



THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE ■

Voice Control:

How can new audio technologies build trust and engagement?

Ville Kinnunen

Radio Producer, Finnish Broadcasting Company Yle

Visiting Research Fellow, Polis, Department of Media and Communications, LSE



Contents

Summary

1	What are voice assistants?	5
1.1	What is voice assistance used for?	6
2	What publishers are doing now	7
2.1	<i>The Evening Standard</i>	7
2.1	<i>The Guardian</i>	8
2.3	Sport Social	9
2.4	BBC	10
3	Distributing audio content in VAT	11
3.1	Deconstruct long pieces	11
3.2	Turn audio to text for search engines	13
3.3	Immerse your listener	14
3.4	What if The New Wave of Voice is just hype?	16
3.4.1	Talking to machines	18
3.4.2	Voice will win if listening becomes habitual	19
3.5	Public broadcasters' foothold	21

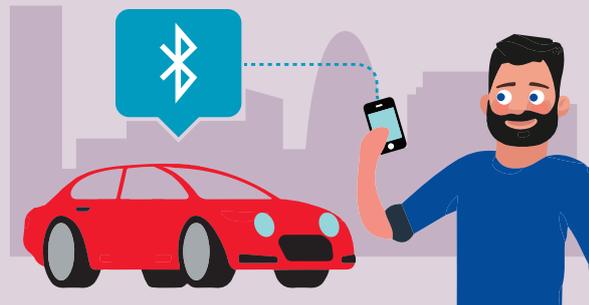
4	Young people and other “hard to reach” audiences	23
4.1	News moments during the day	24
4.1.1	Could AI help to find “news moments”?	27
4.2	Finding the right tone	28
4.2.1	Personality	29
4.3	But we cannot force the youth to come to us?	30
5	What will happen to radio?	31
5.1	Will a robot take my job?	31
5.2	Hourly news	32
5.3	Cherish trust	34
5.4	Boost your humanity	35
6	Conversational news assistants and editorial ethics	37
6.1	Who answers when you ask for more?	37
6.2	How much control over content can you give away?	39
6.3	Problems with material on third-party services	40
6.4	Voice can be stolen	41
6.5	Stick to the truth	43
7	References	44
8	Resources	47
9	Interviews	49



Summary

Voice is changing the way we use machines to access information. We already talk to several different kinds of machines to make them work for us. Talking is becoming the easiest way to command our devices due to the advance of Voice Assistant Technology or VAT, which means the ability of a machine or an application to understand natural spoken language and act according to our intentions.

Voice assistance on smart speakers, smartphones, smartwatches and in smart cars is opening up possibilities for voice-delivered news and audio content. Broadcasters and content producers may be able to use automation and machine learning to find the right forms of content and the right moments of day to reach and secure new audiences.



Audio content is already abundant, and in order to be found and consumed, it needs to be searchable, discoverable, relevant and entertaining.

People seem to be wearing headphones wherever they go. Voice assistance and automation with new technologies can be the way for traditional broadcasters to reach young people and other hard to reach audience segments. At the same time traditional broadcasters need to be aware of the constant changes of technology, platforms and forms of content in order to stay on top.



The new wave of audio is, however, not a shortcut to young people's hearts. Newsrooms are beginning to realise that younger generations use media mainly because they are bored, not just because they want to stay informed. Broadcasters and publishers have to make content so irresistible and easy to access that audiences will build new habits in consuming it with their voice assistants.

“ Human capabilities will, however, be crucial in audio and voice content production. Editorial jobs will be safe from automation as long as the work requires skills that only humans have, and if journalists remember to make the best of their human qualities. ”

This paper also touches on the editorial ethics of content on conversational voice assistants. Publishers need to find a balance between staying on the platforms and holding on to their own copyright and access to user data. We also need to be aware that the voice of a human can already be cloned at a very high level and potentially be used for malicious purposes, and it is only smart for publishers and broadcasters to be prepared.

This paper was written by Ville Kinnunen, a journalist at Finnish public broadcaster Yle as part of a Polis/Yle Newsroom Fellowship at LSE.

The editor was Professor Charlie Beckett, director of Polis.

The views expressed are those of the author

We welcome feedback to [**c.h.beckett@lse.ac.uk**](mailto:c.h.beckett@lse.ac.uk)



1 What are voice assistants?

The intimacy and personality of voice is growing as technology advances. The most familiar way of using voice assistance is on the smartphone, where Apple's Siri is the pioneer. If you already happen to have a smart speaker at home, you do not have to deliberately pick up your phone and unlock the screen; you just summon your assistant and chat away. The same goes for smartwatches, smart rings and even for the voice-enabled car.

According to NPR, 21 per cent of Americans over 18 years of age owned a smart speaker in December 2018 with a total of 119 million devices (NPR Smart Audio Report, 2019). Of those who had no smart speaker, 26 per cent are planning to buy one (Microsoft Voice Report, 2019). Of owners, 52 per cent had two or more speakers. Smart speakers may be found where the radio used to be: kitchen counters, night stands and living rooms, making sure that you have your assistant's attention day and night.

But is to *have* one also to *use* one? Increasingly, yes. The Reuters Institute for the Study of Journalism's research finds that 14 per cent of US adults say that they regularly use smart speakers. Regular usage in the UK (10 per cent) and Germany (five per cent) has "roughly doubled in the last year" (Newman, 2018).

But what if Amazon's Alexa or Google Home are not yet available in your language, as is the case with Finnish at the time of this report? Then your best bet is Apple's Siri, which has the longest track record of voice assistance in Finnish. Google can also handle basic commands in Finnish but the user experience loses to Siri. Apple's smart speaker, the HomePod, is yet to become available in Finland, so for now Finns will have to rely on their smartphones.



1.1 What is voice assistance used for?

In the Reuters Institute Digital News Project 2018, users identified four key benefits to using voice assistants: voice makes tasks quicker, smoother, more thrilling and more useful. This was consistent across all of the research population in the UK, USA and Germany. Some more advanced voice assistance adopters are also using smart speakers to make their homes smarter, controlling lighting, heating and security.

Primary uses are still the simple ones such as switching on the radio, checking the weather, playing music and searching the web with voice instead of using Google text searches (Newman, 2018).

So, voice is not far in its evolution from ordinary radio listening, but where is the listening to the news? Although around half of smart speaker users say they use the device for news, only 21 per cent in the UK and 18 per cent in the US *really do* use the news briefing functionality on a daily basis. (Newman, 2018):

“ Many users are unaware of the wider range of options around news, including how to access their favourite brand. Others are underwhelmed by existing content, which is mostly re-versioned from radio or print. ”

Newman, 2018: 6



2 What publishers are doing now

Publishers are constantly trying to think of ways for content delivery on voice. In addition to traditional voice publishers like the BBC, voice assistant technology is also in the interest of newspapers.

2.1 The *Evening Standard*

The *Evening Standard* in London has been running content in AI voice assistants since 2017. Their first take in that field was a simple guide “What to do in London,” which they quit after a year and went into audio news bulletins says Chris Stone, the executive producer of video & audio for the *Evening Standard*:

“At the end of 2018 we got involved with Google. They were looking at developing the content that they deliver on smart speaker voice assistant into a curatable feed of short form news items, and that is what we’re doing now.”

The drivers for the *Evening Standard* to venture into audio is the PR value, the potential to build up reach and to gain an early adopter status says Stone:

“It is a growth market and we have a limited period of time to become world leading expert in that market. We were one of a very small number of UK publishers that Google accepted aboard. We saw a really good opportunity to become expert in a new area before a lot of our competitors, to try and get a bit of first mover advantage.”

The *Evening Standard* sees potential to gain users across the Atlantic. They are thinking beyond their status as a large regional newspaper, and Stone says that it is particularly AI voice assistance that makes this growth scenario possible.



2.2 The Guardian

The *Guardian* had produced podcasts for many years, but they launched the *Guardian Voice Lab*, supported by Google, at the end of 2018. The aim was to experiment with interactive audio for the Google Assistant, bringing *Guardian* storytelling to a new platform. They brought together an interdisciplinary team from software and content production to explore the possibilities, says Dave O'Donnell, who was editorial lead on the six month long *Guardian Voice Lab* project that ended in April 2019:

“When the customer requested the Guardian briefing, our system would build an mp3 file, dynamically pulling specific stories from a daily newsletter and using text-to-speech for to serve the dynamic elements, combining the robot voice with template elements voiced by a human.”

O'Donnell is pleased with the combination of human and robotic voice. He has 10 years' experience of producing linear radio, and says that programming with SSML (speech synthesis markup language) can achieve a production level which aesthetically satisfies even the high level audio professionals who worked in the Voice Lab.

The quality was also recognised by the public. The *Guardian Voice Briefing* managed to gather an audience of thousands through marketing in the newspaper's email newsletter, O'Donnell says.

“Sometimes people commented on how the presenter seemed robotic or a little stifled. So, I think a lot of people didn't even realise that there was a robot voice at all. It was just around the time that Google had introduced a new Wavenet voice, which really improved the quality of it. People were mistaking the TTS for a human, not one that they particularly enjoyed listening to, but still they thought it was human.”

Dave O'Donnell is a little regretful about the short life of the *Guardian Voice Lab*. He says that the most useful user data and feedback began flowing in right towards the end of the project, but he is aware of the reasons for it being only an experiment.



2.3 Sport Social

Manchester-based Sport Social publishes football Premier League podcasts. They also have VAT applications by the same name, Sport Social, that offer a wide variety of both human output and machine-read Premier League content. Sport Social is produced by Voiceworks, a Communicorp (UK) subcompany.

Jim Salveson, head of Sport Social at Voiceworks, a UK voice technology company, describes Voiceworks as “a one-stop shop for people who are interested in the audio space” and who want to have their audio branding dealt with:

“As part of that we are also producing, predominantly in sport, a test case to see how easy it is to find an audience and monetise the traditional commercial way.”

Sport Social application, or “skill” in Amazon Alexa and “action” in Google Assistant, is, according to Salveson, the first of its kind in the world to offer daily news for every Premier League team, including match previews with the team’s current lineup:

“What we’re doing differently to other content providers is that we have the whole thing in human voice. We have broadcast journalists to record each of the bulletins individually so you get a human voice delivering them but not only that: we have also replaced Alexa’s voice with a human voice that feels just a little bit warmer, more natural and more conversational.”

Some parts of Sport Social’s daily content is, however, delivered by a robotic voice, and for obvious reasons. The Premier League has 20 teams with an average of 31 players available for each team, so that would mean having to record over 600 player names and numbers. That chore is avoided with robotic voice.

The initial reason why Voiceworks decided to build this kind of service around the Premier League is interesting. Salveson had noticed that, although there are a lot of sports podcasts and sport is good content for audio, the richer VAT content was lagging behind:



“ We used to run a radio phone-in show where people could phone up and talk to a former football professional about the weekend games, and the show would usually air 24 hours after the game. It was a format that worked brilliantly 10 years ago. Now, I have noticed that people were not that interested in talking to our guys about the game, because they can now have that gratification instantly via social media. ”

What he describes are clear signs of a change in both media and human behaviour. If something can be done right now, it will happen:

“ You can watch a football goal in the Premier League within 15 seconds on your smartphone because it would be on Twitter somewhere. You do not have to wait until 10 o'clock for Match of the Day to watch the goal replay or listen to the hourly sports news on radio, where you have to take the chance that there might be news about your football team or there might not. ”

2.4 The BBC

The BBC launched their first full voice skill for Alexa at the end of 2017, and it has been a success, says Mukul Devichand, executive editor for BBC Voice and AI:

“ I believe it's the largest custom skill in the world, and it basically is the centre point for all of the traffic for the BBC has in this space today. ”

In October 2019, the BBC launched a new skill for Alexa, which offers an interactive audio bulletin, the first of its kind in the UK, says Devichand. The user can skip forward to the next part of the bulletin but is also able to use the key phrase “Alexa, more from the BBC,” on some of the stories. The request for “more” will guide the listener to explainers and in-depth stories that are curated by the BBC editors.

The BBC are also developing their own voice assistant that goes by the working name *Beeb*, which is a nickname for the BBC in the UK. The BBC says that Beeb “will not be a hardware device in its own right but is being designed to work on all smart speakers, TVs and mobiles”.



3 Distributing audio content in VAT

As described earlier, publishers are very much aware of both the potential of reach but also the necessity of being involved in VAT. Nobody wants to risk missing out or getting left behind.

3.1 Deconstruct long pieces

Much of the content on offer at the moment is either short briefings or longer podcasts. The recently-launched BBC's interactive news service for Alexa might well be a glimpse of the future for other publishers as well, because the listener, or user, is to some extent in control of the length of the pieces, ie, the ability to skip to next and asking for more.

Length plays a very important role in whether a piece of audio is consumed or forgotten. In 2018, Newman found that short flash briefings are the most used and most suitable audio forms for smart speakers. Podcasts are not a big hit on smart speakers but very popular on the smartphone.

So how to prevent your precious podcast from slipping under TL;DL (too long; didn't listen)? According to David Caswell, Executive Product Manager of BBC News Labs, deconstructing and tagging content is the primary solution to enhance reach, discoverability and effectiveness. That is also the way that the BBC's interactive news skill for Alexa works. For that, the BBC is slicing up existing content:

“ We are taking the existing output and in a semi-automated way breaking that up into little pieces. Then, with some sophisticated tagging, we are able to reassemble that in different ways for different audience segments. But we are still more in the deconstruction stage than we are in the reconstruction. ”



Automation and the use of Artificial Intelligence, (AI) and Machine Learning, (ML), can be used to create audio content that will be more personal, meaningful and efficient to produce. The BBC have been working on a concept named Internet Fit Radio since 2016. They are developing the means to insert metadata into audio for it to become searchable and discoverable and therefore easier to distribute in a multitude of ways says Caswell:

“ Internet Fit Radio is about deconstructing all the parts so that they can be combined in different ways for different audience members and different audience segments. ”

Jim Salveson of Sport Social says that what he calls “the Golden Age of Podcasts” we now live in will move towards shorter pieces:

“ Voice will develop into people being able to listen to small sections of podcast rather than listen to half an hour podcasts. They will get to find the five fundamental minutes they want from inside the longer podcasts. It will be the second wave of the podcast. ”

Length is also a key factor, if and when newsrooms want in on people’s playlists. How that can be achieved, especially when newsrooms want their content into personal Spotify, Apple Music or Tidal playlists, is a technical issue. But within those playlists the news content must be in a correct rhythm with the music, and that means short pieces, such as 30-second briefings. Dave O’Donnell, who was the editorial lead of the Guardian Voice Lab project, says content must fit into people’s behaviour:

“ Gimlet are now re-versioning podcasts that previously were 20 to 40 minutes and lifting six–seven minute highlights, that are then put into the Daily Drive in Spotify. That is an attempt to cater existing content for another experience more specifically to the way people are living. ”



Spotify launched Your Daily Drive in June 2019 especially for commuters. It inserts the bits from podcasts into the users' musical playlist. This kind of cooperation became possible when Spotify acquired Gimlet Media in early 2019. Of course, agreements can and will be made without business acquisitions.

O'Donnell sees this kind of playlist integration to be a good way to reach audiences, because people are accustomed to constantly managing their own playlists. So having a publisher "help" in that will probably not annoy the users, especially when it is done at their will.

3.2 Turn audio to text for search engines

One key to success is discoverability. Audio content needs to be found just as easily as text, and that will happen soon, says Mukul Devichand, Executive Editor for BBC Voice and AI:

“ Google AI in short time will be able to search any audio it has access to dynamically because of speech to text technology. And that is the sort of disruption that will change a lot of audio publishing. ”

David Caswell says that the BBC's massive output has a lot more potential for getting attention:

“ Our huge daily output is far more than any audience member consumes. They miss enormous amounts of content that we produce every day that they would be rabidly interested in. They just don't know that it exists. We can use the interactive service to make audience members aware of what we have already produced that they can consume. There is years, years and years of good benefit to get from that. Even if it takes us some time to develop new techniques in the AI-based storytelling. ”



Dave O'Donnell, who developed voice content for the *Guardian*, sees Spotify as a significant place for informative audio to be discovered:

“*Playlist formatting that people have become very accustomed to through Spotify and elsewhere, is absolutely a way that broadcasters can bring audiences from those platforms to explore more on their own platforms.*”

3.3 Immerse your listener

The beauty of voice is that you do not have to invest money in visuals, because the images are formed inside the listeners' head. When I talk to my listeners in person they tell me, “you are the man from radio, right? You don't look at all like you sound”. Radio programmes about nature are among the most engaging in Finland, because of the rich imagery that is instantly formed with sounds.

That evocation of imagery is one of the possibilities for building engaging content to VAT. Podcast makers have already realised this. The rapidly increasing success of podcasts has improved the quality and awareness of sound design elements, such as different ambiances and sound effects, where even a subtle effect is enough, because podcasts are very often listened to on headphones.

Google has also realised the immersive potential. They have included sound effects into the technical requirements of audio in Google News (Google Developer Guidelines, 2019):

Creative elements

Be sure to make audio-first or audio-rich content that's meant to be listened to, instead of read. See the following points about creative elements:

- We recommend to include more than one voice in a story.
- When possible, use natural or ambient sound to enrich the story. For example, a report about wildfires might include the sound of a fire crackling.



Jim Salveson predicts that immersive audio content has huge potential. For example, the Alexa skill, Chompers, which is created to help children brush their teeth. When the user says, “Alexa, start Chompers,” the skill will immerse the teeth-brushing children in jokes and silly songs that will keep them brushing for the recommended two minutes.

As with many other applications and games, Chompers also has a “streak” or continuity challenge which is kept alive by regular brushings. Maybe that could also be a useful approach for news? By keeping the streaks alive in the news app the user could be rewarded.

Salveson makes the point that immersion can be better achieved with skills that are entirely human voiced instead of ones with mixed human and machine voicings:

“*In the future, you could hear children’s toy companies voiced by children and makeup companies voiced by women. Car companies will probably have male voices, you might have BMW voice their skill with a slightly German accent. Whatever it might be, I think it all becomes a part of that individual brand’s audio identity.*”

Once immersion is reached, it can also be broken fairly easily, warns Salveson:

“*The voice skills that focus on children are all about adventures or stories, taking them on a journey somewhere. What you don’t want then is to have a different voice, like Alexa’s voice, you want to make it as immersive as possible, so that they stay within the story.*”

But what about adults? They have the right to be immersed as well, says Salveson.

“*I guess you can kind of extract that out to adults, although we don’t interact quite the same way. If you have an immersive skill that feels like you are talking to one individual, if you have that broken by a different piece of content, or the sudden Alexa voice, it breaks your illusion and gives you that opportunity to leave the skill or do something else.*”



For the time being, Salveson sees human voice as the only way to a truly immersive audio experience.

“ *In order for this voice-first world to reach its full potential, you need to be able to feel a conversation, to deliver emotion, to deliver emphasis in the right places. That could be through human voice, which could mean that people have to revoice every piece of content. But certainly at the moment the voice delivery within the system becomes much smarter with a human voice than an automated voice.* ”

One of the beautiful aspects of radio is that you can listen to it while doing something else. In Finland it is very common to listen to FM radio that is embedded into hearing protection while working with heavy machinery or doing other noisy jobs. Australian research (Shenker, 1999) finds that FM earmuffs are a good way to increase worker comfort and efficiency.

The immersive aspect of audio can help to concentrate on the work, and there are already people who have made it a habit to listen to talk radio while doing repetitive and flow-demanding work such as coding.

3.4 What if The New Wave of Voice is just hype?

Are we looking at yet another wave of technological hype? Critics point out that while Amazon and Google are selling their speakers in large volumes, they keep very quiet on the data for the actual use of the speakers (Blackie, 2019).

My view is that voice does have a better chance in bringing added value to news than, say, virtual reality. That is because voice assistant technology is more affordable; many people have it ready in their smartphones; speech is natural for humans in communication; and there is a massive body of present and archived content for voice news. And VAT does not require you to wear a pair of clumsy goggles.

But what if voice really does fail? What if radio really does die and does not get replaced by voice assistants? Mukul Devichand from the BBC is cautiously optimistic:



“ I don’t know for sure, but certainly the big tech companies seem to be betting on [voice] being as big a change as these other platforms. They’re spending a lot of money on it and giving away devices, they obviously think that this is going to be a big deal. And that means that every publisher, and certainly a large public service broadcaster, needs to be aware of this change, to make sure that the public broadcasting system is ready for this change and can preserve its values and its audiences. ”

Devichand emphasises that voice assistants must not be merely seen as the New Radio but a much bigger change:

“ AI assistance generally is not just this year’s or next year’s change. This is a shift in the underlying ecology of the internet, and voice assistants are an audible element of that shift. ”

Dave O’Donnell, formerly editorial lead on the Guardian Voice Lab, says a future where we talk less to our devices is difficult to imagine:

“ Voice has the potential to be fundamentally disruptive to how print based media works. Therefore, they need to have a presence in voice assistant technology spaces, in order to be ready for the future. But, for example, a lot of money was ploughed into video with the promise that video was where everything was going to be, so I think perhaps there is a cautiousness with which news organisations are now pursuing voice. ”

O’Donnell says that the next generation has already learned to control their devices with voice, which is a very good sign for voice technology. But how will we know that it has happened? When will it be safe to say that VAT is a “big thing”?



3.4.1 Talking to machines

We live in a time of voice control natives. Many people, especially children, speak to their smartphones, tablets and smart speakers as naturally as they would speak to other people. Teenagers may speak even more naturally to machines than they do to adults, because the machine is not an authority and, for the time being, the devices give little backchat. Voice is becoming more and more popular in searching, as predicted already a couple of years ago (Olson, 2016):

“ With voice search, spoken language connects people to what they’re searching for with an immediacy, convenience and intimacy that text-only search could never provide. ”

Voice commands are used with what is referred to as Voice Assistant Technology or VAT, which means the ability of a machine or an application to understand natural spoken language and act according to our intentions. Smart devices constantly listen to hear wakewords, such as “Alexa” or “Hey Siri”. After the wakeword, the device listens to the user speak and converts spoken audio into text, which is then processed by Natural Language Understanding (NLU) to find out what the user wants. After that, a Dialogue Manager decides what, if anything, needs to be said back to the user. Finally, the response is worded out by a text-to-speech generator (Breslin, 2019).

People are increasingly using voice to make their machines do things, especially as it often is the easiest way. And at this early stage of voice control, there is also an element of novelty to it. That is also why Siri, Alexa and friends are exciting to use, as Jonathan Follett, software designer and tech author puts it:

“ The user interface is flattened down – in the best case – to a single voice command. It will take just one or a few seconds, time during which your mind and body need not think about the request anymore (Follett, 2019). ”



But, as Follett adds, that ideal is not always reality:

“ Talking might be easy, but getting the correct results rarely seems to be. Talking with machines can cause us a great deal of frustration: Alexa, do this. Siri do that. No, don't do THIS, please do THAT instead (Follett, 2019). ”

Follett predicts things will change for the better as Natural Language Processing (NLP) advances. NLP is the way that computers process and analyse human speech and that suggests it is also the most important way that Artificial Intelligence (AI) communicates with us. As an example Follett takes Google Duplex, an AI service launched a year ago which makes telephone calls on behalf of its users, for example, to make reservations and appointments. Google Duplex is one of the most advanced AI tools available to the public, and Google are now expanding it gradually outside of the US, first in New Zealand (Google New Zealand blog, 2019).

With Duplex, Google promise to always tell the businesses that it is a machine that is calling them. Business can also opt out of having to talk with Duplex, which is understandable now, when this kind of technological interaction is still a new thing.

3.4.2 Voice will win if listening becomes habitual

A gadget can feel exciting when brought home, carefully unboxed and used for the first couple of times. But what prevents smart speakers and voice assistant applications from joining the long line of forgotten bread makers, egg boilers and sandwich grills?

The answer is habit. A machine is most probably purchased for a real or imagined need, which now needs to be fulfilled time and again for it to find real purpose. If the homemade bread is worse than supermarket bread, farewell bread maker. If the smart speaker isn't smart enough, good riddance.



People seek efficiency in their everyday lives. So, for audio assistance to become habitual, it has to be the easiest option. As noted before, the main reasons for voice use are speed, smoothness, thrill and utility. One additional factor seems to be the potential to reduce screen time. Newman (2018) finds that many people do not want to use screens after a long day working with computers and smartphones:

“Some resent the way in which the internet can distract and waste time by taking people down ‘rabbit holes’. Part of the appeal of voice devices is they act differently. They provide focused information when summoned and, for the moment at least, the lack of a screen means less distraction (Newman, 2018).”

Could audio assistance replace radio as a habit? Inside the home, the potential is evident and the use cases are more plentiful and interactive than with linear radio. That might also be true for the car, as the so-called connected car with seamless Bluetooth usage or other connectivity becomes more prominent. The easiest way to habitual use of voice assistance is the smartphone, because it is nearly always at hand, and VAT is also available on smartphones in countries where smart speakers are not yet available.

Jim Salveson, head of Sport Social at Voiceworks, compares VAT to smartphone apps when building new habits is considered:

“You downloaded [the apps] and had a visual reminder every time that you looked at it: ‘Hey I use that and it was cool, I will use that again!’ The only way to have that kind of user retention with voice is the strength of the content and by giving people reasons to come back on a regular basis.”



3.5 Public broadcasters' foothold

People getting used to voice assistant technology and developing habits are good signs for content publishers. But how do public broadcasters ensure that they are the first choice for voice delivered news?

Many smart speaker owners ask their voice assistant to play a particular radio station. In the UK 60 per cent of smart speaker owners said they have listened to live radio on the speaker, and in the US 41 per cent of respondents had done so (Newman, 2018). Live Digital radio listening at home grew 33 per cent from Q3/2018 to Q3/2019 in the UK, according to RAJAR (10/2019).

But, when VAT users specifically ask for news briefings instead of linear radio, they will have to choose from an abundance of sources. As Bullard (2019) points out, there are hundreds of flash briefings available for Alexa, and “there are even updates delivered entirely by synthetic voices”.

Presently, the most common way to get news from voice assistants is to ask a basic question like “Alexa, what’s the news?” That brings up a flash briefing, inside which the user is able to skip backward or forward or choose to cancel the briefing.

Google is testing a new “action” called News on Google Assistant that curates a stream of news from many different news sources. That mode of operation could be a major challenge for newsrooms aiming to become the preferred choice, because “[t]hese feeds cater to voice by removing the need for a listener to ask for a specific station or source each time they want an update” (Bullard, 2019).



Trusted brands will have an advantage in VAT-delivered news, and that is not only due to having been on the air “forever” as with the BBC and Yle, but also being available on the new platforms right from the start:

“ Broadcasters like the BBC and NPR in the US have had a significant advantage. Because of their leading offline position they were awarded prominent status on these platforms at the start. (...) Though a list of options is presented during setup, qualitative interviews also suggest that most users don’t know how to change these via the Google, Amazon, or Apple apps. In the UK, we also found little desire to change the default setting, which is BBC News on all three platforms (Newman, 2018). ”

Five key issues to think about when planning a voice strategy

- Audio needs to bring added value to access content
- Voice should be the easiest way
- Content creation requires audio professionals
- Audio content must be searchable (speech to text)
- Long pieces should be sliceable for playlist curation



4 Young people and other “hard to reach” audiences

Traditional news media no longer seem relevant to many people under 35 according to the Reuters Institute for the Study of Journalism:

“ Young people are less likely to go directly to news apps and websites, and spend much of their time online with social media and entertainment services. Against this background, it has proved hard to attract attention to traditional news content – which is often seen as a chore – or to news brands that often feel irrelevant to their personal lives. ”

Galan, L. Osserman, J. Parker, T. Taylor, M. 2019, p. 15

The institute’s research finds that news experience for young people needs to feel as easy and accessible as Facebook and Netflix. That means simplicity of use, but not dumbing down the content, which young people dislike. Additionally, news brands need to tell stories in ways that fit the moments when young people are open to the news. This means creating formats that are native to mobile and social platforms. Thirdly, coverage needs to change both in addressing news issues and how content and brands are present on third party platforms (Galan et al., 2019).

To interest youth, news need to be relevant, timely and in the right format, and presented with a tone that resonates and is trustworthy. But isn’t that true for audiences of all ages? Yes, but we need to remind ourselves here that younger people do not go to news as a habit in the same way as previous generations. To build habits, the content needs to be worth their time even more than with older audiences.



That is also true for any other hard-to-reach, hard-to-habitualise audience segments. Age does not always define our ways to use media and to consume news. Therefore, we need to broaden our thinking beyond age and start figuring out how to be discoverable, relevant and, all the more often, entertaining. Mukul Devichand from the BBC:

“ I am convinced that we do need to have destinations that work for all audiences, including younger audiences. We’ve also got to think about how we bring people to those destinations and sometimes serve people where they’re at. Getting that balance right is going to be tricky. ”

4.1 News moments during the day

Many people spend all of their time with their smartphone, even when on the toilet or in bed. The latest waterproof phones even allow you to take them for a swim. We have seen Grand Slam level judo matches, where world class judokas have been disqualified after having a smartphone drop onto the tatami from inside their judo outfit (International Judo Federation, 2019).

The Reuters Institute’s research confirms that 18–34 year olds across Europe as well as in the US and Japan spend a large majority of their waking hours with their smartphone:

“ They use them for communication, for media, for games, for dating – and for news. Across all markets, our survey data reveal that the smartphone is the main device used for accessing news for the vast majority of under 35s (69 per cent). ”

Newman, N. Fletcher, R. Kalogeropoulos, A. Nielsen, R., 2019

The younger they are, the more they get the news from social media, messaging apps being the most used. People towards 35 years get their news mostly from directly accessing the publishers, but Facebook and messaging apps are also big players there.



Four typical key news moments for younger groups

DIRECT



DEDICATED

Finding time to focus on the news, like a novel or a TV series
Less common; suits evenings or weekends

Mindset: more introspective; deepening understanding.



UPDATED

Getting the key news updates you need efficiently
Suits mornings; preparing for the day

Mindset: more something I feel I need to do.



TIME-FILLER

Not about the news per se; something to do while doing something else

Constant: on the train, break, when time to fill

Mindset: more something I do to distract/amuse.



INTERCEPTED

A notification or message intercepts what was being done
Can happen anytime and anywhere

Mindset: passive recipient.

INDIRECT

The above illustration (Reuters, 2019: 55) is a good place to start thinking about how and where to usher our voice news content and in what sort of portions, ie, finding the news occasions.

One example of *interception* is breaking news. When something happens, the best way to publish that is to send a notification to smartphone lock screen. To get there, we need to make the user install our app and to allow



notifications. That applies also to content on third-party platforms, where we also are present, but in order to get to lockscreen for example with an Instagram notification, the user needs to be not only aware of us there but also to be subscribed to notifications about our content.

Those subscriptions and notification allowances can be achieved through advertising and education but also more organically by offering relevant content often enough, and here also we are hoping to make new habits. Then we have the chance to be upgraded to the *time-filler* step in that ladder.

When our content fills the users' time adequately often enough or resonates so deeply that it leaves a trace, we might gain a position in the *updated* ladder. That is a very good place to be, because we are then on our way to a proper habit and being the first choice for the user. Then, they might also come to us to consume other than news content at the highest ladder *dedicated*.

Newman (2018) finds that VAT is used during the whole day: smart speakers at home in the morning and in the evening, and across all devices at every time of day. He concludes that voice could have a much broader role in news consumption, but at the time of his research the news offer in voice did not meet the needs for most users. According to Newman, VAT is effective on short flash briefings but does not make the most of longer articles, podcasts, search requests or breaking news notifications: "Content in all these other formats needs to be better optimised for VAT to get similar levels of takeup [compared to flash briefings]."

The *Evening Standard* started their voice content creation with that same anticipation of consumption moments. Chris Stone found that the vast majority of the content is consumed in a two hour time window in the morning. That gives a good balance to their printed newspaper which is aimed at the evening commute says Stone:

“ So it's basically breakfast news. I think there is a clear use case for people waking up getting ready for work and then saying "OK Google, play me the news." ”



4.1.1 Could AI help to find “news moments”?

People are granting their devices access to a vast amount of information about their daily lives. Some do this willingly, some discover this data collection themselves by accident, some realise it by the advice of someone else, while others never access their gathered data or even know about it. One example is motion sensor data, or more commonly referred to as pedometer, as motion sensing is most often used to count the steps of the user during the day. Smartphones can also be set to monitor sleep with the aid of accessories like smart watches or smart rings.

So, could publishers use that data of motion and awakesness to more timely push content to users? The BBC’s David Caswell is doubtful:

“ You can totally use accelerometers in phones and all that to find new ways to distribute news to a small portion of the audience, who happens to want to get news in that way. But how do you do that sustainably in an editorial resource point of view? You would need to do the news for radio, the web, augmented reality, virtual reality, and things you have not yet thought of. Where do you stop? ”

Instead, Caswell suggests that we take the deconstructive approach:

“ The question is how can we produce news and create information about our society in a way that we can make it accessible to people as they want to access it? Somebody gets the news from the smartphone at work during the day, somebody gets it in a bulletin at night. Somebody gets it in little interstitial experiences a few minutes each with a snacking behaviour. Somebody else just gets notified on certain criteria. ”

Caswell assures us that good content will more likely be found if tagged properly and given all the correct metadata. He says that we need to accommodate multiple human behaviour modes during the day, even within the same individual, because people’s behaviour changes as the day goes and their energy levels change:



“ *The sustainability is in doing it once in little pieces and then reassembling it for the most behaviours.* ”

One day there might be a voice assistant that wakes up not only when asked, but once it senses that its owner is awake. Would that be useful or annoying? I believe we will find that out sooner than later, as somebody is bound to make that kind of functionality available soon on VAT.

4.2 Finding the right tone

The tone of voice in news content is something that the Finnish Broadcasting Company Yle has actively looked into only in the last three to five years. Now, the idea is starting to yield results. Tone and perspective, that is to say whether or not we talk to our audiences at their level, are especially important in reaching younger audiences.

Chris Stone of the *Evening Standard* finds the tone of voice is important in making the audience return to their content. Stone says that if people find it worthwhile to come back to the *Evening Standard's* content on voice assistants, there is a good chance that they will find other content through them as well.

The Finnish Broadcasting Company Yle have found that sports is often the way to reach audiences that normally would not come to their content. And this goes well beyond the question of event broadcasting rights. Web articles about ice hockey, Formula 1 and E-sports are very popular reads among men under 45.

Maybe voice content about sports could be one way into younger men's headphones? It certainly has the potential to be relevant and engaging, says Jim Salvesson of Sport Social:

“ *Emotion is one of the most important things particularly with sport: if you take emotion out of sport, you take a lot of the meaning out of it. Therefore, human voice is a key asset, because at the moment a lot of the content on voice assistance is delivered with absolutely zero emotion.* ”



Simplifying information in the wrong way by dumbing down the news is counter-productive, because you will get caught, with embarrassing consequences. Mukul Devichand from the BBC points out that young people are very smart:

“Both globally and in the UK you’re talking about a more educated population and a more well traveled group of people than with earlier generations. They understand what is going on in different parts of the world without even being there. Young people are also very keen on good quality and trustworthy advice in key areas of their lives.”

4.2.1 Personality

Children are often very heavy users of video content. They can easily spend hours daily watching videos; the younger ones on TikTok and teenagers more on IGTV and YouTube.

The appeal of YouTubers is largely based on their authenticity, human appeal and resonance with their audiences’ lives. So, is it too simplified to say that young people avoid public broadcasters’ content because they lack those qualities? Chris Stone from the *Evening Standard* says audio must match this “intimacy”.

“I think what TV and YouTube and podcasts teach us is that people respond well to hosts. The hosts of a YouTube channel or the hosts of a podcast are effectively leaders of a community, and they speak to and draw people into that community. That has definitely something to do with the intimacy of human voice that lends itself to trustworthiness.”

Personality and charisma are things that publishers and broadcasters must have though that is easier said than done.



4.3 But we cannot force the youth to come to us?

You might ask: “What’s all this fuss about news not reaching youth? They are citizens in their own right and will keep themselves informed”. But public service broadcasters both in the UK and Finland have an obligation to serve people of all ages to enable them to fully engage and participate in the democratic process.

Beyond those obligations outlined in our organisations’ charters and strategies, every newsroom wants to do better also in reach and ratings. Nobody wants to wither away with their audiences over time. We need not only to be available, but relentlessly search for new ways to engage audiences of all ages. And while we pay constant attention to the target audiences’ qualities, it automatically benefits more thought out, good quality content that provides added value to everyone, including the journalists who produce the news.

Four key tactics to improve the style of your voice content

- Make your audio immersive to grab attention and create retention
- Reward the headphone listener with subtle elements
- Use human voice whenever you can
- Go straight to the point or you will be skipped



5 What will happen to radio?

For as long as I remember, the death of broadcast radio has been forecast. First it was digital audio broadcasting (DAB) that was supposed to render FM radio useless. To this day that has only happened in Norway in 2017. Switzerland seems to have decided to end FM broadcasting by 2024, and rumours have it that the UK and Denmark would be next in queue. Along with DAB, 5G is also predicted to topple FM transmission towers, which are undeniably a very expensive way to broadcast.

Whatever the broadcasting technology will be, linear radio will not disappear. There will still be people talking, music playing and news told, even if it was aired on a pair of tins with a string between. The question here is how we can see beyond the broadcasting aspect and embrace the concept of making audio-on-demand? That means we need to know our audiences and their preferences much better than now. This will only be a change for the better, because we need to constantly know what we are doing, to whom we are doing it and how it will be best consumed.

5.1 Will a robot take my job?

Your job will be safe from automation as long as it requires abilities that only humans have. And if journalists make the best of their human qualities, they have nothing to worry about, says David Caswell from the BBC News Lab:

“ I have been working with and looking at this development for quite some time. Journalism and the fundamental editorial tasks such as the application of editorial judgement is not going to be even approximately automated anytime soon. I’m a very optimistic techno-utopian, but even I see no possibility in our lifetime of anything remotely like that happening. ”



His words find support in recent research:

“ [T]here is a great deal of discourse surrounding automation and technology, much of it filled with grandiose and apocalyptic predictions, even proclaiming the “revolution” if not outright “end” of specific industries such as journalism. ”

Lewis, S.; Guzman, A.; Schmidt, T. 2019: 14

Caswell says that journalism is one of the last artisanal jobs, where the handcrafting of specific artifacts is the normal mode of production. However, there are processes in journalism that can be and have already been automated, such as match reports in various sports and quarterly earnings of companies.

Finnish Broadcasting Company Yle’s Voitto the Robot autonomously produced local news on the Finnish 2019 parliamentary elections, and Voitto is continuing to produce sports reports daily. One of the reasons that Yle News Lab is using Voitto the Robot is to “save the journalists” time, to give human journalists more time to think and let mechanical tasks be done by a robot”, says Jukka Niva, the head of Yle News Lab.

Things that can not, at least at this point in time, be automated, are abstract things like story finding and storytelling. But the BBC’s Caswell sees possibilities in there too. He takes machine translation as an example. Translation is a process that cannot fully be understood even by professional translators, they just know how it’s done. But machines learn to translate by inspecting countless numbers of examples of translations, and step by step train themselves how to translate. Maybe that could be applied to journalistic storytelling too?

5.2 Hourly news

One repetitive task in radio is the hourly news, both regionally and nationally. The shortest ones that Yle Radio Suomi have are two minutes and the longest ones 10 minutes long. The production method is very simple: the anchor collects material from various sources, inserts it into the production system, edits the text and audio bits together to fit the desired time slot and then goes into the studio to anchor it live.



Should some of this be automated? Technically it would not be difficult. For the shorter news programs, the content is, nearly always without exceptions, text spoken by the anchor. Voitto the Robot may not be ready for producing sensible text from such a big variation of topics that daily news consist of, but maybe Voitto could anchor the news that humans have written for him?

David Caswell from the BBC News Labs would not introduce robotic voices into present day broadcast radio. He sees it as a violation between the product and the means of delivery, and that would not sound right. Caswell would leave broadcast radio for humans to make and develop automated, machine voiced content to web based radio and audio-on-demand.

In today's world of endless choices, the beauty and ease of radio is that you do not have to choose the content, you only have to choose the station. That is why some people prefer radio over Spotify or other streaming services, and on Spotify they choose curated playlists and "Artist Radios" instead of having to pick the songs themselves. That same kind of user need and behaviour may also apply to news consumption on voice assistants. It is more convenient to ask your platform of choice to "tell me the news" than to request "the latest on the Syrian refugee crisis, the updated schedule of Brexit or Arsenal's lineup tonight.

For radio news production, the above seems like an easy transition from linear radio to voice assistant content: all you have to do is make the linear broadcast searchable through speech-to-text and then apply tags to the metadata, so that the audio content can be curated into different feeds. It is high time to start experimenting this on existing audio content, to find out the right technology and shape the new, additional workflows in the newsroom. This is only wise in preparing for the inevitable transition in distribution and audience behaviour.



5.3 Cherish trust

Replacing human voices on radio would not be wise, if reputation and trust are threatened. Radio is the most trusted medium in Europe according to EBU 2019 research; 59 per cent of EU citizens trust radio and it is growing.

At the same time, 52 per cent of EU citizens trust the written press. In Finland trust in written press is at 69 per cent. In the UK, the trust in print media is very low, at 28 per cent; 27 per cent of Europeans trust the internet as a news source, and that trust continues to fall. News on social networks are least trusted (EBU, 2019).

What explains this considerable gap in trust between media types? The BBC's Mukul Devichand attributes it partly to institutional values:

“ I think that trust, in the case of the BBC, comes from almost 100 years of actually broadcasting radio. There are things that people have known for their lifetimes, so I don't think it's just because of the intimacy of the human voice. I think it's because we've been doing a great job for a very long time. ”

EBU research also finds that in the countries where public service media is trusted, the overall trust in the society is high. The BBC's David Caswell attributes the trust to voice and authenticity.

“ One of the bases of trust is being able to sense whether somebody is authentic or not. You know, it's sort of intuitive, where you make an assessment of a person based on whether they sound like they mean it. ”

The input-to-output ratio of hourly radio news can be further enhanced very easily. In the Yle Areena, the hub for the Finnish Broadcasting Company online, regional news on the half hour have been available for on-demand listening already for years. The process is fully automated and requires no effort whatsoever from the newsroom.



With some additional metadata, those bits of content would be ideal for consumption on VAT. It would also be very good regional and local service. The Finnish Broadcasting Company has a news app named Yle Newswatch. The app is gaining popularity, and less than two years ago the regional newsrooms began publishing local, many times even hyperlocal, news content directly to the app. Users get a notification of their local news, based both on their preferences and the location information provided by their smartphone.

The half hourly broadcast news on Yle Areena could easily be included in the push notifications, and these could be based on either the location or topic, and that information would be tagged into the metadata of the programs. That tagging could be done automatically by speech-to-text.

5.4 Boost your humanity

So, instead of robots taking the jobs of news anchors, automation can help their content spread even further and wider. But, again, what guarantee do we have that it will be relevant, engaging and entertaining? How do we ensure that we are repeatedly chosen as the source for news content?

The answer lies in our human capabilities. We need to bring clear added value compared to robots. One of the simplest ways to test this is to compare the output one-on-one. The tools are available for anyone:

- 1 Record the next radio news with your smartphone's voice recorder**
- 2 Type the content out and insert it into a text-to-speech application like Google Translate, or the slightly more advanced Google Cloud Text-to-Speech or Amazon Polly, that has the best sounding voices**
- 3 Play the original news and the robotic news out at the same time and compare the output. I sincerely hope that your radio station's news anchor does not sound or feel anything like the machine.**



But how is it possible to increase the humanity in the news that have to be made to fill the time slot to the second? How do we not sound like machines when there is evidently little room for improvisation or error in those 120 seconds (minus the start and end jingles)? There is no definitive answer to this. Some news hosts are just incredibly skilled, and with their charisma they can enchant you even with the most boring bits of information. They are the masters of putting human qualities and capacities to work, and from them there is a lot to be learnt.

One way to make even the shortest news programs sound more human is to advise the anchors to not read out what they have written, but instead to narrate the message. This is not easy to explain, but the idea is to tell, not read. The right tone can of course only be achieved through practice, experience and self confidence, and the process takes time.

One key to this is to get into the right speaking mode. The news anchor should not address the public but instead talk to a friend. Maybe the conventions of commanding a smart speaker could help news anchors here? When the red light comes on, the anchor could imagine that someone has just asked their radio a question: “Hey Ville, what’s the news?”. Maybe that could help the anchor to find the human conversational mode and mood?

In my view, this is the most important aspect of creating human relevance into audio content in our daily work. We need to outperform robots, and at the same time make machine learning and artificial intelligence to work with humans for the best possible end product.



6 Conversational news assistant and editorial ethics

In this final section we will look at the structural issues that newsrooms face as they use this new technology. How much control do they have over their content and how can they protect their ethical approach when they are working with third parties, especially the technology companies that provide the platforms and algorithms for Voice? Finally, can newsrooms use Voice to help fight misinformation, a critical part of their public service mission in an era of so-called “fake news”?

6.1 Who answers when you ask for more?

So, you got a smart speaker for Christmas? You have learned the right way to speak to her and she is surprisingly skilled: she reads you the news in the morning and sets the timer for seven minutes when you boil your eggs. You have successfully entered the world of Human–Machine communication.

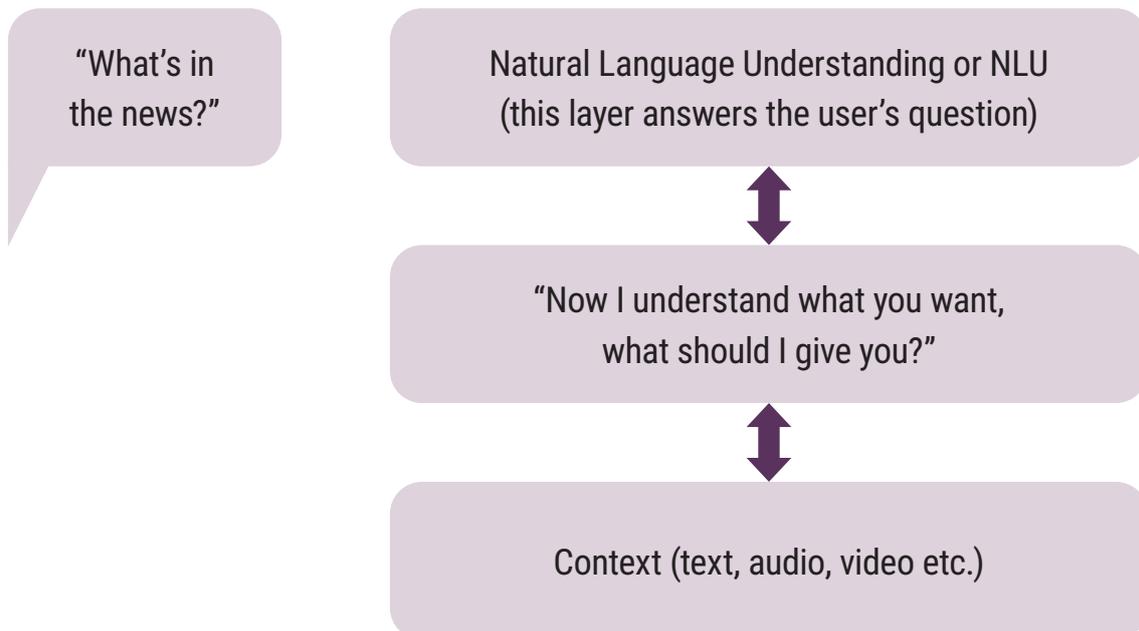
But what if you use an incorrect wording, and your speaker gives you wrong information without you even knowing it? A question like “Why did Turkey attack Syria,” is obviously controversial and a classic area for misinformation.

Can VAT be trusted in news conversations? Mukul Devichand from the BBC says there is a real risk in using technology that the news organisations does not control:

“That is a good question, because in social media, a number of the main platforms are controlled by large technology companies. Their algorithms are not deliberately set out to do anything bad, instead they’re designed to be really good, but we have seen that they’re easily gamed. Everyone from political parties to foreign governments have been trying to manipulate processes.”



The BBC's Mukul Devichand says that he uses a three-level model for understanding Voice, as "a conversational machine".



On the bottom level is editorial, journalistic content, where information is drawn from. It can be text, audio, video or any other format.

The middle level consists of algorithms, with the idea of "now I understand what you want, what should I give you?", based on an understanding of the user data, collected from various places such as social media behaviour, web search history and previously visited sites.

On the top third level is Natural Language Understanding or NLU. It is a form of AI that understands input made in the form of sentences in text or speech format (Rouse et. al, 2019). This layer is the one that will answer the user's question.

NLU is able to pass a statement made by a human, turn it from speech to text and then to infer intention from the speech. "That is a huge deal. That has been very hard for machines to get to", Devichand says.

Mukul Devichand says that broadcasters need to be aware this three-layer model, because AI can have editorial impact on the content. That is, according to Devichand, one of the reasons why the BBC has launched their own Alexa skill that can be requested to tell more, because with their own skill they can assure the user that the information comes from a legitimate source.



But very few publishers have the kind of extensive daily output and massive archives that the BBC or Yle News have. Nor have they got the resources to build highly advanced skills, actions or applications, but instead they have to trust that their content gets picked up and curated into third party feeds.

The *Evening Standard* are relying on their content getting picked up by the Google News feed, so they are making their content mainly into two durations: one 30–60 seconds and the other one two to 10 minutes long, says executive producer Chris Stone:

“Google want enough content to jigsaw a feed together from. We are currently producing a minimum of 55 of those a day.”

David Caswell from BBC News Labs also points out that publishers have to let go of control to some extent:

“Publishers do not have much choice in this because it’s already happening. The competition in audio content in current affairs, commentary and also news is on a very high level.”

The audience relationship was also considered at the Guardian Voice Lab says David O’Donnell:

“By designing for voice, the platforms insert themselves between the user and the news provider. The presence of this intermediary can be seen as a threat by the news organisations, and users might not understand that there are third party actions or skills at all on these platforms.”

6.2 How much control over content can you give away?

Google has fairly strict requirements for content to be eligible for inclusion into Google audio news. The content must meet “journalistic standards” (which Google does not elaborate on), but there is also a sizeable list of requirements: each story must be more than 30 seconds and less than 10 minutes long. Headlines cannot exceed 40 characters and must contain verbs and articles, and words such as “breaking”, “today”, and “this morning” cannot be used.



There is nothing inherently bad in those requirements, but there is a clear potential that publishers may accidentally violate their own codes of journalistic ethics. In Finland, the ethics code says:

“*Decisions concerning the content of media must be made in accordance with journalistic principles. The power to make such decisions must not under any circumstances be surrendered to any party outside the editorial office.*”

Council for Mass Media, 2014

Google’s requirements can be seen, strictly speaking, to cross that line.

However, Steve Henn, who is Content Strategy Lead at Google’s News on its Assistant project, says that their intentions are good:

“[T]o provide a diverse range of news, views, and opinions from as wide a variety of authoritative sources so that users can develop their own critical thinking on a story or subject.”

Bullard, 2019

6.3 Problems with material on third-party services

Having journalistic content available on third-party platforms can raise editorial-ethical questions but also questions about quality. Bullard (2019) points out that Alexa “can stack up several updates from different publishers and play them back-to-back. This can lead to hearing the same story told by different outlets in different ways, possibly at different volumes”.

The BBC pulled its podcasts off the Google Podcasts service in March 2019. Kieran Clifton, who is the director of BBC Distribution & Business Development, then said that:

“*Google has (...) begun to direct people who search for a BBC podcast into its own podcast service, rather than BBC Sounds or other third party services, which reduces people’s choice – an approach that the BBC is not comfortable with and has consistently expressed strong concerns about.*”



The primary reason to pull content, first from Google and then from TuneIn platforms in August 2019, is the unavailability of usage data from the platforms. The BBC's Clifton stresses that data is important in making the right kind of content available to the right audiences, and the BBC is not collecting it for any commercial purposes. At a time of understandable public concerns about data collection it seems right for public service media and any other news organisations to take data collection seriously and seek transparency from the platform providers.

6.4 Voice can be stolen

Misinformation is a critical test of the media's ability to provide credible content for Voice. We have already seen the first successful audio deepfakes. One was used to scam €220,000 from a company in March 2019. The *Wall Street Journal* reported that:

“ The CEO of a UK-based energy firm thought he was speaking on the phone with his boss ...who asked him to send the funds to a Hungarian supplier... whoever was behind this incident appears to have used AI-based software to successfully mimic the German executive's voice by phone. ”

Stupp, 2019

Chris Stone from the *Evening Standard* would rely on the VAT platforms as gatekeepers with possible audio deepfakes, because the skills and actions draw their content from known feeds. Stone says that faking a feed would be difficult and unlikely, but, like others, he sees the need to be aware and prepared against wrongdoers.

Until this year, voice was thought to be, if not impossible, but at least painfully expensive, to copy. But Matthew Aylett, Chief Security Officer at voice company CereProc in Edinburgh says that a mere 30 minutes of voice is enough to steal a voice and make it fully believable. He suggests in an open access government paper that governments and regulators need to introduce dedicated legislation to protect against deepfake videos and audio.



“ At present, anyone misusing the technology is only prosecutable under laws specific to the fraud, scams and theft they commit using deepfakes. ”

Aylett, 2019

He also suggests that genuine audio be watermarked for authenticity:

“ The speech synthesis industry may also consider some form of watermark that can be applied to digital audio files. While this wouldn't prevent a deepfake being created, it would allow for it to be revealed as false or malicious faster and hopefully be contained before mass misinformation occurs. ”

Aylett, 2019

David Caswell would not hurry with audio watermarking. He says that we need to be aware of the threats, however, and be ready to act. Instead, he is more worried about pure machine learning generation of text and audio. As mentioned before, natural language understanding or NLU enables machines to try to act human. Caswell does not like all of what is in store. He is particularly concerned about the GPT-2 Open AI language model developed in California that creates attention-holding nonsense:

“ It is not semantically coherent. There's no thoughtfulness in there at all. It's just machines picking words in a probabilistic way, but at a fidelity that holds your attention. What GPT-2 offers is large amounts of very distracting nonsense, but which sounds good and is close enough to hold the attention away from semantically coherent, editorially curated or produced material. That to me is a bigger worry at this point. ”

The creators of GPT-2 said earlier in 2019 that they would not release their better model, because they are concerned about possible malicious applications of the technology. However, in November 2019, they did release the 1,5B model after finding “no strong evidence of misuse so far”.



6.5 Stick to the truth

Disinformation, misinformation and malinformation are the three types of “information pollution” or “fake news” that are here to stay: *Disinformation* is false information with an intent to harm. *Misinformation* is false information distributed by mistake, and *malinformation* is true information distributed with an intent to harm (Wardle, C. Derakshan, H., 2017).

It is the duty of public service broadcasting to prevent such things from causing harm and to try to replace them with objective, trustworthy information. Therefore, in public broadcast content there must never be a slightest hint of disorderly information or unnecessary controversy. Mistakes will naturally happen, and they will be corrected according to policy.

How do AI, voice assistance and conversational news assistants fit into this? Public broadcasters must be sure to release their content only into platforms where they can be sure of transparent and unaltered usage and distribution. When public service content needs to be made available for distribution in third-party services, appropriate usage needs to be ensured and misuse prevented.

But what control do the original content producers have, if and when their publicly released content gets pulled into third-party feeds? One approach could be to block access to content for parties that do not properly credit the original publishers. The same applies to evident mal-informative use, if any of that should be discovered. With voice content, this is where audio watermarking should be considered for proof of authenticity.

If conversational news assistants become popular, it is essential for public service companies to be present in them and with easily recognisable human voices and audio brands. Only that way can public broadcasters not only fill their duties but also be relevant to the tech-savvy audiences. That is why conversational AI models need to be further researched and applied into daily newsroom work as soon as possible in order to be ready for the change. And what would be more cool than having your radio respond to you, finally after all these years!



References

Aylett, M. (2019), "Fake news: Debunking and defeating deepfake audio", available at: <https://www.openaccessgovernment.org/deepfake-audio/71770/>

Beckett, C. (2019), "New powers, new responsibilities: a global survey of journalism and artificial intelligence" (Polis/LSE), available at: <https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/>

Blackie, R. (2019), "Opinion: The case for smart speakers being overhyped", available at: <https://www.thedrum.com/opinion/2019/02/20/the-case-smart-speakers-being-overhyped>

Breslin, C. (2019), "What are Virtual Assistants?", available at: <https://voicebot.ai/2019/10/05/what-are-virtual-assistants/>

Bruce, A., (2019), "How to polish your voice skill with SSML on Alexa and Google Assistant", available at: <https://www.screenmedia.co.uk/lab/how-to-polish-your-voice-skill-with-ssml-on-alexa-and-google-assistant/>

Bullard, G. (2019) "Smart speaker use is growing. Will news grow with It?" available at: <https://niemanreports.org/articles/reimagining-audio-news/>

Clifton, K. (2019), "BBC podcasts on third-party apps" (<https://www.bbc.co.uk/blogs/aboutthebbc/entries/d68712d7-bd24-440f-94a0-1c6a4cdee71a>)

Clifton, K. (2019) "Changes to access of BBC live radio streams on selected TuneIn platforms in the UK", available at: <https://www.bbc.co.uk/blogs/aboutthebbc/entries/37e4e3f6-fbd2-4c14-8d72-7f7139641582>

Clifton, K. (2019) "Further update on the BBC and TuneIn", available at: <https://www.bbc.co.uk/blogs/aboutthebbc/entries/ba37d6b5-59fd-470d-9da2-dec1b45b8bd5>



Cock, J. (2019), "Routenote: Switzerland will end FM Radio, the 2nd country to do so", available at: <http://routenote.com/blog/switzerland-will-end-fm-radio-the-2nd-country-to-do-so/>

Devichand, M., (2019), "About the BBC blog: Alexa, give me BBC news I can trust", available at: <https://www.bbc.co.uk/blogs/aboutthebbc/entries/f335e227-c3eb-4b8d-ae43-f299bee87a29>

Follett, J., (2019) "Talking to Machines – Language, AI, and the Next User Interface", available at: <https://towardsdatascience.com/talking-to-machines-3da09845c251>

Galan, L. Osserman, J. Parker, T. Taylor, M. (2019) "How Young People Consume News and the Implications for Mainstream Media", available at: <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-08/FlamingoxREUTERS-Report-Full-KG-V28.pdf>

Heilweil, R. (2019) "Alexa, why should we trust you with the news?", accessed online: <https://medium.com/news-to-table/alex-a-why-should-we-trust-you-with-the-news-af2b942797e>

Hirvonen, T. (2019) "Kirjat ahmitaan nyt kuulokkeista, ja se saattaa j risytt   k sityst mme perinteisest  romaanista – "Pit   mietti  valmiiksi yhdeks n cliffhangeria", available at: <https://yle.fi/uutiset/3-11026426>

King, A., "Talk to Transformer", available at: <https://talktotransformer.com/>

Leviathan, Y. Matias, Y. (2019), "Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone", available at: <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>

Lewis, S. Guzman, A. Schmidt, T. (2019) "Automation, journalism, and Human–Machine Communication: Rethinking roles and relationships of humans and machines in news", available at: <https://www.tandfonline.com/doi/full/10.1080/21670811.2019.1577147>

Newman, N. Fletcher, R. Kalogeropoulos, A. Nielsen, R. (2019), "Reuters Institute Digital News Report, available at: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/inline-files/DNR_2019_FINAL.pdf



Newman, N., (2018) "Reuters Institute, The Future of Voice and the Implications for News", available at: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-11/Newman%20-%20Future%20of%20Voice%20FINAL_2.pdf

Olson, C., (2006) "Just say it: The future of search is voice and personal digital assistants", available at: <https://www.campaignlive.co.uk/article/just-say-it-future-search-voice-personal-digital-assistants/1392459>

Radford, A. Wu, J. Amodei, D. Amodei, D. Clark, J. Brundage, M. Sutskever, I. (2019), "Better language models and their implications", available at: <https://openai.com/blog/better-language-models/>

Rouse, M. Haughn, M., (2019), "Natural language understanding (NLU)", available at: <https://searchenterpriseai.techtarget.com/definition/natural-language-understanding-NLU>

Shenker, L., (1999), "Investigation and evaluation of the use of FM radio equipped hearing protection devices in the Australian meat industry", available at: <https://www.mla.com.au/download/finalreports?itemId=2517>

Solaiman, I. Clark, J. Brundage, M. (2019) GPT-2: 1.5B Release, available at: <https://openai.com/blog/gpt-2-1-5b-release/>

Stupp, C. (2019), "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case", available at: <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>

Vincent, J. (2019), "OpenAI has published the text-generating AI it said was too dangerous to share", available at: <https://www.theverge.com/2019/11/7/20953040/openai-text-generation-ai-gpt-2-full-model-release-1-5b-parameters>

Wardle, C. Derakshan, H., (2017), "Information disorder: Toward an interdisciplinary framework for research and policy making", available at: <https://rm.coe.int/information-disorder-report-november-2017/1680764666>



Resources

BBC News, (2019), “BBC to launch digital voice assistant”, available at: <https://www.bbc.co.uk/news/technology-49481210>

European Broadcasting Union EBU: Trust in media (2019), available at: https://www.ebu.ch/files/live/sites/ebu/files/Publications/MIS/login_only/market_insights/EBU-MIS-Trust_in_Media_2019.pdf

Finnish council for mass media: Guidelines for journalists (2014), available at: https://www.jsn.fi/en/guidelines_for_journalists/

Google: News on Google Assistant, developer guidelines (2019), available at: <https://developers.google.com/news/assistant/audio-news/overview>

Google New Zealand Blog: Google pilot aims to provide accurate store hours in New Zealand this Labour Day, available at: <https://newzealand.googleblog.com/2019/10/google-pilot-aims-to-provide-accurate.html>

Google Text-to-Speech, available at: <https://cloud.google.com/text-to-speech/>

International Judo Federation, (2019), Disqualification due to Cellphone on Tatami, available at: <https://youtu.be/-LdOTqHlto0>

Microsoft voice report (2019), available at: https://advertiseonbing-blob.azureedge.net/blob/bingads/media/insight/whitepapers/2019/04%20apr/voice-report/bingads_2019_voicereport.pdf

Mission, values and public purposes of the BBC, available at: <https://www.bbc.com/aboutthebbc/governance/mission>

NPR Smart audio report (2019), available at: <https://www.nationalpublicmedia.com/smart-audio-report/latest-report/>



RAJAR Q3 (2019), Digital Radio listening on the up, available at: <https://radiotoday.co.uk/2019/10/rajar-q3-2019-radio-listening-on-the-up/>

Spotify: Your Daily Drive: Music and news that'll brighten your commute (2019), available at: <https://newsroom.spotify.com/2019-06-12/your-daily-drive-music-and-news-thatll-brighten-your-commute/>

Yle (2018): Voitto-robotti avataan kaikkien käyttöön, available at: <https://yle.fi/uutiset/3-10125596>

Yle's strategy, available at: <https://yle.fi/aihe/artikkeli/2016/01/13/yles-strategy>



Interviews

Interview with David Caswell, Executive product manager of BBC News Labs, 18/10/2019

Interview with Mukul Devichand, executive editor for BBC Voice and AI, 14/10/2019

Interview with Dave O'Donnell, editorial lead of the Guardian Voice Lab project 10/2018–4/2019, 16/10/2019

Interview with Jim Salveson, head of Sport Social at Voiceworks, 17/10/2019

Interview with Chris Stone, executive producer of video & audio, Evening Standard, 15/10/2019



THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE ■

Polis
Department of Media and Communications
The London School of Economics
and Political Science
Houghton Street
London WC2A 2AE

Email: polis@lse.ac.uk

Telephone: 020 7107 5176

blogs.lse.ac.uk/polis

lse.ac.uk/media-and-communications

