacional y generalista. En cambio, nadie podrá competir con un servicio de comunicación de proximidad. Es en este sentido donde creemos que hay espacio muy importante para los medios de proximidad en los asistentes de voz.

Estamos abriendo el espectro para volver a poner en valor los medios de comunicación locales y dejar a un lado la concepción del medio de comunicación local tradicional.

4. ¿La Xarxa Audiovisual Local (XAL) está trabajando en algún proyecto que incluya asistentes de voz? ¿Cual? ¿Cuándo verá la luz?

Estábamos comenzando nuestro proyecto con asistentes de voz, en un proceso de reflexión estratégica sobre la mesa, pero la crisis sanitaria del Covid-19 lo ha parado todo. Teníamos ya un par de proyectos interesantes en los que empezar a trabajar. Ahora tenemos que retomarlos a partir de setiembre. No podemos avanzar demasiado, pero la idea es que este año podamos arrancar con

dos proyectos piloto: uno vinculado a un proyecto de turismo de proximidad y otro vinculado a un servicio de información de una institución.

5. ¿Porqué habéis decidido apostar por tener presencia en asistentes de voz?

Por una razón muy sencilla. Vemos que es necesario que las entidades que ofrecen servicios de comunicación de proximidad, o amplían su rol de difusión y de producción de contenidos o su posicionamiento en el mercado cada vez es más complicado. Por lo tanto, es importante abrir un proceso de reflexión y empezar a pensar en el futuro inmediato que incluye las nuevas tecnologías y, por lo tanto también, los asistentes de voz.

7.3. Entrevista a Geni de Vilar, adjunta a la dirección del área de Medios Digitales de la CCMA

Fecha: 21/11/2019

Nombre del entrevistado: Geni de Vilar

Empresa: Corporació Catalana de Mitjans Audiovisuals (CCMA)

OBJETIVO DE LA ENTREVISTA

Conocer de primera mano cual es la misión y los valores de la CCMA, y su estrategia digital como medio de comunicación.

INFORMACIÓN DEL ENTREVISTADO

4. Nombre y apellidos: Geni de Vilar

- Formación y experiencia: Licenciada en periodismo y Máster en Diseño y Desarrollo de Sistemas Interactivos.
- Cargo dentro de la CCMA: Adjunta a dirección del área de Medios Digitales de la CCMA.

PREGUNTAS

1. ¿Cuál es la misión de la Corporació Catalana de Mitjans Audiovisuals (CCMA) como medio público?

Somos el medio de comunicación público del país – Cataluña -. Nuestra misión fundacional se basa en el compromiso con la calidad y la pluralidad, y lo hacemos con vocación de liderazgo. El eje vertebrador de nuestro servicio es la promoción de la lengua, la cultura y la identidad catalanas. Nuestra misión de difusión de la lengua, la cultura y la identidad catalanas solo la conseguiremos si hacemos contenidos que sean líderes. La gente no vendrá a vernos porque hagamos cosas en catalán, sino porque informamos bien, porque entretenemos. Si no lo hacemos, esta misión no la estaremos cumpliendo.

2. ¿Cuál es la propuesta de valor del medio de comunicación y qué lo diferencia de la competencia?

Nosotros tenemos una clara vocación de servicio público, que los medios de comunicación privados no necesariamente tienen. De hecho, el periodismo tiene ya por sí mismo una vocación de servicio, pero nosotros como medio público lo tenemos de manera fundacional. A nivel de contenidos, tenemos la obligación de informar al país, mientras que la competencia informa porque quiere ganar

dinero (son empresas privadas que venden un producto, la información). Además, nos mueve el compromiso con la lengua y la cultura catalanas.

También tenemos el deber de entretener. Intentamos que sea un entretenimiento de calidad y reflexivo, no sensacionalista. No utilizamos el formato de tele-basura; por ejemplo, apostamos por dos *prime time* documentales que tienen muy buenos niveles de audiencia. Este tipo de acciones son las que nos diferencian de los medios de comunicación privados y del resto de la competencia.

3. ¿Cuánta gente hay actualmente en la CCMA y cómo están distribuidos?

Actualmente, hay cerca de 2.200 trabajadores en plantilla. Para entender como están distribuidos en la organización hay que entender que la transformación digital que ha afectado al medio de comunicación ha hecho que la propia organización varíe y evolucione.

Hasta hace unos años la CCMA se dividía en diversas empresas: TV3, Interactiva, Activa Multimedia (que se encargaban de explotar los contenidos de TV y radio por Internet) y Catalunya Ràdio. Hace 9 años, se produce una fusión por absorción de las empresas tecnológicas de la corporación, que en ese momento eran las únicas que trabajaban el tema digital. En ese momento, TV3 absorbe a esas personas y crea el área de servicios interactivos.

Hace seis años se produce una unificación administrativa y todo queda bajo el paraguas de una misma empresa: la Corporació Catalana de Mitjans Audiovisuals S.L. Es en este momento cuando las personas que trabajaban en la parte digital se reparten. Se generan pequeñas redacciones digitales que se dispersan en los distintos equipos: informativos, meteo, Super 3, deportes... Cada redactor digital lleva 4 o 5 programas y se encarga de diseñar como será su estrategia digital. Además, cada programa tiene también un *community manager*. También se generan equipos de productos digitales, que se encargan de trazar la estrategia digital y el diseño de los servicios (la nueva web, la nueva aplicación...); trabajan codo a codo con los ingenieros y el área técnica.

Desde el departamento de Medios Digitales tenemos el trabajo de definir la estrategia digital y acompañarlos. Además, también llevamos la comunicación de las marcas principales: TV3 y Catalunya Ràdio.

4. ¿Y actualmente, como están organizadas las redacciones de la CCMA?

Impregnamos, cada vez más, la mancha digital dentro de la organización. Todos los programas tienen su propio *community manager* e incluso dentro de algunos departamentos, como en informativos o deportes, tenemos equipos dedicados a hacer noticias exclusivamente para el digital; por ejemplo, generar vídeos nativos para Internet. Estos equipos han visto que gracias a ello pueden conseguir un rendimiento mucho más elevado de los contenidos, pero para ello es necesario una mínima formación, por ejemplo, en SEO o en titulación para web y redes sociales. La formación en este sentido, incluso en periodistas con más de 30 años de trayectoria, es imprescindible.

5. ¿Está costando este proceso de formación?

Es un proceso que sí está costando y, de hecho, no sabemos si acabará. La única constante que tenemos es el cambio. Estamos en un cambio completo de rutinas y eso, en personas con más de 35 años de trabajo es difícil, pero es importante que se reciclen. Es un proceso y somos conscientes de que es importante no forzar. Sabemos que nadie quiere quedarse atrás y no estar actualizados, pero tiene miedo. Es importante hacerlos entender hacia donde están yendo y porqué, y además, han de sentirse acompañados. Nosotros tenemos claro que, si la tecnología los bloquea no vamos a forzar, ya que lo importante es que sea buen periodista; en última instancia ya pondremos nosotros sus historias en formato digital.

En este tipo de procesos transformacionales, el estilo de liderazgo es clave, y los perfiles autoritarios y menos comunicativos y transparentes no son buenos. Es importante que los trabajadores entiendan que la marca TV3 ya no es un canal de televisión, exclusivamente, sino una marca generadora de contenidos que se expresa a través de muchas ventanas y de muchos formatos diferentes. Lo importante es que la empresa se vaya transformando poco a poco.

6. ¿Tiene la CCMA una estrategia de transformación digital?

No tenemos un documento estratégico de transformación digital, como tal, pero si sabemos hacia donde queremos ir. Internet ha de estar en el mismo nivel conceptual que la televisión y la radio. Por eso, trabajamos con una visión claramente *digital first* y *mobile first* en la cobertura de informativos, por ejemplo. Hemos cambiado muchas rutinas que provocan necesariamente cambios en el enfoque editorial de los programas: por ejemplo, ya no han de informar tanto, sino que han de interpretar, puesto que la gente ya se ha informado a través del móvil durante todo el día. Dibujamos una hoja de ruta que marca nuestro rumbo y que revisamos de manera anual. Una hoja de ruta de más de dos años no tendría sentido en un entorno tan cambiante como el digital.

7. ¿Existe un departamento de innovación?

No tenemos un espacio específico dedicado a la innovación. Trabajamos con un modelo diseminado, donde la innovación surge de personas concretas de la empresa; por eso hemos creado una red con personas de perfiles innovadores de los distintos departamentos que tienen relación entre ellos: damos a personas concretas el rol de innovadores y hacemos que estén en red entre ellos. Hemos pasado de un modelo concentrado a uno diseminado; una idea más matriarcal y no tan jerárquica.

7.4. Entrevista a Judith Argila, responsable de Nuevas Plataformas y Redes Sociales, en el departamento de Medios Digitales de la CCMA

Fecha: 07/04/2020

Nombre del entrevistado: Judith Argila

Empresa: Corporació Catalana de Mitjans Audiovisuals (CCMA)

OBJETIVO DE LA ENTREVISTA

Conocer el proyecto que la Corporació Catalana de Mitjans Audiovisuals (CCMA) está llevando a cabo con asistentes de voz, como herramienta de distribución de sus contenidos informativos y de entretenimiento.

INFORMACIÓN DEL ENTREVISTADO

Nombre y apellidos: Judith Argila Blanco

Formación y experiencia: Licenciada en periodismo. Posgrado en Visualización de Información y del Conocimiento en la UPG. Master en Sociedad de la Información y el Conocimiento, especializada en la línea de investigación en Tecnologías Digitales (UOC). Trabajo de Final de Master (TFM) sobre asistentes de voz. Período de investigación en Oxford durante 6 meses en el Reuters Institute.

Cargo dentro de la CCMA: Responsable de Nuevas Plataformas y Redes Sociales, en el departamento de Medios Digitales.

Desde cuando trabaja en el proyecto de asistentes de voz: Agosto 2018.

PREGUNTAS

1. ¿Estamos delante de un cambio de paradigma, donde la voz se va a convertir en la nueva interfaz de comunicación? ¿Dónde quedará la pantalla?

Creo que hemos vivido una época donde la semántica y sintaxis común que hemos creado como sociedad para hablar con las máquinas se ha basado en un lenguaje visual, el lenguaje de las pantallas, el lenguaje de la arquitectura de la información, donde hemos consensuado que un rectángulo en la parte superior de la pantalla implica una cabecera y, por lo tanto, es ahí donde vamos a dar la imagen de la marca. Y que si hay cuatro líneas paralelas, todos entendemos que eso es un menú. Son convenciones visuales que se han creado y se han estandarizado a lo largo de 20-25 años y creo que estamos empezando a mutar hacia un mundo donde esa interfaz visual va a ir

dejando paso a otra interfaz que es auditiva. Está basada en un lenguaje mucho más primigenio para el ser humano, que es la palabra hablada.

No sé hasta qué punto va a ser el cierre de una época para el inicio de otra. Igual que cuando llegó la televisión se hablaba de si la radio desaparecería, ahora no creo que con las tecnologías de la voz vaya a desaparecer la pantalla ni la forma que hemos desarrollado de comunicarnos con las pantallas. Pero sí creo que va a ir a menos. Va a ir dejando espacio a esta nueva interfaz que es la VOZ.

Creo que se van a combinar, al menos durante la época que nos va a tocar vivir, pero también creo que la pantalla como tal va a ir desapareciendo en el sentido de superficie física, como algo integrado en un dispositivo. Creo que iremos a un entorno donde las pantallas van a ser cualquier superficie sobre el que podamos proyectar información. Va a ser un entorno donde la información va a estar presente y a través de voz vamos a invocar qué informaciones necesitamos y la información nos va a ser devuelta, o bien a través de la voz o bien proyectándose sobre la puerta de la nevera o el techo de la habitación, en paneles neutros en las autopistas... Me imagino un sitio donde empiezan a convivir esos dos mundos a largo plazo.

La voz va a ir tomando posiciones porque tiene dos grandes ventajas:

- 1) Es más inclusiva porque no necesita el conocimiento de este lenguaje estándar que hemos creado a lo largo de los años y eso nos permite incorporar muy rápido usuarios de edades muy tempranas que pueden hablar rápidamente con interfaces de voz. Y también nos permite incorporar usuarios analfabetos que, por ejemplo, no han aprendido a leer pero sí hablan y entienden. Esto va a ser interesante en países donde hay altos niveles de analfabetización, pero también para las personas de avanzada edad que quizás no se aclaran con la pantalla de un móvil para programar un recordatorio, pero pidiéndoselo a un asistente sí pueden hacerlo. Tenemos ahí una capa de la sociedad que vamos a poder incorporar.
- 2) Operar por voz resulta mucho más sencillo, ya que no tienes que depender del dispositivo en las manos, ni tener los ojos y todos los sentidos enfocados en un punto que secuestra la atención, como es la pantalla. Con la voz podemos invocar al aire mientras estamos conduciendo, cocinando, haciendo ejercicio... Hay un punto de libertad de la vista que va a tener un peso y que, por lo tanto, va a ayudar al despegue de este tipo de servicios a través de voz.

2. ¿Cómo los medios de comunicación, y en concreto la CCMA, se están preparando para este cambio de paradigma? ¿Qué le puede aportar y qué retos tiene ante sí?

Como servicio público, la CCMA tiende a pensar en términos de qué podemos aportar nosotros como servicio público; qué deberíamos hacer para ponernos al servicio del ciudadano y ofrecerle información, entretenimiento, acompañamiento... a través de estos nuevos entornos. Dentro del departamento de Medios Digitales tenemos un equipo fijo enfocado en desarrollo de contenidos y estrategias de distribución para las plataformas digitales. Es un equipo pequeño pero siempre dedicado a esto.

Actualmente estamos en la fase de cómo utilizar el contenido que se está generando desde las redacciones y reformatearlo para crear una experiencia de usuario que tenga sentido en los entornos de asistentes de voz y aporte valor. Por ejemplo: cómo integramos el directo de Catalunya Ràdio en un entorno servido por la voz, o cómo las ediciones del TN en vídeo pueden integrarse en un entorno de voz, si es que puede integrarse. La primera fase se basa en contenido ya existente. Estamos creando la interfaz de relación que empaqueta ese contenido, lo transporta... Estamos creando los diálogos, pero el contenido es el que ya existía. Una segunda fase, que no estamos abordando, sería la elaboración específica de contenidos para este entorno.

3. ¿Cuándo empezasteis, en la CCMA, a trabajar en el proyecto de los asistentes de voz como plataforma para la distribución y consumo de información y entretenimiento? ¿Porqué lo considerasteis oportuno?

En septiembre de 2018 estrenamos una colección de *flash briefings* para Amazon Alexa. Nosotros acostumbramos a buscar en todos los proyectos un punto medio entre la penetración que tienen los dispositivos o plataformas en mercado (tenemos la obligación, como servicio público, de estar allí donde está la gente). Cuando hay una congregación importante de usuarios alrededor de una nueva tecnología, nosotros debemos tener presencia. Pero luego hay otros casos, como el de los asistentes de voz, en qué sin haber una masa crítica de usuarios a la que servir, nosotros valoramos si como servicio público y como innovación tiene sentido que estemos ahí.

En el departamento de innovación, que llevo 12 años liderando, veíamos que el tema de asistentes había cogido mucha fuerza en los países de habla inglesa, donde se había desarrollado una

capacidad de comprensión del lenguaje importante, y, con la documentación existente y los estudios que hay, le veíamos capacidad de proyección. Entonces decidimos hacer una apuesta.

También hay que tener en cuenta que las industrias evolucionan a pasos distintos. El tema de los asistentes en castellano lleva una velocidad que no es la de las interfaces en inglés y nosotros intentamos adaptarnos a esas necesidades. Lo que no podemos hacer es correr nosotros más que la industria. Al no ser el primer mercado, sino más bien el segundo o el tercero, generalmente las vemos venir. Hay olas de consumo que nos vienen de fuera, pero que nos permiten ver como se van a integrar en nuestro mercado.

4. ¿Qué pasos habéis seguido, desde que empezasteis el proyecto con asistentes de voz?

Lo primero que hicimos fue crear un pequeño equipo de 3 personas, que somos las que actualmente nos dedicamos a los medios digitales. Hay un perfil de desarrollador técnico, un perfil de contenidos y un perfil de estrategia de producto. Empezamos haciendo una inmersión en el mundo de Amazon Alexa para entender qué tipo de sistema habían creado ellos y como nosotros teníamos encaje dentro de ella. Hicimos un poco de auto-aprendizaje con los recursos que ofrece Amazon, para entender qué tipo de *skills* se pueden desarrollar en ese entorno, las características de cada una, etc.

Empezamos optando por un formato de *skill* que se llama *flash briefing* o *news briefing*, que es de muy sencilla implementación. Generamos unos RSS de contenidos para incorporar en estos *flash briefings* y empezar a servirlos. Eran materiales ya existentes que creíamos que iban a tener sentido servidos bajo la lógica de los *news briefing* de Alexa: los boletines informativos de Catalunya Informació, los titulares de los telenoticias, la portada de Mònica Terribas, la sección de Albert Cuesta y la meteo. Amazon nos decía que teníamos que buscar piezas de entre 2 y 7 minutos y buscamos contenidos en ese sentido.

También pensamos en poner cápsulas de recetas de cocina. Aunque no tiene mucho que ver con ponerte al día, Amazon nos decía que el contenido de cocina siempre funciona. Inicialmente, pensamos en ofrecer una cápsula de 1 minuto, con la receta del día de nuestro programa de cocina, el *Cuines*. Pero nos dimos cuenta que este contenido enseña con imágenes todo el proceso de una receta y lista los ingredientes de pantalla, con lo cual en los dispositivos sin pantalla este contenido quedaba descartado.

Durante el proceso de creación de estas *skills* aprendimos algunas cosas que podían resultar menos obvias al principio. Por ejemplo, ellos nos decían que estos contenidos se iban a servir como una fuente de noticias cuando el usuario pidiera: “Alexa, dame las noticias”. Bajo esta petición, nosotros teníamos que encontrar cápsulas de contenido que tuvieran sentido consumidas desde un dispositivo sin pantalla. Por ejemplo, podíamos utilizar un telenoticias pensado para televisión, pero hay muchas partes en qué el relato se está pivotando sobre una imagen de fondo. Entonces, si no es 100% comprensible sin imágenes no es válido para Amazon. Al final, lo que hicimos fue coger las entradas de los telenoticias, los titulares, que es una locución de voz muy rápida y en 4 minutos te ha repasado el arco informativo del día.

Fue un proceso de creación y análisis, donde valoramos opciones de contenidos que desde un punto de vista de usuario pudieran tener valor y sentido, y a la vez que nosotros tuviéramos contenido formateado de una manera que pudiéramos hacer ese encaje.

El siguiente paso fue crear una *skill* propia que no fuera una fuente de noticias integrada dentro del gran paraguas de funcionalidades de Amazon, sino que fuera una especie de aplicación nuestra que el usuario pudiera invocar y donde nosotros sirviéramos, en este caso, el directo de Catalunya Ràdio. Esta *skill* se estrenó en septiembre de 2019.

5. ¿Cómo está avanzando el proyecto? ¿En qué estado está actualmente (abril 2020)?

El último paso ha sido estrenar la incorporación de todo el contenido de programas a la carta, dentro de esta misma *skill*. Es decir, que no solo puedas pedir el directo de una emisora, sino también una sección concreta de un programa concreto.

6. ¿Con qué dificultades os habéis encontrado?

La principal dificultad que tenemos es que, en el momento que salimos del formato *news briefing* y creamos nuestra propia *skill*, tenemos que definir todo el modelo de diálogo que se puede dar. La primera dificultad que nos encontramos es que Alexa habla y entiende castellano y nosotros no solo desarrollamos contenidos 100% en catalán, sino que además nuestra interacción con el usuario es en catalán (es uno de los motivos de ser de la CCMA). Entonces estamos generando un mundo que se tiene que integrar en otro que cuando le habla no le entiende. Ahí hemos tenido el 95% de nuestros problemas.

El gran problema es que cuando un usuario pide: “Alexa, l’ofici de viure”, Alexa entiende cualquier cosa menos “L’ofici de viure”; Alexa no descodifica las palabras y tenemos que hacer un trabajo minucioso por detrás para hacerle entender a Alexa lo que realmente pide el usuario. El gran reto es que como haya una palabra fuera de sitio que Alexa no entienda y nosotros no hayamos previsto el caso, la experiencia de usuario llega a una calle sin salida, y eso puede llevar una gran cantidad de frustración. La experiencia de usuario con Alexa y los contenidos en catalán puede ser altamente frustrante y desgastante.

El segundo problema ha sido ver que, al fin y al cabo, estamos operando en un entorno que han desarrollado unos terceros, y dependemos mucho de las lógicas propias de Amazon. Es decir, no somos 100% amos de la experiencia de usuario. Por ejemplo, si para invocar una *skill* el usuario utiliza una estructura de frase que Amazon no ha previsto, simplemente cambiando el orden de las palabras o el verbo, eso ya no funciona. Y a veces, esas formulaciones que nos obliga a hacer Amazon pueden resultar muy incómodas para un castellano hablante, porque son estructuras que se han traducido del inglés. No se ha hecho una adaptación profunda a la lógica del castellano.

Además, las frases de invocación de los usuarios han de responder a unas estructuras muy concretas para que Alexa sea capaz de descodificarlas. Por ejemplo, cuando un usuario quiere activar un podcast concreto, no basta con decir: “Alexa, quiero escuchar *Un restaurant caníbal a Berlí*”, que es la fórmula que a la gente le saldría de natural. Lo primero que le sucede a Alexa ante esta invocación es que, como hay muchísimos desarrolladores de *skills*, no sabe si eso se supone que lo tiene que contestar ella de forma natural o es una petición que el usuario le está haciendo a alguna de las miles de *skills* que hay desarrolladas.

Por ese motivo, lo que te pide Amazon es que, forzosamente, el usuario cite el nombre de tu *skill* para que ella entienda que eso es una petición que va dirigida a esa *skill*. La fórmula que ellos nos proponen es: “Alexa, pide a Catalunya Ràdio el podcast *Un restaurant caníbal a Berlí*”; una traducción directa del inglés: “Alexa, ask BBC for...”. El problema que vemos en esta formulación es que mencionar a la *skill* para el usuario ya es raro, pero además la manera como se formula también lo es. Lo natural para nosotros sería: vocativo + complemento directo + complemento circunstancial. Es decir: “Alexa, pon el podcast *Un restaurant caníbal a Berlí* en Catalunya Ràdio”.

Lo que implica esto es que las conversaciones que empiezan con un lenguaje natural, acaben desencadenando en un bucle de “lo siento no te entiendo”. Salvar esto es un trabajo de chinos,

porque tienes que conseguir que encajen lo que el usuario hace de forma natural con lo que la máquina realmente entiende. Y ahí es donde nos hemos pasado meses intentando encajar. Hemos intentado trabajar mucho la mejora de la experiencia de usuario, pero obviamente hemos estrenado con el sentimiento de que no es el tipo de experiencia que nos hubiera gustado ofrecer si, por ejemplo, hubiéramos tenido un asistente propio.

Creemos que podemos ofrecer un servicio, al precio de intentar educar mucho al usuario. El usuario tiene que sacrificar una de las principales ventajas de las interfaces por voz que es el lenguaje natural. Estas tecnologías son super útiles y tienen mucho futuro si entienden lo que digo cuando hablo. Si ellas no me entienden a mi y tengo que adaptarme yo a ellas, volvemos a lo de antes: tengo que aprender un lenguaje específico para comunicarme con la máquina, porque esta no entiende el lenguaje natural.

Hay un abismo entre el estado de la tecnología en inglés, donde el procesamiento del lenguaje natural está más desarrollado, y el estado de la tecnología en castellano. Incorporar el catalán en los planes de Amazon ni siquiera está en su *roadmap*.

7. ¿Cuál ha sido la respuesta de la audiencia? ¿Tienes cifras de usuarios?

La sensación general de nuestra *skill* ha sido muy positiva. La comunidad de usuarios que hay usando asistentes de voz en Cataluña es pequeña todavía. Se hablaba de un 4% de penetración de asistentes de voz entre población de internet. Nosotros rondamos los 5.000 usuarios mensuales, aproximadamente; una cifra pequeñita para los consumos que tenemos de televisión y de radio habitualmente. Es una comunidad pequeña, pero la respuesta es muy positiva. Ha habido mucha reclamación para que incorporemos contenidos a la carta, es decir, podcasts.

La principal queja de esta pequeña comunidad de usuarios que utilizan asistentes en Cataluña es que se nos reclama porque solo estamos en Amazon Alexa y no en Google Assistant, su principal competidor de mercado en España.

8. ¿Y por qué no estáis en Google Assistant?

No estamos en Google Assistant porque hace un año y medio que lo intentamos, pero cada vez que enviamos material a certificar Google nos lo rechaza porque dice que nuestro material está en un idioma que no es el de la interfaz. No contempla que el usuario pueda pedir unos contenidos en un idioma, el castellano, y los reciba en otro, el catalán, pese a que la realidad del territorio sea

bilingüe. Google defiende que esto no es una buena experiencia de usuario y ellos protegen la experiencia del usuario.

Hemos estado bloqueados con este tema durante un año y medio y no hemos conseguido avanzar, pero parece que en las últimas semanas hemos hecho mejoras en este sentido, porque estamos trabajando en un formato para paquetizar los *flash briefings* que tenemos en Alexa e intentar que Google entienda que nuestros contenidos están en castellano y pasemos el control de validación. Así que es posible que en breve podamos estar en Google Assistant.

9. ¿Qué tipo de usuarios consumen información y entretenimiento a través de Alexa? ¿A qué usuarios queréis destinar este servicio, desde la CCMA?

No tenemos un usuario ideal, porque somos un servicio público y entendemos que cualquier ciudadano puede ser nuestro perfil.

Nosotros creíamos que a través de estos entornos íbamos a llegar a gente más joven, porque la CCMA a través de tele y radio tiende a tener una audiencia envejecida, y en el entorno digital bajamos esa media un poco, entorno a los 45 años. Pensábamos que con el tema de asistentes podíamos llegar a capas de la sociedad un poco más jóvenes, pero a sorpresa ha sido que no. Que no son necesariamente los adolescentes y jóvenes entorno a los 20 años los que están metidos en el mundo de asistentes de voz. Es más bien un perfil tecnológico que roza franja de los 35-50.

10. ¿Qué posibilidades ofrece este nuevo canal, los asistentes de voz, a los medios de comunicación?

Una gran oportunidad es la personalización de los contenidos a los usuarios. Es un entorno en el que a través del diálogo podemos llegar a servir al usuario contenido muy específico, de una forma sencilla. Hasta ahora, cuando quieres buscar una determinada entrevista en el entorno web, son necesarios unos circuitos de búsqueda que implican un trabajo. Para mí, lo ideal sería que llegara un momento en que nuestros contenidos estuvieran tan bien documentados, estructurados, indexados e informados, a nivel de metadatos, que pudiéramos hacer extracciones de material muy particular a través de las palabras clave naturales que te de el usuario. Por ejemplo, que le pudieras decir al asistente: “Ponme la entrevista que hizo Mònica Terribas a Rosalía donde hablaban de calamares”, y que nosotros a través de metadatos pudiéramos ofrecerla.

Yendo más allá, lo ideal sería que el propio asistente pudiera construir un relato, que no fuera la pieza informativa de un determinado medio de comunicación. Es decir, que según la información que consultara el usuario, el asistente pudiera construir un diálogo propio a petición específica de un usuario. Es un modelo mucho más personalizado donde el asistente, e incluso la marca informativa, se convierte en una voz que puede responder a tus necesidades de información o entretenimiento como si fuera una conversación natural. Llegar a un punto donde se construyeran contenidos *ad hoc*.

También una oportunidad interesante sería integrar la información a nivel geográfico. Para que, por ejemplo, cuando alguien esté viajando por territorio catalán pudiera pedir al asistente: “Dame información sobre este territorio”.

Se abren, también, puertas en el tema infantil, con todo lo que tenga que ver con educación y entretenimiento. Por ejemplo, que los niños pudieran hablar con sus personajes preferidos y que estos personajes pudieran enseñarlos, por ejemplo, a lavarse los dientes.

La tecnología todavía está muy verde para todos estos escenarios y, además, hace falta mucho trabajo de documentación. Aunque hay mucho trabajo hecho de años anteriores, pero pensado para televisión hace 30 años, hay que ver como todo este trabajo puede conectar con las nuevas interfaces de voz; hay que conectar dos mundos. Hay trabajo para años, pero muchos escenarios y muchas posibilidades.

11. ¿Podemos encontrar en los asistentes de voz un nuevo modelo de negocio? ¿Cómo se dibuja? ¿Qué dificultades y qué oportunidades presenta?

Hay uno, hoy, que sería el básico: integrar cápsulas publicitarias convencionales en los contenidos, como hacemos en otros entornos, o a través de *branded content* (cápsulas de contenido patrocinadas).

Otro de los modelos de negocio y por el que Amazon empieza a apostar es el modelo *freemium* de *skills*, en los que la descarga de la *skill* es gratuita pero luego hay funcionalidades extra por las que si quieres pagas. Este segundo escenario es válido para los medios privados, pero nosotros en la CCMA no. No usamos nunca un modelo en el que se tenga que hacer pagar al usuario, puesto que somos un servicio público.

12. ¿Queda muy difuminada la marca del medio, en el contexto de los asistentes?

Hemos trabajado mucho en el nuevo diseño del *branding* sonoro de la *skill* de Catalunya Ràdio, para posicionar a través del sonido. Hemos creado nuevas sintonías con un diseño específico para asistentes de voz, adaptando las clásicas con un toque digital y sintético. También hemos trabajado mucho los momentos en los que el usuario recibe estos impactos sonoros.

Además, hemos creado una voz corporativa para dar la bienvenida al usuario en este nuevo entorno. Cuando tu pides el directo de Catalunya Ràdio, por ejemplo, no entras directamente al directo, sino que hay una transición donde está este sello sonoro y la voz que te da la bienvenida. Nos hemos asegurado de empaquetar la experiencia con diseño sonoro de marca.

En los *flash briefings*, sin embargo, la marca sí queda más diluida. *En las skills* estamos trabajando para que sea justo lo contrario.

- A nivel periodístico, ¿adaptar los contenidos a los asistentes de voz ha provocado cambios en las rutinas periodísticas? ¿Qué cambios?

A nivel de producción de contenidos no, porque no se han hecho contenidos específicos. Lo que sí ha cambiado han sido las rutinas de producción. Una cápsula que antes nadie cortaba porque nosotros ofrecíamos el telenoticias entero, ahora tiene que haber alguien que los corte. Y que esa pieza suelta quede a disposición del asistente.

13. ¿Toma mucha más importancia el formato que antes? ¿Los periodistas, entonces, han de estar mas centrados en la forma que en el contenido?

Nosotros aún no hacemos contenidos específicos, pero cuando se elaboran contenidos específicos hay una parte de las rutinas periodísticas que no cambia: buscar el tema, los protagonistas, hacer la investigación... Una vez tienes eso, el formato va a ser distinto según el medio donde lo vayas a emitir: televisión, radio, redes sociales o, en este caso, asistentes de voz.

El formato siempre ha tenido importancia, pero estamos tan adaptados a ellos que creemos que hacer periodismo es hacer ese formato, y no. El formato siempre ha sido importante, tanto que hemos integrado de manera intrínseca lo que significaba contar historias. Igual hace falta un ejercicio de consciencia de que otros formatos son posibles. Con las redes sociales, que trajo una fragmentación de las narrativas muy bestia, las redacciones ya vieron venir este golpe. Ya han

atravesado esta primera ola y ya asumen que hay formatos distintos. Ya hay consciencia de que no todo es una cápsula para televisión o para radio.

7.5. Entrevista a Jeremy Walker y Chris Dix, *Head of Technology Transformation and Voice Management* y *Head of Technology Transformation and Product Architecture* de la BBC

Fecha: 15/05/2020

Nombre de los entrevistados: Jeremy Walker y Chris Dix

Empresa: British Broadcasting Corporation (BBC)

OBJETIVO DE LA ENTREVISTA

Conocer el proyecto que la British Broadcasting Corporation (BBC) está llevando a cabo con asistentes de voz, como plataformas de distribución de información y entretenimiento.

INFORMACIÓN DE LOS ENTREVISTADOS

Nombre y apellidos: Jeremy Walker y Chris Dix.

Cargo dentro de la BBC:

Jeremy Walker – Head of Technology Transformation and Voice Management.

Chris Dix – Head of Technology Transformation and Product Architecture.

PREGUNTAS

1. Are we facing a paradigm shift, where voice is going to become the new communication interface? Where will the screen be?

The screen is likely to be complementary, used in combination with voice. The conversation experience does have its limit, so the screen can assist with navigation and will still be very important.

Voice is the most valuable input, but it is half of the picture. It helps to unlock the value of the picture, but other things are also very important around personalization and recommendations.

2. What future do you see in voice technologies, for the media?

Current assistants are instructions based, but they could become more of a companion for audiences. The more it knows about audiences the more it can offer the right content on the right interface at the right time. Voice assistants could therefore become an interface to allow machine/computation to understand an audience member better.

3. What possibilities does this new channel, voice assistants, offer to the media?

There are opportunities for the BBC to use voice assistants in education (as a theme, rather than a product). Learning process is a feedback loop, voice interface gives you the ability to supercharge that loop. It can also extend concentration time.

4. Can we find in it a new business model for media? How do you draw it? What difficulties and opportunities do you present, in this regard?

There have not been many successful uses cases yet, but voice may also assist the commercialization of content. For example, a person with a low-attention span who wants to have their attention grabbed quickly and be told why something matters to them on a personal level.

Online ad serving currently works by making a judgement call based on previous click throughs to build data profiles. But the technology that enables voice and voice content discovery really helps to understand the kind of content that resonates on a much more meaningful level. Conversational interface allows a much more meaningful interaction and therefore a much better profile built of the consumer – so deeper personalization and (for commercial organizations unlike the BBC in the UK) commercialization of content serving.

5. How are the media, and in particular the BBC, preparing for this paradigm shift? What can it bring you? What challenges do you face? And what can you offer to the audience with interactive voice service?

One of the key challenges we face in general is the necessity of platforms for audiences to be able to access our content. Often users of BBC content (whether it is news or audio, for example) will be intermediated by a platform, such as Google or Apple aggregating news articles or podcasts. These platforms often have a great deal of market power and differing objectives from the BBC. The same is true for voice assistants – BBC services and content are available through a range of platforms. As explained below, voice assistants offer another way for audiences to access the depth and breadth of BBC services.

6. When did you start working on the project of voice assistants as a platform for the distribution and consumption of information and entertainment?

Summer 2017.

7. What have been the steps you have taken, from the beginning, in this interactive voice service project?

We considering building our proposition on other services but we realised that there is an issue around how platforms are mediating our content, so we looked at how we might deal with that business and consumer challenge. It is technically possible to create deeper experiences for audiences with voice, but we could not do that on other platforms – audiences would not have as rich an experience of BBC services. So we started to validate idea that we could do it ourselves. It also allows us to have direct relationships with audiences meaning that the typical issues with powerful platforms that we face (e.g. access to data, brand attribution) are not at issue.

8. About integrating your content into Alexa and other voice assistants... What phase is this project in? What results and conclusions have you drawn?

We provide a range of content on Alexa through BBC “skills”. Further information can be found here: <https://www.bbc.co.uk/usingthebbc/account/the-bbc-on-voice-devices/>

9. What difficulties have you encountered?

It is not possible to create the functionality and depth of user experience we would like. In addition, the access to data and the attribution is limited. Attribution and prominence in particular.

10. Why did you decide to develop your own voice assistant, named “Beeb”? What features does it have?

The BBC produces and hosts an extremely wide variety of content ranging from television programmes, news, podcasts, radio, live events etc. A voice assistant will help users to access this content.

It is clear that voice assistants are increasingly important, with one in five adults in the UK using a voice-enabled smart speaker last year. However, in this area we are increasingly forced to rely on huge US-based tech companies to distribute our content to audiences.

By creating Beeb, we can build an assistant that is trusted and puts audiences at its very heart. It will serve up what people want to watch or listen to - from across the BBC (from BBC iPlayer (our live and video on demand service) to BBC Sounds (our audio service), News to Sport – on smart-speakers, our website, apps and TVs - and free of commercial interests.

It will allow us to have the creative freedom to experiment and try out new ways of connecting with people through conversation. We can move quickly rather than needing someone else’s permission

to build in a certain way. It'll allow us to be much more ambitious and more experimental in the content that teams can create and the ways in which we help people to find the best bits of the BBC for them.

11. Is it intended to be direct competition from Alexa or Google Assistant, for example?

No. We will continue to distribute content via other voice assistants and smart speakers. The BBC assistant is being built to work on any voice assisted device (smart-speaker, mobile, TV or anything else) so ultimately we'd like to make it as widely available as possible. However, we would need the device manufacturers to let people have access to it.

We're looking at what services are in the public service interest for us to deliver – there are plenty of services which we don't need to provide, but our focus is on content and discovery. We've delivered PS values in radio, TV, early online era and later, how do we safeguard PS values in the assistant enabled environment.

12. One of the features of "Beeb" is that you will be able to understand the different accents, which is not the case with Alexa, for example. Why do you think it's necessary? How are you or have you worked on it?

The BBC's has an obligation to reflect, represent and serve the diverse communities of the all of the UK's nations and regions. BBC license fee payers come from all across the UK and must be able to access the services that they pay for. This means ensuring that people throughout the UK are able to use the voice assistant and feel that it serves them. We have worked on this by conducting audio collection from across the UK. We have even asked our staff to record audio to help train our assistant.

13. In your view, what phase are interactive voice platforms in relation to understanding human language?

For adults it is reasonably advanced (in some languages). For children there are still many problems and for children assistants cannot record data sets so the learnings are difficult to come across. Depending on the language and type of assistant there are different levels of maturity, but is a lot more work to be done in relation to children. Google, Apple and Microsoft are constantly leapfrogging each other in terms of closeness to human parity.

14. Another of the technologies that will incorporate "Beeb", named Songbird, will allow users of BBC.com access thousands of articles in audio format. Why? What's the point?

Accessibility – everyone should be able to have access to our written content whether they can read or not, this includes people with different levels of literacy and disabilities. There is a rapidly growing market for this – e.g. Apple is introducing this service.

7.6. Entrevista a Mónica Lablanca, directora de Cugat Mèdia

Fecha: 16/06/2020

Nombre del entrevistado: Mónica Lablanca

Empresa: Cugat Mèdia

OBJETIVO DE LA ENTREVISTA

Conocer la experiencia de Cugat Mèdia, como medio de comunicación de proximidad que ha apostado recientemente por los asistentes de voz.

INFORMACIÓN DE LOS ENTREVISTADOS

5. Nombre y apellidos: Mónica Lablanca

6. Formación y experiencia: Licenciada en Periodismo. Máster en Dirección y Administración de Empresas Audiovisuales y posgrado en Periodismo Digital.

7. Cargo en Cugat Mèdia: _directora

8. Desde cuando trabaja en proyectos con asistentes de voz: 2020

PREGUNTAS

1. ¿En qué asistentes de voz podemos encontrar Cugat Mèdia, actualmente?

Básicamente en Alexa, aunque estamos trabajando en otros como Siri.

2. ¿Porqué Cugat Mèdia decide apostar por los asistentes de voz?

Cugat Mèdia tiene como objetivo hacer accesible la información a todo su público potencial. Es con la voluntad de dar cumplimiento a este objetivo que trabaja cualquier canal, formato o soporte que se convierta en un hábito de consumo de información por parte de nuestro público. Uno de los otros objetivos es ser referente de innovación y hacer de la tecnología una aliada para conseguir llegar a nuestro público, ofreciendo formatos adaptados a sus hábitos de consumo, unos hábitos que evolucionan rápidamente, igual que la tecnología.

Apostamos por los asistentes de voz después de analizar las tendencias de usos de los medios en Estados Unidos y donde, por ejemplo, se desprende que el 50% de la población dice haberse informado a través de asistentes de voz.

3. ¿La población de Sant Cugat es usuaria de este tipo de tecnología?

De momento no tenemos datos concretos. Aún así, sí tenemos datos del Observatorio Sociológico, que cada cuatro meses se hace en la ciudad, y donde se desprende que la ciudadanía santcugatena responde a un perfil muy tecnológico en el consumo de información, uso de móvil, redes sociales...

4. ¿Cuándo dais el paso y empezáis a trabajar con asistentes de voz?

En el mes de enero, establezco que uno de los objetivos del área de tecnología sea formar parte de los asistentes de voz, ya sea como radio en directo como a la carta, y el máximo de servicios posibles que podamos dar. Es con la crisis del Covid-19 que, viendo las tendencias, decidimos ponerlo en marcha y hacer así un paso más para facilitar el acceso a la información en un momento tan esencial como la pandemia y la necesidad informativa existente.

5. En qué formato está presente Cugat Mèdia en Alexa?

Hemos desarrollado nuestra propia *skill*.

6. ¿Qué respuesta ha tenido de los usuarios, a día de hoy? ¿Disponéis de datos?

Aún no tenemos datos, pero en breve podremos analizar algunos que nos podrán ofrecer una radiografía real de su uso.

7. ¿Qué queréis conseguir gracias a este nuevo canal de distribución de la información?

Facilitar el acceso a la información y adaptarnos a los nuevos usos y hábitos de consumo. Cugat Mèdia es un medio que avanza, que responde con eficiencia a las necesidades comunicativas y adapta las nuevas tecnologías para conseguir sinergias y eficiencia del contenido informativo.

8. ¿Qué pueden aportar los asistentes de voz a un medio local y de proximidad como Cugat Mèdia?

Mucho. Los medios de comunicación de proximidad tradicionalmente han estado considerados medios de segunda. Su uso a su consulta siempre ha respondido a una necesidad informativa concreta y ha requerido de la pro actividad del oyente o usuario. En este caso, los asistentes de voz ayudan a hacer un paso adelante para democratizar el medio de comunicación, su acceso y hacer más eficiente el contenido. Cugat Mèdia ha dejado de ser un medio de comunicación convencional para ser una productora de contenidos que distribuye por diferentes canales, formatos y plataformas para dinamizar y hacer más eficiente el contenido trabajado.

9. ¿Conocéis más experiencias con asistentes de voz en medios de proximidad?

No tenemos constancia de ningún otro.

