

ALEXA, DÓNA'M LES NOTÍCIES!

Disseny d'un model pilot de servei informatiu públic per l'assistent de veu d'Amazon

RESUM: Aquest treball d'aplicació professional presenta un model de servei informatiu de TV3 i Catalunya Ràdio pensat per a ser integrat en aquells dispositius tecnològics que serveixen la informació a través d'assistents de veu - en particular, a l'assistent Alexa, d'Amazon. El treball utilitza la metodologia del disseny basat en usuari per identificar usuaris model i contextos d'ús, i a partir d'aquí estableix quins continguts s'han de servir, defineix la veu de la marca, desenvolupa un arbre de conversa i fa una previsió de les mètriques de seguiment que permetran avaluar l'èxit del servei.

PARAULES CLAU: assistents de veu, Alexa, Amazon, serveis informatius, TV3, Catalunya Ràdio, CCMA, Servei Públic, UX, UI, interfícies naturals, interfícies de veu, comunicació home-màquina, assistent virtual, periodisme conversacional.

KEYWORDS: Voice-controlled digital assistants, Alexa, Amazon, news service, TV3, Catalunya Ràdio, CCMA, public service, UX, UI, Natural User Interfaces (NUI), voice interfaces, human-machine communication, virtual assistant, conversational journalism.

Autora: Judith Argila
Directora: Sílvia Martínez
Treball Final de Màster d'Aplicació Professional
Màster en Societat de la Informació i el Coneixement
UOC 2018-2019

Índex

1. Introducció.....	5
1.1. Context.....	5
1.2. Objectius.....	5
1.3. Preguntes.....	6
1.4. Justificació.....	7
2. Marc contextual.....	8
2.1. Consum digital d'informació periodística.....	8
2.2. L'ús dels assistents de veu, una tendència a l'alça.....	9
2.3. L'ús d'assistents de veu a Espanya.....	10
2.4. Els mitjans de comunicació als assistents de veu.....	11
3. Marc teòric.....	16
3.1. Dissenyats per encaixar en l'user journey.....	16
3.2. Push vs. Pull: repensar la informació en un entorn públic.....	18
3.3. La veu com interfície d'usuari.....	19
3.4. "Perdona... no he entès la teva pregunta".....	20
3.5. La percepció de la veu com un individu.....	22
3.6. La promesa de la informació conversacional personalitzada.....	25
3.7. Reptes pendents de l'acadèmia i de la indústria.....	25
4. Definició del Projecte.....	27
4.1. Missió i valors.....	27
4.3. Audiència.....	28
4.4. Planificació.....	28
5. Disseny del Servei I: objectius, <i>persona</i> i escenaris.....	29
5.1. Objectius d'audiència i posicionament de marca.....	29
5.2. Definició d'usuaris: buscant el perfil real de la <i>persona</i>	29
5.4. Definició de contextos d'ús: ubicant la <i>persona</i> al seu escenari.....	34
5.6. Política de privacitat i tractament de dades.....	36
5.7. Opcions de rendibilització comercial.....	37
6. Disseny del servei II: continguts, veu i interacció.....	40
6.1. Identificació dels continguts i del tipus de servei.....	40
6.2. Estructura i tipus de continguts.....	42
6.2. Definició de la veu de la marca.....	44
6.2. Model d'arbre de conversa.....	46
6.6. Tests qualitius, mètriques quantitatives i avaluació.....	47
7. Conclusions.....	50

8. Bibliografia.....	52
9. Annexos.....	55

*“We are lying on the moon
it’s a perfect afternoon
Your shadow follows me all day
making sure that I’m okay
- and we are million miles away”*

Karen Lee Orzolek: «The moon song», a *Her*, 2013

1. Introducció

1.1. Context

En el darrer any, les audiències digitals de la Corporació Catalana de Mitjans Audiovisuals han vist una caiguda moderada del tràfic provinent de xarxes socials – en concret, del tràfic provinent de Facebook. Per primer cop en els darrers cinc anys, el consum de notícies ha començat a perdre usuaris provinents d'aquesta plataforma.

Segons el que mostren els estudis, aquesta situació no és un cas aïllat: com veurem més endavant, des de l'aplicació dels canvis en l'algoritme de Facebook, la visibilitat de les notícies en aquests entorns ha caigut de manera general, privant als mitjans d'una important font de tràfic. Davant aquest escenari, com a ens de servei públic que té el mandat de fer arribar la informació a la ciutadania és necessari fer una reflexió sobre cap a on s'està movent el consum digital d'informació per tal d'adaptar les nostres estratègies de distribució. D'una banda, hi ha la necessitat d'arribar a l'usuari allà on es trobi; d'altra, és necessari que diversifiquem les nostres fonts de tràfic per no tenir una dependència alta de cap plataforma, especialment d'aquelles controlades per altres empreses tecnològiques, i evitar així ser vulnerables a una modificació sobtada i brusca del consum.

Amb aquesta voluntat i d'aquest escenari neix aquest treball de final de màster d'aplicació professional. La irrupció dels assistents de veu, amb el seu ràpid creixement, representa en primer lloc una oportunitat per explorar la diversificació de fonts de distribució, *per se*. En segon lloc, ens permetria oferir continguts en un entorn on l'usuari pot estar en relació directa amb la nostra marca, obtenint els continguts en el moment que ell ho desitja, sense dependre de la mediació d'un algoritme de tercers que pot condicionar la nostra visibilitat. I en tercer lloc, que el format del contingut en aquests entorns estigui basat en àudio representa una avantatge per un mitjà com la CCMA, que ja té una gran experiència de la seva producció informativa en aquest format. Poder donar una nova sortida a un nou entorn als continguts ja produïts a la Corporació és una forma de fer-la, en tots els sentits, més eficient.

1.2. Objectius

El principal objectiu d'aquest treball és el **disseny d'un model de servei informatiu públic adaptat a les noves interfícies de veu**, per tal que els continguts informatius de TV3 i Catalunya

Ràdio puguin **arribar a la ciutadania i aportar-hi valor** en aquests nous entorns de consum informatiu.

La voluntat és dissenyar un model pilot que pugui ser posat a disposició de l'audiència amb l'inici de la temporada 2019-2020, i mantenir-lo sota observació constant i en fase d'ajust durant almenys els primers dos mesos.

1.3. Preguntes

Per tal de garantir un bon disseny d'aquest servei, caldrà prèviament analitzar i donar resposta a una sèrie de preguntes contextuals:

- Quins models d'interfícies de veu hi ha disponibles al mercat actualment pels parlants de llengua castellana i catalana?
- Quins són els dispositius on s'integren les interfícies de veu actualment a Catalunya i quines són les seves especificitats tècniques? (si tenen o no pantalles, si són de sobretaula o integrats a electrodomèstics, vehicles..., si tenen o no connexió pròpia a la xarxa, etc.).
- Quins serveis informatius de parla castellana o catalana hi ha disponibles actualment a través d'assistents de veu a Espanya?

Un cop aclarits aquests elements inicials, s'hauran d'abordar les preguntes que permetran definir el model de servei des d'una triple vessant d'experiència d'usuari, gestió de recursos i avaluació:

- Quins han de ser els objectius d'audiència i posicionament de marca als que vol donar resposta el servei?
- Quin és el perfil o perfils d'usuaris potencials d'aquest servei ("persones")?
- Quins són els contextos d'ús potencials d'aquest servei ("escenaris")?
- Quin format han de tenir els continguts informatius per donar resposta a les necessitats i contextos de consum dels usuaris potencials?
- Quins continguts específics de TV3 o Catalunya Ràdio encaixen en aquest entorn? És necessària una producció o post-producció específica?
- Quines necessitats jurídiques relacionades amb protecció i explotació de dades hem de tenir presents?
- Quines necessitats o possibilitats comercials hi ha lligades a aquest servei?
- Com són els diagrames complets d'experiència d'usuari en aquest servei?
- Quins tests podem implementar per posar a prova la validesa del servei?

- Quines mètriques ens permetran realitzar un seguiment i avaluació de la consecució dels objectius definits pel servei? Amb quina periodicitat s'han d'analitzar?

1.4. Justificació

El disseny d'experiències d'usuari a través d'interfícies de veu és un camp molt recent, encara inexplorat pels equips de disseny i desenvolupament de l'àrea de Producte dels Mitjans Digitals de la Corporació Catalana de Mitjans Audiovisuals - a la qual pertanyo. La seva especificitat i radical diferenciació de les interfícies visuals en les que s'ha basat el disseny d'experiència d'usuari en les darreres dues dècades fan necessària una immersió conceptual prèvia abans de poder abordar el disseny d'un projecte informatiu per aquest tipus d'entorn, que garanteixi la presència de la CCMA en aquest nou entorn on la ciutadania pot consumir informació.

Aquest treball pretén donar resposta a ambdues necessitats: en primer lloc, realitzant una exploració teòrica i conceptual d'aquest nou llenguatge d'interacció home-màquina basat en veu, i en segon lloc, fent una proposta ferma i concreta per l'aplicació d'aquest llenguatge en els serveis informatius de TV3 i Catalunya Ràdio. Així, podríem dir que és un treball orientat a servir, en primer lloc, a la CCMA - però sobretot, i a través d'ella, a la ciutadania de Catalunya i el seu dret d'accés a la informació.

2. Marc contextual

2.1. Consum digital d'informació periodística

Abans d'abordar la definició d'un nou projecte periodístic que doni servei a la ciutadania catalana hem d'entendre quina és la realitat del consum digital d'informació al 2018 – no només en termes de continguts sinó, sobretot també, en termes de dispositius, plataformes i hàbits de consum. Per fer-ho, utilitzem el major estudi de tendències de consum que existeix en aquest àmbit: el *Digital News Report*, realitzat pel Reuters Institute en base a més de 74.000 usuaris a través de 37 països – incloent 2.023 usuaris d'Espanya (Newman, Fletcher, Kalogeropoulos, Levy, & Kleis Nielsen, 2018).

La primera gran conclusió de l'estudi és que estem vivint un canvi de cicle respecte els darrers 7 anys, on la distribució de notícies a través de xarxes socials i agregadors s'havia convertit en una gran tendència de la indústria, guanyan terreny any a any. Al 2018, per primer cop, aquesta tendència «s'ha aturat – o fins i tot està començant a remetre, amb el model de subscripció augmentant a un gran nombre de països». La principal causa és el canvi d'algoritme de Facebook, que ha disminuït la visibilitat de les notícies als *feeds* dels usuaris, però l'informe apunta que pot haver-hi altres factors que influencien aquest canvi, com la desconfiança generada per les *fake news* o per la falta de control sobre el propi consum que imposen els algoritmes d'aquestes plataformes.

Al mateix temps, estem experimentant un augment del consum d'actualitat a través d'aplicacions de missatgeria com WhatsApp, on s'ha triplicat el consum de notícies des de 2014 (superant a Twitter a molts països). En aquests entorns els usuaris troben «espais més privats (i menys susceptibles a generar confrontació)» on comunicar-se i compartir notícies. Per la seva banda, la televisió es manté com una gran font d'informació per molta gent, però manté un lent decliu d'audiència, any a any, generant dubtes sobre quines estratègies haurà de desenvolupar per atraure audiències més joves.

Finalment, l'estudi també destaca, entre d'altres, el ràpid creixement dels assistents de veu. «Els assistents digitals activats per veu com Amazon Echo i Google home segueixen creixent ràpidament, obrint noves oportunitats per les notícies en format àudio. L'ús s'ha més que duplicat a Estats Units, Alemanya i el Regne Unit, amb la meitat dels usuaris d'aquests serveis utilitzant-los pel consum de notícies i informació» (Newman, Fletcher, Kalogeropoulos, Levy, & Kleis Nielsen, 2018).

2.2. L'ús dels assistents de veu, una tendència a l'alça

Per a calibrar la necessitat d'un projecte periodístic basat en assistents de veu, fem abans un repàs a les estadístiques de consum que aquests nous interlocutors tecnològics representen.

Segons l'estudi anual de tendències digitals publicat per Mary Meeker's, una referència en la indústria tecnològica, l'any 2016 el 20% de les cerques a Google fetes amb un telèfon ja estaven basades en veu, enlloc de text (Yurieff, 2017). Si aquest percentatge ja era elevat, l'any 2017 Comscore parlava de que el 50% dels usuaris que fan servir telèfons mòbils interactuaven amb ells a través de la veu d'alguna manera i preveïen incrementar aquest ús en el futur; el 30%, de fet, ho feien a diari (Engleson, 2017). En altres paraules: un de cada tres usuaris està emprant la veu per obtenir informació dels seus telèfons, i la perspectiva pels propers anys és a l'alça.

Al mateix temps, entre el 2017 i el 2018 hem vist un alça en la venda d'altaveus dedicats a assistents de veu d'un 197%. Amazon i Google copen, a dia d'avui, el top del mercat d'aquestes vendes, segons un anàlisi de les vendes realitzades per Strategy Analytics al 3r quadrimestre del 2018; entre tots dos tenen un 55% de quota de mercat, seguits pels assistents xinesos, Baidu, Alibaba i Xiaomi, que lluiten pel lideratge al mercat xinès (Keynes, 2018).

Els assistents de veu, de fet, no només ens permeten fer cerques d'informació, sinó que, segons la seva sofisticació i el dispositiu en què estiguin incorporats, podran ajudar-nos a enviar missatges, realitzar trucades, programar alarmes o controlar les llums de casa, entre d'altres. Segons un estudi recent del Pew Research Center (Pew Research Center, 2017), de fet, el 47% dels adults a EEUU afirma utilitzar ja la veu per interactuar amb algun dispositiu electrònic. Si bé el més habitual segueix sent invocar-los a través del telèfon mòbil (42%), seguit de lluny per l'ordinador o la tablet (14%), comença a apuntar el seu ús directament en altaveus intel·ligents (8%), com la família d'Amazon Echo o Google Home.

En el futur proper, és previsible que l'ús d'assistents de veu es diversifiqui en famílies molt més extenses de dispositius. En motiu de la darrera International CES Tech Conference, Amazon va anunciar que el seu assistent Alexa s'integrarà properament en televisors Hisense, en miralls de bany Kohler, ordinadors HP, Asus i Acer, televisors Panasonic i Garmin i cotxes Toyota o Lexus, mentre que Google va anunciar integracions amb televisors LG, auriculars de Sony i pantalles intel·ligents de Lenovo (Wakabayashi & Wingfield, 2018). Els assistents aniran impregnant la porositat de la nostra vida digital abans que ens n'adonem.

Si baixem a desgranar per sectors demogràfics aquesta tendència a l'alça en l'ús de la veu, trobem que, significativament, la franja d'edat més propera a l'ús dels assistents és la dels **millennials** - aquells nascuts entre 1981 i 1997. Alguns estudis han identificat aquest col·lectiu com el que actualment utilitza més la interacció amb la veu per parlar amb els seus dispositius: en concret, un 46% se'n declaren usuaris, en comparació al 14% de la franja entre 46 i 60 anys (Jordan, 2017). Que els millennials abracin la nova tecnologia és rellevant perquè en estudis digitals, el comportament d'aquest col·lectiu s'ha considerat sovint un indicador de tendències que a mig termini acaben popularitzant-se en altres capes demogràfiques.

L'estudi de Pew Research aboca llum també sobre quins són els motius que fan que els usuaris en general optin per utilitzar els assistents de veu: el 55% diu que, majoritàriament, perquè els hi permet "interactuar amb els dispositius sense fer servir les mans". Aquesta és una línia d'argumentació molt important que retrobarem més endavant, quan ens endinsem en el marc teòric del disseny d'assistents de veu. Hi ha altres raons esgrimides pels usuaris que ahonden en aquesta idea: el 22% diuen que els fan servir perquè parlar "és més natural que escriure", o perquè aquests assistents són "més fàcils de fer servir per part dels nens" - un 14% . La segona gran línia d'argumentació és més disruptiva: el 23% dels enquestats confessa que els fa servir *principalment* "perquè són divertits" (Pew Research Center, 2017).

2.3. L'ús d'assistents de veu a Espanya

El primer factor a considerar quan analitzem l'adopció dels assistents de veu en el cas espanyol és l'idioma. El desenvolupament dels assistents de veu en els seus inicis va estar fortament vinculat a l'anglès, cosa que va potenciar la seva popularitat en països de parla anglesa o que tenen l'anglès com a llengua vehicular. Google va estrenar la seva opció de cerca per veu en castellà al març de 2009, quasi fa una dècada, i Siri va arribar a la península parlant castellà a l'octubre de 2011. Google Assistant i Google Allo, que trascendien la cerca per veu per apropar-se més al concepte d'un assistent, van estrenar-se a Espanya al maig de 2017 - i no ha sigut fins l'octubre de 2018 que Alexa, l'assistent d'Amazon, ha començat a entendre i parlar el castellà.

Segons l'estudi "The Future is Voice Activated", realitzat per iProspect a Espanya amb una base de més de 1.000 entrevistats, més del 55% dels usuaris declara haver fet servir algun cop assistents de veu en algun dispositiu, i aquest percentatge s'enfilava per sobre del 66% a la franja dels millennials (Reason Why, 2018).

Si analitzem des de quins dispositius s'estan invocant aquests assistents de veu, el focus és encara a **telèfons mòbils** - especialment pel que fa a Google Assistant i Siri. És remarcable que aquest ús d'assistents de veu a mòbil és sis punts superior a Espanya que a la resta de la mitjana europea: un 53% dels usuaris de mòbils asseguren haver utilitzat l'assistent de veu almenys un cop al mes, i d'aquests, el 30% assegura fer-lo servir com a mínim un cop per setmana (*Speak Easy: Global Edition - JWT Intelligence, 2017*), cosa que es podria considerar ja un ús habitual. Dit d'una altra manera: des de la perspectiva d'un mitjà de comunicació, una de cada tres persones de la nostra audiència potencial està fent servir setmanalment l'assistent de veu al seu telèfon.

Per a què el fan servir? Declaren que bàsicament el fan servir per a fer cerques d'informació i executar tasques funcionals, però caldria veure fins a quin punt això és rellevant quan, cal dir-ho, aquestes són les principals opcions que han ofert els assistents de veu a Espanya en els darrers anys. Més interessants són els motius que argumenten com a raons per preferir l'assistent a la cerca habitual: estalvi de temps, rapidesa i comoditat. Les principals queixes? La falta de comprensió o mala interpretació d'algunes ordres.

En el cas de Catalunya, no hem localitzat estadístiques específiques sobre la penetració o l'ús dels assistents de veu. Si paga la pena mencionar, però, que no hi ha cap assistent de veu que a dia d'avui entengui o parli el català. Per tant, la comunitat catalana ha d'adreçar-se a aquests assistents emprant el castellà, l'anglès, o algun dels altres idiomes que estan oferint (en el cas d'Alexa, p.ex.: alemany, francès, italià o japonès). L'idioma de conversa no determina, però, l'idioma dels continguts als que pot accedir l'usuari. En el cas d'Espanya, p.ex., Alexa està oferint continguts en idiomes regionals, com el vasc o el català.

2.4. Els mitjans de comunicació als assistents de veu

Trushar Barot, Knight Visiting Nieman Fellow a Harvard i Digital Launch Editor a la BBC, descrivia al 2017 la intel·ligència artificial basada en veu com una "gegantina plataforma en flames sobre la qual la indústria informativa ni tan sols sap que s'hi està aixecant" (Barot, 2017). Barot ressalta les enormes diferències de mentalitat amb que la indústria tecnològica (especialment Google, Microsoft, Amazon i Facebook) i la indústria informativa s'estan apropant a aquest fenomen: pels primers, és un factor clau i focus absolut de desenvolupament; pels segons, és observada des de la distància - en el millor dels casos.

La realitat és que quan se'ls hi pregunta als consumidors per quina raó volen adquirir un altaveu intel·ligent una de les principals respostes apela directament als mitjans de comunicació: el 77%

afirma que vol escoltar notícies o informació a través del seu assistent de veu (The Smart Audio Report, 2018).

En paraules d'Amy Webb, professora de predicció estratègia a la New York University's Stern School of Business i reconeguda analista tecnològica, "parlar a les màquines, més que escriure en elles, no és només un truc temporal per cridar l'atenció. Humans parlant a màquines (...) representa el proper moviment més important en el nostre ecosistema informatiu. La veu és la propera gran amenaça pel periodisme" (Webb, 2017).

Amenaça, repte o plataforma en flames, els assistents de veu ofereixen als mitjans de comunicació la possibilitat d'arribar a les seves audiències en nous formats narratius - i també, per què no, en vells formats, que es redescobreixen ara transportats a noves situacions de consum informatiu: el butlletí de les notícies del matí, demanat amb un cop de veu des de la dutxa, o els ingredients de la recepta del programa de cuina d'ahir, demanats al assistent mentre tallem les cebes curosament sobre el marbre de la cuina, ens eviten manipular dispositius tàctils mentre estem en situacions on les nostres mans estan mullades, brutes o senzillament, ocupades.

Des de l'expansió dels assistents de veu integrats a altaveus intel·ligents diversos mitjans de comunicació han fet un pas ràpid per a oferir contingut en aquests entorns. Segons l'entorn, aquestes opcions d'integració varien - però comparteixen una idea comuna: l'assistent parteix d'unes capacitats natives de base, que es poden ampliar a base d'afegir *skills*, en llenguatge de l'Alexa d'Amazon, o *actions*, en el cas de Google Assistant. En tots dos casos es tracta de funcionalitats creades per un tercer que amplien les capacitats natives de l'assistent.

En aquest ecosistema de "funcionalitats" afegides, Amazon ofereix als mitjans dues opcions d'integració:

- **Skills ad hoc**, desenvolupades totalment pel mitjà de comunicació, que controla la paraula d'invocació, el tipus d'interaccions que farà l'usuari amb l'assistent, les respostes que s'oferiran, i la funcionalitat que executarà l'*skill* (p.ex.: reproduir una emissora de ràdio, llegir titulars, donar una previsió del temps, avisar d'una última hora...). Per la seva versatilitat en els diàlegs i la seva potencialitat en la funcionalitat, aquest serà el model d'*skill* que utilitzarem per l'elaboració del nostre servei informatiu.
- **Resums informatius**. Amazon s'encarrega de la definició de la interacció amb l'usuari, definint el tipus de conversa que es pot establir, i estableix alhora la única funcionalitat d'aquest tipus d'*skill*, que en aquest cas és oferir a l'usuari un breu resum de notícies

destacades, que necessàriament ha de ser inferior a 7 minuts. És, segurament, el tipus d'integració més senzill: n'hi prou amb la integració d'un *feed* de continguts, sempre i quan aquest *feed* ofereixi continguts informatius de menys de 7 minuts de duració, i el *feed* estigui escrit en el format i el tipus d'etiquetatge que pot descodificar Amazon. Un cop creat aquest *feed*, el mitjà haurà de triar un nom, una imatge, una descripció i unes paraules claus per a la seva *skill*. Alexa oferirà als usuaris els resums informatius com a resposta a la pregunta “*Alexa, ¿cuáles son hoy las noticias?*”

Aquests resums informatius, a la pràctica, funcionen més com a font d'informació que com a funcionalitats en sí mateixes: són les fonts que nodreixen la funcionalitat “resum de notícies” nativa d'Alexa. Això fa que l'usuari no pugui invocar una d'aquestes fonts pel seu nom propi, sinó que els seus continguts necessàriament són executats per Alexa, un darrera l'altra. Per aquells mitjans que ofereixen contingut en vídeo i/o àudio, Alexa executa l'*stream* pertinent. Pels que ofereixen notícies textuais, Alexa llegeix en veu alta els articles amb la seva pròpia veu.

Aquest tipus d'integració, senzilla i ràpida, ha permès una inversió controlada per endinsar-se en els assistents de veu per a molts mitjans. Alexander Bregman, Strategic Partner Development Manager de Google per EMEA, declarava a una entrevista a Journalism.co.uk que Google volia fer servir els flashos informatius matinals com a esquer per atrapar les audiències conforme avançava el dia (Scott, 2018). La idea és que l'usuari identifiqui algun tema del seu interès en el resum del matí - i segueixi demanant informació al respecte conforme avança el dia. «Si un usuari ens demana que el mantinguem informat sobre un tema d'actualitat, li podem oferir més notícies conforme la història es desenvolupa a través de diferents dispositius». D'aquesta manera, els resums informatius poden fer de “passarel·la” cap al consum d'altres continguts d'un mateix mitjà - sigui a través del propi assistent de veu o en altres plataformes.

En el mercat de parla anglesa, que és el més consolidat en qüestió de reconeixement de veu i altaveus intel·ligents, a principis de 2019 Amazon compta ja amb més de 4.000 *skills* categoritzades sota la secció de “Notícies”, malgrat la immensa majoria no corresponen a mitjans de comunicació. Entre els mitjans, trobem moltes emissores de ràdio i *broadcasters* oferint els seus resums informatius, sumaris de butlletins o d'informatius televisius transformats en una sindicació automàtica que es serveix com a font a Alexa (com p.ex, Fox News Channel, la CNN, NBC, NPR, NRK, ESPN...). També hi ha agències de notícies i diaris de tirada nacional i més local que han creat les seves pròpies *skills* (com p.ex., Associated Press, USA Today o Los Angeles Times).

Pel que fa als formats, destaquen els casos de The New York Times, el The Washington Post, la CNN i la BBC. El Times ha estrenat a principis de 2019 un informatiu diari, el “The New York Times Briefing” pensat per oferir un resum breu de 3 minuts sobre l’actualitat diària, i un qüestionari setmanal, d’una durada d’uns 25 minuts, basat en l’actualitat (Hazard, 2018). La *skill* del Washington Post també ofereix els últims titulars, un test d’actualitat, anàlisis dels seus corresponents – i el més novedós: la possibilitat d’activar alertes per rebre les notícies d’última hora al llarg del dia. La CNN també ofereix una *skill* que permet sentir passivament els titulars, o bé demanar informació sobre un tema específic d’actualitat, com p.ex., “les últimes notícies sobre el President Trump”.

A Europa, destaquen projectes com els de la BBC, que ha publicat la seva primera experiència de veu per a infants, anomenada “The BBC Kids skill” (McDonald, 2018), també per Alexa. La *skill* proposa tres jocs diferents als nens, basats en els personatges de la marca infantil de la BBC, CBeebies. Els jocs presenten diferents tipus d’interacció:

- A “Little Monster’s Hide and Seek”, els nens han d’identificar diferents sons per descobrir on s’estan amagant el monstre.
- A “Andy’s Adventure Dance Party” els nens poden cantar i ballar cançons de la sèrie
- A “Go Jetter’s Glitchy Facts” es presenten una sèrie de preguntes en format trivial.

Aquest tipus d’interaccions transcendeixen la distribució de programes de ràdio en directe o a la carta a través dels altaveus intel·ligents, i permeten crear un tipus de vinculació més activa i variada amb els continguts de les marques.

En el marcat espanyol, molt més recent (Octubre 2018), a la categoria de Notícies es poden trobar moltes *skills* d’emissores de ràdio (Cope, La Ser, Onda Cero), agències de notícies (Europa Press), diaris esportius (Sport, As), televisions (Mediaset, TVE, EITB). En qüestió de formats, destaquen els casos de El País i d’ABC. El País, perquè ha volgut transcendir la mera lectura de titulars per oferir a l’usuari unes mínimes opcions de navegació que li permeten triar què vol sentir i aprofundir en una notícia en particular. I en el cas de l’ABC, perquè el diari ha volgut defugir la locució automàtica i robòtica de textos escrits, el conegut com *text-to-speech*, la solució tecnològicament més senzilla pels mitjans escrits, i ha optat per produir un butlletí ad hoc per assistents de veu, «Las noticias de ABC», que dura 2-3 minuts (Vocento Media Lab, 2018).

En resum, podríem dir que els mitjans estan adoptant els següents tipus de continguts:

- Sumaris tallats dels telenotícies de TV o dels butlletins de ràdio
- Lectures automatitzades de notícies escrites

- Seccions d'opinió i anàlisi d'especialistes
- Qüestionaris i jocs sobre continguts informatius
- Alertes d'última hora
- Informació sobre temes específics sota demanda

3. Marc teòric

L'estudi de la comunicació entre homes i màquines (*human machine communication*, HMC) és una línia d'investigació recent desenvolupada a partir dels estudis sobre la interacció home-màquina (HCI), i que té com a principal objectiu entendre la tecnologia com una entitat comunicadora per sí mateixa - i no només com una medidora de la comunicació. La HMC beu de disciplines com la sociologia, l'antropologia, la psicologia aplicada o el disseny industrial, i autors com Licklider, Clark o Hansen han contribuït a la creació del seu corpus de coneixement.

Per abordar el disseny d'aquest servei de notícies a través d'assistents de veu, ens centrarem sobretot, per tant en les actuals investigacions sobre la comunicació home-màquina (HMC, a les seves sigles en anglès), però també posarem molt focus en l'estat actual de les tendències de disseny d'experiències basat en usuari: l'*user experience design* (UX). Aquest terme va ser encunyat pel psicòleg cognitiu i dissenyador Donald Norman, que treballava a Apple a la dècada dels 90, per fer referència a la totalitat d'aspectes de la interacció entre l'usuari i un determinat producte o servei. En paraules de Norman: "Vaig inventar-me el terme perquè vaig pensar que 'interfície humana' i 'usabilitat' eren massa restrictius: jo volia cobrir tots els aspectes de l'experiència de la persona amb el sistema, incloent el disseny industrial, els gràfics, la interfície, la interacció física i el manual".

La combinació d'aquests dos entorns d'investigació ens permetrà definir adequadament tant els aspectes més funcional com els aspectes més conceptuals relacionats amb el nostre servei informatiu per assistents de veu.

3.1. Dissenyats per encaixar en l'*user journey*

El disseny de productes i serveis digitals ha begut en la darrera dècada del que es va anomenar "disseny centrat en l'usuari", una aproximació que proposava abordar els projectes pensant en primer lloc qui és la persona a la que volem donar servei.

La premisa base és que hem de conèixer molt bé quins són els seus principals trets característics (Quina edat té? A què es dedica? Quina és la seva alfabetització digital?), les seves motivacions per apropar-se al nostre producte o servei (Ens busca per entretenir-se? Vol portar a terme una tasca? Què necessita de nosaltres?), i les seves circumstàncies en el moment en que tindrà relació amb el nostre servei (serà a primera hora del matí, al cotxe, mentre porta els nens a

l'escola? Serà al vespre, al menjador de casa, a través de la televisió, amb tota la família al voltant asseguda?). D'aquest perfilat inicial s'acaben definint dos tipus de documents que ens guiaran en la posterior definició del servei:

- **Personas.** Del llatí, “personatge”. Són els individus arquetípics, amb els seus trets, necessitats i circumstàncies, que consumiran el servei. En podem tenir un o diversos, en funció de l'amplitud del nostre públic objectiu. La metodologia proposa que aquesta construcció de personatges ficticis sigui tan realista com sigui possible: cal assignar-los noms, edats, estils de vida, de roba, marques afins, hobbies... Tot allò que ens ajudi a apropar-nos a una personalitat real.
- **User journey.** També traduït com “el camí de l'usuari”, o “el recorregut de l'usuari”, pretén ser una recreació de les activitats i circumstàncies diàries d'aquests usuaris arquetípics als que ens adrecem. Amb la descripció acurada del seu dia a dia, podem identificar amb més claredat quins seran els moments en que aquest personatge es posarà en contacte amb el nostre servei: aquests moments són coneguts com a «touch points». L'anàlisi d'aquests moments, amb els seus condicionants (És un entorn públic o privat? Amb quin dispositiu es porta a terme el contacte? De quin marge de temps disposa l'usuari?), acaba de completar la informació que necessitem per a una bona definició del servei.

En el cas dels assistents de veu, els primers mitjans de comunicació que estan dissenyant aplicacions per aquests entorns comencen a pensar en aquests termes per definir els seus serveis informatius. Elizabeth Johnson, senior editor de CNN Digital i líder del desenvolupament d'aplicacions informatives per aquestes plataformes, parla d'aquest tipus d'aproximació al disseny en un article publicat pel Nieman Journalism Lab (Barot, 2017); és el que a la CNN han convingut en anomenar internament “The Bagel Time Philosophy”:

“Hem assumit que alguns usuaris tindran l'aparell [l'altaveu intel·ligent, en aquest cas] a la cuina (...). Això ens porta a preguntar-nos: què és el que estaran fent probablement els usuaris a la cuina pel matí? L'esmorzar. Quant temps triga en preparar-se un bagel? Cinc minuts. Així que aquest és probablement el temps que l'usuari tindrà per escoltar-nos, així que ens hem d'assegurar que el podem posar al dia de les notícies en menys de cinc minuts”.

Així doncs, els desenvolupadors s'enfoquen en dissenyar aplicacions i serveis pensats per a ser consumits en un moment determinat del dia a dia (p.ex.: resums informatius a primera hora del matí, contes infantils per anar a dormir al vespre, consulta del trànsit des del cotxe mentre tornem

a casa de la feina...) o bé en un moment determinat de la nostra vida on tenim una necessitat temporal específica (p.ex.: aprendre un idioma, fer una dieta, passar la ITV...).

Segons dades aportades per Google (Kleinberg, 2018), el 72% de les persones que han adquirit un altaveu intel·ligent afirmen que l'utilitzen com una part més de les seves rutines diàries. Alhora, el 52% el tenen al menjador, 25% a l'habitació, i 22% a la cuina. Les enquestes realitzades per HigherVisibility (Heltzman, 2017) revelen que la majoria dels usuaris actuals d'assistents de veu (que recordem, no només estan integrats als altaveus intel·ligents, sinó també a telèfons mòbils i altres dispositius) els fan servir especialment en aquells moments on *no poden o no els hi és convenient* tocar la pantalla d'un dispositiu; p.ex., mentre condueixen (52%), veuen la televisió (7.5%), cuinen (5.5%), fan exercici (2.3%) o realitzen altres tasques (21.3%). Aquestes dades també ens ajudaran a definir de forma correcta l'usuari journey, i entendre per tant en quin entorn i situacions l'usuari podrà fer ús del nostre servei.

3.2. Push vs. Pull: repensar la informació en un entorn públic

Un factor especialment a tenir present des de la perspectiva de disseny centrat en l'usuari és el fet que molt d'aquests contextos d'ús de l'usuari tenen lloc en entorns públics - o millor dit, en entorns oberts on la resposta de l'assistent pot ser escoltada *per altres usuaris*, de manera voluntària o involuntària. A diferència dels ordinadors personals, els telèfons mòbils o les tablets, que generalment ofereixen una experiència de consum d'informació individual, l'experiència de veu dels assistents torna a ser comunal, pública, de grup - com ho havien estat les experiències de consum radiofònic o televisiu.

En aquest sentit, els assistents de veu ofereixen una oportunitat per a pensar en serveis que portin un retorn al sentit del entreteniment col·lectiu en família, per exemple, amb iniciatives que involucrin a diversos membres del nucli familiar o grups d'amics (com per exemple, jocs d'enginy, acudits, relats infantils...).

Aquest escenari porta a una especial reflexió sobre les tecnologies *push* i *pull*, enteses com a dos mètodes de fer arribar informació a l'usuari - on el que canvia és, bàsicament, l'actor que inicia la transacció informativa: en els serveis informatius *push*, és el mitjà el que inicia la comunicació cap a l'usuari, sent pro-actiu en l'enviament d'informació, que interromp a l'usuari sigui quin sigui el seu context. En els serveis de tipus *pull*, en canvi, és l'usuari el que pren la iniciativa d'anar a demanar al mitjà una informació.

En aquells serveis que requereixen d'un *pull* de l'usuari per a ser activats ("Alexa, dóna'm les notícies"), és l'usuari qui prèviament a la petició d'informació està avaluant quin és el seu entorn i si és un bon moment per a rebre aquella informació. Però què passa amb les tecnologies *push*, tan emprades en la indústria informativa per alertar als usuaris d'informació rellevant d'última hora? Què passa si forcem l'enviament d'una alerta informativa de veu a l'usuari en un moment on ell es troba en una conversa delicada - o en un moment on hi ha nens presents, que podrien no necessitar rebre aquella informació...?

El mitjà nord-americà Washington Post, que, com Amazon, pertany a l'imperi de Jeff Bezos, està estudiant maneres de fer aquestes entregues *push* informatives de manera que l'usuari mantingui una part de control sobre el moment exacte en que escoltarà l'àudio. *"Per exemple, si hi ha una notícia d'última hora, el Post farà que l'Echo de l'usuari [l'altaveu intel·ligent d'Amazon] dringui i emeti una llum verda. En aquest moment l'usuari pot preguntar: "Alexa, què m'he perdut?", o "Alexa, quines són les meves notificacions?". Els usuaris hauran de triar explícitament accedir a les alertes del seu dispositiu - o bé triar desactivar les alertes de manera temporal amb un mode de no-molestar"* (Bishop, 2017).

3.3. La veu com interfície d'usuari

Conceptualment, però, el gran salt en el disseny de serveis d'informació per assistents de veu no té a veure tant amb la metodologia aplicada o les circumstàncies de consum de l'usuari (que en aquest sentit, no són excepcionals respecte a altres models de servei ja existents). El gran salt té a veure amb el tipus *d'interfície* amb què es produeix la interacció home-màquina: els assistents de veu representen un canvi de paradigma, que es desplaça de la tradicional interfície visual - a la interfície auditiva.

En termes acadèmics, les *Natural User Interfaces (NUI)* es defineixen com aquelles interfícies que resulten més naturals per a l'ésser humà: aquelles que podem operar d'una manera més lògica, sense coneixement previ de la tecnologia, aplicant raonaments i comandaments naturals en la nostra comunicació humana. Són aquelles que generen menys *fricció* comunicativa, menys necessitat de traducció. Els assistents de veu actuals fan ús de NUIs basades, en aquest cas, en veu.

Aquest tipus d'interfícies són conegudes com les *Voice User Interface (VUI)*, i han sigut possibles gràcies a les millores recents en el reconeixement de veu i, molt especialment, en el processament del llenguatge natural. Shawn DuBravac, de CTA, en una presentació prèvia al

certamen CES de 2017, va afirmar que el 2017 representava un punt d'inflexió en el reconeixement de veu, ja que les màquines havien arribat a assolir la pràctica paritat amb els humans, amb una capacitat per transcriure adequadament la veu parlada en el 94 dels casos. "Ens embalem cap a una nova era de computació sense rostre" (Speak Easy: Global Edition - JWT Intelligence, 2017).

La interacció amb la informació a Internet ha viscut una era basada en el clic, generada per l'explosió de la teranyina de pàgines webs, una era basada en el *touch*, impulsada pel boom de les pantalles tàctics a tauletes i telèfons intel·ligents, i sembla que ara ens adentrem a l'era de la veu - on la pantalla passa a ser prescindible. La veu, a banda de ser més eficient perquè ens permet comunicar-nos alliberant els nostres ulls i les nostres mans, és sovint també la única opció viable per alguns casos com la conducció.

La veu també és un dels mètodes més humans i més bàsics de comunicació humana. Per molts usuaris, bé sigui per motius generacionals, educacionals o de recursos, la capacitat d'escriure i llegir és una barrera d'accés a les tecnologies operades a través d'interfícies visuals - per no parlar de la complexitat operativa i del coneixement de convencions gràfiques necessàries que requereixen algunes interfícies visuals. En el cas de les interfícies de veu, la corba d'aprenentatge de la interacció és molt ràpida: parlar a la màquina resulta molt més intuïtiu per un ampli espectre de la població.

Jason Amunwa sintetitza el pes que pot tenir aquest canvi de paradigma amb la següent frase: "Dissenyar per la veu podria ser el catalitzador que ens ajudés a recuperar l'objectiu original de la UX: tractar a les persones com a persones de nou" (Amunwa, s.d.).

3.4. "Perdona... no he entès la teva pregunta"

Tot i l'entusiasme del sector, el disseny de VUIs es troba encara en una fase molt incipient. Les interaccions no acaben de ser pura, neta i naturalment converses obertes - sino que en molts casos l'assistent no entén l'usuari i aquest ha de fer un exercici d'adaptació a la tecnologia, que és precisament el que es busca defugir amb les NUIs. I és que reconèixer la pregunta que formula una veu (i passar-la a text per poder operar la cerca, que és el que fan els assistents de veu) és més senzill que interpretar-la correctament. Les dades varien segons l'estudi: alguns apunten que percentatge de respostes correctes oscil·la segons l'assistent i la categoria en un rang que va del 12 al 98% (segons dades de Statista; (Moreno, 2017). Segons l'IQ test sobre altaveus intel·ligents realitzat anualment per Loup Ventures, Google Assistant encerta el 81%, Alexa el 64% i Siri el

52.3% (Munster & Thompson, 2018). Però la conclusió és la mateixa: el reconeixement de veu està molt evolucionat, però a la interpretació semàntica li falta maduresa.

Una de les línies de treball per minimitzar les respostes incorrectes dels assistents és apostar per la estandarització de les comandes veu. De la mateixa manera que en altres dècades es van estandaritzar certes icones visuals, arquitectures de la informació i representacions gràfiques per a identificar funcionalitats específiques, el sector podria treballar per una estandarització de les VUIs. Això, al seu torn, facilitaria la compatibilitat entre aparells i assistents de veu. Luis Esteban, CEO de iProspect España, afirma que “el repte de la indústria està en que les cerques de veu s’estandaritzin, i que cada vegada més dispositius digitals puguin connectar-se amb facilitat i amb independència de l’assistent de veu que s’utilitzi” (Marketing Directo, s.d.).

Un dels principals reptes tant de reconeixement de veu com d’interpretació per part dels assistents són, òbviament, les diferents llengües - i les seves varietats dialectals. Aquest pot ser un dels grans impediments per l’expansió i adopció dels assistents de veu arreu del món. És interessant en aquest sentit una iniciativa recent impulsada per Mozilla: el projecte “Common Voice”, que busca crear un corpus lingüístic auditiu per a cada idioma del món, que “ensenyi a les màquines com parlen les persones” (Borchert, 2018). El projecte busca superar les barreres de cost i accés que impliquen pels petits i mitjans desenvolupadors accedir als corpus lingüístics - i que per tant limiten la innovació a les grans empreses que poden costejar-s’ho. A Common Voice, qualsevol internauta pot proposar un idioma, i la comunitat de parlants respectiva s’encarregarà d’alimentar aquell corpus a base d’hores de veu cedides al projecte. Els desenvolupadors d’assistents de veu més modestos podran així accedir gratuïtament a aquest corpus, sense restriccions, sobre el que opera una llicència Creative Commons 0, i desenvolupar assistents per llengües més minoritàries que l’anglès, el castellà o el xinès. El projecte compta actualment amb 30 idiomes, incloent el català.

El reconeixement de la llengua pròpia de l’usuari i la comprensió semàntica de la pregunta no són, però, les úniques causes de fricció en la interacció actual amb els assistents de veu. Veba Mohan, dissenyadora freelance, diagnostica que el disseny de la interfície està en una fase madura pel que fa als assistents que funcionen en base al *tap to talk* (“toca i parla”): és a dir, per Siris i Google Assistents, que responen davant d’una activació expressa de l’usuari. Però això difereix de l’experiència de mans lliures, on l’assistent ha de reaccionar davant una invocació explícita (Mohan, s.d.). Mohan detecta en concret els següents aspectes a tenir en consideració alhora de dissenyar VUIs:

- La importància de la representativitat de l'aparell. Que els assistents es presentin com una personificació humanoide o amb una forma clàssicament d'aparell, com ara un altaveu, canvia les expectatives de "humanitat" que n'esperarà l'usuari. Aquí també podríem encabir el nom que els desenvolupadors han decidit per invocar a l'assistent, que en alguns casos pren forma humana (com «Alexa», d'Amazon) i en d'altres es manté en una neutralitat funcional descriptiva (com «OK Google», per Google Assistant).
- Quan la veu es converteix en la interfície, què passa amb els elements visuals que han permès identificar les marques fins ara a les interfícies digitals visuals? Els aspectes definitoris de les marques hauran de ser definits, en molts casos, a través de la veu. Com sonarà la teva marca? Com Adele? Com Snappy Dog?
- La facilitat d'accés de les VUI, que poden ser operades per a qualsevol persona que es trobi a prop de l'aparell amb la simple invocació d'un nom, planteja problemes de seguretat pel que fa a l'accés a informació confidencial, privada o potencialment perjudicial per les criatures.
- La tria d'una opció en llistes molt llargues llegides en veu alta pot ser tediosa - en comparació en l'escombrat visual ràpid que permet el text. La solució és desgranar la llista en blocs de característiques compartides, i oferir-li diferents opcions de triar a l'usuari en una conversa, de manera que cada resposta de l'usuari limiti el rang d'opcions. Cal avançar-se i recrear el procés mental de tria que seguiria l'usuari si estigués cercant lliurement en una llista tancada. "Vols caminar o conduir?" "Vols que estigui obert en diumenge?" "Que acceptin mascotes?" "La teva millor opció és X - però si vols alguna cosa més barata, tria Z".
- Cal evitar els silencis incòmodes. Quan la veu és la interfície, la importància del feedback sonor és importantíssima perquè l'usuari sàpiga si les seves instruccions han estat enteses o no. Si l'assistent necessita temps per fer una tasca, ha d'informar-ho. "Un moment, si us plau, estic reservant la taula. De mentre, voldries...?"
- Aspectes d'accessibilitat i inclusivitat. Si bé les VUI poden ser un gran aliat en la inclusivitat de col·lectius no letrats o amb dificultats lectores, deixen al marge un altre col·lectiu: el dels sord-muts. Si les interfícies de veu passen a ser la norma, haurem de treballar en noves maneres de garantir-ne la inclusivitat.

3.5. La percepció de la veu com un individu

Al 1966 es va establir el terme "Efece Eliza" per identificar la proximitat emocional que sentim els humans quan interactuem amb dispositius tecnològics que integren aspectes humanoides - i la facilitat amb què els hi assignem un caràcter orgànic que en realitat no poseeixen. En concret,

l'efecte Eliza, segons Wikipedia, defineix “la tendència a assumir, inconscientment, que els comportaments informàtics són anàlegs als comportaments humans” (Wikipedia, s.d.).

L'efecte Eliza és una de les vessants que ens apropen a entendre la complexa relació que establim els humans amb les màquines conforme aquestes més simulen comportar-se com a humans. Aquesta relació ha sigut el focus d'investigació d'algunes de les línies de recerca més interessants en els estudis de relació home-màquina.

En el cas dels assistents de veu, que no només es comuniquen amb nosaltres emprant llenguatge natural i establint diàlegs que imiten la interacció humana, sinó que a més, ho fan a través d'un suport *exclusivament* humà com es la veu, aquests temes prenen de sobte una absoluta rellevància. Com afecta aquesta relació amb l'ús que l'usuari fa de l'assistent? I anant un pas més enllà, com afecta a la *interpretació i descodificació* que l'usuari farà dels continguts oferts a través de l'assistent de veu?

Els estudis de Google sobre assistents de veu revelen que el 41% dels propietaris d'altaveus intel·ligents afirmen que “els hi fa sentir com si estiguessin parlant amb un amic o amb una altra persona” (Kleinberg, 2018). “Speak Easy”, un informe de tendències globals que analitza sis mesos d'ús d'assistents de veu a 9 països diferents (incloent Estats Units, el Regne Unit i Espanya), explicava que el 43% dels entrevistats asseguraven que estimaven tant el seu assistent de veu que desitjaven que fos una persona real - especialment al mercat asiàtic (*Speak Easy: Global Edition - JWT Intelligence, 2017*). Aquest estudi recull les dades d'un anàlisi neurològic que es va realitzar per mesurar les diferències de resposta emocional del cervell humà davant d'una dada donada per Alexa (com a representant dels assistents de veu) en comparació a la mateixa dada entregada en format text.

Un dels estudis recents més complets sobre aquesta relació és “*Voices in and of the machine: Source orientation toward mobile virtual assistants*”, realitzat per Andrea L. Guzman. Analitzant les entrevistes als subjectes de l'estudi, on s'intentava perfilar la percepció que els usuaris tenen dels assistents de veu (integrats a telèfons mòbils, en aquest cas), Guzman conclou que la conceptualització dels usuaris envers aquestes assistents no és unívoc, sino que pot adoptar dues formes: “alguns participants descriuen l'agent en els seus telèfons com una entitat que poden sentir però no veure (l'assistent és reconegut com una entitat separada), mentre d'altres perceben la tecnologia que empen com la veu del propi telèfon (l'assistent és interpretat com el propi aparell). (...) Quan les tecnologies poseeixen una veu sense cos i estan dissenyades amb

elements socials i un cert nivell d'intel·ligència, el locus [la ubicació] i la natura de l'interlocutor digital no és uniforme al cap de la gent" (Guzman, 2019).

Guzman desenvolupa estudis previs sobre la percepció de l'usuari respecte la màquina (amb *què* o *qui parla*) i com aquesta influència la interpretació que es fa dels missatges que rep. Especialment rellevant és el descobriment que la gent respon als trets i atributs humans que s'hagin programat en aquella tecnologia, incloent el *gènere*, la personalitat, i fins i tot la nacionalitat - que ajuden a construir una imatge mental de la personalitat de l'assistent al cap de l'usuari. Algun dels participants a l'estudi descrivien a l'assistent com "una petita persona al meu telèfon" o, fins i tot, "aquella dona boja que viu al telèfon" - fent referència a les a vegades surrealistes converses que tenen lloc quan l'assistent no és capaç de descodificar la petició de l'usuari i ofereix respostes inconnexes.

És interessant constatar també que aquesta percepció humanoide depèn molt, com apuntàvem prèviament, del fet que l'assistent tingui un nom propi (com Siri, Alexa o Cortana) o no en tingui (com Google Assistant). Aquesta diferència fa que alguns usuaris percebin els assistents com una entitat independent amb personalitat pròpia ("una veu dins del meu telèfon"), o l'entenguin en canvi com una funcionalitat integrada, pròpia de l'aparell.

Per a ponderar-lo, però, s'ha de tenir en consideració que l'estudi de Guzmán es va realitzar al 2015, en un moment de transició en els assistents de veu: una etapa en que l'ús dels assistents estava encara lligat bàsicament a l'ús dels telèfons mòbils. Com canviarà la percepció de l'usuari sobre aquestes "entitats humanoides" ara que aquests assistents no estan només integrats en el dispositiu mòbil, sinó que són "veus" que responen des de l'aire, sense necessitat de tocar cap botó o subjectar cap aparell electrònic a les mans? Com modifica això la nostra sensació d'estar parlant amb un interlocutor humà / robotoide - del que no percebem pràcticament ni la carcassa?

Finalment, més enllà de la personalitat de l'assistent hi ha un altre element social que també hem de tenir present: "Siri i altres agents similars han estat programats amb trets humans (...) però d'igual manera i importància, han sigut programats per reproduir un rol específic en les funcions que poden dur a terme i en el tipus de missatges que ofereixen a l'usuari: el rol d'assistent" (Guzman, 2019).

És significatiu, en aquest cas, i digne de desenvolupar en futurs estudis, el fet que tots els assistents del mercat, a dia d'avui, tinguin noms i veus de dones.

3.6. La promesa de la informació conversacional personalitzada

Les interfícies de veu, que basen la interacció dels mitjans amb l'usuari en una conversa natural, porten implícit en la seva essència la promesa d'una connexió més profunda amb els nostres oients que no només ens permeti informar-los, sinó també respondre les seves preguntes específiques, oferint una personalització que no ha conegut el periodisme fins ara. La possibilitat d'interrompre el relat de les notícies per preguntar-li a l'assistent "Perdona, qui ha dit això?", o "I quins efectes té això al meu país?", o "Aquestes declaracions no s'havien fet ja abans?" pot sonar a ciència ficció en l'actualitat, però les interfícies de veu ofereixen la lògica perquè aquest tipus de periodisme fet a mida comenci a ser possible.

Aquest tret és, de fet, el que fa singular als assistents de veu - i els diferencia d'altres mitjans de reproducció d'àudio tradicional, com poden ser les emissores de ràdio, la ràdio a la carta o la distribució en format podcast. En paraules d'Amy Webb: "La raó per la qual Echo i Google Home s'estan diferenciant de la ràdio o els podcast és per la seva habilitat per pausar i senzillament preguntar alguna cosa al voltant d'aquella història o de qualsevol altra tema (...) Aquesta manera única de sentir les notícies podria impactar la indústria dels mitjans profundament. Pels diaris o les ràdios, serà molt difícil competir en interacció conversacional ja que cap d'ells té la tecnologia o els recursos per crear els seus propi sistema d'anada-i-tornada amb els oients. Un podria argumentar que això no és una gran avantatge respecte teclejar les preguntes a la cerca de google mentre escoltem les notícies [a la televisió o a la ràdio], però el factor *comoditat* ha impulsat les ventes en la indústria tecnològica durant anys" (Webb, 2017).

Més enllà de la tecnologia, conceptualment tenim el repte de dissenyar aquestes converses de manera que, en primer lloc, identifiquem correctament les peticions de l'usuari (ningú vol mantenir una conversa amb un interlocutor que repeteix sistemàticament una disculpa perquè no ha entès les nostres peticions); i, en segon lloc, ens hem d'assegurar que incloem la suficient variabilitat a la conversa de manera *natural* per tal de mantenir l'atenció de l'usuari. El llenguatge natural, doncs, no ha de ser només el llenguatge de la petició, sinó també el llenguatge del mitjà.

3.7. Reptes pendents de l'acadèmia i de la indústria

La informació personalitzada en format conversa no és l'únic repte que proposen les VUI. En l'estat de desenvolupament actual, podem apuntar ja algunes de les línies importants que queden pendents de madurar tant des d'un punt de vista conceptual com a nivell de definició d'interfície. Alguns d'aquests aspectes són:

Discovery. Com farem saber a l'usuari *d'skills* que existim si no ens podem mostrar prèviament en un aparador? En un entorn de veu, dissenyat perquè l'usuari obtingui només aquells continguts que ha demanat de forma explícita - com accedirà a aquells continguts que no sap que existeixen prèviament? Aquest és un problema que pot afectar especialment les marques més petites, que són localitzades sovint pel producte que ofereixen - però no pels seus noms. Els portals webs dels assistents, que mostren les *skills* assistents endreçades per categories, poden servir com a aparador d'opcions, sens dubte, però caldrà investigar més a fons quina és la lògica de visibilitat de les *skills* en aquests entorns. Què determinarà la jerarquia d'opcions que ofereixen els assistents? Acords comercials? Algoritmes interns? Podria ser la cerca per veu la mort del SEO - o una transformació?

Taxonomies. En aquesta nova lògica de cerca, els mitjans de comunicació hauran de posar esforços en un sistema d'etiquetatge i identificació dels continguts segurament molt més extensiu i consistent, que permeti servir el contingut a les diferents interfícies. Quins elements hauríem de identificar i etiquetar en cadascun dels nostres continguts? Marconi, d'Associated Press, parla de temàtiques, persones, organitzacions, localitzacions geogràfiques i dates (Barot, 2017). En el futur, potser identificarem també quins són els dispositius idonis pel consum d'aquella peça, el context d'usuari més adient, i potser, fins i tot, per quins estats emocionals és recomanada.

Identitat de marca. Ho havíem avançat: quan la veu es converteix en la interfície, què passa amb els elements visuals que han permès identificar les marques fins ara a les interfícies digitals visuals? Serà necessari un exercici de construcció de la identitat sonora de la marca basat en aspectes com el gènere i l'edat, però també l'accent, la cadència, el ritme, el to, la inflexió de veu o el llenguatge que s'empra. La utilització de veus conegudes del mitjà, en aquest sentit, pot redoblar la seva importància.

Audiències. Quines seran les mètriques que té sentit mesurar en les VUI, més enllà d'usuaris únics i temps de consum? La quantitat de temps que parlem amb la màquina, p.ex., o el número de vegades que tornem a la *skill*, són indicadors que ens podrien parlar de la fidelització en entorns de veu. Però podríem anar més enllà i començar a mesurar, per exemple, la satisfacció amb l'experiència d'usuari basant-nos en el to de veu en que ens interpel·la l'usuari...?

La indústria i l'acadèmia hauran de treballar plegades per definir les respostes a aquests plantejaments.

4. Definició del Projecte

4.1. Missió i valors

Com a treball d'aplicació professional, la missió, visió i valors que impulsen aquest TFM són els de la Corporació Catalana de Mitjans Audiovisuals, definits de la següent manera a la seva web corporativa (CCMA, 2014):

- La CCMA té la missió d'oferir a tots els ciutadans de Catalunya, en compliment del mandat del Parlament, un servei públic audiovisual de qualitat, compromès amb els principis ètics i democràtics i amb la promoció de la cultura i de la llengua catalana.
- La producció i difusió d'aquest servei es gestionarà amb criteris d'eficiència i buscant la màxima acceptació per part del públic.
- També ha d'impulsar l'oferta de continguts a través dels nous mitjans de comunicació social i de les noves tecnologies, com la ràdio i la televisió digital, internet o telefonia mòbil.
- Totes aquestes funcions s'han d'acomplir sent un model de referència de qualitat i innovació.
- La programació de servei públic dels canals de la Corporació Catalana de Mitjans Audiovisuals ha de ser plural per tractar de satisfer les necessitats del conjunt de la població. Tots els canals i els serveis considerats de servei públic són de lliure accés.
- La llengua institucional per prestar el servei públic de comunicació audiovisual de la Corporació Catalana de Mitjans Audiovisuals és el català.

En aquest marc, per tant, la missió d'aquest projecte serà oferir un servei públic d'informació, plural i de lliure accés. Els valors en que s'arrela són el compromís amb la qualitat, l'eficiència, la innovació i la llengua catalana.

4.2. Abast

En un entorn de tan ràpida evolució i amb tanta falta d'experiència d'usuari com és el dels assistents de veu, el disseny d'un Servei Informatiu ha de ser necessàriament un projecte temporal: una aposta acotada, amb data de finalització, que permeti posar en pràctica una visió específica i recollir feedback i dades d'audiències sobre els que re-pensar el servei i re-dissenyar-lo, si és necessari. Aquest projecte es planteja, per tant, dissenyar un servei inicial que abasti la temporada setembre 2019 - juny 2020, i plantejar-ne la seva reformulació després.

4.3. Audiència

Aquest projecte està orientat, en termes generals, a donar servei a una audiència de parla catalana, siguin residents a Catalunya, a Espanya o a l'estranger. Els continguts dels serveis informatius de la CCMA volen oferir una informació generalista de qualitat a tota la ciutadania, sense distinció de sexes, edats o altres criteris demogràfics.

Dit això, tenint present el caràcter disruptiu i poc estès dels dispositius que integren els assistents de veu, considerem que seria raonable definir aquest servei en el seu estat preliminar pensant sobretot en una audiència adulta, especialment sensibilitzada amb la tecnologia i la informació. El consumidor de notícies en assistents de veu és un usuari que ha realitzat una cerca pro-activa d'un servei de notícies dintre d'un entorn tecnològic; no és un usuari que s'exposa de manera involuntària o casual a la informació. El seu interès, per tant, és explícit i busca cobrir una necessitat informativa correcta; per tant, ens hem d'assegurar que el contingut que oferim dona resposta a aquesta necessitat.

Per últim, aspirem a arribar a una audiència catalano-parlant que no té perquè ser consumidora habitual dels continguts de TV3 i Catalunya Ràdio en l'entorn audiovisual tradicional, però que és susceptible d'establir relació amb la marca si aquesta satisfà les seves necessitats en aquest nou entorn tecnològic. És, per tant, una oportunitat d'arribar a noves audiències.

4.4. Planificació

Amb l'objectiu de finalitzar el disseny del projecte al gener de 2019, i implementar-ho al setembre de 2019, plantegem el següent calendari de projecte:

- 15 OCTUBRE. Validació dels objectius, preguntes, abast i audiència del projecte.
- 22 OCTUBRE. Recerca documental i redacció del marc teòric.
- 05 NOVEMBRE. Recerca documental i redacció del marc contextual.
- 26 NOVEMBRE. Definició dels objectius del Servei, persones i escenaris.
- 17 DESEMBRE. Identificació dels formats i les peces necessàries.
- 24 DESEMBRE. Identificació de les necessitats jurídiques i comercials.
- 31 DESEMBRE. Diagrames d'experiència
- 14 GENER. Eines d'Avaluació i Conclusions.
- 21 GENER. Tancament del Disseny i entrega.

5. Disseny del Servei I: objectius, *persona* i escenaris

5.1. Objectius d'audiència i posicionament de marca

Ens plantegem el disseny d'un servei informatiu per assistents de veu que doni resposta als següents grans objectius:

- Oferir continguts en català que garantitzin l'accés a la informació pública de qualitat en els principals entorns de consum informatiu de la població catalana
- Rendibilitzar el consum del fons audiovisual informatiu de TV3 i Catalunya Ràdio ja existent, donant una nova possibilitat de consum a través de noves plataformes a continguts produïts inicialment per la ràdio i la televisió
- Posicionar les marques informatives de TV3 i Catalunya Ràdio com a pioneres en innovació i experimentació de consum audiovisual en noves plataformes

De manera més específica, el disseny del servei haurà de permetre als seus usuaris:

- Informar-se sobre l'actualitat general, especialment de Catalunya
- Informar-se en català
- Informar-se de manera ràpida, a qualsevol moment del dia, a petició
- Informar-se de forma natural a través dels assistents de veu

5.2. Definició d'usuaris: buscant el perfil real de la *persona*

Per a qui dissenyem aquest servei? La metodologia de disseny de serveis basats en usuari proposa la creació de *personas*, en la seva accepció llatina: la creació, basada en estudis quantitius o qualitius, d'uns usuaris prototípics que sintetitzen el perfil, les necessitats, motivacions i habilitats dels usuaris als que volem adreçar el nostre servei.

Seguint les recomanacions de la UOC recollides en el seu «Design Toolkit» per a l'aplicació de metodologies de disseny basat en usuari (UOC, s.d.), optem per aplicar el mètode de la *persona* per sintetitzar en una fitxa específica, que no ha d'extendre's més d'una o dues pàgines, aquells factors que són especialment rellevants pel nostre tipus de servei i descriu així el nostre usuari tipus. «Aquests perfils ens serveixen per a modelar patrons d'actuació dels usuaris, comportaments, actituds, aptituds, objectius, entorns de treball, eines que fan servir, reptes que han de superar, entre d'altres» (UOC, s.d.).

En el nostre cas, hem escollit els següents trets distintius per elaborar el perfilat de l'usuari tipus:

- **Perfil personal**
 - Nom. Malgrat ser un nom totalment fictici i aleatori, ens ajudarà a identificar el prototip de persona amb una persona real, i adreçar-nos a ell o ella de forma ràpida en el procés de disseny.
 - Fotografia. De nou, malgrat ser fictícia, ens ajuda a atorgar corporeïtat a l'usuari.
- **Demografia**
 - Edat. Ha de correspondre al rang d'edat al qual estem adreçant el servei. És rellevant perquè de l'edat es derivarà una certa capacitació tecnològica i hàbits diaris de consum informatiu.
 - Ocupació. Malgrat no és un condicionant pel servei, la professió ens ajuda a donar corporeïtat al personatge – i sí ha de ser coherent amb el perfil professional i/o de formació i estudis al qual estem adreçant el servei.
 - Vivenda compartida. La dada és rellevant en tant que estem dissenyant un servei informatiu que es consumeix principalment a la vivenda, i per tant, l'existència d'altres persones pot fer que aquestes es converteixin en consumidors involuntaris – o passius del nostre servei.
- **Capacitació**
 - Telèfons mòbils. Malgrat la interacció amb Alexa es realitza majoritàriament per veu, la configuració inicial de l'assistent requereix la instal·lació d'una aplicació al telèfon mòbil. La capacitació de l'usuari per dur a terme aquest pas és imprescindible com a pas previ a la utilització del nostre servei.
 - Cerca d'informació. L'habilitat i capacitació fent cerques d'informació serà rellevant pel nostre usuari a l'hora de fer una possible interrogació a la informació que li presentarem.
 - Assistents de veu. L'experiència en interlocució amb assistents (o la manca d'ella) és un factor determinant a l'hora de dissenyar una interlocució més o menys complexa pel nostre servei.
- **Consum informatiu habitual**
 - Marques. Les marques habituals de consum informatiu ens ajuden a perfilar, d'una banda, la nostra possible competència, i d'altra, les preferències editorials i de tot de l'usuari al que ens estem adreçant. Obtindrem aquesta relació de les enquestes que realitzarem als usuaris potencials.
 - Plataformes. Les plataformes habituals de consum informatiu ens ajuden a entendre els hàbits digitals de consum d'aquell usuari, la seva capacitació tecnològica i la seva

preferència pel que fa a formats i plataformes. Obtindrem aquesta relació de les enquestes que realitzarem als usuaris potencials.

- **Aproximació al servei**

- Motivacions. Aquesta és la informació més rellevant de tot el perfil, doncs posa en relleu quin és el principal motiu que porta a l'usuari a consumir-nos. Ens hem d'assegurar que el nostre servei dona satisfacció a aquesta necessitat o motivació, o el servei deixarà de tenir sentit. Obtindrem aquesta relació de les enquestes que realitzarem als usuaris potencials.
- Reticències. De la mateixa manera, el nostre disseny haurà d'intentar evitar generar les possibles incomoditats o friccions que generen reticències en el nostre usuari potencial, o podríem trobar-nos que malgrat cobrim les seves motivacions, les reticències pesin més i s'acabi per evitar l'ús del nostre servei. Obtindrem aquesta relació de les enquestes que realitzarem als usuaris potencials.

Per acotar la dimensió d'aquest model de pilot informatiu i ser capaços d'implementar-ho en els terminis previstos, optem per perfilar de moment un únic usuari tipus - entenent que una evolució posterior del servei podria requerir de entre 3 i 5 usuaris diferents, idealment, per adaptar-nos a tot el rang de personalitats que poden fer ús del servei.

Per fidelitat a la missió de la Corporació, adreçarem el servei a un consumidor de parla catalana, major d'edat, interessat en la informació general. Per perfilar els elements demogràfics del nostre target tipus, ens basem en els estudis sobre consum d'altaveus intel·ligents publicats per Comscore (Engleson, 2017), que ens diuen que les llars on són més presents els aparells són aquelles amb uns ingressos més elevats, i amb un membre de la família entre 35 i 49 anys. L'estadística també mostra una lleu predominància de llars amb fills (108 vs. 95), però la diferència és poc destacable i optarem per no incloure aquest tret com un element definitori de l'usuari tipus, ja que suposaria una exclusió gairebé de la meitat dels usuaris potencials.

Per perfilar la seva aproximació al servei, marques habituals, plataformes i ocupació, necessitem realitzar enquestes qualitatives a usuaris que complissin amb el target 35-49 i amb rentes mitjes – altes, i que poguessin ser per tant usuaris potencials d'assistents de veu. Per aquest treball, triem una mostra de 13 usuaris tipus que aconsegueixen aquests requisits i recollim les seves respostes en un formulari de Google Forms (veure Annexos). De les seves respostes obtenim una síntesi molt interessant de quines serien les seves principals motivacions i reticències a l'hora d'utilitzar un servei informatiu a través d'assistents de veu.

Pels usuaris interrogats, les **motivacions** més mencionades tenien a veure amb tres grans àmbits temàtics, amb una freqüència molt similar. Els usuaris ressaltaven:

- Rapidesa, immediatesa
- Comoditat, «poder fer altres coses mentre ho escolto»
- L'aprofundiment sota petició. «Poder preguntar pel context de la notícia, ampliar informació», «poder demanar informació de diferents països ràpidament»

Pel que fa a les **reticències**, n'hi ha una que sobresurt respecte la resta:

- La seva limitada comprensió. «Que m'entengui», «l'idioma», «que no em doni la info que busco perquè no entèn les meves preguntes»

En segon terme, sembla que existeix la percepció que l'assistent serà qui decidirà la font d'informació o els criteris de cerca, restant autonomia i capacitat de decisió a l'usuari, i això genera incomoditat i suspicàcia.

- Falta de control sobre fonts i cerques. «No poder controlar el criteri de cerca», «La dificultat d'establir criteris de cerca o fonts d'informació», «La font».

Finalment, es citaven preocupacions més individuals però igualment indicatives:

- La privacitat. «Sota cap circumstància empraria un assistent de veu pel risc a la privacitat que representen».
- La seqüencialitat del consum. «No poder decidir entre diverses opcions simultànies, atès que hauria d'esperar necessàriament a que acabés cada locució».
- La incomoditat a l'hora de consumir peces més llargues.
- La potència de l'altaveu
- La falta de costum

Veiem que la algunes d'aquestes reticències (privacitat, potència, control sobre les fonts) tenen més a veure amb la tecnologia dels assistents de veu que amb el possible servei informatiu – i per tant, en alguns casos seran difícils d'abordar des del disseny del servei. Intentarem, no obstant, tenir-les present en tan quant puguin afectar el nostre consum. Dit d'una altra manera: no podem fer que Alexa t'entengui millor en termes generals, però sí podem intentar que el nostre servei informatiu t'entengui mentre interlocutis amb ell.

Pel que fa a les habituals plataformes de consum informatiu que han esmentat els usuaris de l'enquesta, veiem que presenten majoritàriament un consum informatiu en plataformes digitals. En

concret, el rànquing de les 3 plataformes més emprades és: webs (100%), xarxes socials (85%) i ràdio (77%). Pel que fa a marques informatives, les 4 més comuns han estat: El País (54%), Ara (46%), Rac1 (31%) i 324 (31%). Mantenim 4 marques en aquest cas perquè la tercera i la quarta tenien el mateix pes en el rànquing.

Només ens queda perfilar la capacitació tecnològica del nostre usuari tipus. A falta de dades estadístiques per el nostre perfil concret (usuari català interessat en informació general, entre 35 i 49), ens basarem en les dades generals d'ús de tecnologia recollides al informe «Digital 2018» per Espanya (We are social & Hootsuite, s.d.). L'estudi recull que un 96% de la població espanyola adulta és usuària habitual de telèfons mòbils, i que entre aquests usaris, l'activitat més habitual és l'ús d'un cercador (60%). Basant-nos en aquestes dades, deduem un alt nivell de capacitació en l'ús bàsic d'un telèfon mòbil, i un nivell mig d'habilitats de cerca. Pel que fa a l'ús d'assistents de veu, tenint present que Alexa va aparèixer al mercat espanyol fa 3 mesos i encara no han aparegut estudis estadístics, considerarem un coneixement inicial d'aquesta tecnologia.

Aquesta combinació de trets estadístics i qualitatius ens permet dibuixar la següent *persona*, entesa com a un personatge fictici que encarna el nostre usuari tipus i sintetitza els seus hàbits, interessos, habilitats tecnològiques i circumstàncies:



Figura 1: *Fitxa de persona*. Font: elaboració pròpia a partir de les enquestes a usuaris

5.4. Definició de contextos d'ús: ubicant la *persona* al seu escenari

Ara que tenim perfilat el nostre usuari arquetípic, *l'Anna*, passarem a identificar quins són els possibles escenaris d'ús del servei en el context del seu *user journey* - o dit d'una altra manera: passem a identificar en quin moment del dia d'aquesta vida arquetípica i en quin context l'usuari es posarà en contacte amb el nostre servei.

Aquesta identificació ens permet definir amb claredat aspectes importants que condicionaran la posterior definició del nostre servei, tal com:

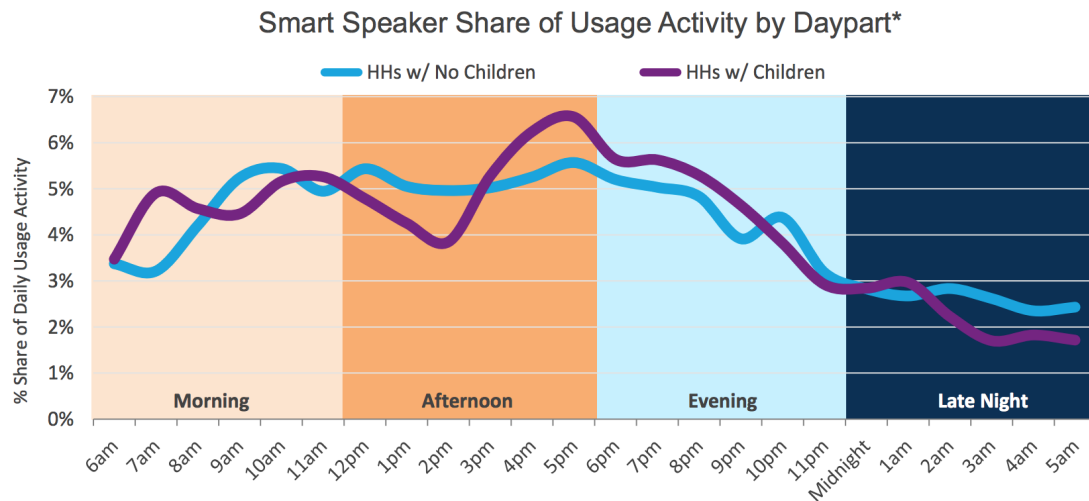
- **Dispositiu.** L'aparell físic a través del qual s'establirà el contacte amb el servei. En aquest cas, sabem d'avant mà que es tracta d'un altaveu intel·ligent.
- **Ubicació.** L'espai físic on tindrà lloc el punt de contacte amb el servei. És un espai públic o privat? És sempre el mateix espai? Hi habiten altres persones? En quin punt de l'espai està exactament ubicat el dispositiu? Hi té accés directe l'usuari? Hi ha soroll ambiental?
- **Hora.** El contacte de l'usuari amb el servei tindrà lloc a una hora regular, de manera aproximada? Serà variable? Quina hora és? Quin tipus d'informació podem oferir a aquella hora?
- **Context.** Quines són les circumstàncies de l'usuari quan entra en contacte amb el servei. Està cercant pro-activament informació per satisfer una demanda, necessària per executar una tasca? Està compartint temps amb d'altres i busca una via d'entreteniment comuna? Està realitzant tasques rutinàries i busca un acompanyament de fons?
- **Disponibilitat de temps.** De quina quantitat de temps disposa l'usuari per a fer ús del servei? És un ús constant o és interromput?

Els altaveus intel·ligents estan dissenyats per a ser ubicats a la llar, i necessiten estar endollats a una presa elèctrica i connectats a una xarxa wi-fi per poder operar. Això ens descarta d'entrada l'ús del servei informatiu en mobilitat, i ens garanteix una certa estabilitat de connexió. En l'entorn domèstic, podem trobar el dispositiu ubicat al menjador, la cuina, el despatx, el quart de bany o una habitació. Tenint present que el nostre usuari tipus no viu sol, i que la majoria d'aquests espais acostumen a ser espais comuns, hem de comptar amb la presència de més persones o de **soroll ambiental** durant l'actuació del servei.

La presència de tercers serà especialment delicada en el cas que aquests siguin **menors**. Els anàlisis de consum d'altaveus intel·ligents publicats per Comscore (Engleson, 2017) mostren que una mica més de la meitat dels posseïdors d'un altaveu intel·ligent tenen fills. Haurem d'incloure

algun tipus d'advertiment sobre el contingut que oferim al servei si aquest pot ser no apte per menors.

Households with children have more defined peaks of smart speaker activity in the early morning and late afternoon



COMSCORE

Source: comScore Connected Home, U.S., October 2017
* Usage activity includes background calls made by smart speakers.

© comScore, Inc. Proprietary. 14

Figura 2: Quota d'ús d'altaveus intel·ligents al llarg del dia. Font: Comscore Analytics, 2017: «The future of voice, from Smart Speakers to Smart Homes».

Pel que fa a l'hora i el context de consum, l'anàlisi de Comscore (Engleson, 2017) ubica els pics més alts d'activitat entre **les 10h i les 11h del matí, a les 17h** (el pic més alt de tots en el cas de llars amb nens) i després un consum sostingut **fins a les 21h**, on es precipita. Malgrat aquests lleugers pics, el consum és molt sostingut al llarg del dia; de fet, el més rellevant és la caiguda de consum a l'hora de l'esmorzar (8 a 10h) i a l'hora del migdia (1 a 3h) en el cas de les llars que tenen nens – que significativament mostren un pic de consum molt més alt a primeríssima hora del matí. Aquestes caigudes semblen indicar que els adults fan menys ús dels assistents en els moments que han d'atendre els nens, especialment en els àpats – tot i que és només una suposició. No hem d'oblidar, a més, que aquesta mètrica fa referència a qualsevol tipus d'ús del altaveu, no exclusivament al consum informatiu – que no té perquè ser necessàriament tan sostingut. En tot cas, l'estadística ens ajuda per entendre que durant els àpats el consum decau a les llars amb criatures, i al vespre el consum decau en general.

Un estudi específic del Reuters Institute sobre el futur dels assistents de veu i les oportunitats que aquests representen, basat en entrevistes i focus groups a Estats Units, Regne Unit i Alemanya,

perfilava una corba de consum relativament similar – amb l'afegit de distingir quin tipus d'activitats eren més comuns en cada moment del dia (Newman, 2018):

Smart speaker use peaks early and late in the day

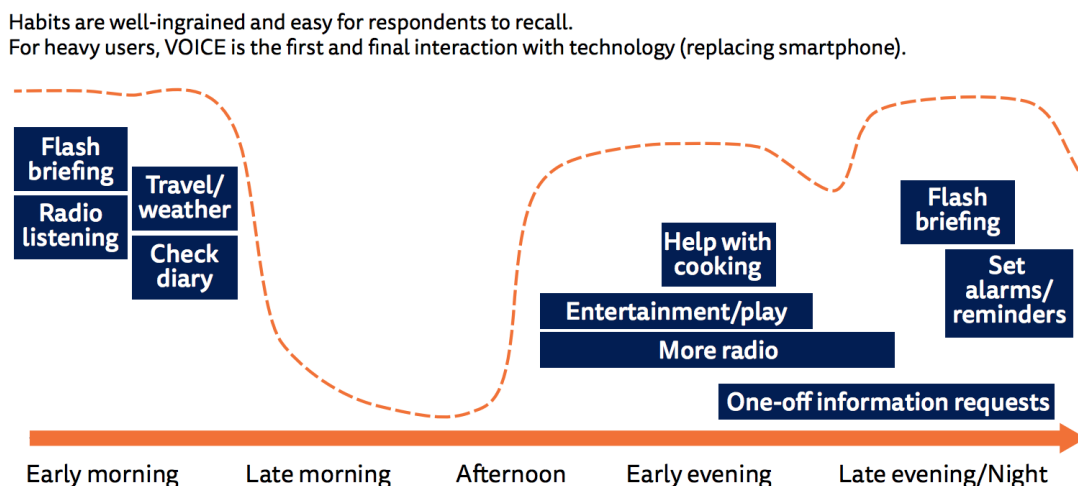


Figura 3: *Pics d'ús dels altaveus intel·ligents a primera i última hora del dia.* Font: Reuters Institute for the Study of Journalism, 2018: «The future of Voice and the implications for News».

Per entendre millor el context i la disponibilitat de temps, hem de recórrer a les enquestes qualitatives, on alguns usuaris apunten que el que volen és «estar al dia mentre fan altres coses», i d'altres «obtenir informació addicional sobre alguna cosa». Per tant, hi ha un doble context d'ús: un més passiu, que entèn el servei com una actualització general que l'acompanya mentre realitza altres tasques – i un més actiu, on l'usuari pot estar enfocat totalment en la consulta d'informació, interrogant i obtenint respostes a allò que l'importa. Aquests contextos poden tenir una durabilitat variable, però definirem una franja de 4 minuts màxim pel context passiu (el que recomana Amazon per les *skills* d'actualització informativa, els 'flash briefings'), i una duració oberta pel context actiu, on serà l'usuari qui posi fi a la consulta.

5.6. Política de privacitat i tractament de dades

La privacitat és un aspecte delicat que aixeca suspicàcies en els potencials usuaris dels assistents de veu – que han de saber, com a mínim, en tot moment, quin tipus d'informació s'està recollint sobre ells. Arran de l'aprovament per part del Parlament Europeu del Reglament General de Protecció de Dades (RGPD) el 25 de maig de 2018, la CCMA va ajustar les seves polítiques de privacitat i tractament de dades per assegurar la màxima transparència envers l'audiència: l'usuari pot consultar a quines de les seves dades està tenint accés la CCMA, quin tractament en fa, i pot optar per no cedir-les, arribat el cas. Uns mesos després, al desembre de 2018, el govern

espanyol va aprovar la Llei orgànica 3/2018 de 5 de desembre de Protecció de dades i garantia dels drets digitals, que substitueix a l'anterior LOPD vigent fins ara i adapta al dret espanyol el model establert pel RGPD. Aquesta nova llei contempla, entre d'altres, sistema d'exclusió publicitària, vetlla pels drets d'accés, rectificació i supressió, i també pel dret a l'oblit, inclòs el dret a l'oblit en cerques d'internet i xarxes socials i el dret a la portabilitat en xarxes socials. Aquests drets treballen en conjunt per garantir el dret i la llibertat dels usuaris a l'hora de saber quina informació es té sobre ells i demanar-ne la total supressió, si fos necessari (Grup Carles, 2019).

En el cas del nostre servei informatiu, que s'integra en una plataforma al·liena com és Amazon, l'escenari jurídic es complica – perquè passen a haver-hi dos actors que potencialment poden recollir dades d'usuari. Les dues empreses, quan operen al nostre territori, estan obligades a complir les esmentades lleis.

Com podem abordar aquest escenari? D'una banda, la pròpia plataforma d'Amazon demana explicitar en cadascuna de les *skills* un accés a la política de privacitat de l'empresa on consti les dades que es recullen a través de l'aplicació. En el cas del nostre servei informatiu, on oferim un contingut pre-establert, igual per tothom, amb accés gratuït i directe (i per tant, sense cap tipus d'identificació de l'usuari), no és necessari recollir cap dada de caràcter personal. Qui si recull dades és, no obstant, Amazon, que facilita als desenvolupadors informació estadística sobre el volum d'usuaris que han habilitat una aplicació, el volum de consum que ha tingut aquella aplicació, i des de quins dispositius s'està elaborant el consum.

Donades aquestes circumstàncies, i després de consultar-ho amb els assessors jurídics de la CCMA, s'opta per incloure un breu paràgraf al actual avís legal de la CCMA que ho expliciti:

«La CCMA ofereix una selecció dels seus continguts a algunes plataformes dissenyades per tercers que són controlades a través d'assistents de veu, com per exemple, Alexa, d'Amazon. En aquests casos, la CCMA no utilitza «cookies» ni cap tipus de tecnologia que reculli informació sobre l'usuari; són les pròpies plataformes les que poden recollir informació sobre el consum dels nostres continguts que, en tot cas, sempre serà informació anònima, relativa a quin tipus de continguts tenen més consum i en quins dispositius es consumeixen.»

5.7. Opcions de rendibilització comercial

Com a proveïdora de servei públic, la rendibilitat comercial no és una condició indispensable per a la Corporació Catalana de Mitjans Audiovisuals a l'hora de desenvolupar nous serveis – però sí

que és una opció desitjable. Sent fidels a la seva missió i valors, intentarem buscar solucions comercials que maximitzin l'eficiència en la producció del nou servei, però garantint-ne sempre un accés gratuït.

En el cas concret de l'explotació comercial de serveis que es despleguen en plataformes tecnològiques de tercers, de les quals la CCMA no té el control del disseny de l'experiència d'usuari, l'elecció d'un format o un altre vindrà, a més, condicionada pels suports publicitaris que ofereixen aquestes plataformes.

En plataformes com YouTube, per exemple, s'ofereix als mitjans la possibilitat de servir clips de vídeo que es reproduïen en format de pre-roll, mid-roll o post-roll durant el consum del vídeo. Aquestes posicions publicitàries estan identificades com a tals i majoritàriament serveixen anuncis per subhasta, provinents d'una gran repositori publicitari gestionat per YouTube i que tria un anunci o un altre en funció de les metadades del vídeo, del perfil del consumidor, del preu que estigui pagant l'anunciant, etc. Els beneficis econòmics d'aquest anunci són repartits entre el mitjà de comunicació, propietari del vídeo, i YouTube, gestor de la publicitat.

Altres plataformes com TuneIn o iVoox operen amb un sistema similar (oferint formats publicitaris que s'emeten associats a l'àudio dels mitjans de comunicació, en funció d'uns criteris de subasta), però mantenint els beneficis de la publicitat exclusivament per a la plataforma tecnològica. El model YouTube, sobra dir-ho, és més equitatiu en termes de negoci i genera un al·licient de cara al mitjà per tal de seguir produint i explotant continguts en aquella plataforma.

En el cas d'Alexa, la plataforma no ofereix a dia d'avui cap tipus d'integració publicitària pel mercat espanyol, malgrat sí que n'ofereix al mercat d'Estats Units. El format? Ha anat variant des de l'estrena del servei.

Amazon va posar en proves a principis de 2018 un model de rendibilització en el que, enlloc de mostrar anuncis aleatoris quan l'usuari feia servir una *skill*, Alexa proposava recomanacions de producte quan l'usuari feia una cerca específica d'un producte de consum. Aquest model publicitari, malgrat podia obrir les portes a Amazon per treure rendiments comercials, no contemplava en l'equació als desenvolupadors de *skills*. Finalment, però, va ser discontinuat sense massa explicacions per substituir-lo per un altre model: la comercialització de contingut premium de pagament dins de les pròpies *skills* (Levine, 2018). D'aquesta manera, *skills* que són

inicialment gratuïtes ofereixen continguts addicionals i opcionals de pagament – com per exemple, jocs interactius, contingut exclusiu, nous sets de preguntes per un Trivial...

Aquest model, que s'apropa als models comercials de les apps dels telèfons mòbils, permet als creadors de les *skills* conservar una part important dels ingressos: el 70% (Sterling, 2018). És, en aquest sentit, més similar a la filosofia de comercialització de YouTube, que entén que els beneficis s'han de repartir entre creadors de continguts i plataformes per tal que el model sigui sostenible i interessant per ambdues parts.

Per facilitar aquestes noves transaccions econòmiques dins les *skills*, Amazon va desenvolupar l'anomenat Amazon Pay per Alexa *skills*: un sistema que identifica el perfil d'Amazon de l'usuari (amb la seva informació bancària, domicili, etc.) dins de les *skills* particulars, de manera que per efectuar les compres l'usuari no ha de tornar a identificar-se o accedir a un entorn d'interfície visual (p.ex, la pantalla del mòbil) per poder acabar el procés. D'aquesta manera, es manté durant tot el fluxe de compra l'experiència en veu – facilitant la integració de compres dins d'*skill* a tots els desenvolupadors (Zanuldin, 2018). Aquest model està disponible també a alguns països europeus des de setembre de 2018, però encara no ho està a Espanya.

Al setembre de 2018, Amazon anunciava pel mercat d'Estats Units un nou format de venda, el que ells han anomenat “**consumibles**” (Fisseha, 2018): *“Un consumible és un producte dins de la skill que els clients poden comprar, utilitzar, i després comprar de nou. Com a afegit a les compres i subscripcions, els consumibles donen als desenvolupadors més maneres d'entregar experiències premium als seus clients. Amb els consumibles, pots vendre productes que són rellevants en aquell precís moment pels clients de la teva skill”*

Aquests models no estan disponibles a Espanya encara, i per tant, no els podem incorporar al disseny del servei, però tot i així el cert és que no tenen una fàcil adaptació als models de rendibilització comercial de la CCMA. El mandat d'accés gratuït als continguts bloqueja les possibilitats d'oferir contingut addicional de pagament – i encara més tractant-se d'un producte informatiu, on el dret a la informació està per sobre de qualsevol rendiment econòmic. Serà interessant fer un seguiment per veure si s'ofereix en el futur algunes de les següents opcions:

- Formats publicitaris audiovisuals que puguem incloure abans, durant o en el tancament de l'emissió de les càpsules.
- La possibilitat d'oferir contingut patrocinat per un anunciant.

6. Disseny del servei II: continguts, veu i interacció

6.1. Identificació dels continguts i del tipus de servei

Per identificar els continguts informatius que servirem i de quina manera, recuperem i analitzem en primer lloc les motivacions que presenta la nostra persona tipus:

«Vull estar al dia d'una forma ràpida, immediata i còmode, mentre faig altres coses»

La premisa inicial és que vol "estar al dia", d'una manera generalista: això ens porta a fer una selecció de continguts amb criteri periodístic, seleccionant els temes del moment com ho faríem per un butlletí informatiu radiofònic o televisiu - sense plantejar fer una agenda periodística especial o alternativa per aquests dispositius, ni de fer un filtratge temàtic per oferir una agenda vertical. El treball per tant de cerca, verificació d'informació i selecció de titulars pot ser, per tant, reaprofitat de l'actual treball de les redaccions. Donarem una selecció de 6 titulars, que cobriran les àrees de política, societat, economia, internacional, cultura i esports - endreçada amb criteri periodístic.

Pel que fa a la manera d'entregar aquests titulars, el primer condicionant és la rapidesa: hem de poder oferir un resum informatiu general de consum ràpid, que ofereixi un sumari de l'actualitat de manera sintètica. Plantegem, d'entrada, un sumari de titulars que duri 1 minut, aproximadament. Aquesta és una durada clarament breu per un sumari i que ofereix un bon argument de venda pel consum ("L'actualitat, en 1 minut"). La majoria *d'skills* de mitjans de comunicació que hem consultat (veure el punt 2.4. *Els mitjans de comunicació als assistents de veu*) oferint resums informatius d'una durada al voltant dels 4 minuts; tenint present que el nostre sumari està previst que pugui ser ampliat a petició de l'usuari, creiem que 1 minut és un bon punt de partida. Veurem més detalls de la durada de la peça al punt 6.2. *Estructura i tipus de continguts*.

El segon condicionant és que sigui immediat: que el més destacat de l'actualitat es serveixi de manera actualitzada en el moment que fa la petició l'usuari. Aquesta voluntat ens condiona a l'hora de plantejar l'entrada al servei: evitarem una configuració o personalització en la primera interacció, així com una introducció llarga o una careta. L'objectiu és servir el cor del contingut tan aviat com sigui possible. Alhora, ens haurem de segurar que podem oferir informació actualitzada en aquells moments en que l'usuari és més procliu a fer la consulta - que segons hem definit als nostres usuaris, són entre les 10h i les 11h del matí, a les 17h i després un consum sostingut fins

a les 21h. Aprofitant les rutines actuals de les redaccions, procurarem servir contingut actualitzat cada hora, d'acord amb els butlletins informatius radiofònics i les rodes de notícies del 324.

El tercer condicionant és la comoditat: hem de procurar que el consum del contingut sigui una experiència agradable i còmode, que no requereixi un esforç especial, sigui d'escolta o de comprensió, i que pugui ser consumit mentre l'usuari està executant altres tasques. Això ens porta a plantejar un servei que no requereixi obligatòriament en cap cas de visualitzar una pantalla, ni de realitzar accions complexes per part de l'usuari. Malgrat puguem oferir diferents experiències d'aprofundiment en les notícies, una d'elles ha de ser absolutament passiva per l'usuari.

«Vull poder ampliar informació sobre els aspectes de la notícia o els temes que jo triï, quan jo triï»

El segon requeriment és el que ens plantejarà un major repte de disseny. La nostra persona vol ser informada de manera ràpida, immediata i còmoda - però alhora, vol poder *interrompre* el flux informatiu per preguntar coses concretes o demanar més informació o context sobre un aspecte de la notícies.

Per tant, hem de dissenyar un model de servei que sigui, al mateix temps, autònom (que pugui ser consumit de manera passiva) i fluïd, i que contempli possibles interrupcions amb ramificacions del relat per aquells que vulguin aprofundir.

Aquest segon escenari ha de tenir present l'actual estat del reconeixement de veu en català i de la intel·ligència artificial - així com de les capacitats tecnològiques de la CCMA per oferir una experiència en aquest entorn. Sobretot tenint present que la principal reticència del nostre usuari tipus és, precisament, la por a la incomprensió: «Em genera frustració que l'assistent no entengui el que li pregunto – o no entengui el meu idioma». L'experiència de la CCMA en periodisme conversacional redunda aquesta por: es va realitzar un pilot de bot pels serveis informatius integrat a Facebook Messenger, que permetia la conversa oberta amb els usuaris, i una major part dels missatges naturals que aquests enviaven eren malinterpretats o no eren descodificats. Això va generar un volum alt de frustració i decepció d'expectatives amb el servei que caldria evitar. La solució, en aquests casos, és optar per oferir opcions tancades on l'usuari decideixi - però sense permetre la pregunta oberta. Aquesta solució, però, ens portar a enfrontar la segona principal reticència de l'usuari: «M'incomoda haver d'esperar pacientment a sentir totes les opcions per poder triar – o estar sentint peces molt llargues».

Amb aquests requisits, plantegem oferir a l'usuari la possibilitat de parar la locució automàtica de titulars en qualsevol moment per ampliar informació d'aquella notícia - però ho farem oferint sempre les mateixes alternatives d'ampliació, de manera que puguem identificar amb claredat quina és la petició que ens fa l'usuari i a l'hora tenir preparada la informació necessària per satisfer-la. D'aquesta manera, mantenim la seva llibertat per triar el moment i el tema sobre el que vol ampliar informació, però rebaixem les seves opcions de quin tipus d'informació pot demanar per assegurar-nos que podrem entendre tecnològicament aquesta petició i podrem donar-li resposta editorialment. Amb aquests condicionants, les opcions de tria que oferirem seran:

- Context previ d'aquesta notícia
- Ampliació de la notícia
- Notícies relacionades amb aquesta notícia

6.2. Estructura i tipus de continguts

En termes de continguts, doncs, el nostre servei ha d'oferir una selecció base de 6 notícies multitemàtiques, amb un doble format:

- Format de consum passiu: oferirà els titulars corresponents a les 6 notícies, en una locució que en conjunt durarà aproximadament 1 minut.
- Format de consum actiu: oferirà un segon bloc de continguts ampliat, que inclourà 2 peces extres per notícia (context i ampliació; 12 peces addicionals, per tant), més un conjunt de notícies relacionades per cada notícia. En aquesta fase pilot, establim un màxim de 3 notícies relacionades, que és un volum d'informació acotable, i esperem a les primeres analítiques per avaluar si l'usuari demanda més o menys informació; aquestes 3 notícies relacionades per cadascuna de les 6 peces generen 18 peces addicionals. Crearem en total, doncs, 30 unitats d'informació addicionals a les 6 inicials per cada bloc horari.

L'estructura dels continguts per tant seria:

- Titular notícia 1 (10')
- Context (20')
- Ampliació (20')
- Notícia relacionada 1 (10')
- Notícia relacionada 2 (10')
- Notícia relacionada 3 (10')

Al tractar-se d'un servei pilot on volem oferir un producte molt acotat a les necessitats de l'usuari,

apostem per una selecció i creació manual dels continguts – enlloc d'optar per una automatització de continguts ja existents. Per tant, necessitarem una persona dedicada a la producció diària d'aquest servei, que dediqui una part de la jornada a elaborar els continguts, locutar-los i emmagatzemar-los digitalment per tal que els equips tècnics puguin integrar-los en la nostra *skill* de servei informatiu.

Pel que fa als tipus de continguts, definim amb més detall què volem oferir en cada cas:

- **Titular.** Serà breu, directe i clar, amb una durada aproximada de 10 segons. Donarà la informació imprescindible per entendre la notícia. El redactarem a partir del material treballat per les redaccions d'Informatius pels butlletins horaris de Catalunya Informació i el canal 324 – de manera que optimitzem els recursos periodístics de la CCMA i no creem equips paral·lels de rastreig, verificació i elaboració de notícies, ni agendes editorialment incoherents amb la línia de les redaccions centrals.
- **Context.** Explicarem els antecedents de la notícia i en quin marc general s'emmarca, què ha passat fins ara que expliqui el fet, quin és el seu pes en el conjunt del tema que es tracta, qui són els seus protagonistes. Tindrà una durada aproximada de 20 segons.
- **Ampliació.** Entrarem al detall de la notícia per explicar aspectes concrets, reaccions generades entre els seus protagonistes i altres actors relacionats, quines són les possibles conseqüències. Tindrà una durada aproximada de 20 segons.
- **Notícies relacionades.** En aquest cas no farem una elaboració de nous continguts, sinó que utilitzarem els continguts elaborats prèviament per les redaccions informatives i que estan indexats en l'entorn digital informatiu de la CCMA, al portal www.324.cat. Farem ús de l'actual sistema d'etiquetes temàtiques del portal, que identifica els temes d'actualitat i els agrupa sota una mateixa etiqueta (p.ex., "judici procés"), per fer una tria de les 3 notícies relacionades que oferirem en cada cas. Per cada notícia, farem una locució del titular – que durarà, de nou, uns 10 segons aproximadament.

Per acotar l'experiència en aquest primer pilot, durant el consum de les notícies relacionades l'usuari no podrà interrompre la locució per demanar una ampliació o el context d'aquella nova notícia. Aquests titulars relacionats han d'operar més com una informació addicional de context sobre la notícia principal, que com un nou fil de consum d'una nova notícia. Un cop el servei en marxa i amb les primeres analítiques i feedback per part dels usuaris, ens podem plantejar l'evolució del servei en aquesta direcció.

Amb aquest plantejament, per tant, el consum màxim d'una de les 6 notícies *core* s'extendria

1m20s, aproximadament. Si l'usuari consumís la totalitat de la informació oferta cada hora, el total de l'experiència s'allargaria uns 8 minuts. Ho considerem una durada encara molt acotada, que s'encabeix dins de l'objectiu principal d' "estar al dia d'una forma ràpida, immediata i còmode".

6.2. Definició de la veu de la marca

Com hem vist anteriorment, la definició de la veu de la marca és un dels elements més importants en el cas dels assistents de veu a l'hora de transmetre una impressió sobre la nostra marca. En primer lloc, en entorns de consum d'àudio és habitual la creació del que es coneix com una «**marca sonora**», una breu sintonia que pot durar entre 2 i 5 segons i que ens servirà per crear un recurs auditiu que l'audiència identifiqui immediatament amb el nostre servei. En un entorn sense imatges, la creació d'una marca sonora serà vital per donar presència a la nostra marca i capacitat de ser recordada.

En segon lloc, la personalitat de marca vindrà definida també per la veu de la persona que realitzi la locució del servei, així com pel tipus de llenguatge que emprem, el to de la comunicació, i la utilització d'un registre lingüístic o d'un altre.

La veu dels serveis informatius de la CCMA es caracteritza per ser una veu *sòbria, continguda, templada, formal, molt neutre*. És una veu de servei públic pensada per adreçar-se a un públic generalista molt ampli que consumeix informació a través dels mitjans tradicionals, abraçant un ampli espectre d'edats, gèneres i interessos. En el cas del servei per assistents de veu, però, nosaltres podem acotar molt més aquest públic al que ens adrecem: per començar, estem dissenyant un servei per una persona que té entre 35 i 49 anys, i que ens consumirà possiblement en l'intimitat de la seva llar. En segon lloc, estem dissenyant un servei de veu per uns assistents que han generat una personalitat pròpia – fins al punt que molts usuaris els consideren d'alguna manera "individus", entitats amb un perfil humanoide (veure el capítol ""). En tercer lloc, estem dissenyant un servei *interactiu*, al que l'usuari s'hi pot adreçar per demanar informació addicional. No podem deixar aquest aspecte de banda a l'hora de definir el to de la comunicació, que per ser natural, necessàriament s'ha d'apropar més al to d'una conversa.

Amb aquests elements, podem perfilar una mica la veu de la nostra marca informativa en aquests entorns, per fer-la més *propera, íntima i informal*, però haurem de mantenir la neutralitat del to per no allunyar-nos de la nostra missió de servei públic. La proposta que fem consisteix en mantenir el to sobri en els titulars, per mantenir-los directes i escuets, i treballar la proximitat i el to més informal en les peces de context i ampliació, que és el moment en que el nostre usuari ha decidit

establir una conversa amb el servei. Volem que la nostra marca soni propera, com si fos la veu d'un amic ben informat que t'explica les notícies d'una manera planera i còmplice.

Veiem com aplicar aquesta treball de to amb un exemple real:

Titular del butlletí horari de Catalunya Informació:

“Segons ha pogut saber Catalunya Informació, Esquerra Republicana ha plantejat a Rubén Wagensberg que ocupi un lloc destacat a l'ajuntament de Barcelona. Wagensberg, que ara és diputat pel Parlament, destaca per la seva implicació en diverses entitats socials de la capital catalana”.

Titular del servei informatiu per assistents de veu:

“Esquerra Republicana proposa a Rubén Wagensberg que ocupi un lloc destacat a l'ajuntament de Barcelona”.

Context:

Wagensberg és un activista, polític, professor i productor musical de Sants nascut al 86 que ara està com a diputat per Esquerra al Parlament de Catalunya. És molt conegut per la seva implicació en moviments socials de Barcelona; de fet, va ser l'impulsor de la campanya «Casa nostra és casa vostra», que va promoure la gran manifestació «Volem acollir» ara farà un parell d'anys.

Ampliació:

Amb aquesta proposta, Esquerra Republicana intenta reforçar les seves files de cara a les properes eleccions municipals a Barcelona, que tindran lloc a finals de maig. Wagensberg és un perfil d'activista social que podria esgarrapar vots de Barcelona en Comú, el partit d'Ada Colau.

Notícies relacionades:

- *Maragall es confirma com a candidat a les primàries d'ERC per encapçalar la candidatura de Barcelona a les municipals*
- *Colau està disposada a fer un debat electoral a la presó de Lledoners si finalment Forn és candidat pel PdCat a l'alcaldia de Barcelona*
- *Els diputats Carrizosa, de Ciutadans, i Wagensberg, d'Esquerra Republicana, s'enganxen al Parlament per l'ús de termes com “supremacistes” o “feixistes”.*

6.2. Model d'arbre de conversa

Identificades les peces de continguts que haurem d'elaborar, cal treballar ara els textos que presentaran el servei, oferiran ajuda o informaran a l'audiència d'algun error en el servei. Aquests textos d'acompanyament són especialment rellevants en un servei com aquest que és un model de consum nou per a tots els usuaris, i on per tant és important que reforcem les comunicacions pedagògiques i el feedback de l'estat del servei, perquè l'usuari no tingui en cap moment la sensació d'haver-se perdut, de no entendre què està passant o de no saber com continuar.

Amb aquesta filosofia, introduïrem els següents textos de servei:

Missatge d'ajuda / introducció. Aquesta locució s'escoltarà el primer cop que l'usuari obri la *skill*, o cada vegada que la invoqui utilitzant el terme «Ajuda».

- *Hola! Et porto els titulars de les notícies més destacades del moment, acabats d'elaborar per les redaccions d'informatius de TV3 i Catalunya Ràdio. Si en algun moment vols ampliar els titulars que t'ofereixo, pots interrompre'm i demanar-me que t'expliqui el context de la notícia, que t'ampliï la informació amb més detalls, o que et doni altres notícies relacionades. Només cal que em diguis «Context», «Ampliació» o «Relacionades». I si vols que continuï amb la lectura habitual dels titulars, pots dir-me «Titulars» o «Continua». Si vols sortir del servei, digues «Suficient, gràcies». Pots tornar a sentir aquest tutorial de nou en qualsevol moment dient «Ajuda».*

Careta d'entrada. Aquesta locució anirà precedida de la marca sonora del servei, i s'escoltarà com a preludi als titulars del dia.

- *Bon dia, avui és 12 de gener i aquestes són les notícies.*

Careta de tancament. Aquesta locució anirà procedida de la marca sonora del servei, i s'escoltarà un cop finalitzi la lectura dels titulars del dia – o bé un cop finalitzi la lectura de la informació addicional demandada per l'usuari, i sempre que l'usuari no hagi fet cap nova demanda. Aprofitem aquest tancament per donar un to marcadament informal i còmplice, de proximitat amb l'usuari.

- *Amb això estàs al dia per ara. T'espero en una estona, amb més notícies.*

Missatges d'error. Aquestes locucions s'oferiran en cas que hi hagi algun impediment tècnic que impedeixi oferir el servei. Identifiquem dos tipus principals d'errors: un error de comprensió d'avant d'una ordre donada per l'usuari, o un error de falta tècnic de falta de dades o de connexió.

- *Sembla que tinc algun problema de connexió i no puc oferir-te les notícies ara mateix... Sisplau, torna-ho a intentar més tard.*
- *Perdona, no he entès què volies dir-me. Pots repetir-ho? Si encara no t'entenc, demana'm «Ajuda».*

La combinació d'aquests missatges de servei i dels continguts informatius ens permet crear ja una recreació de com hauria de ser l'arquitectura del nostre arbre de conversa, que es pot consultar a: <https://sitemap.mockflow.com/view/Mf798bce174595f512fe08a617207e62b1547313554526>. Per a posar a prova la interacció de la conversa, hem preparat una simulació escrita del diàleg basada, a mode de test, amb Facebook Messenger. La simulació es pot consultar en aquesta adreça web: <https://www.messenger.com/t/1125538687485749>.

Al tractar-se d'un model experimental, però, la validació final no la podrem fer fins que no posem el servei en marxa i el provem amb usuaris reals.

6.6. Tests qualitius, mètriques quantitatives i avaluació

Com a projecte pilot, és importantíssim que puguem tenir accés a l'analítica suficient com per avaluar les diferents característiques del servei – i incorporar així correccions en posteriors iteracions de desenvolupament. Per això definirem unes fases de test i avaluació que s'han de realitzar en dues fases: prèviament a l'estrena del servei, i un cop aquest ja sigui accessible per l'audiència.

Un cop tinguem la primera versió de la *skill* desenvolupada, abans de posar-la a disposició de l'audiència preveiem una fase de test amb un grup reduït d'usuaris d'uns quinze dies. Amb aquest període de temps podem donar marge perquè els usuaris no només provin el servei, sinó que l'integrin de manera natural a les seves rutines informatives – i obtenir feedback així no només de la qualitat d'una primera experiència, sinó de la utilitat a mig i llarg termini del servei, així com de la qualitat en diferents tipus d'entorns i escenes.

Per fer el pilot, buscarem un grup de control de 10 usuaris tipus que encaixin amb la nostra *persona*, i els hi demanarem que s'instal·lin la *skill* al seu altaveu intel·ligent i en facin ús segons les seves demandes naturals durant aquells 15 dies. Passat aquest període, recollirem dades qualitatives basant-nos en:

- Entrevistes personals en profunditat amb el grup de control
- L'històric de les seves converses amb la nostra *skill*. Aquesta informació no és accessible

pels desenvolupadors de *skills*, però sí pel propi usuari propietari d'un altaveu intel·ligent. Amb estricta confidencialitat, i en pro de millorar el servei, els hi demanarem que ens cedeixin aquest històric per poder analitzar la qualitat de les interaccions.

Amb la informació recollida aplicarem les correccions necessàries sobre el prototip d'*skill*.

Un cop el servei estigui disponible pel gran públic, passarem a una fase de recollida de dades més quantitativa, basada ja en l'estadística de tots els usuaris del servei. Les mètriques que ens interessa analitzar tenen a veure amb d'avaluació de diferents aspectes: l'interès dels continguts, la qualitat de l'experiència d'interacció, i la fidelitat respecte la marca. Veiem quins indicadors podem observar en cada cas:

Interès dels continguts

- Top Continguts. Per quines notícies es demana més informació addicional?
- Consum de l'ampliació. Quin percentatge d'usuaris demana ampliació d'una notícia?
- Consum del context. Quin percentatge d'usuaris demana ampliació d'una notícia?
- Consum de les notícies relacionades. S'escolten les 3? Només 1?
- Percentatge de consum del butlletí base. Els usuaris s'escolten tot el butlletí base, les 6 notícies core, o paren la locució a mig servei? Fins a on arriben?

Qualitat de l'experiència d'interacció

- Quin és el percentatge d'usuaris que interlocuten amb el servei?
- Quin és el percentatge de peticions que l'assistent no es capaç d'interpretar?
- Quin és el punt de sortida més habitual del servei? El final de la locució de les notícies base? Les notícies relacionades? La secció d'Ajuda?

Fidelitat cap a la marca

- Usuaris únics. Quantes persones fan ús del servei?
- Sessions. Quantes vegades es fa ús del servei?
- Durada mitja per sessió. Quant de temps es manté un usuari consumint continguts en una mateixa sessió?
- Freqüència del consum. Amb quina regularitat torna un mateix usuari a consumir informació?

Per un servei estabilitzat, n'hi hauria prou amb consultar aquesta analítica un cop al mes. En

aquest cas, que es tracta d'un pilot experimental, proposem l'anàlisi un cop per setmana durant els primers dos mesos – de manera que els equips de desenvolupament puguin respondre amb agilitat si cal fer alguna correcció severa. Un cop estabilitzat el model, es pot realitzar l'anàlisi de manera mensual fins al final de la temporada al juny del 2020. Arribat a aquest punt, caldrà fer una avaluació general del pilot i replantejar, si esau, el possible futur del servei.

7. Conclusions

Aquest treball final de màster d'aplicació professional ha nascut de la necessitat de diversificar les plataformes digitals on la CCMA serveix contingut informatiu, per tal de donar resposta a la frenada en el consum de notícies que ha tingut lloc a xarxes socials en el darrer any, especialment pel que fa al consum provinent de Facebook, que està experimentant un canvi de cicle. En aquest context, l'alça en la venda i l'ús dels altaveus intel·ligents amb assistents de veu s'ha vist com una oportunitat per explorar un nou entorn de distribució informativa, aliniada al mateix temps amb dos dels grans objectius de la missió de la CCMA: donar servei allà on els usuaris consumeixen informació, i ser líder en innovació informativa en català.

Són molts els mitjans que s'han llençat a desenvolupar continguts informatius per aquests entorns, especialment en els mercats de parla anglesa – però també, més recentment, a Espanya. El nostre anàlisi d'aquestes iniciatives posa de manifest que malgrat no existeix una línia estratègica clara i definida en el sector, tant diaris com ràdios, televisions i agències de notícies han fet un pas endavant posant a prova diferents formats, sent els més habituals els resums de titulars *text-to-speech*, especialment en el cas dels diaris i les agències de notícies, i l'oferta de butlletins radiofònics i sumaris de telenotícies, directament extrets dels directes d'emissions de televisió i ràdio. Els mitjans que han fet una aposta més clara d'apatació a la interfície conversacional dels assistents de veu han optat per oferir continguts interactius, on l'usuari pot aprofundir en determinades notícies a petició, o bé realitzar qüestionaris per posar a prova el seu coneixement sobre l'actualitat. En aquest sentit, creiem que aquests formats inicials evolucionaran en els propers anys per apropar-se a serveis molt més personalitzats, amb narratives creades a mida i més properes al periodisme conversacional.

Un dels principals valors d'aquest TFM ha estat perfilar l'usuari tipus pel servei informatiu que la CCMA vol oferir en aquests entorns. Aquesta *persona* arquetípica respòn al perfil d'un usuari entre 35 i 49 anys, amb renda mitja-alta, amb interès general per la informació i un consum habitual de notícies en entorns digitals, especialment a webs i xarxes socials. La seva principal motivació per consumir notícies amb assistents de veu és la comoditat i rapidesa que això suposa, i està interessat en poder fer puntualment, quan ho consideri, un aprofundiment en temes que l'interessen. La seva principal reticència és que l'assistent no entengui el que li demana, o que la interacció sigui llarga i tediosa. El nostre usuari no té molta experiència en la interacció amb assistents de veu, així que hem d'optar per un servei que sigui molt accessible.

Aquestes necessitats ens porten a identificar el primer repte important a l'hora de definir serveis

informatius en aquests entorns: hem de posar molt de focus en la claredat i la simplicitat a l'hora de definir les nostres interaccions, de manera que evitem al màxim la possible frustració davant d'una incomprensió. És vital oferir en un primer moment unes instruccions d'ús de manera senzilla i pedagògica, oferint a l'usuari una entrada amable al consum – i a partir d'aquí, fer un seguiment exhaustiu de les analítiques, combinat amb entrevistes a un petit grup de control, per recollir informació que ens permeti corregir o reforçar determinats aspectes del servei. En aquests entorns d'innovació on no hi ha estàndards definits encara, l'estratègia de desenvolupament a aplicar ha d'estar necessàriament basada en la iteració del cicle experimentació – avaluació – correcció.

És important recalcar que en cap dels estudis consultats ni en les enquestes amb els usuaris potencials que hem emprat en aquest TFM s'ha detectat una voluntat o petició explícita de *continguts* informatius diferents als que ofereix a data d'avui TV3 i Catalunya Ràdio. El focs sempre ha estat en el canvi en el model d'entrega, en la nova manera d'interactuar amb aquestes notícies, però no en el contingut o l'abordatge editorial de les notícies en sí. Això ens porta a concloure que per aquest pilot inicial podem aprofitar perfectament la feina realitzada a diari per les redaccions d'informatius, sense haver de crear una agenda de continguts específica per aquest nou entorn. Això suposa un estalvi potencial dels recursos elaborats a la producció de les peces, que es poden enfocar totalment en treballar el canvi en el model d'interacció.

Per últim, el segon gran repte que hem identificat és la definició de la personalitat de la marca en aquestes noves interfícies. Pels mitjans de comunicació, la creació d'aquests nous diàlegs directes amb la seva audiència suposa la construcció d'una nova veu que marcarà la percepció de la marca, com l'ha marcat en les darreres dècades la seva identitat i presència gràfica. Aquesta nova personalitat vindrà definida tant pel llenguatge i el to amb que s'expliquin les notícies, com per elements més humans com són el gènere, l'entonació o l'accent. La CCMA haurà d'abordar també sens falta aquesta caracterització de la marca si vol seguir tenint una presència coherent amb la resta del seu desplegament.

8. Bibliografia

- Amunwa, J. (s.d.). The UX of Voice: The Invisible Interface. Recuperat 17 desembre 2018, de <https://www.dtelepathy.com/blog/design/the-ux-of-voice-the-invisible-interface>
- Barot, T. (2017). The future of news is humans talking to machines. Recuperat 3 desembre 2018, de <http://www.niemanlab.org/2017/09/the-future-of-news-is-humans-talking-to-machines/>
- Borchert, K. (2018). Parlez-vous Deutsch? Rhagor o Leisiau i Common Voice. Recuperat 17 desembre 2018, de <https://blog.mozilla.org/blog/2018/06/07/parlez-vous-deutsch-rhagor-o-leisiau-i-common-voice/>
- CCMA. (2014). Missió i principis de la Corporació Catalana de Mitjans Audiovisuals. Recuperat 3 desembre 2018, de <https://www.ccma.cat/corporatiu/missio-i-principis/>
- Engleson, S. (2017). The Future of Voice From Smartphones to Smart Speakers to Smart Homes. Recuperat 30 desembre 2018, de <https://www.comscore.com/Insights/Presentations-and-Whitepapers/2017/The-Future-of-Voice-From-Smartphones-to-Smart-Speakers-to-Smart-Homes>
- Fisseha, M. (2018). Now You Can Sell Consumables to Enrich Your Voice Experience and Make Money with Alexa Skills : Alexa Blogs. Recuperat 30 desembre 2018, de <https://developer.amazon.com/es/blogs/alexa/post/aab44ff4-c19a-4862-8014-35c0e8846743/introducing-consumables-a-new-way-to-make-money-with-your-alexa-skill>
- Grup Carles. (2019). Coneix la nova Llei orgànica de protecció de dades i garantia dels drets digitals | Grup Carles. Recuperat 20 gener 2019, de <http://www.gcarles.com/nova-llei-organica-de-proteccio-de-dades-i-garantia-dels-drets-digitals/>
- Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior*, 90, 343-350. <http://doi.org/10.1016/J.CHB.2018.08.009>
- Hazard, L. (2018). "Here's what else you need to know today": The New York Times launches a flash audio briefing and other voice stuff for Alexa » Nieman Journalism Lab. Recuperat 19 gener 2019, de <http://www.niemanlab.org/2019/01/heres-what-else-you-need-to-know-today-the-new-york-times-launches-a-flash-audio-briefing-and-other-voice-stuff-for-alexa/>
- Heltzman, A. (2017). How Popular is Voice Search? Recuperat 17 desembre 2018, de <https://www.highervisibility.com/blog/how-popular-is-voice-search/>
- Jordan, E. (2017). Future of Retail 2017 - The Rise of the Connected Consumer. Recuperat 30 desembre 2018, de <https://www.walkersands.com/walker-sands-2017-future-of-retail-report-the-rise-of-the-connected-consumer/>
- Keynes, M. (2018). Strategy Analytics: Smart Speaker Shipments Hit Record Level in Q3 2018 as Installed Base Approaches 100 million | Strategy Analytics Online Newsroom. Recuperat 30 desembre 2018, de <https://news.strategyanalytics.com/press-release/intelligent-home/strategy-analytics-smart-speaker-shipments-hit-record-level-q3-2018>
- Kleinberg, S. (2018). 5 ways voice assistance is shaping consumer behavior. Recuperat 20 desembre 2018, de <https://www.thinkwithgoogle.com/consumer-insights/voice-assistance->

consumer-experience/

- Levine, B. (2018). Amazon VP reaffirms: 'No plans' for ads on Alexa devices - MarTech Today. Recuperat 30 desembre 2018, de <https://martechtoday.com/amazon-vp-reaffirms-no-plans-for-ads-on-alexa-devices-228928>
- Marketing Directo. (s.d.). El uso de los asistentes de voz contempla un crecimiento exponencial en los próximos años | Marketing Directo. Recuperat 6 gener 2019, de <https://www.marketingdirecto.com/digital-general/digital/el-uso-de-los-asistentes-de-voz-contempla-un-crecimiento-exponencial-en-los-proximos-anos>
- McDonald, A. (2018). BBC launches first children's voice experience. Recuperat 17 desembre 2018, de <https://www.digitaltveurope.com/2018/09/03/bbc-launches-first-childrens-voice-experience/>
- Mohan, V. (s.d.). Designing the Invisible. Recuperat 17 desembre 2018, de <https://uxplanet.org/designing-the-invisible-260291a02b71>
- Moreno, G. (2017). Asistentes de voz, nunca en público. Recuperat 17 desembre 2018, de <https://es.statista.com/grafico/9445/asistentes-de-voz-nunca-en-publico/>
- Munster, G., & Thompson, W. (2018). Annual Smart Speaker IQ Test. Recuperat 6 gener 2019, de <https://loupventures.com/annual-smart-speaker-iq-test/>
- Newman, N. (2018). *The Future of Voice and the Implications for News*. Oxford. Recuperat de https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-11/Newman - Future of Voice FINAL_2.pdf
- Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Kleis Nielsen, R. (2018). Reuters Institute Digital News Report. Recuperat 13 gener 2019, de <http://www.digitalnewsreport.org/>
- Pew Research Center. (2017). Nearly half of Americans use digital voice assistants, mostly on their smartphones. Recuperat 17 desembre 2018, de <http://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/>
- Reason Why. (2018). La voz recupera su protagonismo. Recuperat 17 desembre 2018, de <https://www.reasonwhy.es/actualidad/digital/uso-y-penetracion-asistentes-digitales-de-voz-en-espana-2018-03-15>
- Scott, C. (2018). Hey Google, how can publishers engage audiences using voice-activated devices? Recuperat 17 desembre 2018, de <https://www.journalism.co.uk/news/hey-google-how-can-publishers-engage-audiences-using-voice-activated-devices-/s2/a719237/>
- Speak Easy: Global Edition - JWT Intelligence*. (2017). *J. Walter Thompson Intelligence*. Recuperat de <https://www.jwtintelligence.com/trend-reports/speak-easy-global-edition/>
- Sterling, G. (2018). Amazon's Alexa finds its business model: in-skill purchases - Marketing Land. Recuperat 30 desembre 2018, de <https://marketingland.com/amazons-alexa-finds-its-business-model-in-skill-purchases-248948>
- The Smart Audio Report*. (2018). Recuperat de <https://www.edisonresearch.com/the-smart-audio-report-from-npr-and-edison-research-spring-2018/>

- UOC. (s.d.). Design Toolkit: Diseño centrado en las personas. Recuperat 17 desembre 2018, de <http://design-toolkit.recursos.uoc.edu/es/disenio-centrado-en-las-personas/>
- Vocento Media Lab. (2018). Contexto, proceso y dudas en la creación de “Las Noticias de ABC”, el nuevo formato de boletín.... Recuperat 19 gener 2019, de <https://medium.com/@VocentoLab/contexto-proceso-y-dudas-en-la-creación-de-las-noticias-de-abc-el-nuevo-formato-de-boletín-c4fa779514f9>
- Wakabayashi, D., & Wingfield, N. (2018). Alexa, We're Still Trying to Figure Out What to Do With You. Recuperat 17 desembre 2018, de <https://www.nytimes.com/2018/01/15/technology/virtual-assistants-alexa.html>
- We are social, & Hootsuite. (s.d.). Digital in 2018 in Southern Europe Part 1 - West. Recuperat 20 gener 2019, de <https://www.slideshare.net/wearesocial/digital-in-2018-in-southern-europe-part-1-west-86864268>
- Webb, A. (2017). AI Is Journalism's Next Big Threat (or Opportunity). Recuperat 17 desembre 2018, de <https://niemanreports.org/articles/ai-is-journalisms-next-big-threat-or-opportunity/>
- Wikipedia. (s.d.). ELIZA Effect. Recuperat de https://en.wikipedia.org/wiki/ELIZA_effect
- Yurieff, K. (2017). Voice is replacing typing in online search and other internet trends. Recuperat 30 desembre 2018, de <https://money.cnn.com/2017/05/31/technology/mary-meecker-internet-trends/index.html>
- Zanuldin, K. (2018). Make Money with Amazon Pay for Alexa Skills (Now Generally Available) : Alexa Blogs. Recuperat 5 gener 2019, de <https://developer.amazon.com/es/blogs/alexa/post/64090337-7957-419a-a8fd-b3fb15d30760/amazon-pay-for-alexa-skills>

9. Annexos

9.1. Enquesta a usuaris per completar el perfil tipus de *persona*

L'enquesta per perfilar un usuari potencial del servei es va distribuir entre un grup de 18 persones, de les quals **13** corresponien a la franja d'edat entre els 35 i els 49, que és la que vam utilitzar finalment per aquest projecte. Les respostes dels altres 5 enquestats s'han descartat.

El resum de les preguntes i les respostes que vam obtenir per la franja triada van ser:

- *Vius sol?* Identifiquem si l'altaveu s'emprarà en un entorn on viuen altres persones.
 - *No:* 69.2%
 - *Sí:* 30.8%
- *Ocupació. A què et dediques?*
 - *Arquitecte i músic*
 - *Auxiliar Administratiu*
 - *Enginyer informàtic*
 - *Gestor d'equip comercial*
 - *Infermer*
 - *Periodista (4)*
 - *Professora (2)*
 - *Productora i docent*
 - *Responsable de disseny digital*
- *En quines de les següents plataformes consumeixes notícies de manera habitual? Amb aquesta pregunta volíem comprovar si hi havia consum de notícies en entorn digital, i fins a quin punt hi havia un ús habitual de determinades plataformes.*

Consum de notícies per plataforma

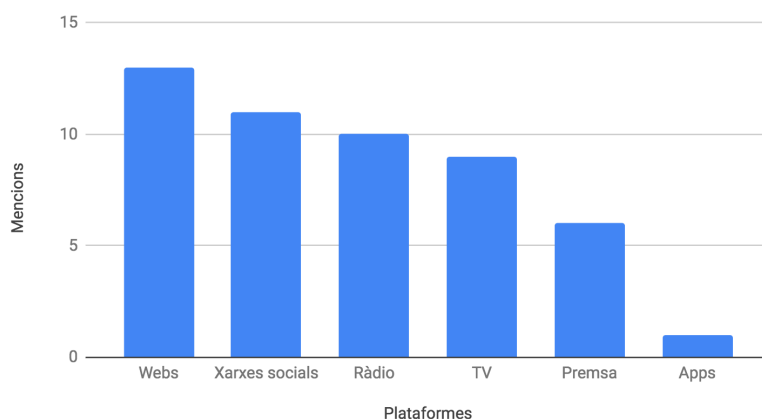


Figura 1: **Consum de notícies per plataforma.** Font: elaboració pròpia

- **Quines capçaleres informatives consumeixes? Citen 5, màxim.**

Mencions per mitjà

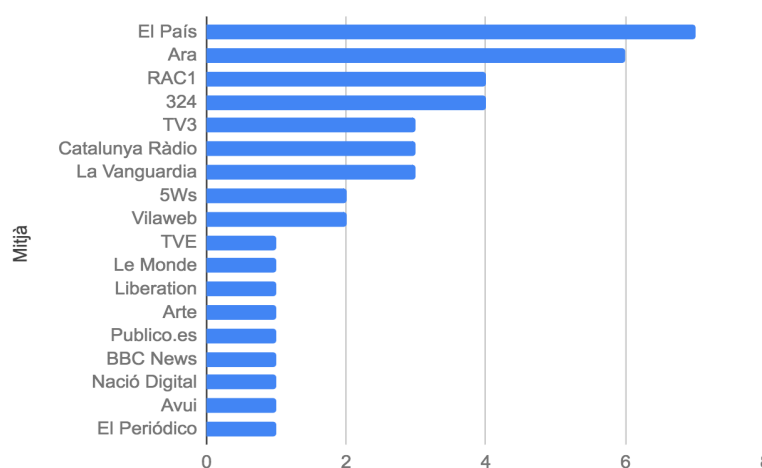


Figura 2: **Consum de marques informatives.** Font: elaboració pròpia

- **Quina seria la teva principal motivació a l'hora de consumir notícies a assistents de veu?**
 - *És tan còmode!*
 - *Realment no em motiva gens*
 - *Sentir els butlletins actualitat i poder demanar informacions de diferents països ràpidament*
 - *Política. Ciència. Història i Arqueologia. Educació. Cultura.*
 - *Rapidesa*
 - *Comoditat*
 - *Estar al dia i poder preguntar pel context de la notícia (ampliar informació)*

- *Immediatesa, comoditat i especificació (notícies temàtiques, internacionals...).*
- *Estar informat, per exemple, quan no puc mirar el mòbil o el portàtil. Un cas seria mentre cuino o mentre em dutxo*
- *Rapidesa*
- *Preguntar per dades concretes*
- *Si estic fent una altra cosa que em té ocupades les mans. Per exemple: conduir o cuinar.*
- *Poder escoltar les notícies mentre faig altres tasques*
- *Quin seria el teu principal impediment a l'hora de consumir notícies a assistents de veu?*
 - *Que he de buscar les fonts de notícies*
 - *No poder controlar el criteri de cerca, ni poder decidir entre vàries opcions simultànies, atès que hauria d'esperar necessàriament a que acabés cada locució.*
 - *Que m'entengui...*
 - *Que sota cap circumstància empraria un assistent de veu pel risc a la privacitat que representen i per la impracticitat que suposen.*
 - *Actualització de la informació*
 - *Idioma ??*
 - *Que no em doni la info que busco perquè no entèn les meves preguntes*
 - *En principi cap.*
 - *Crec que la connexió a Internet, perquè a casa és inestable, però sí anés bé, crec que no tindria cap impediment*
 - *Mandra*
 - *Que no estaria acostumat*
 - *Privacitat: no m'agrada tenir un dispositiu que sempre té una "escolta latent activa".*
 - *No tindria clar la seva fiabilitat*

Es pot trobar el formulari amb l'enquesta disponible a: <https://goo.gl/forms/TBVsWaK541r2qcSC2>

Es pot trobar el recull brut de les respostes a: https://docs.google.com/spreadsheets/d/1_hnyD6O-ynO7LXqdso9R8cufV4YDpsINBFMIGHcFRv4/edit?usp=sharing